参考資料５

Chapter 5

Dose–response assessment and derivation of health-based guidance values

Second edition
(2020)

5. DOSE–RESPONSE ASSESSMENT AND DERIVATION OF HEALTH-BASED GUIDANCE VALUES

This chapter updates Chapter 5 of Environmental Health Criteria 240 (EHC 240), which was originally published in 2009. It was developed through an expert consultation and further advanced following comments received through a public consultation in December 2019.

For abbreviations used in the text, the reader may refer to the list of abbreviations at the front of this chapter. Definitions of select terms may be found in the glossary in Annex 1 of EHC 240 (http://www.inchem.org/documents/ehc/ehc/ehc240_annex1.pdf).

List of abbreviations

ADI	acceptable daily intake
ADME	absorption, distribution, metabolism and excretion
AIC	Akaike information criterion
ARfD	acute reference dose
AUC	area under the concentration–time curve
BBDR	biologically based dose–response
BBMDS	Bayesian Benchmark Dose System
BIC	Bayesian information criterion
BMD	benchmark dose
BMDL	95% lower confidence limit on the benchmark dose
$BMDL_{05}$	95% lower confidence limit on the benchmark dose for a 5% response
$BMDL_{10}$	95% lower confidence limit on the benchmark dose for a 10% response
BMDS	Benchmark Dose Software
BMDU	95% upper confidence limit on the benchmark dose
BMR	benchmark response
C_{max}	peak plasma concentration
CSAF	chemical-specific adjustment factor
DDEF	data-derived extrapolation factor
DNA	deoxyribonucleic acid
DRM	dose–response modelling
ED	exposure dose (concentration)
EFSA	European Food Safety Authority
EHC	Environmental Health Criteria
FAO	Food and Agriculture Organization of the United Nations
HBGV	health-based guidance value
IPCS	International Programme on Chemical Safety
IQ	intelligence quotient
JECFA	Joint FAO/WHO Expert Committee on Food Additives
JMPR	Joint FAO/WHO Meeting on Pesticide Residues
LOAEL	lowest-observed-adverse-effect level
mADI	microbiological acceptable daily intake
mARfD	microbiological acute reference dose
MD	maximum dose tested
MIC	minimum inhibitory concentration
MIC_{50}	minimum inhibitory concentration for 50% of strains of the most sensitive relevant organism
MIC_{calc}	90% lower confidence limit for the mean MIC_{50}

MOA	mode of action
MOE	margin of exposure
MRL	maximum residue limit
N	number of subjects
NOAEC	no-observed-adverse-effect concentration
NOAEL	no-observed-adverse-effect level
OECD	Organisation for Economic Co-operation and Development
PBDE	polybrominated diphenyl ether
PBPK	physiologically based pharmacokinetic
PBTK	physiologically based toxicokinetic
PFOA	perfluorooctanoic acid
PMTDI	provisional maximum tolerable daily intake
POD	point of departure
PTMI	provisional tolerable monthly intake
PTWI	provisional tolerable weekly intake
RIVM	National Institute for Public Health and the Environment (the Netherlands)
SD	standard deviation
TDI	tolerable daily intake
TEQ	toxic equivalent
TMI	tolerable monthly intake
Tox21	Toxicology Testing in the 21st Century
ToxCast	Toxicity Forecaster
TWI	tolerable weekly intake
UF	uncertainty factor
USA	United States of America
USEPA	United States Environmental Protection Agency
VICH	International Cooperation on Harmonisation of Technical Requirements for Registration of Veterinary Medicinal Products
WHO	World Health Organization

List of contributors

Scott Auerbach
National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, United States of America (USA)

Susan Barlow (*co-lead author*)
Brighton, East Sussex, United Kingdom

Virunya Bhat
NSF International, Ann Arbor, Michigan, USA

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

Bernard Bottex
European Food Safety Authority, Parma, Italy

John Bowers
Center for Food Safety and Applied Nutrition, United States Food and Drug Administration, College Park, Maryland, USA

Roni Bronson
Health Products and Food Branch, Food Directorate, Bureau of Chemical Safety, Health Canada, Ottawa, Ontario, Canada

Clark Carrington
Gaithersburg, Maryland, USA

Jose Cortinas
European Food Safety Authority, Parma, Italy

Stuart Creton
Food Standards Australia New Zealand, Canberra, Australia

Allen Davis
National Center for Environmental Assessment, United States Environmental Protection Agency, Cincinnati, Ohio, USA

Lutz Edler
Dudenhofen, Germany

Lynne Haber
University of Cincinnati, Cincinnati, Ohio, USA

Thor Halldorsson
Faculty of Food Science and Nutrition, University of Iceland, Reykjavik, Iceland

Akihiko Hirose
National Institute of Health Sciences, Kawasaki-ku, Kawazaki, Kanagawa, Tokyo, Japan

François Pouzaud
French Agency for Food, Environmental and Occupational Health & Safety (ANSES), Maisons-Alfort, France

David Richardson
University of North Carolina, Chapel Hill, North Carolina, USA

Susanne Rudzok
Unit of Toxicology of Active Substances and their Metabolites, Department of Pesticides Safety, German Federal Institute for Risk Assessment (BfR), Berlin, Germany

Josef Schlatter
Zurich, Switzerland

Klaus Schneider
Forschungs- und Beratungsinstitut Gefahrstoffe GmbH, Freiburg, Germany

Matthew Wheeler (*co-lead author*)
National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA

Gerrit Wolterink
Centre for Nutrition, Prevention and Health Services, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands

Janet Zang
Center for Food Safety and Applied Nutrition, United States Food and Drug Administration, College Park, Maryland, USA

5.1 Introduction

For a given chemical, a health-based guidance value (HBGV) represents a range of exposures (either acute or chronic) that are expected to be without appreciable health risk. As EHC 240 focuses on risk assessment of chemicals in food, the derivation of HBGVs for oral (dietary) exposures is discussed in this chapter. A key step in establishing an HBGV is determining a point of departure (POD), which is a reference point or range on a dose–response curve that is derived from experimental or observational data. This reference point is the lower bound on dose (upper bound on response) that corresponds to an estimated or predetermined low-effect level or no-effect level of the dose–response curve. There are a variety of techniques that can be employed to derive a POD, all of which fit under the umbrella of dose–response assessment.

Historically, rather simple approaches were followed to derive a POD when utilizing dose–response methodology for risk assessment, such as using the no-observed-adverse-effect level (NOAEL) or lowest-observed-adverse-effect level (LOAEL) of the observed dose–response curve. This approach, hereinafter referred to as the NOAEL approach, is mostly based on stepwise statistical comparisons of responses between the control group and the different dose groups and can be considered to be a simple conceptual model that describes the (mean) response by a mathematical step function, where the steps correspond to the sequence of the experimental doses.

Towards the end of the last century, a more advanced modelling approach, the so-called benchmark dose (BMD) approach, was introduced for dose–response assessment (Crump, 1984). This approach considers all available dose–response data together in a comprehensive data analysis that allows interpolation between tested doses to estimate a BMD. The BMD is a dose associated with a specified incidence of risk for a health effect or a specified change in biological response, which is defined as the benchmark response (BMR). The BMD and, more importantly, its statistically derived 95% lower confidence limit, known as the BMDL, can be derived from dose–response modelling (DRM). The BMDL is then usually used as the POD for establishing an HBGV.

The BMD approach, rather than the NOAEL approach, is the approach now preferred by the Joint Food and Agriculture Organization of the United Nations (FAO)/World Health

Organization (WHO) Expert Committee on Food Additives (JECFA) and the Joint FAO/WHO Meeting on Pesticide Residues (JMPR) for deriving a POD.

In what follows, generally accepted practices for dose–response assessment are outlined, and the derivation of the POD from each of the two methods mentioned above is considered. The process of establishing HBGVs to protect against toxic effects is described. In instances where an HBGV cannot be established, the POD can also serve as a starting point to estimate the margin of exposure (MOE), which is the ratio of the POD for the critical effect to the theoretical, predicted or estimated dietary exposure. Application of the MOE approach is described in section 5.5.

5.2 Dose–response assessment

5.2.1 Basic concepts of dose–response assessment

Hazard characterization is one of the four basic steps in risk assessment. In this chapter, the term dose–response assessment refers to a key step in the process of hazard characterization (see Chapter 2, section 2.4). It is used as a general term to include the entire procedure for establishing HBGVs (as described in section 5.4), including data selection, developing a POD (whether by modelling or using other approaches) and application of uncertainty factors, if appropriate. The term dose–response modelling or DRM refers specifically to modelling approaches.

Dose–response assessments establish relationships between exposure and the adverse health outcomes under study. They are used to link a database of adverse effects to target human populations that are exposed and at risk and for which an HBGV should be established. This section gives a short overview of current practices in dose–response assessment, including generic issues regarding the two basic components, dose and response.

One of the primary steps when performing a risk assessment is the determination of the presence or absence of a dose–response relationship. This is facilitated by using biochemical, toxicological and (if appropriate) pharmacological information from the hazard identification step (the first step in risk assessment) and the outcome of a statistical analysis of available and relevant dose–response data. The existence of a dose–response relationship is usually supported by statistically significant differences in response between dosed and

control groups or dose-dependent trends in response. In cases where the response is not statistically significant, it may be considered that the level of exposure is without biologically significant adverse health effects. However, it should also be borne in mind that the power of the study to detect an adverse effect might be too low.

Dose–response data, which usually provide the basis for risk characterization (the fourth step in risk assessment), may come from in vivo experimental studies in laboratory animals, experimental studies in humans or epidemiological studies in humans. These types of data have been used in dose–response assessments by JECFA and JMPR in three main ways:

1) establishment of an HBGV, such as an acceptable daily intake (ADI), tolerable daily intake (TDI) or acute reference dose (ARfD);
2) estimation of the MOE between a defined POD and the level of human exposure; and
3) quantification of the magnitude of the risk at specified levels of human exposure.

These approaches and variants of them are discussed below, based on EHC 239, *Principles for modelling dose–response for the risk assessment of chemicals* (IPCS, 2009), together with more recent considerations and practical experience in JECFA and JMPR.

Whereas the focus has traditionally been primarily on experimental animal studies, dose–response relationships are also critical in the assessment of in vitro data and for studies that attempt to define the relationships of different steps in a postulated mode of action (MOA). Owing in part to governmental mandates and a desire to reduce experimental animal testing and be more efficient and human relevant in toxicological assessment, dose–response data from in vitro studies are increasingly being considered in risk assessment (Thomas et al., 2019). In vitro data can be subject to DRM using the BMD approach described in this chapter (Sand et al., 2017). To date, uses of in vitro data in risk assessment include MOA investigations (Li et al., 2017) and comparative biological potency assessment (Bhat et al., 2017).

One unique consideration when modelling in vitro data is the manner or scale in which the data are generated. Relative to in vivo data or single end-point in vitro data, high-throughput in vitro data, such as those generated by the USA's Toxicology Testing in the 21st

Century (Tox21) collaboration (Tice et al., 2013) and the United States Environmental Protection Agency's (USEPA) Toxicity Forecaster (ToxCast) programme (Judson et al., 2010), contain additional complexities that are a by-product of experimental design and data pre-processing and that must be accounted for prior to modelling. It is therefore recommended that an expert be consulted when modelling high-throughput in vitro data.

There are a number of inherent challenges to the use of in vitro data in risk assessment that have been discussed elsewhere (Crump, Chen & Louis, 2010). One significant challenge that is central to risk-associated BMD modelling is the establishment of a BMR that reflects a level of change that is considered adverse. Not only is the qualitative relationship between in vitro effects and adverse in vivo effects often unknown, but even when a qualitative relationship to adversity can be established, the level of change in an in vitro system that reflects an adverse change is also often unknown. Because of this challenge, in vitro data modelling has employed BMR values that are based on statistical thresholds as opposed to toxicologically derived response thresholds.

While scientific efforts are ongoing to address the challenges of in vitro toxicity testing, it is likely that, in the future, risk assessors will encounter greater quantities of in vitro data. Risk assessors are encouraged to consider the use of in vitro data within appropriate contexts and to engage with scientific researchers to clarify their needs and facilitate use of the data. Guidance on DRM of in vitro data is not included in this chapter at this time, as the methodology is still evolving.

5.2.1.1 Dose

In this chapter, the term "dose" is used exclusively to describe any type of chemical exposure. This includes experimental dose as well as what is typically called external concentration or incidental exposure.

For both laboratory animal and human studies, it is critical, when performing dose–response analyses for food risk assessment, to have a clear concept of what type of dose has been used in the study that produced the available dose–response data. There are different types of dose that arise from scientific investigations, and these have different interpretations and should be distinguished. Although highly

interrelated, each of them can be used to express a relevant dose–response relationship.

There are primarily three types of dose that exist simultaneously in the exposed organism:

1) the administered dose or external dose;
2) the absorbed dose or internal dose; and
3) the target or tissue dose.

The administered or external dose denotes the amount of a chemical administered to experimental animals or humans in a controlled experimental setting by some specific route at some specific frequency or through incidental exposure in occupational or epidemiological studies. In the terminology used by JECFA and JMPR, the external dose is often referred to as exposure or intake (see Chapter 6). Oral administration is the most relevant route in the context of food risk assessment, and, as such, this chapter focuses on oral exposure. Data from other routes of exposure would need appropriate conversions requiring special consideration, in particular data comparing toxicokinetics following different routes of exposure.

The absorbed or internal dose is the amount of chemical that is systemically available and can be regarded as the fraction of the external dose that is absorbed and enters the general circulation. It is affected by the absorption, distribution, metabolism and excretion (ADME) of the chemical and can be derived from suitable toxicokinetic mass balance studies. The analytical method used in the toxicokinetic studies will determine whether the dose refers to the parent compound alone or to the parent compound plus its active metabolites (see Chapter 4, section 4.2). Biomarkers of body burden, such as concentrations in whole blood, plasma or urine, are sometimes available in epidemiological studies and can be used as the internal dose and for the dose–response assessment of human data.

The target or tissue dose is the amount of chemical present in a specific tissue of interest. This type of dose is relevant when a response at the target tissue of interest is selected as the critical effect for the risk characterization (e.g. total mercury and methylmercury concentrations in umbilical cord tissue were used by JECFA as biomarkers of fetal exposure to methylmercury; FAO/WHO, 2007). As for the absorbed or internal dose, the analytical method used in toxicokinetic studies will determine whether the tissue dose refers to

the parent compound, to active metabolites alone or to the parent compound plus metabolites (see Chapter 4, section 4.2).

The description of dose should reflect the magnitude, frequency and duration over which it applies. The aim and the overall type of the study that provides the dose–response data should be recognized: was it an acute, short-term or long-term experimental study or an observational study? Acute and short-term studies are typically the first choice for the establishment of an HBGV for acute exposure (e.g. an ARfD). Long-term (and sometimes short-term) studies are usually the first choice for an HBGV for chronic exposure (e.g. an ADI or TDI). The three types of dose identified above can apply to any of these studies, and the principles of dose–response assessment apply to all three forms. Considering frequency of dosing, one may differentiate between a daily bolus (oral gavage) administration and repeated administration (in food or feed), which will affect the absorption and toxicokinetics of the chemical under examination. It is well known that the gavage dose needs to be adjusted when laboratory animals are dosed for less than 7 days per week (e.g. it was common in the past to dose laboratory animals for only 5 working days per week). It should also be noted that controlled experiments in laboratory animals often adjust the dosing to account for the contemporary (often the weekly average) body weight.

The dose metric for internal and tissue doses is typically a concentration at a specific time point or total concentration integrated over time, such as the area under the plasma concentration–time curve (AUC). Internal and tissue doses can also be expressed as peak body burdens or body burdens averaged over a given period of time (e.g. nanograms per kilogram of body weight) or tissue concentration (e.g. nanograms per kilogram of tissue).

External dose can be expressed using a variety of metrics, including a simple single dose (e.g. milligrams per kilogram of body weight when establishing an HBGV for acute exposure) or a daily intake or daily exposure (e.g. milligrams per kilogram of body weight per day when establishing an HBGV for long-term exposure). When the chemical is administered via food, feed or drinking-water, concentrations (e.g. milligrams per kilogram of food or feed or milligrams per litre of drinking-water) are the original dose metric. In such cases, the concentrations in food, feed or drinking-water can be converted into an external dose, either before DRM or after DRM but before an HBGV is established. Conversions are preferably made by

using measured intakes of food, feed or drinking-water together with contemporaneous body weight measurements to calculate the actual exposure, which is then expressed as "X mg/kg food/feed or X mg/L water, equal to Y mg/kg body weight per day". Alternatively, if such measurements are not available, default conversion factors (see Table 1 in WHO, 2015, which is an updated version of the dose conversion table in Annex 2) are used to estimate exposure, which is then expressed as "equivalent to" rather than "equal to".

External dose or external exposure is frequently the dose metric that is used in observational epidemiological studies where the actual exposure is rarely known precisely. Its estimation often requires various assumptions, such as when dietary exposure is estimated retrospectively from dietary survey data, using either individual and subject-specific data or aggregated data from subjects grouped into a single exposure group. In general, these exposures are less precise than those obtained in controlled experiments. Sometimes exposure is measured by the biomonitoring of blood or tissue concentrations. Dose–response assessment for such data usually raises the issue of conversion of the biomarker of internal exposure into an external dose. An additional problem that has arisen (e.g. with the dioxin database) is that measurements of the biomarker were made many years after what was believed to be the period of highest exposure (FAO/WHO, 2002a); in such cases, concentrations in blood or tissue at the time of the original exposure were back-calculated using, for example, half-lives of the chemicals or physiologically based toxicokinetic (PBTK) modelling.

PBTK models (or physiologically based pharmacokinetic [PBPK] models, as they are often termed) quantitatively describe the ADME of the chemical of interest (and, to the degree needed, of its metabolites). IPCS (2010) provides guidance for developing and documenting PBPK models and applying them in risk assessment contexts. A key part of that guidance is that the design and testing of the model should be tied to the needs for the use of the model in the risk assessment. These needs, and the specifics of the chemical's ADME and MOA, determine what should be included in the model, such as the specific tissues described by the model and the level of detail needed for the description of metabolite kinetics. IPCS (2010) notes a variety of potential uses for PBPK modelling in risk assessment, including estimating internal dose metrics, extrapolating across species, evaluating human pharmacokinetic variability, route-to-route extrapolation and cumulative risk assessment. MOA

evaluation is integral to PBPK/PBTK modelling, with the MOA determining the appropriate dose metric for the model and PBPK/PBTK modelling having the potential to help in identifying the appropriate dose metric from different alternatives. Development and application of PBPK/PBTK models can be data and labour intensive, but their use in risk assessment is growing as more experience is acquired. In addition, the development of standard methods and documentation, such as described by the IPCS (2010) guidance, is providing greater confidence in the results.

Sometimes the doses used in an experimental animal study are transformed to the equivalent human exposures prior to DRM. A basic approach for such transformation is allometric scaling, based on proportional differences in body size between the experimental animal species and humans (USEPA, 2011; Nair & Jacob, 2016; IPCS, 2018 [section 4.4.1]). Information about the relationship between external dose and internal dose (e.g. based on AUC in blood) derived from chemical-specific toxicokinetic data can be incorporated into chemical-specific adjustment factors (CSAFs; see section 5.4.2 and IPCS, 1994, 2005). Although CSAFs are usually thought of as a replacement for uncertainty factors, they can also be considered to be an adjustment to dose based on relative internal doses. Regardless of the approach used to adjust dose, it is preferable to make such interspecies adjustments prior to conducting DRM. When the adjustment is the same regardless of dose, the ultimate result is the same regardless of whether it is made prior to or after modelling. However, if the adjustment varies with dose (e.g. if the experimental animal body weight varies with dose or if nonlinear toxicokinetics are captured in a PBPK/PBTK model), then conducting the adjustments prior to the modelling can often improve the overall fit of the models to the data. Regardless of when adjustments are made for interspecies differences, application of uncertainty factors related to database deficiencies should be conducted after DRM.

When quantitative information is also available on a chemical's toxicodynamics, such information can be used to develop a biologically based dose–response (BBDR) model. A BBDR model includes a quantitative description of the chemical's toxicokinetics and toxicodynamics and thus describes the dose–response relationship from the external dose to the internal dose to the tissue response. The toxicodynamic part of the model may be relatively simple or may be as complicated as a fully elaborated stochastic

model for carcinogenesis. Owing to the large amount of data and effort needed to develop and verify BBDR models, their use in risk assessment is much rarer than that of PBPK/PBTK models. BBDR models can be thought of as a quantitative expression of a set of biological hypotheses that, when rigorously tested against critical experiments, becomes a credible tool for extrapolating from experimental results into exposure realms that are difficult or expensive to reproduce in controlled experiments. To date, BBDR models have mainly been used in research contexts, such as for evaluating mechanisms of toxicity, interspecies extrapolation and identification of data gaps, but they have also found limited application in a regulatory context (reviewed by Andersen & Dennison, 2001; Crump, Chen & Louis, 2010). BBDR models are currently quite expensive to construct in terms of both resources and time and thus would be expected to be developed fully only for exposures and toxicities of the highest concern.

5.2.1.2 Response

Toxicological, pharmacological and microbiological responses generally relate to an observation or effect seen following exposure in vivo or in vitro. Possible end-points cover a broad range of observations, from early responses, such as biochemical alterations, to more apical responses, such as cancer and developmental defects. The discussions in the rest of this section are focused on toxicological responses, whereas pharmacological and microbiological responses used in establishing HBGVs primarily for veterinary drugs will be discussed in sections 5.4.3.2 and 5.4.3.3, respectively.

For the purpose of determining a POD, it is important to consider the entire hazard identification data set and to identify the critical toxicological effects considered to be treatment-related adverse responses. Adverse responses are defined as a change in the morphology, physiology, growth, development, reproduction or lifespan of an organism or subsystem (e.g. subpopulation of cells) that results in an impairment of functional capacity, an impairment of the capacity to compensate for additional stress or an increase in susceptibility to other influences (IPCS, 2004). To discriminate between adverse and non-adverse effects, consideration should be given to whether the observed effect is an adaptive or local response, transient or reversible, of minor magnitude or frequency, a specific response of an organ or system, or a response secondary to general

toxicity (WHO, 2015). A more detailed discussion on adversity can be found in Chapter 4.

When establishing HBGVs, only effects that link to adverse responses should be chosen for dose–response assessment. In other words, PODs should be based on effects linked to adverse responses. Such responses are sometimes species or tissue specific and have different degrees of variation across individuals. When identifying the critical adverse effect, it is also important to consider whether an adverse effect observed in experimental animals is relevant to humans and whether it is caused by the parent compound or its metabolites.

Most observed responses of interest fall into one of the following categories:

1) *Continuous responses*: These generally relate to a measurement that is associated with each individual subject and can take on any value within a defined range of outcomes (e.g. changes in body weight). These types of data can be reported as either individual values or summary statistics (i.e. typically the arithmetic mean, some measure of variability, most often the arithmetic standard error or standard deviation, and the number of subjects per dose group).

2) *Quantal responses*: Also referred to as binary or dichotomous responses, these generally relate to an effect that is either observed or not observed in each individual subject (laboratory animal or human). For each dose, the number of subjects responding out of the number of subjects available is recorded, and the proportion of subjects experiencing the event is reported (e.g. the proportion of animals with a tumour in a cancer bioassay).

3) *Counts*: These generally relate to a discrete number of items measured in a single experimental unit (e.g. number of papillomas on the skin). Count data arise in a variety of contexts, especially in epidemiological studies. In some situations, count data can be handled as continuous measurements (e.g. when the count is large); however, in these situations, it is recommended that a variance stabilizing transformation technique be used (e.g. the Box-Cox transformation; Sakia, 1992).

4) *Ordinal categorical responses*: These generally take on one value from a small set of ordered values (e.g. severity grades of histopathology). Ordinal data are an intermediate type of data and reflect (ordered) severity categories – that is, they are qualitative data but with a rank order. When the categories are non-ordered, they are called nominal data, but these are rare for response data and are often aggregated to quantal data on presence or absence of a defined outcome.

5) *Hierarchically nested responses*: Nested (e.g. clustered) data occur when the observation is nested within a larger group. Results from developmental toxicity studies are a common example of this type of data: here, units (i.e. fetuses or pups) are nested within litters. Nested data require the estimation of correlation within the clusters.

6) *Multivariate responses*: These occur when multiple end-points are measured for a single experimental unit, such as when genomic data consist of thousands of gene responses over different concentrations. Modelling each gene independently would ignore correlations and other features in the data. Multivariate data require more observations than univariate ones, and the underlying model assumptions should be carefully considered and explicated. For example, the model should be validated on a wide variety of end-points and describe the data in a toxicologically meaningful way.

Although most data usually fall into one of the above categories, it is possible for data to fall into multiple categories. For example, multivariate responses may have both quantal and continuous measurements, categorical measures may have been categorized by grouping a continuous measurement into bins and, as noted above, counts can be treated as continuous in some situations. Additionally, any type of response may be nested. When using dose–response data, it is important to understand the type of data and whether any simplifying assumptions have been used that may result in a large loss of information (e.g. continuous responses that have been dichotomized into simple, quantal measures).

There are some differences in the way in which each of these different types of data are handled in DRM. However, as a general rule, the goal of DRM is to describe the distribution of the response

by modelling the mean and variance of the response as a function of dose, possibly incorporating covariates.

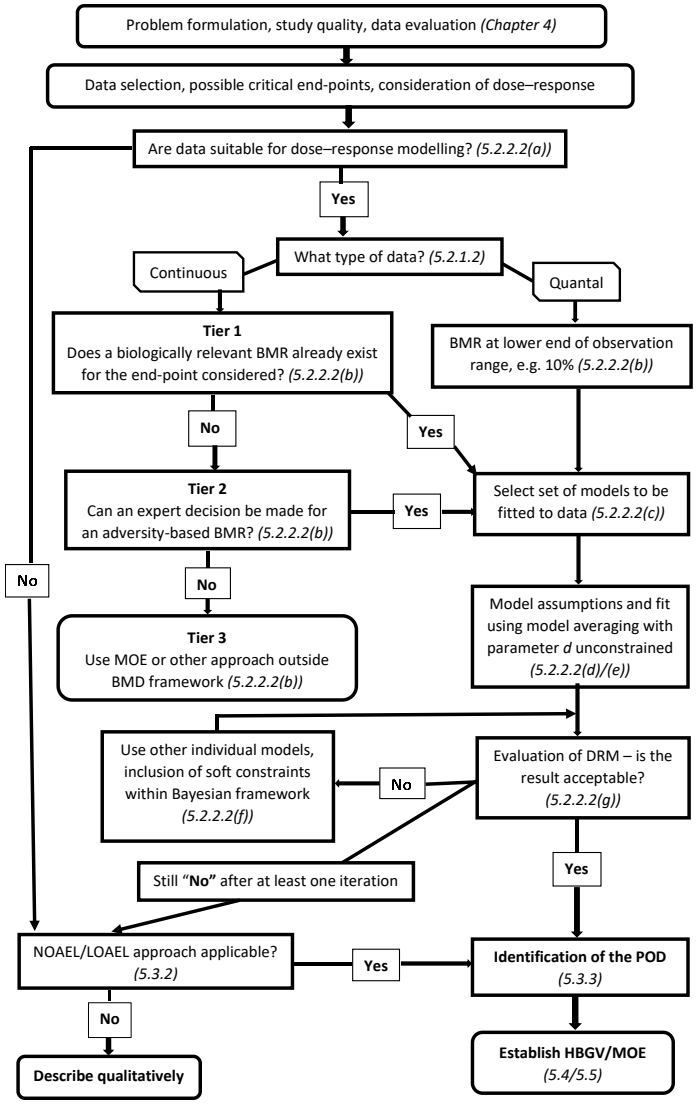
5.2.1.3 *Use of dose–response assessment to develop risk assessment advice*

Dose–response assessment may be used to develop risk assessment advice in two main ways, as summarized in Fig. 5.1:

1) *The BMD approach*: In this approach, DRM is performed by fitting a suite of mathematical models to the data in order to define the dose–response relationship, typically between the zero dose and the maximum tested/observed dose. In this setting, DRM is used to determine a dose with a specified response (the BMR). For a given BMR, the dose associated with that response, the BMD, and its 95% lower confidence limit, the BMDL, are estimated. The BMDL is then used as a POD for establishing an HBGV or for calculating an MOE. This approach is preferred by JECFA and JMPR, as the BMD approach takes all of the dose–response data into account and quantifies the uncertainty in the data.

2) *The NOAEL approach*: There will be situations in which multiple pairwise comparisons of the data for different doses can be used to identify the NOAEL or sometimes the LOAEL, which is used as a POD for the observed dose–response data. In the past, this approach was used routinely by JECFA and JMPR to establish HBGVs in order to protect against effects that are considered to show a threshold and for substances that can be controlled and are intentionally added to food. Today, however, the NOAEL approach should be used only in very limited circumstances, typically related to a very small number of dose groups or a lack of a dose–response relationship. For example, when there is only one dose group and a control group or when there is no adverse effect, even at the highest dose tested, this approach can be used. The uncertainty around the NOAEL or LOAEL should always be described. The disadvantages of the NOAEL approach are discussed below (see section 5.3.2) and have also been discussed elsewhere (Haber et al., 2018).

Fig. 5.1. Flow chart for the dose–response assessment processes used to develop risk assessment advice described in this chapter



BMD: benchmark dose; BMR: benchmark response; DRM: dose–response modelling; HBGV: health-based guidance value; LOAEL: lowest-observed-adverse-effect level; MOE: margin of exposure; NOAEL: no-observed-adverse-effect level; POD: point of departure

5.2.2 DRM for the BMD approach

Before modelling can be conducted, the data need to be evaluated as described in Chapter 4, including consideration of study quality, biological relevance, relevance to humans and the existence or absence of a dose–response relationship. The data should then be evaluated for their suitability for modelling (e.g. whether there is a dose–response trend and identifying the end-points to focus on for modelling), as described below in section 5.2.2.2(a).

Note that the step-by-step discussion here for DRM describes the modelling for a single data set, and the same process needs to be followed for each data set considered. Once a data set is selected for modelling, the family of models is chosen based on the data type (as described in sections 5.2.1.2 and 5.2.2.2(c)). The appropriate degree of change (the BMR) that is the basis for the POD is also determined, based on the nature of the data and the biology of the end-point. Specialized software is then used with a variety of flexible mathematical models to optimize parameters, fit the data and estimate the BMD and the confidence interval (BMDL–BMDU, where BMDL and BMDU are the 95% lower and upper confidence limits on the BMD, respectively). Finally, the results are documented, and uncertainty is characterized.

Covariates may be needed in DRM to account for sources of variation other than dose, such as sex and time. In such cases, the dose–response relationship is dependent upon the covariate and results in a range of PODs. Covariates are often important when accounting for variability, and inclusion in the model can improve the estimate of the DRM. For example, at its eighty-third meeting, JECFA noted that there was a difference between the incidence of liver cancer in hepatitis B–positive and hepatitis B–negative subjects who were exposed to aflatoxin. As a result, the model used reflected the potencies for both groups (FAO/WHO, 2017a).

The most sensitive relevant effect level is identified, based on the results of the modelling. The dose–response assessment then considers the totality of the data, including study quality and human relevance, to derive the overall POD for establishing the HBGV or estimating the MOE (IPCS, 2009).

For all the key decision points in the process, the rationale for the decisions should be explained and documented, including information on the underlying assumptions and uncertainties.

5.2.2.1 Software for BMD modelling

Many of the dose–response models require specialized software to fit the models to the data. There is no single preferred software package for dose–response analyses. It is important that the software used for dose–response estimation be thoroughly tested, and the source code should be made publicly available to allow for reproducibility and transparency. The version of any particular software used for the analyses should be clearly stated. For most BMD analyses, the software packages known as PROAST (https://proastweb.rivm.nl/) and Benchmark Dose Software (BMDS) (https://www.epa.gov/sites/production/files/2020-08/bmds320.zip) meet these requirements. For more detailed guidance on the use of these two publicly available software packages, the reader is referred to the application guides at https://www.rivm.nl/en/proast for PROAST and https://www.epa.gov/sites/production/files/2020-09/documents/bmds_3.2_user_guide.pdf (USEPA, 2020) for BMDS. The European Food Safety Authority (EFSA) also provides a platform for a web-based application of PROAST (https://shiny-efsa.openanalytics.eu/app/bmd); the EFSA platform is freely available but does require an account sign-in for access.

5.2.2.2 Key steps in DRM

DRM can be performed following the process summarized in Fig. 5.1. The following describes the steps of this process for evaluating a single data set.

(a) Suitability of data for modelling

The first step is to consider whether the data are suitable for modelling. These considerations include the following:

- Are there sufficient dose groups (e.g. there may be only one dose group plus a control group)?
- Is there a biologically or statistically significant trend in the end-point?
- Is there a clear dose–response relationship (e.g. there may be no adverse effect, even at the highest dose tested, or there may be a non-monotonic dose–response relationship)?
- Is the response at the first non-zero dose in the range of the BMR?

If the answer to any of the above questions is no, then the criterion for suitability of the data is not met. Further guidance on whether data sets are suitable for modelling is provided in section 5.3.1 and elsewhere (USEPA, 2012; Edler et al., 2014).

(b) Selection of a BMR

The second step requires the selection of a degree of change, known as the BMR, that defines a level of response that is measurable, considered adverse and relevant to humans or to the model species (i.e. extra risk). The breaking point between adverse and non-adverse is called the critical effect size (Dekkers, de Heer & Rennen, 2001), and this can also be used in DRM instead of a clearly adverse BMR (Slob, 1999, 2017).

For quantal data, the BMR needs to be a value within the observed range of experimental response and near the lower end of this range. For example, the first dose might exhibit a 0% increase in response, and the second dose might exhibit a 7% increase; the BMR should be close to the second dose. The chosen BMR should allow the BMD to be associated with a low level of exposure-related effects. When the BMD is based upon experimental animal data (e.g. tumour incidence), an extra risk level of 10% is often used as a default, on the basis of typical sample sizes in most experimental animal studies. However, there may be reasons to deviate from this default, and other BMR values may be used with a sound scientific rationale provided (Haber et al., 2018). Various studies have estimated that the median of the upper bounds of extra risk at the NOAEL is close to 10%, suggesting that the BMDL for a 10% response, or $BMDL_{10}$, may be an appropriate default for quantal data (Allen et al., 1994; Fowles, Alexeeff & Dodge, 1999; Sand, Portier & Krewski, 2011).

For continuous data, a biologically meaningful BMR depends on the type of end-point and therefore varies. Ideally, it is set numerically so that the BMR reflects the onset of a human-relevant adverse effect, meaning that a response above the BMR is considered adverse. A default value of 5% is sometimes used. A reanalysis of a large number of studies from the United States National Toxicology Program involving continuous data showed that the BMDL for a BMR of 5% (the $BMDL_{05}$) was, on average, close to the NOAEL derived from the same data set, whereas in most individual data sets, the $BMDL_{05}$ and NOAEL differed within one order of magnitude (Bokkers & Slob, 2007). Similar observations have been made in studies of fetal weight data (Kavlock et al., 1995).

A tiered approach should be followed in the order given below when setting the BMR for continuous data (see Fig. 5.1):

- ***Tier 1: Does a biologically relevant BMR already exist for the end-point considered?***
  If a biologically relevant BMR is already established based on prior consensus for the end-point considered, then it should be used. Sources for BMR values considered adverse include Dekkers, de Heer & Rennen (2001) and WHO (2015).

- ***Tier 2: Can an expert decision be made for an adversity-based BMR?***
  If a biologically relevant BMR does not already exist, then it should be considered whether an expert decision can be made (involving the collaboration of risk assessors, statisticians, toxicologists and clinicians) to determine the quantitative definition of adverse, considering the type and severity of the effect, the background variability and the MOA leading to the effect. The BMR may be defined using any of the methods that are available in the literature, taking biological relevance into account. The nature of the value determines the appropriate definition of the BMR (e.g. critical effect, relative deviation, standard deviation, point or hybrid approaches). If it is a per cent change from background, this methodology defines a BMR using the relative deviation definition (USEPA, 2020). In cases in which the adverse effect size is defined as a cut-point, this methodology defines a BMD using the point or hybrid definition of the BMR (USEPA, 2020). For example, JECFA used a β_2-microglobulin cut-point to determine a provisional tolerable monthly intake (PTMI) for cadmium (FAO/WHO, 2011b). Note that this tier assumes that a level of adversity can be identified, even though the minimal degree of adversity may not be known. Thus, BMRs may also be represented by a range rather than by a point; the corresponding BMD confidence intervals require case-by-case consideration when choosing uncertainty factors.

- ***Tier 3: Use the MOE or other approach outside of the BMD framework***
  The BMD framework is predicated on the assumption that the response poses some known adverse outcome; if no definite BMR can be established based on adversity, the assessment should be viewed as falling outside of the scope of the BMD framework. However, this does not prevent dose–response

assessment, which may include DRM. The toxicologist or the risk assessor can then choose quantitative values of response that may be considered from this assessment; these may include, but are not limited to, investigating fold changes, looking at differences from background and estimating response from a dose–response curve. Although this analysis may produce a range of plausible values (i.e. lead to MOE calculations), it is stressed that these values should not be used for establishing an HBGV.

In all cases, the rationale for the decision made on the BMR and associated uncertainties should be explained and documented.

(c) Model selection

The third step consists of selecting a set of models to fit to the data. Although biological considerations may motivate the choice of one or several empirical models (e.g. mathematical models from the general suite of models described below), the level of biological detail in such models is minimal. Thus, their credibility for interpolating and extrapolating a data set derives mainly from their fit to the data, as evaluated statistically.

When selecting a set of models, appropriate software should be used. For quantal (dichotomous) and continuous outcomes, standard suites of models are available in the PROAST/EFSA and USEPA BMDS software programs, and these models should be used as a default set (see Appendix 5.1 for more details). In cases where these models fail to describe the data, alternative models may be considered, but the justification for the use of these alternative models must be provided.

For other types of outcome data – for example, count, multivariate, ordinal and nested (e.g. litter) data – there is not a standard set of models. When applicable, models should be selected from the literature; in some cases, however, the model may not have been applied in a risk context. Regardless of how a set of models is developed, all model choices should be clearly documented, and a rationale should be given for the models included, as well as for potentially relevant models excluded. Some of the criteria to consider include, but are not limited to, the confidence interval width for parameters of interest, evaluation of residuals with regard to fit and dispersion, and visual fit.

When a model is used that is not part of the standard packages, copies of the software should be made available for download, and the source code should be provided and properly documented, for archival purposes; this ensures that the analysis can be reproduced even if the software changes in the future.

Further guidance on the models to be used, depending on the data types to be modelled, is provided below.

Dose–response models that do not imply an underlying biology. A general family of dose–response models, for quantal and continuous data, is specified as

$$\mu(\text{dose}) = a \cdot \{1 + (c - 1) \cdot F(\text{dose}^d, b)\}$$

where parameters a, b, c and d refer to the background response, potency of the substance, maximum response and steepness (i.e. influences the slope), respectively. Further, the function $F(\text{dose}^d, b)$ is an increasing function that goes from 0 to 1 over the dose range. For more information on this modelling strategy, the reader is referred to Aerts, Wheeler & Cortinas Abrahantes (2020).

Dose–response models for continuous data. The dose–response model for continuous data describes how the magnitude of response changes with dose and is typically defined as the central tendency of the observed data in relation to dose. The dose–response model may be linked to other summary statistics related to the data (e.g. response quantiles). For an example of these other approaches, the reader should refer to Wheeler, Shao & Bailer (2015), who use dose–response models to describe the quantiles of the data.

The models below describe the relationship between dose and the magnitude of a response. When linked to a statistical distribution (see below), these equations describe the relationship between dose and the central tendency in a population.

As an example, for continuous data, a common dose–response model is the exponential model; by setting $F(\text{dose}^d, b) = 1 - e^{-b \cdot \text{dose}^d}$, one arrives at

$$\mu(\text{dose})_F = a \cdot \{c - (c - 1) \cdot e^{-b \cdot \text{dose}^d}\}$$

which is described in Slob (2002). Slob & Setzer (2014) found that most continuous dose–response data were adequately described by

the exponential model or the Hill model. In what follows, we assume that only models contained within the general family of dose–response models that includes the exponential and Hill models are used. Use of models outside of this family should be well justified.

When modelling continuous data and considering the background response at zero dose, the appropriate approach is to account for the background in the model with a parameter that needs to be estimated from the data (IPCS, 2009). In some cases, dose–response data are adjusted by subtracting the (mean) control value from each individual observation before modelling and removing the background parameter from the dose–response model. However, this procedure does not account for the fact that the background response level is also not known with certainty, and it should not be used.

Dose–response models for quantal data. To model quantal data, one is interested in modelling the observed frequency of the response. As with the case for continuous data, DRM estimates the central tendency of these frequencies, which can be interpreted as the probability that the outcome will be observed in a population.

Like the continuous data case in the example above, the family of models can be used with

$$\mu(\text{dose}) = a + (1 - a) \cdot F(\text{dose}^d, b)$$

The maximum response for quantal end-points is 1; it is considered that if the dose is large enough, the response will reach the value 1. As for the case of continuous data end-points, the function $F(\text{dose}^d, b)$ is in principle a function going from 0 to 1 over the dose range. For example, the log-probit model is obtained by considering the function $F(\text{dose}^d, b) = \Phi(\ln(b) + \ln(\text{dose}^d))$. This implies

$$\mu(\text{dose})_F = a + (1 - a) \cdot \Phi(\ln(b) + d \cdot \ln(\text{dose}))$$

where $\Phi(\cdot)$ is the cumulative distribution function for the standard normal distribution.

Dose–response models for counts. Count data most frequently occur in observational epidemiological studies in which the number of adverse events in each group is the outcome of interest (Breslow & Day, 1980). For example, one may count the number of cancers observed in a given population (often expressed as a rate per 100 000) that is grouped by exposure. When modelling count dose–response

data, the statistical expectation of the count is modelled relative to some unit (e.g. time). For example, in Poisson models, a dose–response model can be used to describe the rate of cancer from exposure to aflatoxin per person-year. This rate, which is defined in all count models, is termed *R*; as *R* varies over exposed dose, *R* is defined as a dose–response model in this context (Frome, 1983). For example, at its eighty-third meeting, JECFA estimated the rate, *R*, of liver cancer in populations exposed to aflatoxin (FAO/WHO, 2017a).

For count data, there is an increased need for collaboration between the subject matter experts and the modeller. Typically, models come in two forms: additive models and relative rate models. Additive models assume covariates, and the dose–response function enters the model as a linear combination; that is,

$$R_a\,(\text{dose}, X) = f_a(\text{dose}) + g_a(X)$$

where $f_a(\text{dose})$ and $g_a(X)$ are the functions of dose and covariates X, respectively. Relative rate models assume that dose and covariates affect the rate multiplicatively; that is,

$$R_m\,(\text{dose}, X) = f_m(\text{dose}) \times g_m(X)$$

where $f_m(\text{dose})$ and $g_m(X)$ are the functions of dose and covariates X, respectively.

In many cases, models used for counts are not the same as those preferred in the continuous and quantal data setting. For count modelling, as generic software does not exist, the model forms should be supported by the literature or scientific understanding. For example, at the eighty-third meeting of JECFA, for aflatoxin, all evaluated models were based upon the peer-reviewed literature (FAO/WHO, 2017a).

Dose–response models for ordinal categorical measures. In some circumstances, the response is grouped into distinct outcomes. For example, a pathologist may give a non-neoplastic lesion a grade from 1 to 5, based on an assessment of severity. Modelling such data using dose–response models is complex; consequently, the dose–response models used are often specific to the given circumstances. However, some general considerations should be taken into account. For a full discussion of these considerations, the reader should refer to Agresti (2013). Many different dose–response models are available and can be represented by an extension of the logit, complementary

log-log or probit link functions (USEPA, 2020; PROAST: https://proastweb.rivm.nl); other dose–response models may be considered. Chen & Chen (2014) describe DRM for ordinal categorical responses.

Dose–response models for nested data. For nested data, complex assumptions are necessary, and individual animal data are needed to allow for the application of modified models that handle the cluster effect. For example, in developmental toxicity studies, because the dam is the unit of treatment, one expects greater similarity in the response within a litter than across litters. One aspect of the cluster effect is the intra-cluster correlation, which varies among dose groups and conceptually represents the degree of correlation in effect due to treatment condition or experimental variability or error. This value must be accounted for to accurately represent the dose–response model.

Dose–response models for multivariate responses. As multivariate responses are often situation specific, modelling should be done on a case-by-case basis, referring to models available and accepted in the literature. For example, for clustered developmental toxicity studies with multiple end-points, Geys, Molenberghs & Ryan (1999) and Molenberghs & Ryan (1999) developed models that consider multiple malformation end-points, and Budtz-Jørgensen (2006) and Mbah et al. (2014) developed general latent structural equation dose–response models to establish BMDs for families of chemicals and for sets of end-points. Another example of this data type arises when utilizing toxicogenomic data. Although such data have not been used for risk assessment purposes to date, such developments can be expected in future. As further developments with dose–response models for multivariate responses are ongoing, it is anticipated that these models will be used increasingly.

(d) Model assumptions, model fitting and estimation of parameters

Model fitting makes statistical assumptions about the data and the selection of a dose–response model. Before fitting a model to the data, a distribution for the response should be assumed. Appropriate distributional assumptions accounting for the data type should be used. For example, continuous response data are usually assumed to be normally or log-normally distributed; if the data exhibit right skew, a log-normal assumption may be more appropriate than a normal assumption. For all data types, assumptions, such as distributional assumptions, should be reviewed prior to the analysis,

and the rationale for the choices made should be clearly communicated. Given those assumptions, modelling finds parameter values (see section 5.2.2.2(c)) for the dose–response model that are optimal. In those situations in which one distribution is not preferred, model averaging over all distributions (described below in section 5.2.2.2(e)) may be an alternative.

Two methodologies available for model fitting are the frequentist and the Bayesian approaches (see Appendix 5.1 for further details). Historically, software designed for DRM used frequentist methodologies; however, recent advances in numerical mathematics and software development (Shao & Shapiro, 2018; USEPA, 2020) allow the use of Bayesian methods for DRM. Regardless of the type of analysis, both methodologies require an understanding of some basic principles of modelling before outputs from the analysis can be properly interpreted. Although a full description of the methodologies is beyond the scope of this chapter, some basic remarks may be helpful here.

Given a specific dose–response model, the general objective of model fitting is to best describe the dose–response data. Therefore, one would search those parameter values (e.g. *a*, *b*, *c* and *d* described in section 5.2.2.2(c)) that lead to a function or curve that describes the data well, using some statistical criterion that defines a good fit. Parameter estimation can proceed using either Bayesian or frequentist methods. For Bayesian analyses, the prior[1] should be reasonably diffuse over values of the target parameter considered relevant to the analysis (described in more detail in Appendix 5.1). To mitigate against possible biases, a sensitivity analysis of the effect of the priors should be clearly documented. When frequentist methods are used, care should be taken that the parameters are estimable given the data. In many cases, the data provide little information on the parameter. For example, for continuous data, the Hill model is sigmoidal, and some data may not suggest a sigmoidal shape; in these cases, the data give no information on the steepness parameter *d*. When the data do not inform the value of the parameter, parameter bounds add information and should be considered in order to mitigate the possibility of biologically implausible responses.

[1] In Bayesian analysis, a prior is information included in the analysis that explicitly encapsulates information outside of the experiment. Prior information may include information that prevents models from unrealistic sharp changes in the DRM.

Parameter constraints are not necessary when using model averaging or Bayesian methods in general. However, fitting should always be with regard to the biological plausibility. Here, biological plausibility is defined in relation to the end-point of interest – that is, the BMD and the dose–response curve estimate; when using model averaging (see section 5.2.2.2(e)), estimates are generally more stable and thus plausible. When the fit or its corresponding bounds result in biologically implausible estimates of the target parameter (e.g. the POD being a very small number of molecules), alternative approaches should be investigated, or data should be considered too poor in quality to allow the derivation of a POD.

For either approach (frequentist or Bayesian), the optimal parameter estimates can be thought of as the "best guess" of the model's parameters for the observed data (see Appendix 5.1 for more information). Using those estimates, a "best-fitting" dose–response curve can be calculated. Although the estimate is optimal in relation to the observed data, it does not reflect the statistical uncertainty in the data – that is, the scatter typically observed in experimental studies. To reflect this, confidence intervals are computed on the model parameters, including the BMD. These intervals can be used as a bound to plausible dose–response curves that are consistent with the observed data – that is, lower and upper bounds on the dose–response model. The upper-bound curve allows the visualization of the BMDL for a given BMR. This reflects the statistical uncertainty (sampling error) associated with the data. In most cases, both frequentist and Bayesian methods are fit for purpose. Although Bayesian methods are generally preferred in this guidance, frequentist methods can be used if deemed appropriate. Discussion of the strengths and weaknesses of these approaches is provided in Appendix 5.1.

(e) Model uncertainty and model averaging

Historically, one or more dose–response models were fitted to the data, but a single model was chosen, based upon some criterion (e.g. lowest BMDL), to estimate the POD. Although sampling error is accounted for in individual model fits and is quantified by confidence bounds, additional uncertainty arises from the fact that estimates from dose–response models may differ. That is, many of these models may reasonably reflect the observed data but produce different BMD estimates. As no model is assumed to be the true model, this variance over model estimates is known as model

uncertainty. Model averaging is an appropriate method to address model uncertainty in DRM.

Model averaging allows estimation of the dose–response relationship and derived statistics, such as the BMD and derived confidence interval, using all model fits through a weighted average (see Kang, Kodell & Chen, 2000; Wheeler & Bailer, 2007; Shao & Gift, 2014; and references therein). Weights are computed using Bayesian or frequentist methods based upon a criterion of fit (e.g. Akaike information criterion [AIC] for frequentist approaches and posterior[1] model probabilities for Bayesian methods).

(f) Model parameter constraints

In certain situations – for example, when it is deemed biologically appropriate – parameters *a*, *b*, *c* and *d* in the dose–response model family (described in section 5.2.2.2(c)) may be constrained. The effects of constraints on parameters *a*, *b* and *c* are often minimal for BMD modelling. For example, *b* is often constrained to be greater than zero to reflect the assumed positive response in the dose–response curve, which is a prerequisite for modelling most data; this is an example of a constraint that is necessary to make when modelling data. However, in other situations, the use of constraints is subject to rigorous debate and may not be necessary in practice. For example, a constraint on the steepness parameter *d*, a key parameter for describing the effect of dose on the outcome, can have a substantial impact on estimates. The parameter *d* can be unconstrained, which may reduce bias in BMD estimation, but this may lead to unreasonable BMDL estimates (e.g. equal to zero). This guidance supports an approach that should be used when there is an absence of other biological knowledge.

For a parameter such as *d*, the following strategy is recommended to minimize the impact of this decision on constraining the dose–response model's parameters for estimation of the BMD and the respective confidence interval:

[1] The posterior distribution is a probability distribution that describes the uncertainty of the dose–response model given the data.

- Bayesian model averaging, which penalizes estimates of *d* that approach the constraint. For example, BMDS does such a penalization.

- Bayesian fitting of a single model with a soft constraint on *d* (i.e. prior over the parameter that places low probability on regions normally constrained). Note that in this case, the steepness parameter is not constrained.

As noted previously, model averaging is recommended. It has an additional advantage specific to this issue, because unconstrained models are used. It often mitigates the need to choose between options, because it gives near zero weight to the problematic curves with infinite slope. In the few cases where model averaging results are deemed inappropriate, a Bayesian unconstrained model with soft constraints can be considered.

(g) Evaluation of DRM

There is a variety of methodologies applicable to determining whether the dose–response model adequately describes the data and whether there is an adequate fit of the model to the data. Although an individual technique cannot be recommended for every case, it is important that known model fit criteria be used and documented (see Appendix 5.1). For example, commonly applied criteria for model fits for model averaging include examination of visual fit, bootstrap goodness-of-fit statistics (for frequentist model averaging) or Bayes factors comparing the model-averaged posterior probability to the null mean model with diffuse beta priors. For individual models, one can compare models using the AIC or Bayesian information criterion (BIC) and evaluate them using analysis of deviance and Pearson χ^2 goodness-of-fit tests. The goodness of fit can also be evaluated by comparing a meaningful difference of the AIC of the fitted models with suitable reference models, such as the null model, to ascertain the existence of a dose–response trend, and the full model, which just consists of the data points to ascertain the general goodness of fit. Care should be taken to investigate the fit in the low-dose region, as this region is more relevant to estimating the BMD.

(h) Reporting of results of BMD modelling

In reporting the summary results of BMD modelling in JECFA and JMPR reports, the BMRs and software packages used should be stated, and the effects selected for modelling and the ranges of

BMDLs and BMDUs estimated by the different acceptable fits should be given. When using model averaging or Bayesian approaches, the individual model weights and Bayes factors should also be reported.

In the detailed JECFA and JMPR monographs, the following should be reported:

- a summary table of the data for the end-points considered and the BMD analysis. For quantal end-points, both the number of responding animals and the total number of animals should be given for each dose. For continuous end-points, the mean or individual responses, associated standard deviations or standard errors, and sample sizes should be given at each dose;
- the value of the BMR chosen, and the rationale for the choice;
- the software used, including the version number;
- all assumptions in the model-fitting procedure. If the assumptions deviate from the recommended guidance, the rationale for deviating should be provided, as well as the results using the recommended defaults;
- a table presenting the models used in the model averaging, including each model's weight and individual BMDL and BMDU. Values should be reported with two significant figures. If model averaging was not used, a table with the individual models used (including the null model and the full model) should be provided; for Bayesian analyses, Bayes factors, and for frequentist analyses, *P*-values and AICs, or the values of the log-likelihoods of the models and the values of the AICs, should be reported. If applicable, supplementary information (e.g. software outputs) should be given;
- a plot of the fitted average model or, if model averaging was not used, a plot of all the models fitted to the data for the critical end-points. In the case of nested families – that is, models whose parameters are supersets of other simpler models – a plot of the selected model for each family should be provided; and
- a conclusion regarding the selected BMDL (or BMD confidence interval) to be used as a POD.

5.2.3 Modelling observational data from epidemiological studies

Observational studies in humans refer to those studies in which treatment or exposure is not under the control of the investigator. Experimental studies refer to those studies in which assignment to

treatment or exposure is under the control of the investigator. This section focuses on observational epidemiological studies.

Observational data derived from epidemiological studies may provide quantitative information on dose–response relationships used in risk assessment. Two advantages of using human epidemiological data are that no extrapolation between species is required and that exposure scenarios represent the real world.

Major designs in observational epidemiological studies include the cohort design and the case–control design. In the cohort design, exposure is assessed over time, and the study population is followed over time to ascertain information on outcomes possibly depending on the exposure. In the case–control design, two groups differ in an outcome of interest but are otherwise comparable. These two groups are compared with respect to a supposed cause of the outcome (e.g. an exposure to a specific substance). Exposure is either assessed contemporaneously or reconstructed for some earlier period. For the purposes of this chapter, when using such data, it is assumed that a judgement on the causal association and suitability for analysis has been made; however, it should be noted that the association observed may not always be causal.

5.2.3.1 Study design

The focus of the discussion on observational epidemiological study design is on three key conditions: exchangeability, positivity and consistency.

(a) Exchangeability

Exchangeability of treatment groups under comparison with respect to an outcome measure means that the groups' outcomes would be the same whenever they were subjected to the identical exposure history (Hernan & Robins, 2006). In experimental epidemiological study designs (e.g. intervention studies with vitamins or other dietary supplements), comparability of exposed and unexposed subjects is obtained through random assignment to exposure or treatment groups. Randomization leads (in expectation) to balance between exposure groups in other factors that affect risk of disease. However, in observational epidemiological studies, people are not randomly assigned to an exposure group. Without the benefit of randomization, exposure groups may differ with respect to factors other than the agent of primary interest. If these factors are also

related to the disease of interest, then the observed effect of the agent on disease risk may be mixed with the effects of these other disease risk factors. This is referred to as confounding. Failure to account for confounding can lead to bias in an estimate of association between exposure and disease. Confounding may be controlled through restriction, matching, standardization or conditioning in a regression model. Exchangeability also implies no selection bias, which may occur if other factors that affect risk of the disease of interest are correlated with exposure status in the study sample owing to the way in which the sample was collected. If selection into a study (e.g. through self-selection of participants) is related to exposure and the disease of interest, then an association may be induced in the study sample where it did not exist in the underlying total population. In practice, it is useful to explore exchangeability by assessing the sensitivity of the results to a variety of potential confounders and functional forms; we note that here, unlike in experimental settings, as discussed in section 5.2.2.2(e), model averaging is unable to solve the problem of omitting confounders.

(b) Positivity

Positivity, also termed the experimental treatment assignment assumption, means that there is a non-zero probability of exposure or treatment at each level under comparison across levels of the covariates in the population under study (Hernan, 2012). Experimental studies involve settings where exposure is under the control of the investigator. If the investigator administers doses of an agent in a controlled experimental setting, then the investigator can ensure non-zero probability of exposure by design. In observational studies, this is not ensured. It is possible that exposure or treatment assignment is fully determined by some of the confounders. For example, in a clinical setting, patients who receive treatment may differ from those who do not; and treatment assignment may be fully determined by some of the confounders. Yet epidemiologists need to have variation in exposure across covariates if they wish to estimate the effect of the exposure. In observational studies, there may not be positivity across confounders, and there may be a need to make some interpolation or smoothing over regions of non-positivity.

(c) Consistency

Consistency refers to an unambiguous description of how contrasting levels of exposure would be assigned (Cole & Frangakis,

2009). This is needed to clearly specify the causal contrast of interest. A well-defined exposure is a necessary condition for causal inference. In a controlled experimental setting in which the investigator administers doses of an agent, there is, for example, less measurement error than in observational studies and a clear definition of the exposure or dose administration regime. In observational studies, however, there are often obstacles to assessment and estimation of exposure that may relate to uncertainty of times of exposure, measurement and quantification of magnitudes of exposure. Thus, in observational studies, exposure should be defined as unambiguously as possible to ensure consistency of the estimated effect.

The determination that a given study or studies are suitable for analysis of dose–response relationships that may be used for setting HBGVs involves expert judgement regarding these conditions.

5.2.3.2 *Analyses*

Although most of the methodology described in this chapter can in principle be applied to estimating the BMD and corresponding confidence interval using observational data from epidemiological studies, there are important methodological considerations that may require adaptations. These relate to the fact that DRM has, to a large extent, been designed around the use of data from controlled laboratory animal experiments. The types of information available from observational epidemiological studies often differ from the types of information derived from experimental animal studies.

For BMD modelling, the lack of a controlled experimental setting is the most important difference between observational studies and controlled experimental studies. The observational setting means that adjustment for several covariates is often needed when doing a BMD analysis. PROAST/EFSA BMD software does allow analysis of covariates. If the currently existing BMD software is not designed to deal with such multivariable modelling requirements, this is not problematic, as many existing statistical packages (STATA, SAS and R, to name a few) can be used for such purposes. Even if existing BMD software were to be updated to allow for the handling of several covariates, another issue is that access to individual participant data from human studies is severely restricted by data protection requirements – that is, sending or sharing individual participant data containing sensitive health and sociodemographic information is often not compatible with data protection regulations.

Although these are important concerns, the problem of data sharing can be overcome by modelling aggregated (or quantile) data. For that purpose, confounder-adjusted summary statistics that reflect the underlying dose–response curve must be generated. This can be done by dividing the exposure variable into enough quantiles (quartiles, quintiles, deciles or finer subdivision) and then, using multivariable analyses, generating the expected confounder-adjusted response in each quantile using the lowest quantile of exposure as the point of comparison (Wheeler et al., 2017). Such an approach is compatible with how epidemiological data are frequently analysed and reported. The loss of information when using aggregated quantile data is generally considered non-substantial if the numbers of quantiles generated are sufficiently large to allow for proper evaluation of the underlying dose–response curve. The only specifications needed for such an approach are that, for each quantile, the authors report the response (e.g. mean response, proportional hazard or excess risk), its standard error, the number of subjects and the median exposure. For quantal outcomes where relative risk estimates are used, it is also important for authors to provide information that allows for the extraction of the absolute risk in each quantile. This approach is essentially comparable to how controlled animal experiments are analysed, where the use of summary statistics, not individual data, is accepted.

Another difference between modelling data from human observational studies and modelling data from experimental studies is the lack of a well-defined control group, or "zero dose". For human observational studies, the equivalent of a zero dose would be the lowest quantile that is used as the point of comparison. The exposure level and the background response for that point depend on the number of quantiles generated to describe the dose–response curve. The exposure level for that point may also differ across study populations, which highlights the need to model more than one study, if possible, to establish whether consistent results can be obtained. Extrapolation beyond the observed data should generally not be done without clear justification.

In human observational studies, the exposure range is often narrower than that which can be created in experimental settings (i.e. laboratory animal studies often use much higher doses than those to which humans would normally be exposed); in contrast, the sample size is usually much larger. A narrow exposure range has the implication that the full theoretical (e.g. sigmoidal) dose–response

curve is often not observed. Instead, the dose–response curve depends on the level and range of exposure in the observed population, and it can be either nonlinear (at the two extremes of the sigmoidal curve) or approximately linear. The use of linear models can therefore in some cases be justified at the expense of using more complex nonlinear models. In such cases, the benefit of using BMD analyses is confined to determining the BMD and corresponding confidence intervals based on a predefined response that is considered biologically relevant (the BMR). Furthermore, the high variability in human observational studies, relative to the controlled settings in an experimental animal study, means that the same default BMR frequently applied in laboratory animal studies is not necessarily applicable. The BMR used in human settings should be based on what is considered normal or abnormal from a clinical point of view or acceptable or unacceptable from a public health point of view.

5.3 Determining the POD: NOAEL/LOAEL or BMDL

A thorough review of the entire hazard data set is a prerequisite for determining an appropriate POD. The NOAEL and BMD approaches allow estimation of the response of an organism given an exposure, but do not implicitly define a methodology by which a POD can be determined. For that, additional considerations are necessary. As the POD is used to define a dose associated with an estimable risk, this risk level must be defined a priori to estimate the POD. What level of risk may be estimable is determined by the quality of the data set.

5.3.1 Data selection

Regardless of which of the two approaches is used, NOAEL or BMD, data selection is central to both. In section 5.2, the types of dose–response data one may encounter in dose–response assessment were discussed without consideration of the appropriateness of a given data set for use in risk assessment. In this section, considerations on selecting appropriate experimental animal data, experimental human data and human observational data are given. In selecting data for use in risk assessment, due consideration needs to be given to matching, as far as is possible, the pattern of potential human exposure – that is, the route and duration of exposure (as a fraction of a lifetime) and the pattern of exposure (e.g. intermittent bolus dosing or dietary administration). Further, converting administered laboratory animal doses to estimated human equivalent

doses by allometric scaling prior to modelling may also be appropriate.

When considering which data to use from a set of available toxicity studies on a substance, it is not necessary to undertake DRM for each observed end-point in each study. Whether the NOAEL or BMD approach is used for risk assessment, the aim is to determine the lowest levels of exposure producing an adverse effect that is relevant to humans. Therefore, a first step would be to exclude studies in which the onset of adverse effects occurs at much higher doses than those from the other studies (assuming these other studies are of good quality) on end-points that are not markedly different in severity. In certain cases, even though an end-point is appreciably less sensitive than the most sensitive one, it might still be desirable to determine its BMDL/NOAEL (e.g. if irreversible neurotoxicity occurred at a higher dose than for a more sensitive end-point, such as an effect on acetylcholinesterase), to enable its risk characterization. End-points clearly not showing a dose–response relationship on visual inspection of the data can also be omitted. If the existence of a dose–response relationship is unclear, the EFSA platform (https://shiny-efsa.openanalytics.eu/app/bmd) allows DRM for multiple parameters in one run. Although the run is time-consuming, it provides a better scientific basis to identify those end-points for which a dose–response relationship is present and those for which it is not. Then, based on the toxicological impact together with the apparent magnitude of the response, a selection of end-points as candidates for DRM can be made. After selecting the potentially relevant end-points, the suitability of each dose–response data set for dose–response analysis is considered.

For DRM, it is recommended to have at least three or four different doses (including controls) and different levels of effect associated with the different doses. Even when these data requirements are not met, the BMD approach retains the advantages outlined in section 5.2. The BMD approach can also be used for combined analysis of two or more similar studies – for example, two studies of a chemical with the same design from the same laboratory that tested different portions of the dose–response curve (Allen et al., 1996). However, if the BMD approach is not deemed appropriate, the NOAEL approach (see section 5.3.2) can sometimes be used for the combined analysis of studies that meet strict criteria, in order to derive an overall NOAEL (see, for example, FAO/WHO, 2004).

In some experimental animal studies, it may be difficult to derive a POD when the number of animals per dose group is very small – for example, when the critical effect is seen in an experimental animal such as the dog, with only four animals of each sex per dose group. In such cases, the uncertainty around the NOAEL is likely to be high because of the insensitivity of the test. The BMD approach is better for evaluating sparse dose–response data, and a high level of uncertainty would be clear from a very wide BMDL–BMDU interval, indicating that the study data do not provide sufficient information to derive a reliable BMD. In such cases, if the data are used to derive a POD, the high level of uncertainty should be made clear in the narrative.

It is also important to recognize that there may be some useful data sets for which BMD modelling may not apply. For example, it may not be possible to fit an acceptable model to the data. In other cases, insufficient data may be available for modelling, but the data are reported in such a manner that a NOAEL can be identified, albeit with some uncertainty. In such cases, the risk assessor may need to weigh the utility of the study against the associated uncertainties.

Data on effects on humans following exposure to a substance can be extremely valuable in setting the laboratory animal data into context and, when available, should always be evaluated, even if they are not used to establish an HBGV. Not only may a human study sometimes allow identification of end-points (PODs) for use in risk assessment, but other important information may be gained, such as the nature of the adverse effect, its pattern of onset and duration, and individual variability in sensitivity. Even if the human data are insufficient to be used quantitatively, they may identify important data gaps not addressed by the laboratory animal data.

Human data may be available from several sources, including epidemiological studies of effects in human populations exposed to the chemical, direct administration to volunteers (e.g. of a food additive), monitoring of those exposed following normal use of the chemical, exposures from accidental or deliberate poisonings, and exposures from use of the same substances as human pharmaceuticals. Such studies often involve single or short-term exposures that can be particularly relevant, directly or indirectly, for establishing ARfDs.

5.3.2 The NOAEL approach for deriving a POD

The reliability of the NOAEL approach is dependent on the sensitivity of the test method. The statistical linkage determines whether there is a statistically significant effect (e.g. at the 5% level) compared with background (e.g. the control group) for each dose separately. Also of importance to establishing statistical significance is the application of the appropriate statistical test. This entails consideration of the data type (e.g. continuous versus quantal) and the distribution characteristics of the data (OECD, 2012; Hamada, 2018).

Establishing statistical significance can be in the form of trend analysis or in the more typical manner of a pairwise analysis between treated and control groups. The no-statistical-significance-of-trend test (Tukey, Ciminera & Heyse, 1985) is often used for trend analysis for continuous data, whereas the Cochrane-Armitage test is commonly used for quantal data (Hothorn, 2016). When the response is not statistically significant, it may be considered that this level of exposure is without biologically significant adverse health effects, although whether the power of the study to detect an adverse effect might be too low should also be considered. Therefore, decision-making solely based on statistical hypothesis testing should take account of the sample sizes, the statistical error probabilities and the problem of multiple comparisons. Given the typical laboratory animal studies used in toxicology, the minimum effect size that can be detected by a statistical test may be larger than 10% (extra risk), or even much higher, particularly in the presence of high background response or when sample sizes are of the order of 10 per group or smaller. Thus, in general, a NOAEL should be regarded as a dose at which the effect is somewhere between 0% and an upper bound of 10% or more. The upper bound for the effect size at the NOAEL can also be calculated (EFSA, 2017). The selection of the NOAEL identifies the highest dose that does not produce a statistically significant effect compared with the control. The NOAEL approach tends to give lower HBGVs for studies that have a higher power to detect adverse effects, which in effect "penalizes" better-designed studies. This emphasizes the importance of adherence to testing guidelines in order to ensure that the data are suitable for risk assessment purposes.

If a NOAEL is not apparent from the data (e.g. in cases where the lowest dose tested produces a small but measurable response), then the LOAEL may be used. Derivation of an HBGV from a

LOAEL requires the application of an additional uncertainty factor (see section 5.4.2).

The NOAEL approach is limited by the following characteristics of the study design:

- *Group size*: The power to detect a NOAEL at some dose is directly dependent on the sample size chosen at that dose. The larger the group size, the more power there is to detect a small effect.

- *Dose selection*: The NOAEL must be one of the doses in the study. If the true NOAEL is higher than the NOAEL indicated in the experimental study, the distance between the two can be expected to be limited (related to the dose spacing used). In contrast, if the true NOAEL is lower than the NOAEL indicated in the study, the distance between the two is dependent on the magnitude of the first test dose.

- *Experimental variation*: Experimental variation comprises biological (e.g. genetic) and other uncontrollable variation between subjects, experimental conditions (e.g. time of feeding, location in room in which the animals are housed, time of sacrifice or interim measurements) and measurements. Larger experimental variation will result in decreased statistical power and a higher NOAEL derived from the study, regardless of the location of the true NOAEL.

It is for the above reasons – together with the fact that the NOAEL approach does not take account of all the available data, allow for variability in the estimate of the dose–response relationship, take account of the slope of the dose–response curve or allow quantitative characterization of the uncertainties – that the BMD approach is preferred over the NOAEL approach.

5.3.3 The BMD approach for deriving a POD

The detailed steps involved in DRM to develop a POD have been set out in section 5.2.2.2. The use of model averaging without constraining parameters is recommended for calculating the BMD and its confidence interval (BMDL–BMDU). For laboratory animal data sets, the BMDL is generally used as the POD. Using the lower bound accounts for the experimental variability inherent in a given

study and ensures (with 95% confidence for the experimental context) that the selected BMR is not exceeded at the BMDL.

In some cases, the entire BMD confidence interval (BMDL–BMDU) can serve as the POD (IPCS, 2018). For human data sets, the BMD may be used as a POD instead of the BMDL in some cases, but this may underestimate the risk. When the POD for risk assessment of a chemical could be based on more than one data set, the DRM process needs to be reiterated for each relevant data set, the results need to be summarized and the most appropriate POD needs to be identified.

5.4 Establishing HBGVs

5.4.1 Introduction

HBGVs provide quantitative information from risk assessment, which, together with information on exposure, enables risk managers to make decisions concerning the protection of human health. HBGVs developed by JECFA and JMPR for substances found in food are the quantitative expression of the range of oral exposure (either acute or chronic) that would be expected to be without appreciable health risk. In general, HBGVs would not be established for substances that produce effects for which there is concern that there may be no biological threshold – for example, directly DNA-reactive substances that are mutagenic and carcinogenic. In such cases, an MOE approach could be used (see section 5.5).

It should be noted that in this chapter, the terms "margin of safety" and "margin of exposure" are both used. They are not synonymous. A margin of safety is defined as the margin between an HBGV and the actual or estimated exposure dose or concentration. An MOE is defined as the ratio of the NOAEL or BMDL for the critical effect to the theoretical, predicted or estimated exposure dose or concentration.

For substances intentionally added to food, such as food additives, and for residues of pesticides and veterinary drugs in food, the HBGVs are the ADI for chronic (lifetime) exposure and the ARfD for acute exposure. A range of exposure scenarios can be developed to compare with an ADI or ARfD, so that risk managers can assess whether the general population and more highly exposed subgroups are protected. JECFA and JMPR establish ADIs and ARfDs based on all the known facts at the time of their evaluations.

Data packages should include metabolism and excretion studies designed to provide information on the potential bioaccumulative properties of the substances under consideration. Substances that have long half-lives and accumulate in the body are not suitable for use as food additives (FAO/WHO, 1962) and may require additional considerations when establishing HBGVs if they occur in food (e.g. contaminants).

At the time of its first meeting, JECFA recognized that the amount of an additive used in food should be established with due attention to "an adequate margin of safety to reduce to a minimum any hazard to health in all groups of consumers" (FAO/WHO, 1957). The second JECFA meeting (FAO/WHO, 1958), in outlining procedures for the testing of intentional food additives to establish their safety for use, concluded that the results of laboratory animal studies could be extrapolated to humans and that:

> some margin of safety is desirable to allow for any species difference in susceptibility, the numerical differences between the test animals and the human population exposed to the hazard, the greater variety of complicating disease processes in the human population, the difficulty of estimating the human intake, and the possibility of synergistic action among food additives.

These conclusions formed the basis for establishing the ADI, not only for food additives, but also for residues of pesticides and of veterinary drugs.

The ADI is defined as an estimate of the amount of a food additive or residue, expressed on a body weight basis, that can be ingested daily over a lifetime without appreciable health risk. It is expressed in amount (e.g. milligrams) per kilogram of body weight and as a numerical range from 0 to an upper limit, which is considered to be the range of acceptability of the substance. This is done in order to encourage the lowest level of use that is technologically feasible. ADIs are normally expressed numerically using only one significant figure. The use of more than one significant figure might be taken to imply a greater degree of precision than that which can be achieved when assessing the hazard from the wide range of factors that influence toxicity.

When appropriate, JMPR and JECFA establish ARfDs (see section 5.4.6 for more detail). The ARfD is defined as (FAO/WHO, 2002b):

> The estimate of the amount of a substance in food or drinking-water, normally expressed on a body-weight basis, that can be ingested in a period of 24 h or less, without appreciable health risk to the consumer.

The ARfD is expressed as a single value, in milligrams of the chemical per kilogram of body weight. ARfDs are normally expressed numerically using only one significant figure, for the same reasons as for the ADI.

For food contaminants that are generally unavoidable, JECFA has used the term "tolerable" for HBGVs. "Tolerable" was considered more appropriate than "acceptable", as it signifies permissibility for the dietary exposure to contaminants associated with the consumption of otherwise wholesome and nutritious food. The terms used have included TDI, provisional maximum tolerable daily intake (PMTDI), provisional tolerable weekly intake (PTWI) and PTMI, the unit of time being that most appropriate to the half-life of the substance. Historically, JECFA has used the term "provisional", as there is often a paucity of reliable data on the consequences of human exposure at low levels, and new data may result in a change to the tolerable level. However, as any HBGV would be revisited if new data indicated the need for a change, and as the word maximum is redundant, it is recommended that the terms "provisional" and "maximum" no longer be used – that is, using only the terms TDI, tolerable weekly intake (TWI) and tolerable monthly intake (TMI), as appropriate. Tolerable intake values are expressed as an amount (often in micrograms) per kilogram of body weight, as a single value and not a range, and normally using only one significant figure.

JECFA and JMPR establish HBGVs based on the most appropriate BMDL or NOAEL/LOAEL. In general, this will be the lowest relevant BMDL or NOAEL/LOAEL in the most sensitive species. The entire database is reviewed to identify those effects relevant to the exposure duration of concern – that is, a single exposure, exposure over a lifetime or repeated exposure that may be less than lifetime (see section 5.4.6 for further information on effects associated with acute exposure). Only those effects that are adverse or clear indicators of adversity (e.g. a clinical chemistry measurement) are used to establish HBGVs (see section 5.2.1.2). If an effect is clearly not relevant to humans, based on the weight of the evidence, as described in the International Programme on Chemical Safety (IPCS) guidance on the MOA human relevance framework (IPCS, 2007; Meek et al., 2014), it is not used to establish an HBGV. When

relevance cannot be excluded, it is assumed. When there are unique species- or strain-specific differences in biology, it may be concluded that the BMDL/NOAEL in a particular species or strain should not be used to establish an HBGV (e.g. developmental toxicity of emamectin in CF-1 mice, which is a consequence of a strain-specific deficiency of the adenosine triphosphate–binding cassette transporter p-glycoprotein; WHO, 2012). The relevant effect with the lowest BMDL/NOAEL for the most relevant sex, species and strain is considered the "critical effect", and the BMDL or NOAEL (or, in exceptional cases, the LOAEL) for the critical effect is usually used as the POD to establish an HBGV.

Calculation of the HBGV is as follows:

$$\text{HBGV} = \frac{\text{POD}}{\text{UF}}$$

where UF is the uncertainty factor and POD is a BMDL, NOAEL or LOAEL.

Occasionally, an adverse effect with a higher BMDL/NOAEL than the effect with the lowest BMDL/NOAEL may need to be considered if, for example, it is a severe effect that may require use of a higher uncertainty factor.

5.4.2 Uncertainty factors

In the past, the term "safety factor" rather than "uncertainty factor" was often used, particularly in the establishment of ADIs. The term "uncertainty factor" is now preferred. Comparable terms used by other bodies are "adjustment factor" and "assessment factor", although some reserve the use of "adjustment factor" for data-derived factors. Application of the factors is intended to provide an adequate margin of safety for the consumer, considering sensitive human population subgroups, including infants (above 12 weeks of age) and young children.

Extrapolation is a necessary part of all risk assessments, except in rare cases where it is possible to determine a POD from adequate human data representative of the potentially exposed population (including sensitive subgroups) that has had a level of exposure similar to that which is of concern. Otherwise, the POD needs to be extrapolated, where possible, to a human-equivalent exposure, ideally using quantitative species- and chemical-specific information. Where

possible, chemical-specific data on interspecies or intraspecies (interindividual) differences in toxicokinetics and toxicodynamics should be used to derive CSAFs to use in the overall uncertainty factor. In practice, suitable data for this purpose should be identified; however, if, as is often the case, such data are not available, default factors can be used to establish an HBGV (e.g. ADI, TDI, ARfD).

In the absence of quantitative species-specific toxicokinetic or toxicodynamic information on a particular substance, an overall default uncertainty factor of 100 is used to convert a POD from a study in experimental animals into an HBGV for human exposure (IPCS, 1999). The default 100-fold uncertainty factor represents the product of two separate 10-fold factors: a factor of 10 to cover interspecies differences between the experimental animal species and humans and a factor of 10 to cover intraspecies (interindividual) differences between humans. Each of the default factors of 10 takes into account differences in toxicodynamics (the effect of the chemical on the body) and in toxicokinetics (the fate of the chemical in the body). The interspecies uncertainty factor for toxicodynamics is designed to cover differences between the average responses in the experimental animals used in the study identified to derive the POD and those in average humans. The intraspecies (interindividual) uncertainty factor for toxicodynamics is designed to cover the differences in responses in humans between those who are least sensitive to the substance and those who are the most sensitive. Similarly, the interspecies uncertainty factor for toxicokinetics is designed to cover the variability in kinetics between that observed in the experimental animal species and that in average humans, and the intraspecies (interindividual) uncertainty factor for toxicokinetics covers the variability in kinetics between individual humans (IPCS, 1999). If the POD is derived from a human study, precluding the need for interspecies extrapolation, a default 10-fold uncertainty factor can be considered. For example, when establishing an ARfD for chlorpyrifos, JMPR used an uncertainty factor of 10, as the ARfD was based on inhibition of acetylcholinesterase in humans (FAO/WHO, 1999a). These default uncertainty factors were to some degree initially selected arbitrarily, but subsequently they have been validated by scientific data and practical experience. If allometric scaling has been carried out at the dose adjustment stage (see section 5.2.1.1), then remaining uncertainties for interspecies differences should be accounted for by a smaller uncertainty factor of 2.5, applied in addition to the allometric scaling factor.

In addition to the 10-fold default uncertainty factors to account for interspecies and intraspecies differences in toxicodynamics and toxicokinetics, other uncertainty factors may also be used, for example, 1) to allow for deficiencies in the database, such as poor data quality or absence of a particular type of study, 2) to account for the use of a shorter-duration study in the absence of a long-term study or 3) to account for the use of a LOAEL rather than a NOAEL for the POD. Selection of the size of the additional uncertainty factors to account for these situations is a matter for expert judgement, but these uncertainty factors are generally in the range of 2–10.

Some basic principles for applying appropriate uncertainty factors and their associated numerical values for HBGV derivation (adapted from IPCS, 1990, 2005; WHO, 2015) include (but are not limited to) the following:

- When relevant human data serve as the basis of the POD, the interspecies uncertainty factor to extrapolate from laboratory animals to humans is not applied.

- If the POD from human data is from a subpopulation that does not adequately characterize the population of interest or there are no data to examine human variability in toxicokinetics or toxicodynamics, then the default uncertainty factor of 10 for intraspecies differences should be applied.

- If the available data are of sufficient quality for risk assessment, but the data are for a less-than-lifetime exposure duration, then an additional uncertainty factor can be applied. The actual value of the uncertainty factor is a matter of expert judgement and depends on the available data (WHO, 2015).

- The steepness of the dose–response curve may warrant an increased uncertainty factor, particularly if the effect is severe. For example, an additional uncertainty factor of 5 was applied by JMPR (WHO, 2012) in establishing an ADI for emamectin benzoate to take account of the steep dose–response curves in animal studies and the irreversible histopathological effects in neural tissue at the LOAEL.

- If in vitro data are considered for the derivation of a POD, the uncertainty around a number of factors should be considered, including the relevance of the model to human biology and in

vitro–to–in vivo extrapolation modelling to account for toxicokinetic considerations.

The choice of the numerical value of an uncertainty factor depends on the quantity and quality of relevant data. When adequate data exist, derivation of CSAFs (IPCS, 2005) to describe interspecies differences or human variability in either toxicokinetics or toxicodynamics is preferred to reliance on the default 10-fold uncertainty factor for each. CSAF derivation can include PBPK/PBTK modelling (IPCS, 2010; as discussed previously in section 5.2.1.1). One or more of the four default subfactors can be replaced by the appropriate CSAF (Table 5.1).

Table 5.1. Values for default uncertainty subfactors that can be replaced by CSAFs to derive composite uncertainty factors

Source of uncertainty	Default subfactor		Combined
	Toxicokinetics	Toxicodynamics	
Interspecies variation	4.0	2.5	10
Human interindividual variation	3.16	3.16	10

Source: IPCS (2005)

For example, JECFA used comparative body burden data rather than external dose data in its calculation of a PTMI (now called TMI) for dioxin-like substances; this allowed the usual 100-fold uncertainty factor to be subdivided and replaced by chemical-specific values, as there was no need to include a factor for interspecies differences in toxicokinetics or toxicodynamics (FAO/WHO, 2002a). Detailed guidance on the derivation and application of CSAFs in risk assessment is given elsewhere (Meek et al., 2002; IPCS, 2005; Bhat et al., 2017). Many factors should be considered in deriving a CSAF, some of which depend on the respective subfactor, such as identification of the active chemical moiety (i.e. parent compound or a metabolite), knowledge of the relevant dose metric, relevance of the study population and route of exposure, and adequacy of the study design. For CSAFs for interspecies toxicodynamic differences, key data also include comparative potency data on the MOA for the critical effect, a causal late key event or a surrogate indicator effect.

If data are insufficient to derive a CSAF, the broader concept of data-derived extrapolation factors (DDEFs), which include

categorical factors (Dorne & Renwick, 2005), should also be considered. DDEF methodology and examples are described elsewhere (USEPA, 2014). JMPR has used such categorical DDEFs in the derivation of HBGVs for several carbamate insecticides that inhibit acetylcholinesterase (FAO/WHO, 2009a). These compounds do not require metabolic activation, their reaction with the pharmacological target (acetylcholinesterase) is rapidly reversible, the magnitude of the pharmacological effect is proportional to peak plasma concentration (C_{max}) rather than to plasma concentration integrated over time (AUC) and excretion is rapid. C_{max} has lower variability than clearance, as it depends mainly on the rate and extent of gastrointestinal absorption. This reduced variability in toxicokinetics would justify a 2-fold reduction in the respective subfactors, leading to an overall composite factor of 25 (5 × 5), derived from subfactors of 2 × 2.5 for extrapolation from laboratory animal studies and 1.58 × 3.16 for interindividual differences (for details, see section 2.6 of FAO/WHO, 2009a).

If data are inadequate to derive a CSAF or DDEF, the default interspecies and intraspecies uncertainty factors can be applied. Regardless of whether a CSAF, DDEF or default uncertainty factor is applied, the supporting data and rationale should be clearly stated.

5.4.3 ADIs

5.4.3.1 General considerations

The primary aim of toxicological assessments of food additives and residues of veterinary drugs or pesticides is to determine the amount of substance to which humans can be exposed daily, for up to a lifetime, without adverse health effects. This is achieved by comparing predicted dietary exposure with the ADI. In establishing the ADI, an uncertainty factor (see section 5.4.2) is applied to the POD to provide a conservative margin of safety to account for the inherent uncertainties in extrapolating from experimental animals to humans and for variability in human populations. When results from two or more laboratory animal species are available, the ADI is based on the most sensitive relevant animal species – that is, the species that displayed toxicity at the lowest dose – unless metabolic, toxicokinetic or toxicodynamic data are available establishing that data in the other species are more appropriate for humans.

The ADI is established based on toxicological, pharmacological or microbiological information, as appropriate. Evaluations depend

on studies performed with a chemical substance of defined identity, purity and physical form. In particular, the ADI is valid only for substances that do not differ significantly in identity and quality from the material used to generate the data on which JECFA's or JMPR's decision is based (see Chapter 3). For food additives, JECFA establishes specifications for the product in commerce, and these need to be compared with those for the product tested toxicologically. When specifications have been previously established – for example, by the Joint FAO/WHO Meeting on Pesticide Specifications – these should form the basis of consideration for the evaluation.

In order to establish an ADI, the following information should be available:

- Information on the metabolism and toxicokinetics of the substance in species used for toxicity testing. At a minimum, data should be provided in the rat, in other species, such as the dog, when possible, and in humans, when available. The information should include rate and extent of absorption by the oral route, distribution to and any accumulation in tissues, rates and routes of elimination, extent of metabolism, any information on enzymes involved, saturation of toxicokinetic processes, and any auto-induction or auto-inhibition.

- The chemical nature of the substance in the diet. For example, pesticides and veterinary drugs may undergo chemical changes and are frequently metabolized by the tissues of plants or animals that have been treated with them. Even when a single chemical has been used, the residues may consist of a number of derivatives with different toxicological effects, the exact nature of the residues varying among animals and plants and in different crops and products. It is also possible that processing and cooking of food can change the chemical nature of the substance of concern. This is regularly considered when assessing pesticide residues (FAO, 2016), but only occasionally (e.g. diflubenzuron; FAO/WHO, 2016a) when considering residues of veterinary drugs (Boobis et al., 2017). Stability and generation of reaction and breakdown products during food processing may also need to be considered for food additives.

- Information on the toxicity or pharmacology, as appropriate, of the substance and any of its derivatives (e.g. those occurring as residues) that might occur in the diet. This includes data from

acute, short-term, long-term, reproductive and developmental toxicity studies in laboratory animals and data on their possible genotoxicity and carcinogenicity, together with information on MOA, as appropriate. When specific information on a relevant metabolite is not available, a reasoned case for its toxicological effects should be presented (see WHO, 2015).

- Information on any effects of chemical forms to which consumers are exposed via the diet (parent compound and relevant metabolites) on the intestinal microbiota of humans.

5.4.3.2 *Metabolite considerations*

There may be occasions when the specific effects of concern for a compound are due entirely to a metabolite (or degradation product). In this situation, the activity of the parent compound would be discounted in establishing the ADI; the ADI would instead be based on the toxicological properties of the metabolite (or degradation product), with application of an appropriate uncertainty factor.

JMPR considers that metabolites of pesticides would, under certain conditions, be included in the ADI for the parent compound. Generally, if the metabolites in food commodities are qualitatively the same as and quantitatively similar to those observed in laboratory test species, the ADI would apply to such metabolites as well as to the parent compound. If the metabolites are not identical or not present at the same order of magnitude, separate studies on the metabolites may be necessary. When one or several pesticides are degradation products of another pesticide, a group ADI may be appropriate for the pesticide and its metabolites (e.g. oxydemeton-methyl, demeton-*S*-methyl sulfone and demeton-*S*-methyl; FAO/WHO, 1989).

The toxicological ADI for a veterinary drug established by JECFA is usually based on the toxicity of the parent drug, assuming that all metabolites have similar or lower potency. However, it may sometimes be necessary to calculate an ADI for individual metabolites. Although most compounds have been evaluated as individual substances, there are instances (e.g. streptomycin/dihydrostreptomycin; FAO/WHO, 2002c) where an ADI has been established as a group ADI (see section 5.4.5) using relative potency factors as necessary.

5.4.3.3 Toxicological and pharmacological ADIs

The majority of ADIs are established based on toxicological effects. If pharmacological effects (i.e. effects on the target pharmacological system) are relevant and occur at doses in the same range as or lower than those at which the toxicological effects occur, the ADI may need to be established based on the pharmacological effect. This is relevant primarily to veterinary drugs and has been discussed elsewhere (EMA, 2012; Boobis et al., 2017).

Calculation of a toxicological or pharmacological ADI uses the same formula, as follows:

$$\text{ADI} = \frac{\text{POD}}{\text{UF}}$$

where UF is the uncertainty factor and POD is the BMDL, NOAEL or LOAEL for a toxicological or pharmacological effect appropriate for assessing risk from lifetime exposure.

5.4.3.4 Microbiological ADIs

Antimicrobial risks associated with residues of veterinary drugs have been systematically evaluated by JECFA for several years, and, where appropriate, the ADI is established based on a microbiological end-point (e.g. spiramycin and spectinomycin; FAO/WHO, 1998, 1999b). At its 2017 meeting, JMPR agreed that it should consider, as part of its toxicological assessment of residues of pesticides, their possible effects on the human intestinal microbiota, using the approach developed by JECFA (FAO/WHO, 2017b). There will also be occasions when this is relevant to the evaluation of food additives (Roca-Saavedra et al., 2018).

A decision-tree approach that complies with Guideline 36 of the International Cooperation on Harmonisation of Technical Requirements for Registration of Veterinary Medicinal Products (EMA, 2019; VICH, 2019) has been developed by JECFA for assessing the effects of residues of veterinary drugs on the microbiota of the human intestinal tract. The same decision-tree is now also being applied to residues of pesticides by JMPR. The decision-tree is used to determine the need to establish a microbiological ADI for the substance under review.

The decision-tree approach starts by determining whether there may be microbiologically active residues entering the human colon. This is done in three steps, in which the questions are:

1) *Step 1*: Are residues of the substance, and (or) its metabolites, microbiologically active against representatives of the human intestinal flora?

2) *Step 2*: Do residues enter the human colon?

3) *Step 3*: Do the residues entering the human colon remain microbiologically active?

If the answer is "no" to any of these three questions, then no microbiological ADI would be necessary, and the toxicological ADI would be used.

However, should such residues be present, then two end-points of public health concern – i) disruption of the colonization barrier and ii) an increase in the populations of resistant bacteria – need to be considered:

4) Is it possible to provide scientific justification to eliminate testing (i.e. eliminate the need for a microbiological ADI) for either one or both end-points of concern?

5) If the answer to question 4 is "no" for either end-point, the reference point for the end-point(s) of concern as established in question 4 should be identified. The most appropriate reference point should be used as the POD to determine the microbiological ADI.

A microbiological ADI, based on either of the two end-points mentioned above (i.e. disruption of the colonization barrier or increase in populations of resistant bacteria), is established using either in vitro or in vivo data (see VICH, 2019, for details). The typical situation is that the microbiological ADI is derived from in vitro minimum inhibitory concentration (MIC) data that are obtained by exposing pure or mixed cultures of microorganisms from human faeces to the veterinary drug. In vivo data can be derived from germ-free rodents implanted with human faecal flora (human flora–associated rodents), from conventional laboratory animals or from studies on human volunteers.

The following formula is used to derive a microbiological ADI (mADI) from in vitro MIC data:

$$\text{Upper bound of mADI} = \frac{\text{MIC}_{calc} \times \text{Volume of colon content}}{\text{Fraction of oral dose available to microorganisms} \times 60 \text{ kg}}$$

where:

- The MIC_{calc} represents the 90% lower confidence limit for the mean MIC_{50} (the minimum inhibitory concentration for 50% of strains of the most sensitive relevant organism) of the relevant genera for which the substance is active.

- The volume of colon content value of 500 mL is based on the volume of colon content measured in humans. Previously, this factor in the equation was "Mass of colon content" (220 g/day), but "Volume of colon content" (500 mL) is now considered more appropriate (FAO/WHO, 2016a, 2018).

- The fraction of an oral dose available to microorganisms is ideally based on in vivo measurements for the substance administered orally. Alternatively, if sufficient data are available, the fraction of the dose available for colonic microorganisms can be calculated as 1 minus the fraction of an oral dose excreted in urine. Human data are preferred; in their absence, non-ruminant animal data are recommended. In the absence of data to the contrary, it should be assumed that metabolites have antimicrobial activity equal to that of the parent compound. The fraction may be lowered if there are quantitative in vitro or in vivo data to show that the compound is inactivated during transit through the intestine.

- 60 kg is the standard human body weight.

The following formula is used to derive a microbiological ADI (mADI) from in vivo data:

$$\text{Upper bound of mADI} = \frac{\text{POD}}{\text{UF}}$$

In these cases, the uncertainty factor (UF) is used in an entirely different way than when applied to an ADI based on toxicological or pharmacological data. When determining a microbiologically based ADI, the uncertainty factor is used to account for uncertainty about the amount and relevance of the data available for review. For example, where microbiological effects are studied directly in humans, an uncertainty factor of 1 may be used. Generally, uncertainty factors considered appropriate for microbiological end-points are in the range of 1–10, depending on the quantity and quality of the data.

In establishing a microbiological ADI based on in vivo data, the following need to be considered, and the value of the microbiological ADI needs to be adjusted accordingly:

- Uncertainty factors for in vivo studies should be applied as appropriate, taking into consideration the class of substance, the protocol, the numbers of donors and the sensitivity of the measured outcome variables.

- As microbiological end-points used in in vivo evaluations reflect the reference point for impacts on the intestinal ecosystem, and not the host species itself, it is not necessary to include an uncertainty factor for interspecies differences.

Where both toxicological (or pharmacological) and microbiological ADIs have been determined, these are compared, and the lower is established as the ADI for the substance.

5.4.3.5 Numerical ADI not needed

There are occasions when the establishment of an ADI in numerical terms is not considered to be appropriate. This occurs most commonly when JECFA is assessing food additives, but it can occur occasionally in the assessment of residues of veterinary drugs by JECFA or of pesticides by JMPR. The situation arises when the estimated dietary exposure to or toxicity of the substance is so low that maximum anticipated exposure would be well below any numerical value that would ordinarily be assigned to the upper bound of the ADI. Under such circumstances, the term "ADI 'not specified'" is used by JECFA for food additives. This term is defined to mean that, on the basis of the available data (chemical, biochemical, toxicological and other), the total daily dietary exposure to the substance arising from its use at the levels necessary to achieve the

desired technical effect (e.g. according to Good Manufacturing Practice for food additives, Good Practice in the Use of Veterinary Drugs for veterinary drugs or Good Agricultural Practice for pesticides) and from its acceptable background in food does not, in the opinion of JECFA or JMPR, represent a hazard to health. For that reason, and for the reasons stated in the individual evaluations, the establishment of an ADI in numerical form is not deemed necessary.

A food additive meeting this criterion must be used within the bounds of Good Manufacturing Practice – that is, it should be technologically efficacious and should be used at the lowest level necessary to achieve this effect, it should not conceal inferior food quality or adulteration, and it should not create a nutritional imbalance (FAO/WHO, 1974). That the background occurrence of the chemical must be taken into account in the evaluation of its safety was articulated by the WHO Scientific Group on Procedures for Investigating Intentional and Unintentional Food Additives (WHO, 1967).

The thirty-second JECFA recognized that in some instances it might be inappropriate to establish an ADI for residues of a veterinary drug. When it has been determined that establishing an ADI is unnecessary because of a large margin of safety, the recommendation of a maximum residue limit (MRL) is also unnecessary. For example, at the fortieth meeting of JECFA, an ADI "not specified" was established for bovine somatotropin and its analogues (FAO/WHO, 1993). JECFA noted the lack of activity of the recombinant somatotropins and insulin-like growth factor-1 after oral dosing as well as the low amounts and non-toxic nature of the residues of these substances, even at exaggerated doses. JECFA concluded that these results provided an extremely large margin of safety for humans consuming dairy products from animals treated with the recombinant somatotropins and therefore warranted the establishment of an ADI "not specified". This was reconfirmed at the seventy-eighth meeting of JECFA (FAO/WHO, 2014a).

JMPR also recognizes that in some instances it might be inappropriate to establish an ADI. When it has been determined that establishing an ADI is unnecessary because of a large margin of safety, the recommendation of an MRL is also unnecessary. Although it is rare that the toxicological profile of a pesticide is such that an ADI is unnecessary, the situation has arisen at least once (e.g. the

2012 JMPR concluded that an ADI for ametoctradin was unnecessary; WHO, 2013).

5.4.3.6 Temporary ADIs and temporary MRLs

Temporary ADIs are occasionally established by JECFA in two situations. One situation is when the safety data for a new substance are deemed adequate to establish an ADI, but some non-toxicological information (e.g. some chemical specification information) is missing. For example, temporary ADIs were established for the food additives lutein esters from *Tagetes erecta* and spirulina extract because the chemical specifications were tentative (FAO/WHO, 2014b, 2019). The other situation is when the key safety data are adequate to establish an ADI, yet there are minor deficiencies in toxicological information. An example of the latter situation is the temporary ADI established for the food additive polydimethylsiloxane, pending results of studies to elucidate the mechanism and relevance of the ocular toxicity observed in the submitted toxicology studies (FAO/WHO, 2009b). In both situations, there is confidence that the use of the substance is safe over the relatively short period of time until the missing information is provided. A temporary ADI thus may be established pending the submission of appropriate data to resolve the corresponding issue or data limitation on a predetermined timetable established by JECFA. When establishing a temporary (numerical) ADI, JECFA always uses a higher than usual uncertainty factor, usually by a factor of 2.

JECFA and JMPR may also recommend temporary MRLs for residues of veterinary drugs or residues of pesticides. This may be done for reasons similar to those given above in relation to temporary ADIs or for additional reasons, such as the lack of availability of reliable analytical methods or the need for additional information on the nature or quantification of residues.

In both cases (temporary ADI or temporary MRL), the additional data required for the establishment of an ADI or for recommendation of an MRL are clearly stated in the JECFA and JMPR reports. The additional data required must be submitted by the date specified in the report and should be evaluated by the relevant committee (JECFA or JMPR) at its next meeting.

At the reassessment of a substance with a temporary ADI, the options are as follows: 1) to establish a full ADI, 2) to extend the temporary ADI or 3) to not extend the temporary ADI (i.e. the ADI

is withdrawn). The same options are available with temporary MRLs. For example, the thirty-sixth JECFA established a temporary ADI and temporary MRLs for the veterinary drug levamisole and requested additional toxicological and residue data for re-evaluation by JECFA (FAO/WHO, 1990). Based on the additional data provided, the forty-second JECFA established an ADI. However, JECFA withdrew the temporary MRL for levamisole in milk, as no additional data were made available. Similarly, JECFA withdrew the MRL in eggs because of high amounts of residues (FAO/WHO, 1995).

5.4.3.7 Short-term studies as basis for ADI

When no adverse health effects are seen in long-term studies, an uncertainty factor of 100 may be applied to the POD derived from short-term studies in which higher doses have been used at which an effect has been noted, to establish an ADI (e.g. Brilliant black; FAO/WHO, 1981). Typically, acceptable short-term studies need to be at least 3 months in duration.

Occasionally, short-term toxicity studies have been used as the basis to establish an ADI in cases in which no long-term studies were available. In such cases, the evaluation is not based solely on short-term studies, but is supported by other data, such as biochemical and genotoxicity studies and possibly reproductive or developmental toxicity studies. An example is rosemary extract, for which JECFA established an ADI based on the highest NOAEL from five 90-day studies (FAO/WHO, 2016b). An additional uncertainty factor in respect of the lack of long-term studies may or may not be applied, depending on the nature of the substance and the weight of evidence from the supporting data.

5.4.3.8 Special consideration: allergenicity

There have been a limited number of veterinary drugs for which no numerical value or range could be determined for the toxicological ADI. For example, in the case of ampicillin, based on considerations of allergenicity, the most sensitive toxicological end-point, JECFA did not establish an ADI, as there were insufficient data with which to identify a toxicological NOAEL (FAO/WHO, 2018). However, JECFA had previously concluded that the risk of allergenicity from the structurally related antibiotic benzylpenicillin would be minimal if daily intake in the diet was kept below 0.03 mg per person (0.0005 mg/kg body weight for a 60 kg adult) (FAO/WHO, 1999b). This was

referred to as a toxicological guidance value. JECFA concluded that because ampicillin is much less allergenic than benzylpenicillin, the microbiological ADI of 0–0.002 mg/kg body weight should be protective of potential allergenicity from residues of ampicillin in the diet. Further details can be found in the IPCS guidance on assessing immunotoxicity, including allergenicity (IPCS, 2012).

5.4.4 Tolerable intakes

The principles for establishing tolerable intakes, such as TDIs, TWIs (formerly PTWIs) and TMIs (formerly PTMIs), are the same as for acceptable intakes, as described in section 5.4.3. JECFA has considered the presence of food contaminants on many occasions since 1972, when mercury, lead and cadmium were first assessed (FAO/WHO, 1972). These food contaminants have included, in addition to heavy metals, environmental contaminants such as dioxins, mycotoxins, impurities arising in food additives, solvents used in food processing, packaging material migrants and residues arising from the use of animal feed additives or the non-active components of veterinary drug formulations. Each of these classes of food contaminants possesses its own unique characteristics and evaluation requirements. In addition, JECFA has considered substances that are no longer approved as veterinary drugs, but which might be used illegally and hence should be considered as adulteration of food (e.g. malachite green; FAO/WHO, 2009c). JMPR has on occasion considered metabolites or degradation products that can also occur as contaminants (e.g. aniline; FAO/WHO, 2015) as well as substances no longer approved for pesticidal use (e.g. aldrin/dieldrin; FAO/WHO, 1994), but where previous or illegal use can lead to contamination of food. Thus, the principles for the evaluation of such chemicals should pertain to classes or groups of contaminants rather than to food contaminants as a whole. Guidelines for the evaluation of classes of contaminants are provided in various sections of this EHC monograph.

For contaminants, epidemiological studies are often available that can form the basis for establishing tolerable intakes. If sufficient information is available to perform a dose–response assessment, the POD can be defined from epidemiological studies, and uncertainty factors can then be applied according to the principles outlined in section 5.4.2. JECFA often applies the concept of CSAFs when establishing tolerable intakes for contaminants, particularly in assessing interspecies and interindividual differences in

toxicokinetics. JECFA may also use PBTK modelling for this purpose (e.g. acrylamide; FAO/WHO, 2011a,b).

For contaminants that accumulate in the body over time, consumption of food containing above-average levels of a contaminant on any day may exceed the proportionate share of its weekly (TWI) or monthly (TMI) tolerable intake. JECFA's assessment considers such daily variations, but its real concern is prolonged exposure to the contaminant, because of its accumulation.

5.4.5 Group ADIs/tolerable intakes

If several substances that produce similar toxic effects are to be considered for use as food additives, pesticides or veterinary drugs, or occur as contaminants (e.g. dioxins), it may be appropriate to consider the group of substances together in establishing an ADI or tolerable intake (TDI, TWI, TMI), to limit their overall dietary exposure. For this to be feasible, the substances should produce the same adverse outcome, by a similar MOA. Either a common ADI or tolerable intake is established for the group or one member of the group is identified as the index compound, and exposure to each member of the group is adjusted according to its potency relative to that of the index compound.

When establishing a group ADI or tolerable intake, flexibility should be used in determining which POD to select. As a conservative approach, a group ADI or tolerable intake could be based on the substance with the lowest POD. The relative quality and duration of studies on the various substances should be considered when establishing the group ADI or tolerable intake or when choosing an index compound. In addition, choice of index compound should include consideration of the extensiveness of the toxicological database. Relative potency can be based on the POD, but other DRM techniques could also be used for this purpose. In the case of dioxins and dioxin-like substances, toxic equivalents (TEQs) are agreed by international consensus (FAO/WHO, 2002a). When the POD for one of the substances differs appreciably from the others in the group, that substance may need to be treated separately.

When considering a substance that is a member of a series of substances that are very closely related chemically (e.g. fatty acids, acids and their simple salts, or structurally related metabolites of a known toxic parent compound), but for which toxicological information is limited, it may be possible to base its evaluation on the

group ADI or tolerable intake established for the series of substances or, in the case of a pesticide metabolite, on the ADI for the parent compound. This procedure can be followed only if a great deal of toxicological information is available on at least one member of the series. Confidence in the procedure is enhanced when the known toxic properties of the various substances in a series fall along a well-defined continuum. Interpolation, but not extrapolation, can be performed. The use of this procedure represents one of the few situations in which structure–activity relationships have been used in safety assessments by JECFA and JMPR for chemicals in food.

In some instances, group ADIs can be established primarily based on metabolic information. For example, the safety of esters used as food flavouring agents can be assessed based on toxicological information on their constituent acids and alcohols, provided it is shown that the esters are quantitatively hydrolysed in the gut.

5.4.6 ARfDs

5.4.6.1 General considerations

JMPR routinely evaluates the acute, in addition to the chronic, effects of exposure to pesticide residues in food and has developed a proposed test guideline for a single-dose oral toxicity study (FAO/WHO, 2001a) and guidance on establishing ARfDs for pesticides (FAO/WHO, 1999a, 2001b, 2002d, 2004; Solecki et al., 2005). The guidance provided in these documents for agricultural pesticides should be of value in general considerations of the necessity of establishing an ARfD, as well as in the specific end-point considerations in establishing an ARfD. JECFA has also developed guidance on the need and process for establishing ARfDs for residues of veterinary drugs, for both toxicological and microbiological end-points (WHO, 2017). The text that follows relates mainly to residues (of pesticides or veterinary drugs), but JECFA uses similar principles for other types of substances (e.g. deoxynivalenol: FAO/WHO, 2011a; cyanogenic glycosides: FAO/WHO, 2011c) when the establishment of an ARfD is considered necessary.

The ARfD of a chemical refers to the amount of a substance that can be ingested in a period of 24 hours or less with reasonable certainty of no harm (see section 5.4.1). Because the ARfD is compared with dietary exposure data for a 24-hour period, this will provide a conservative risk assessment for rapidly reversible effects

(e.g. acetylcholinesterase inhibition by carbamates) where the ARfD would be applicable to a single eating occasion.

The decision as to whether the establishment of an ARfD is necessary should be based on the hazard profile of a substance, as well as on specific end-points that may be particularly relevant to effects resulting from acute exposure. Most of the scientific concepts applying to the establishment of ADIs or tolerable intakes apply equally to the establishment of ARfDs (e.g. consideration of the scientific quality of studies, selection of the critical effect). When assessing the need for an ARfD, the entire database should be reviewed, including all available information on the effects of human exposure (Solecki et al., 2005), using a weight-of-evidence approach to determine whether adverse effects seen in repeated-dose toxicity studies might be relevant to single exposures.

Usually a single ARfD is established for the whole population. However, in exceptional cases, two values may be required (e.g. one for the general population and one for a subgroup of the population); this most often occurs when the critical effect is developmental toxicity, and only the developing fetus is at risk, in which case an ARfD would be set for women of childbearing age. In some cases, it may also be necessary to establish an additional ARfD for significant metabolites if they occur on crops and are therefore included in the residue definition (e.g. if these metabolites are likely to show an acute toxicity profile that is different from that of the parent compound) or when metabolites formed in plants or animals are not observed in experimental animal metabolism studies.

If, during the establishment of an ARfD, it becomes apparent that a previously established ADI or tolerable intake is higher than the ARfD, the ADI or tolerable intake should be reconsidered. Such a situation can occur for a number of reasons (e.g. the availability of additional studies, or substances producing more severe effects when given by gavage than in the diet) (FAO/WHO, 2001b). In such a case, even when there is no obvious basis to revise the ADI or tolerable intake, it is recommended that the lower value (i.e. the ARfD) be used as the ADI or tolerable intake.

A different approach is used for the establishment of an ARfD based on an effect on the human gut microflora, which closely follows the approach outlined in section 5.4.3.4 for establishing a microbiological ADI. Of the two potential microbiological end-points

of concern, JECFA considers disruption of the colonization barrier to be the more relevant for acute exposure. It is considered that a single exposure to a substance is unlikely to provide the selective pressure necessary to change the susceptibility of the bacterial population within the microbiome (i.e. antimicrobial resistance).

The key difference between establishing a microbiological ARfD and establishing a microbiological ADI is the estimate of concentration of the substance in the lumen of the colon. In acute dietary exposure to a microbiologically active substance, the dose is ingested as a one-mealtime event and transits down the gastrointestinal tract into the colon, which would contain no other amount of the same substance. In the case of chronic dietary exposure to a microbiologically active substance, there is an assumption of ingestion of the substance every day – that is, each day the substance ingested in a meal enters the gastrointestinal tract, which already contains the same substance as a result of ingestion in a meal from the day before. It can be considered to be present at a "steady state" over a lifetime. Thus, the exposure of intestinal bacteria to a microbiologically active substance in vivo from a single exposure will be lower than that occurring as a result of regular ingestion of the same substance. In addition, available data show that a meal does not transit through the gastrointestinal tract as an intact bolus as is conservatively assumed in calculations of a microbiological ADI. There is a sequential process of stomach and small intestine loading, transit and emptying, leading to a colonic entry that occurs as a series of small pulsed doses over time. Hence, substances contained within a single meal will not enter the colon as a single bolus dose but in a gradient, suggesting that the use of a dilution correction factor in the numerator of the equation for determining the ARfD for microbiological effects would be appropriate, conservatively a factor of 3 (three meals per day).

5.4.6.2 Practical cut-off value for ARfDs

Bearing in mind practical considerations, such as the maximum quantity of a particular food likely to be consumed in a single sitting, a value above which the formal establishment of an ARfD is unnecessary can be proposed (comparable to ADI "not specified"). Hence, when the acute toxicity of a substance is so low that maximum anticipated exposure would be well below any numerical value that would ordinarily be assigned to the ARfD, an ARfD is considered "unnecessary". This practical cut-off value (upper limit) for an ARfD

should be considered with reference to the potential range of dietary exposures to an acutely toxic substance. JMPR has estimated that an ARfD of 5 mg/kg body weight would serve as a conservative value to cover all eventualities for agricultural pesticides, based on practical considerations on consumption and maximum residue levels in foods. JECFA has estimated that a lower ARfD would be adequate for this purpose, but in the interests of harmonization decided to adopt the cut-off proposed by JMPR. An ARfD cut-off at 5 mg/kg body weight would equate to a POD of 500 mg/kg body weight per day in a laboratory animal study, when default uncertainty factors are applied. Thus, if no acute toxicity is seen at doses below 500 mg/kg body weight, then there would be no necessity to establish an ARfD.

5.4.6.3 Biological and toxicological considerations

JMPR has given detailed consideration to the use of particular toxicological end-points that are most relevant to establishing ARfDs (reviewed by Solecki et al., 2005), with a focus on interpreting effects that have been problematic when deciding whether an effect is relevant to an acute exposure to residues of agricultural pesticides in foods. More recently, JECFA has published guidance on the establishment of ARfDs for residues of veterinary drugs, covering toxicological, pharmacological and microbiological effects (WHO, 2017). Much of this guidance is relevant to residues of pesticides and possibly to other chemicals found in food. The following are key points for consideration when evaluating the database regarding the potential for acute toxicity:

- In the absence of data to the contrary, all indications of acute toxicity observed in repeated-dose toxicity studies should be considered as potentially relevant to establishing an ARfD.
- Weight should be given to observations and investigations at the beginning of repeated-dose toxicity studies.
- The BMDL/NOAEL/LOAEL from the most sensitive species should be used as the POD for acute effects unless there is evidence to demonstrate that it is not appropriate for a human risk assessment.
- Isolated findings showing no specificity or clear pattern are not necessarily indications of acute toxicity.

In determining the appropriateness of using doses and end-points from short- or long-term toxicity studies to establish an ARfD, a weight-of-evidence evaluation should be conducted that considers all relevant data. This evaluation includes what is known about the MOA for toxicity and the pertinent biology of the system that is affected. One of the main challenges is to evaluate whether those effects are also likely to occur at the same doses following an acute exposure. A conservative assumption would be that this is the case.

Toxicological information from interim results or consideration of progression of a lesion in repeated-dose toxicity studies may provide insights into the relevance of end-points for establishing ARfDs. For example, if interim data indicate that the response is minimal and becomes pronounced or severe after increasing exposure duration, then repeated exposures are probably the determining factor in the response. Interpretation of the relevance of end-points should also consider toxicokinetic information that would raise concern for acute toxicity, such as slow elimination kinetics or toxicities dependent on the maximum plasma concentrations (C_{max}) achieved, as well as information on the acute toxicity of chemicals with a similar structure. If structure–activity considerations, such as compound class, suggest the likelihood of acute effects but adequate data supporting such effects are not available, then consideration should be given to using the upper bound of the ADI or the tolerable intake as a conservative surrogate for a health-protective acute guidance value.

5.4.6.4 *Stepwise process for establishing ARfDs*

The following stepwise process for establishing ARfDs is recommended:

- *Step 1*: Evaluate the total database for the substance, and establish a toxicological (and, if relevant, pharmacological) and antimicrobial profile for the active substance.

- *Step 2*: Consider the principles for not needing to establish an ARfD:

  - No findings indicative of effects elicited by an acute exposure are observed at doses up to 500 mg/kg body weight; and/or
  - The substance is not an antimicrobial, or the substance is an antimicrobial, but the answer to step 1, 2 or 3 of the decision-tree described in section 5.4.3.4 is "no"; and/or
  - No substance-related mortalities are observed at doses up to 1000 mg/kg body weight in single-dose oral studies; and/or
  - If single-dose mortality is the only trigger, the cause of death is confirmed as not being relevant to human exposures (although this will rarely be possible); and/or
  - For veterinary drugs, dietary exposure to residues through consumption of an injection site does not exceed 5 mg/kg body weight, when they are used in accordance with Good Practice in the Use of Veterinary Drugs – that is, at the established regulatory withdrawal period according to the approved label use.

- *Step 3*: If a decision is taken at this stage not to establish an ARfD, an explicit statement should be provided that "it was unnecessary to establish an ARfD" and the reasons clearly explained.

- *Step 4*: If the above criteria indicate the need to establish an ARfD, then the most appropriate data and end-point should be selected, and the BMDL/NOAEL/LOAEL for that end-point should be identified and used as the POD.

- *Step 5*: An ARfD should be derived by application of appropriate uncertainty factors to the POD.

For substances with a toxicological or pharmacological POD for acute effects below the cut-off of 500 mg/kg body weight per day, a toxicological ARfD should be determined. For substances with antimicrobial effects, a microbiological ARfD should be determined.

For veterinary drugs, when there could be high exposure from the injection site, the database should be assessed for potential acute effects at doses above the cut-off value of 500 mg/kg body weight per day. This may necessitate establishing a particularly high ARfD, to ensure protection from exposure to injection site residues.

An end-point from a repeated-dose toxicity study should be used if the critical effect of the substance has not been adequately evaluated in a single-dose study. Guidance on potentially relevant end-points can be found in Solecki et al. (2005). This is likely to be a more conservative approach and should be stated as such. This does not mean that an uncertainty factor other than the default value should be applied in establishing the ARfD. A refinement of such a POD (e.g. in a special single-dose study) may be considered if the acute exposure estimation exceeds such a potentially conservatively established ARfD. This will be necessary for only a very limited number of substances, according to a retrospective analysis (Solecki et al., 2010). The Organisation for Economic Co-operation and Development (OECD) has developed guidance for the establishment of an ARfD that includes, as Annex II, "Guidance for conducting a single exposure toxicity study" (OECD, 2010), based on the guidance developed by JMPR, to inform investigators should a specific study be considered necessary as a basis for refinement of the ARfD.

Some veterinary drugs designed to act on the physiology of target animals (e.g. mammals) are likely to have an MOA that is also relevant for humans, producing effects that would not be desirable in the consumer. Therefore, pharmacological effects (i.e. those effects caused by the pharmacological MOA of the molecule) are relevant for consumer safety and, hence, for the establishment of an ARfD.

Pharmacological effects (i.e. interaction with molecular targets such as receptors) were not considered in the context of the ARfD by Solecki et al. (2005) or OECD (2010). Such effects do not automatically raise an acute health concern, but need to be considered for acute and chronic health effects in the same way as for toxicological effects. In practice, this may lead to the same numerical value for the ADI and ARfD. For the evaluation of pharmacological effects, careful consideration should be given to the MOA of the substance. In some cases, the MOA can involve several different effects on physiological systems. For example, stimulation of adrenergic receptors can have acute effects on airways, blood pressure and heart rate. In such cases, studies may be needed with observations at appropriate time points to cover the range of effects arising from the MOA. Particular attention should be paid to the appropriateness of the observation times, as they may not be the same as for other toxicological end-points. For example, if the plasma levels peak at 2 hours after oral administration, then it would make

no sense to take measurements of acute pharmacological effects at 24 hours after dosing.

5.4.6.5 Uncertainty factors for ARfDs

The process of determining a toxicological ARfD is essentially the same as that for determining a toxicological ADI or tolerable intake (see sections 5.4.3 and 5.4.4, respectively), involving the identification of the appropriate POD and application of an uncertainty factor. As explained above, when the effect under consideration is due to reversible short-term interaction of the substance with a pharmacological target (e.g. a receptor or ion channel), then the concentration of the substance rather than total exposure usually determines the magnitude of the effect (i.e. the C_{max} is likely to be more relevant than the AUC). Similarly, if the effect of concern is due to a topical effect, such as direct irritation, then the local concentration at the site of action is more relevant than the total exposure expressed on a body weight basis. In such cases, there will be less interspecies and interindividual variation in toxicokinetics, which should be taken into consideration in the choice of uncertainty factor (see section 5.4.2).

If human data are available but are not sufficient to be used directly to establish the ARfD, they might nevertheless be of value in determining quantitative differences in the toxicokinetics or toxicodynamics of the substance (e.g. data on the production and degradation of a toxic metabolite) between experimental animals and humans, enabling the calculation of data-derived CSAFs.

The use of uncertainty factors that are lower than the default values (see section 5.4.2) might be appropriate if the end-point used to establish the ARfD is of minimal adversity and the POD is from a repeated-dose toxicity study (e.g. increased organ weight with minimal pathological change, or reduced feed consumption and body weight gain observed in the first days of dosing), as the use of such an end-point to establish an ARfD is very likely to be conservative. Again, the choice of uncertainty factor used must be fully justified.

Situations where additional uncertainty factors might be used are the same as those when establishing an ADI or tolerable intake.

5.4.6.6 ARfD calculation

When a toxicological or pharmacological effect serves as the basis of the POD, the ARfD is determined as follows:

$$\text{ARfD} = \frac{\text{POD}}{\text{UF}}$$

where:

- POD is the BMDL, NOAEL or LOAEL;
- UF is the default uncertainty factor of 100 when extrapolating from data in experimental animals to humans or 10 when using data from a human study, a CSAF, a DDEF or some other combination of factors, as justified in the assessment.

When a microbiological effect based on in vitro data (e.g. MIC, no-observed-adverse-effect concentration [NOAEC]) is used to determine the microbiological ARfD (mARfD), the following equation is used:

$$\text{mARfD} = \frac{(\text{MIC}_{\text{calc}} \text{ or other POD}) \times \text{Correction factors} \times \text{Volume of colon content}}{\text{Fraction of oral dose available to microorganisms} \times 60 \text{ kg}}$$

where:

- Correction factors (where appropriate) take into account considerations that are not used for the microbiological ADI, but may be appropriate for the microbiological ARfD. These include a factor of 3 to allow for temporal dilution during gastrointestinal transit and for dilution by consumption of additional meals, as explained above (see section 5.4.6.1). Additional factors may be considered to take into account the inoculum effect on MIC determinations, pH effects on MIC and possibly other physicochemical-specific factors of the growth conditions used in testing (e.g. incubation atmosphere, growth substrates/factors that affect growth and metabolism of the tested organisms; Cerniglia & Kotarski, 1999, 2005; Maurer et al., 2015). When data from continuous or semi-continuous culture and batch-fed culture are used, the effects of an acute dose (one-time exposure) of

the substance on the intestinal microbiota should be evaluated; however, if this information is not available, then studies of repeated doses or continuous exposure to the substance (i.e. after 1 or a few days of substance added to the test systems) may yield a POD for acute exposure or may provide sufficient information to derive a correction factor.

- Other terms are explained in section 5.4.3.4.

When a microbiological end-point based on a POD from in vivo data is used to establish the microbiological ARfD (mARfD), the following equation applies:

$$\text{mARfD} = \frac{\text{POD}}{\text{UF}}$$

In establishing an ARfD based on a microbiological end-point, the following need to be considered, in addition to the factors described in section 5.4.3.4 for microbiological ADIs, and the value of the microbiological ARfD needs to be adjusted accordingly:

- Were the assumptions and uncertainty factors used in determining the microbiological ADI applicable for a single exposure? For example, were multiple doses administered to the animals?

- If the experimental design used repeated dosing to determine a microbiological ADI for chronic exposure, particular attention should be paid to observations and investigations at the beginning of the studies (i.e. after 1 or a few days) as the basis to establish an ARfD, or as the basis for a correction factor.

Where both toxicological (or pharmacological) and microbiological ARfDs have been determined, these are compared, and the lower is established as the ARfD for the substance.

5.4.6.7 Different ARfDs for population subgroups

It is preferable, especially for clarity of subsequent risk management and enforcement, to establish a single ARfD to cover the whole population. It is important to ensure that any ARfDs established are adequate to protect the embryo or fetus from possible in utero effects. Although an ARfD based on developmental (embryo/fetal) effects would necessarily apply to pregnant women

(or, more pragmatically, women of childbearing age), it is recognized that such an ARfD would not be relevant to other population subgroups and hence may be very conservative for these groups. This will be the case, for example, for children 1–6 years of age for whom specific acute consumption data are available and who can therefore be modelled separately with respect to acute dietary exposure to a substance. Thus, in those situations in which a developmental end-point drives an ARfD for a substance exhibiting no other toxicity at the POD determined for the developmental effect, consideration could be given to establishing a second ARfD for the rest of the population. This second ARfD would be based on another, non-developmental, end-point and would give a higher ARfD for the rest of the population than that for pregnant women. Alternatively, a conclusion might be reached that an ARfD is not necessary for the rest of the population.

5.4.6.8 Dietary exposure considerations in relation to ARfDs

For risk characterization purposes, the ARfD of a substance is compared with the estimated acute dietary exposure to the substance from various foods. This allows risk managers to identify for which food commodities and substance uses regulatory actions may be necessary for public health protection. The methodology for estimating acute dietary exposures is described in detail in Chapter 6.

5.5 MOE approach

There are three general cases where the MOE approach is applied: 1) when assessing a presumed DNA-reactive mutagenic carcinogen, 2) when there are insufficient data for deriving an HBGV and 3) for additives used in infant formula at relatively high inclusion levels.

The MOE is defined as the ratio of the POD (e.g. BMDL or NOAEL) to the theoretical, predicted or estimated exposure dose or concentration (ED):

$$\text{MOE} = \frac{\text{POD}}{\text{ED}}$$

MOEs should be rounded to, at most, two significant figures to avoid spurious precision.

It is important to note that an MOE is not an absolute value but rather a relative comparison of human exposure to a POD derived from laboratory animal or human studies. There are also strengths and weaknesses inherent in the data used for its calculation.

The MOE approach is intended to provide some indication to risk managers as to the level of concern and to help in assessing the need for, and urgency of, further action. The approach provides advice to inform risk managers of how close human exposures are to those anticipated to produce a measurable effect in laboratory animals or humans. In addition, MOEs for different substances or for different risk mitigation strategies can be compared to assist risk managers in prioritizing or assessing risk management actions (Benford et al., 2010).

5.5.1 MOE for DNA-reactive mutagenic carcinogens

JECFA has used the MOE approach most often for substances that are (presumed to be) DNA-reactive mutagenic carcinogens (see Chapter 4, section 4.5). For such substances, it is generally not considered appropriate to establish an HBGV. JECFA has typically used the $BMDL_{10}$ for carcinogenicity derived from laboratory animal studies as the POD for this purpose (e.g. ethyl carbamate: FAO/WHO, 2006; acrylamide: FAO/WHO, 2011a,b). There are some situations in which it has been possible to calculate the potency for a mutagenic carcinogen directly from epidemiological data. For example, at its eighty-third meeting, JECFA (FAO/WHO, 2017a) evaluated the risk of aflatoxin, a known mutagenic carcinogen, and used epidemiological data from Yeh et al. (1989) to derive potency estimates. However, for most such substances, although epidemiological data may exist, they do not often enable determination of potency. Hence, in order to provide risk managers with some indication of the level of health concern, the MOE approach is applied.

Note that the MOE approach is used in these situations in preference to linear extrapolation from a BMDL. At its sixty-fourth meeting (FAO/WHO, 2006), JECFA considered the use of linear extrapolation from the BMDL to estimate the risk of cancer at relevant levels of human exposure and concluded that

> calculation of the intake associated with an incidence of 1 in 1 million from the BMDL for a 10% incidence using linear extrapolation is simply

equivalent to dividing the BMDL by 100 000, and this approach is therefore no more informative than calculation of an MOE.

There is not a strict single criterion to decide whether the size of the calculated MOE for a DNA-reactive mutagenic carcinogen is indicative of a human health concern. The following uncertainties need to be considered when interpreting such an MOE: 1) uncertainties regarding human exposure (i.e. whether the exposure is theoretical, predicted or estimated), 2) uncertainties in extrapolating from laboratory animals to humans (i.e. accounting for interspecies differences and human variability), 3) uncertainties related to the quality of the study and 4) additional uncertainties in the process of carcinogenesis (e.g. human variability in DNA repair or cell cycle control). These factors need to be considered when interpreting the MOE with respect to health concerns and need to be clearly described in the report.

5.5.2 MOE for substances with insufficient data

The MOE approach may be needed for contaminants for which an HBGV cannot be established because there are only limited data available. For example, advice may be needed because contamination can be minimized but not eliminated. Similarly, the MOE approach may be appropriate for advice on whole foods and flavouring agents. In contrast, it is possible to request additional data for regulated or registered food additives and pesticides. In order to provide risk managers with some indication of the level of health concern, JECFA has, since its sixty-fourth meeting (FAO/WHO, 2006), applied the MOE approach. On occasion, JMPR has also used this approach.

When interpreting MOEs for substances with insufficient data, consideration needs to be given to the same factors that are used in describing the uncertainties around HBGVs. These include human variability, extrapolation from laboratory animal data to humans, nature of the effect that is the basis of the POD, the assumptions made regarding dietary exposure and database uncertainties (e.g. the extent to which potentially relevant end-points have been assessed). These factors need to be considered when interpreting the MOE with respect to health concerns and should be clearly described in the report.

JECFA has used the MOE approach for contaminants for end-points other than carcinogenicity by a presumed DNA-reactive mutagenic MOA – for example, for polybrominated diphenyl ethers (PBDEs). An HBGV could not be proposed for PBDEs because of

multiple considerations. These included (but were not limited to) the following: 1) the fact that PBDEs represent a complex group of related chemicals, and the pattern of PBDE congeners in food is not clearly defined by a single commercial mixture; 2) data were inadequate to establish a common mechanism of action that would allow a single congener to be used as a surrogate for total exposure or, alternatively, as the basis for establishing toxic equivalency factors; and 3) the lack of a systematic database on toxicity, including long-term studies on the main congeners present in the diet, using standardized testing protocols, that defined dose–response relationships for individual PBDEs of importance (FAO/WHO, 2006).

In the evaluation of flavouring agents, JECFA applies the MOE approach to substances that exceed their respective threshold of toxicological concern values (see Chapter 9).

5.5.3 MOE for additives used in infant formula

The MOE approach is used by JECFA in the evaluation of food additives for use in infant formula, as the ADI concept does not apply to infants up to the age of 12 weeks because they might be at risk at lower levels of exposure compared with older age groups. This is due to special considerations, such as their immature metabolic capacities, the greater permeability of the immature gut, and their rapid growth and development. Therefore, risk characterization for very young infants has to be considered on a case-by-case basis (FAO/WHO, 2014b).

Toxicological testing strategies for food additives to be used in infant formula require approaches that differ from those generally adopted for food additives. For example, evaluation of food additives to be used in infant formula requires consideration of safety studies involving exposure of very young animals. The reproductive and developmental toxicity studies commonly available for evaluations of chemicals in food address the possible impact on neonatal animals arising through in utero and lactational exposure. However, they usually do not include direct oral administration to neonatal animals, and such studies are required for the evaluation of food additives in infant formula. If the food additive is proposed for use in infant formula at relatively high levels (e.g. 0.1% or greater), then conducting toxicological studies in neonatal animals at doses 2 or more orders of magnitude greater than the anticipated human exposure, which is the approach commonly taken for food additives,

may not be feasible. Studies on the effects of direct administration of the food additive to neonatal animals may, therefore, involve use of doses that are only small multiples of human infant exposure. For this reason, the MOEs for some food additives proposed for use in infant formula can be quite low, in the range of 1–10 for infants.

Interpretation of the MOE needs to consider uncertainties or conservatism that may exist in the toxicological POD or in the exposure estimates, as described above. Considerations of particular relevance to infants include:

- the relative maturity of the ADME processes;
- the potential effects of exposure during life stages in experimental animals of relevance to human infants;
- the relevance of the neonatal animal models used in toxicological testing for the human infant;
- the design and outcome of any clinical studies conducted with infants (e.g. total number and age of infants tested, growth, tolerance, types of adverse reaction examined); and
- reports of adverse reactions in post-marketing surveillance, where the infant formula is already in use in some countries.

Specific factors related to the dietary exposure assessments that should be considered for the interpretation of an MOE include the following:

- Formula is the only source of nutrition for the first 12 weeks of life.

- Variability of exposure among infants is small.

- Duration of exposure is for a limited time, and exposure decreases on a body weight basis during the exposure period.

JECFA has concluded that when the above issues are considered, an MOE, based upon appropriate experimental animal studies, in the region of 1–10 could be interpreted as indicating low concern for the health of infants aged 0–12 weeks consuming the food additive in infant formula.

5.6 Conclusions

Chapter 5 outlines how dose–response assessment for various types of data should be performed and how the HBGV or MOE is

established. Several approaches are described. The BMD approach is preferred, as it makes use of all the dose–response data, there is scientific agreement on the models that can be used and there are software platforms available for use in risk assessment. Moreover, model averaging is preferred to single-model selection to account for the uncertainties in data and modelling. In the few cases in which insufficient underlying data are available for modelling, but a NOAEL is reported, the NOAEL can be used as a POD to derive an HBGV. It is a prerequisite that each assessment has to be carefully documented. It has to be transparent, comprehensible and reproducible. A rationale should be given for the chosen approach, especially for the selected critical effect and applied uncertainty factors.

5.7 References[1]

Aerts M, Wheeler MW, Cortinas Abrahantes J (2020). An extended and unified modeling framework for benchmark dose estimation for both continuous and binary data. Environmetrics. 31(7):e2630. doi:10.1002/env.2630.

Agresti A (2013). Categorical data analysis. Hoboken (NJ): John Wiley & Sons.

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 effect levels. Fundam Appl Toxicol. 23(4):487–95. doi:10.1006/faat.1994.1133.

Allen BC, Strong PL, Price CJ, Hubbard SA, Daston GP (1996). Benchmark dose analysis of developmental toxicity in rats exposed to boric acid. Toxicol Sci. 32(2):194–204. doi:10.1093/toxsci/32.2.194.

Andersen ME, Dennison JE (2001). Mode of action and tissue dosimetry in current and future risk assessments. Sci Total Environ. 274(1–3):3–14. doi:10.1016/S0048-9697(01)00744-6.

Benford D, Bolger PM, Carthew P, Coulet M, DiNovi M, Leblanc JC et al. (2010). Application of the margin of exposure (MOE) approach to substances in food that are genotoxic and carcinogenic. Food Chem Toxicol. 48(Suppl 1):S2–S24. doi:10.1016/j.fct.2009.11.003.

Bhat VS, Meek ME, Valcke M, English C, Boobis A, Brown R (2017). Evolution of chemical-specific adjustment factors (CSAF) based on recent international experience; increasing utility and facilitating regulatory acceptance. Crit Rev Toxicol. 47(9):729–49. doi:10.1080/10408444.2017.1303818.

[1] Internet links provided in these references were active as of the date of final editing.

Bokkers BG, Slob W (2007). Deriving a data-based interspecies assessment factor using the NOAEL and the benchmark dose approach. Crit Rev Toxicol. 37(5):355–73. doi:10.1080/10408440701249224.

Boobis A, Cerniglia C, Chicoine A, Fattori V, Lipp M, Reuss R et al. (2017). Characterizing chronic and acute health risks of residues of veterinary drugs in food: latest methodological developments by the Joint FAO/WHO Expert Committee on Food Additives. Crit Rev Toxicol. 47:889–903. doi:10.1080/10408444.2017.1340259.

Breslow NE, Day NE (1980). Statistical methods in cancer research. Volume 1 – The analysis of case–control studies. Lyon: International Agency for Research on Cancer (IARC Scientific Publications No. 32; https://publications.iarc.fr/Book-And-Report-Series/Iarc-Scientific-Publications/Statistical-Methods-In-Cancer-Research-Volume-I-The-Analysis-Of-Case-Control-Studies-1980).

Budtz-Jørgensen E (2006). Estimation of the benchmark dose by structural equation models. Biostatistics. 8(4):675–88. doi:10.1093/biostatistics/kxl037.

Cerniglia CE, Kotarski S (1999). Evaluation of veterinary drug residues in food for their potential to affect human intestinal microflora. Regul Toxicol Pharmacol. 29(3):238–61. doi:10.1006/rtph.1999.1300.

Cerniglia CE, Kotarski S (2005). Approaches in the safety evaluations of veterinary antimicrobial agents in food to determine the effects on the human intestinal microflora. J Vet Pharmacol Ther. 28(1):3–20. doi:10.1111/j.1365-2885.2004.00595.x.

Chen CC, Chen JJ (2014). Benchmark dose calculation for ordered categorical responses. Risk Anal. 34(8):1435–47. doi:10.1111/risa.12167.

Cole SR, Frangakis CE (2009). The consistency statement in causal inference: a definition or an assumption? Epidemiology. 20(1):3–5. doi:10.1097/EDE.0b013e31818ef366.

Crump KS (1984). A new method for determining allowable daily intakes. Fundam Appl Toxicol. 4(5):854–71. doi:10.1016/0272-0590(84)90107-6.

Crump KS, Chen C, Louis TA (2010). The future use of in vitro data in risk assessment to set human exposure standards: challenging problems and familiar solutions. Environ Health Perspect. 118(10):1350–4. doi:10.1289/ehp.1001931.

Dekkers S, de Heer C, Rennen MA (2001). Critical effect sizes in toxicological risk assessment: a comprehensive and critical evaluation. Environ Toxicol Pharmacol. 10(1–2):33–52. doi:10.1016/S1382-6689(01)00068-0.

Dorne JLCM, Renwick AG (2005). The refinement of uncertainty/safety factors in risk assessment by the incorporation of data on toxicokinetic variability in humans. Toxicol Sci. 86(1):20–6. doi:10.1093/toxsci/kfi160.

Edler L, Hart A, Greaves P, Carthew P, Coulet M, Boobis A et al. (2014). 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. doi:10.1016/j.fct.2013.10.030.

EFSA (2017). Update: guidance on the use of the benchmark dose approach in risk assessment. European Food Safety Authority Scientific Committee. EFSA J. 15(1):4658. doi:10.2903/j.efsa.2017.4658.

EMA (2012). Guideline on the approach to establish a pharmacological ADI. European Medicines Agency, Committee for Medicinal Products for Veterinary Use (EMA/CVMP/SWP/355689/2006; https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-approach-establish-pharmacological-acceptable-daily-intake_en.pdf).

EMA (2019). VICH GL36(R2): Studies to evaluate the safety of residues of veterinary drugs in human food: general approach to establish a microbiological ADI. Amsterdam: European Medicines Agency, Committee for Medicinal Products for Veterinary Use (EMA/CVMP/VICH/467/2003; https://www.ema.europa.eu/en/documents/scientific-guideline/vich-gl36r2-studies-evaluate-safety-residues-veterinary-drugs-human-food-general-approach-establish_en.pdf).

FAO (2016). Submission and evaluation of pesticide residues data for the estimation of maximum residue levels in food and feed, third edition. Rome: Food and Agriculture Organization of the United Nations (FAO Plant Production and Protection Paper, No. 225; http://www.fao.org/3/a-i5452e.pdf).

FAO/WHO (1957). General principles governing the use of food additives. First report of the Joint FAO/WHO Expert Committee on Food Additives. Rome: Food and Agriculture Organization of the United Nations; Geneva: World Health Organization (FAO Nutrition Meetings Report Series, No. 15; WHO Technical Report Series, No. 129; http://whqlibdoc.who.int/trs/WHO_TRS_129.pdf).

FAO/WHO (1958). Procedures for the testing of intentional food additives to establish their safety for use. Second report of the Joint FAO/WHO Expert Committee on Food Additives. Rome: Food and Agriculture Organization of the United Nations; Geneva: World Health Organization (FAO Nutrition Meetings Report Series, No. 17; WHO Technical Report Series, No. 144; http://whqlibdoc.who.int/trs/WHO_TRS_144.pdf).

FAO/WHO (1962). Evaluation of the toxicity of a number of antimicrobials and antioxidants. Sixth report of the Joint FAO/WHO Expert Committee on Food Additives. Rome: Food and Agriculture Organization of the United Nations; Geneva: World Health Organization (FAO Nutrition Meetings Report Series, No. 31; WHO Technical Report Series, No. 228; http://whqlibdoc.who.int/trs/WHO_TRS_228.pdf).

FAO/WHO (1972). Evaluation of certain food additives and the contaminants mercury, lead, and cadmium. Sixteenth report of the Joint FAO/WHO Expert Committee on Food Additives. Rome: Food and Agriculture Organization of the United Nations; Geneva: World Health Organization (FAO Nutrition Meetings

Report Series, No. 51; WHO Technical Report Series, No. 505 and corrigendum; http://whqlibdoc.who.int/trs/WHO_TRS_505.pdf).

FAO/WHO (1974). Toxicological evaluation of certain food additives with a review of general principles and of specifications. Seventeenth report of the Joint FAO/WHO Expert Committee on Food Additives. Rome: Food and Agriculture Organization of the United Nations; Geneva: World Health Organization (FAO Nutrition Meetings Report Series, No. 53; WHO Technical Report Series, No. 539 and corrigendum; http://whqlibdoc.who.int/trs/WHO_TRS_539.pdf).

FAO/WHO (1981). Evaluation of certain food additives. Twenty-fifth report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva: World Health Organization (WHO Technical Report Series, No. 669; https://apps.who.int/iris/bitstream/handle/10665/41524/WHO_TRS_669.pdf?sequence=1).

FAO/WHO (1989). Pesticide residues in food – 1989. Report of the Joint Meeting of the FAO Panel of Experts on Pesticide Residues in Food and the Environment and a WHO Expert Group on Pesticide Residues. Rome: Food and Agriculture Organization of the United Nations (FAO Plant Production and Protection Paper, No. 99).

FAO/WHO (1990). Evaluation of certain veterinary drug residues in food. Thirty-sixth report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva: World Health Organization (WHO Technical Report Series, No. 799; http://whqlibdoc.who.int/trs/WHO_TRS_799.pdf).

FAO/WHO (1993). Evaluation of certain veterinary drug residues in food. Fortieth report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva: World Health Organization (WHO Technical Report Series, No. 832; http://whqlibdoc.who.int/trs/WHO_TRS_832.pdf).

FAO/WHO (1994). Pesticide residues in food – 1994. Report of a Joint Meeting of the FAO Panel of Experts on Pesticide Residues in Food and the Environment and a WHO Expert Group on Pesticide Residues. Rome: Food and Agriculture Organization of the United Nations (FAO Plant Production and Protection Paper, No. 127; http://www.fao.org/fileadmin/templates/agphome/documents/Pests_Pesticides/JMPR/Reports_1991-2006/Report1994.pdf).

FAO/WHO (1995). Evaluation of certain veterinary drug residues in food. Forty-second report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva: World Health Organization (WHO Technical Report Series, No. 851; http://whqlibdoc.who.int/trs/WHO_TRS_851.pdf).

FAO/WHO (1998). Evaluation of certain veterinary drug residues in food. Forty-eighth report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva: World Health Organization (WHO Technical Report Series, No. 879; https://apps.who.int/iris/bitstream/handle/10665/42127/WHO_TRS_879.pdf?sequence=1).

FAO/WHO (1999a). Procedures for estimating acute reference dose. In: Pesticide residues in food – 1998. Report of the Joint Meeting of the FAO Panel of Experts on Pesticide Residues in Food and the Environment and the WHO

Core Assessment Group on Pesticide Residues. Rome: Food and Agriculture Organization of the United Nations (FAO Plant Production and Protection Paper, No. 148; http://www.fao.org/fileadmin/templates/agphome/documents/Pests_Pesticides/JMPR/Reports_1991-2006/REPORT1998.pdf).

FAO/WHO (1999b). Evaluation of certain veterinary drug residues in food. Fiftieth report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva: World Health Organization (WHO Technical Report Series, No. 888; https://apps.who.int/iris/bitstream/handle/10665/42182/WHO_TRS_888.pdf?sequence=1).

FAO/WHO (2001a). Proposed test guideline – single-dose toxicity study by the oral route (for use in establishing acute reference doses for chemical residues in food and drinking-water). In: Pesticide residues in food – 2000. Report of the Joint Meeting of the FAO Panel of Experts on Pesticide Residues in Food and the Environment and the WHO Core Assessment Group on Pesticide Residues. Rome: Food and Agriculture Organization of the United Nations (FAO Plant Production and Protection Paper, No. 163; http://www.fao.org/fileadmin/templates/agphome/documents/Pests_Pesticides/JMPR/Reports_1991-2006/Report__2000.pdf).

FAO/WHO (2001b). Further issues related to establishing an acute reference dose. In: Pesticide residues in food – 2001. Report of the Joint Meeting of the FAO Panel of Experts on Pesticide Residues in Food and the Environment and the WHO Core Assessment Group on Pesticide Residues. Rome: Food and Agriculture Organization of the United Nations (FAO Plant Production and Protection Paper, No. 167; http://www.fao.org/fileadmin/templates/agphome/documents/Pests_Pesticides/JMPR/Reports_1991-2006/REPORT2001.pdf).

FAO/WHO (2002a). Evaluation of certain food additives and contaminants. Fifty-seventh report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva: World Health Organization (WHO Technical Report Series, No. 909; http://whqlibdoc.who.int/trs/WHO_TRS_909.pdf).

FAO/WHO (2002b). Pesticide residues in food – 2002. Report of the Joint Meeting of the FAO Panel of Experts on Pesticide Residues in Food and the Environment and the WHO Core Assessment Group on Pesticide Residues. Rome: Food and Agriculture Organization of the United Nations (FAO Plant Production and Protection Paper, No. 172; http://www.fao.org/fileadmin/templates/agphome/documents/Pests_Pesticides/JMPR/Reports_1991-2006/Report_2002.pdf).

FAO/WHO (2002c). Evaluation of certain veterinary drug residues in food. Fifty-eighth report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva: World Health Organization (WHO Technical Report Series, No. 911; https://apps.who.int/iris/bitstream/handle/10665/42560/WHO_TRS_911.pdf;jsessionid=2FF6ED5CF07F1208ADB7CF4C109DB981?sequence=1).

FAO/WHO (2002d). Further guidance on derivation of the acute RfD. In: Pesticide residues in food – 2002. Report of the Joint Meeting of the FAO Panel of Experts on Pesticide Residues in Food and the Environment and the WHO Core Assessment Group on Pesticide Residues. Rome: Food and Agriculture Organization of the United Nations (FAO Plant Production and Protection

Paper, No. 172; http://www.fao.org/fileadmin/templates/agphome/documents/Pests_Pesticides/JMPR/Reports_1991-2006/Report_2002.pdf).

FAO/WHO (2004). Guidance on the establishment of acute reference doses. In: Pesticide residues in food – 2004. Report of the Joint Meeting of the FAO Panel of Experts on Pesticide Residues in Food and the Environment and the WHO Core Assessment Group on Pesticide Residues. Rome: Food and Agriculture Organization of the United Nations (FAO Plant Production and Protection Paper, No. 178; http://www.fao.org/fileadmin/templates/agphome/documents/Pests_Pesticides/JMPR/Reports_1991-2006/report2004jmpr.pdf).

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

FAO/WHO (2007). Methylmercury. In: Safety evaluation of certain food additives and contaminants. Geneva: World Health Organization; 269–315 (WHO Food Additives Series, No. 58; https://apps.who.int/iris/bitstream/handle/10665/43645/9789241660587_eng.pdf;jsessionid=D98C4623A87C7AF429E36D15981C51E6?sequence=1).

FAO/WHO (2009a). Pesticide residues in food – 2008. Report of the Joint Meeting of the FAO Panel of Experts on Pesticide Residues in Food and the Environment and the WHO Core Assessment Group on Pesticide Residues. Rome: Food and Agriculture Organization of the United Nations (FAO Plant Production and Protection Paper, No. 193; http://www.fao.org/fileadmin/templates/agphome/documents/Pests_Pesticides/JMPR/JMPRReport08.pdf).

FAO/WHO (2009b). Evaluation of certain food additives and contaminants. Sixty-ninth report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva: World Health Organization (WHO Technical Report Series, No. 952; http://whqlibdoc.who.int/trs/WHO_TRS_952_eng.pdf).

FAO/WHO (2009c). Evaluation of certain veterinary drug residues in food. Seventieth report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva: World Health Organization (WHO Technical Report Series, No. 954; https://apps.who.int/iris/bitstream/handle/10665/44085/WHO_TRS_954_eng.pdf?sequence=1).

FAO/WHO (2011a). Evaluation of certain food additives and contaminants. Seventy-second report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva: World Health Organization (WHO Technical Report Series, No. 959; https://apps.who.int/iris/bitstream/handle/10665/44514/WHO_TRS_959_eng.pdf?sequence=1).

FAO/WHO (2011b). Evaluation of certain food additives and contaminants. Seventy-third report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva: World Health Organization (WHO Technical Report Series, No. 960; https://apps.who.int/iris/bitstream/handle/10665/44515/WHO_TRS_960_eng.pdf).

FAO/WHO (2011c). Evaluation of certain food additives and contaminants. Seventy-fourth report of the Joint FAO/WHO Expert Committee on Food

Additives. Geneva: World Health Organization (WHO Technical Report Series, No. 966; https://apps.who.int/iris/bitstream/handle/10665/44788/WHO_TRS_966_eng.pdf?sequence=1).

FAO/WHO (2014a). Evaluation of certain veterinary drug residues in food. Seventy-eighth report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva: World Health Organization (WHO Technical Report Series, No. 988; https://apps.who.int/iris/bitstream/handle/10665/127845/9789241209885_eng.pdf;jsessionid=47A218332BEF037F0D930CFA4F4F8B2A?sequence=1).

FAO/WHO (2014b). Evaluation of certain food additives. Seventy-ninth report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva: World Health Organization (WHO Technical Report Series, No. 990; https://apps.who.int/iris/bitstream/handle/10665/150883/9789241209908_eng.pdf?sequence=1).

FAO/WHO (2015). Pesticide residues in food – 2015. Report of the Joint Meeting of the FAO Panel of Experts on Pesticide Residues in Food and the Environment and the WHO Core Assessment Group on Pesticide Residues. Rome: Food and Agriculture Organization of the United Nations (FAO Plant Production and Protection Paper, No. 223; http://www.fao.org/fileadmin/templates/agphome/documents/Pests_Pesticides/JMPR/Report2015/web_JMPR_2015_Report_Draft_Nov_21v2.pdf).

FAO/WHO (2016a). Evaluation of certain veterinary drug residues in food. Eighty-first report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva: World Health Organization (WHO Technical Report Series, No. 997; https://apps.who.int/iris/bitstream/handle/10665/204670/9789240695504_eng.pdf?sequence=1).

FAO/WHO (2016b). Evaluation of certain food additives. Eighty-second report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva: World Health Organization (WHO Technical Report Series, No. 1000; https://apps.who.int/iris/bitstream/handle/10665/250277/9789241210003-eng.pdf?sequence=1).

FAO/WHO (2017a). Evaluation of certain contaminants in food. Eighty-third report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva: World Health Organization and Food and Agriculture Organization of the United Nations (WHO Technical Report Series, No. 1002; https://apps.who.int/iris/bitstream/handle/10665/254893/9789241210027-eng.pdf?sequence=1).

FAO/WHO (2017b). Pesticide residues in food – 2017. Report of the Joint Meeting of the FAO Panel of Experts on Pesticide Residues in Food and the Environment and the WHO Core Assessment Group on Pesticide Residues. Rome: Food and Agriculture Organization of the United Nations (FAO Plant Production and Protection Paper, No. 232; http://www.fao.org/fileadmin/templates/agphome/documents/Pests_Pesticides/JMPR/Report2017/web_2017_JMPR_Report_Final.pdf).

FAO/WHO (2018). Evaluation of certain veterinary drug residues in food. Eighty-fifth report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva: World Health Organization and Food and Agriculture Organization of

the United Nations (WHO Technical Report Series, No. 1008; https://apps.who.int/iris/bitstream/handle/10665/259895/9789241210171-eng.pdf?sequence=1).

FAO/WHO (2019). Evaluation of certain food additives. Eighty-sixth report of the Joint FAO/WHO Expert Committee on Food Additives. Geneva: World Health Organization and Food and Agriculture Organization of the United Nations (WHO Technical Report Series, No. 1014; https://apps.who.int/iris/bitstream/handle/10665/279832/9789241210232-eng.pdf?ua=1).

Fowles JR, Alexeeff GV, Dodge D (1999). The use of benchmark dose methodology with acute inhalation lethality data. Regul Toxicol Pharmacol. 293(3):262–78. doi:10.1006/rtph.1999.1295.

Frome EL (1983). The analysis of rates using Poisson regression models. Biometrics. 39(3):665–74. doi:10.2307/2531094.

Geys H, Molenberghs G, Ryan LM (1999). Pseudolikelihood modeling of multivariate outcomes in developmental toxicology. J Am Stat Assoc. 94(447):734–45. doi:10.2307/2669986.

Haber LT, Dourson ML, Allen BC, Hertzberg RC, Parker A, Vincent MJ et al. (2018). Benchmark dose (BMD) modeling: current practice, issues, and challenges. Crit Rev Toxicol. 48(5):387–415. doi:10.1080/10408444.2018.1430121.

Hamada C (2018). Statistical analysis for toxicity studies. J Toxicol Pathol. 31(1):15–22. doi:10.1293/tox.2017-0050.

Hernan MA (2012). Beyond exchangeability: the other conditions for causal inference in medical research. Stat Methods Med Res. 21:3–5. doi:10.1177/0962280211398037.

Hernan MA, Robins JM (2006). Estimating causal effects from epidemiological data. J Epidemiol Community Health. 60(7):578–86. doi:10.1136/jech.2004.029496.

Hothorn LA (2016). Statistics in toxicology using R. Boca Raton (FL): Chapman & Hall/CRC Press.

IPCS (1990). Principles for the toxicological assessment of pesticide residues in food. Geneva: World Health Organization, International Programme on Chemical Safety (Environmental Health Criteria, No. 104; http://www.inchem.org/documents/ehc/ehc/ehc104.htm).

IPCS (1994). Assessing human health risks of chemicals: derivation of guidance values for health-based exposure limits. Geneva: World Health Organization, International Programme on Chemical Safety (Environmental Health Criteria, No. 170; http://www.inchem.org/documents/ehc/ehc/ehc170.htm).

IPCS (1999). Principles for the assessment of risks to human health from exposure to chemicals. Geneva: World Health Organization, International

Programme on Chemical Safety (Environmental Health Criteria, No. 210; http://www.inchem.org/documents/ehc/ehc/ehc210.htm).

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

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

IPCS (2007). 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. Geneva: World Health Organization, International Programme on Chemical Safety (Harmonization Project Document, No. 4; https://www.who.int/ipcs/methods/harmonization/areas/cancer/en/).

IPCS (2009). Principles for modelling dose–response for the risk assessment of chemicals. Geneva: World Health Organization, International Programme on Chemical Safety (Environmental Health Criteria, No. 239; http://whqlibdoc.who.int/publications/2009/9789241572392_eng.pdf).

IPCS (2010). Characterization and application of physiologically based pharmacokinetic models in risk assessment. Geneva: World Health Organization, International Programme on Chemical Safety (Harmonization Project Document No. 9; http://www.inchem.org/documents/harmproj/harmproj/harmproj9.pdf).

IPCS (2012). Guidance for immunotoxicity risk assessment for chemicals. Geneva: World Health Organization, International Programme on Chemical Safety (Harmonization Project Document, No. 10; http://www.inchem.org/documents/harmproj/harmproj/harmproj10.pdf).

IPCS (2018). Guidance document on evaluating and expressing uncertainty in hazard characterization, second edition. Geneva: World Health Organization, International Programme on Chemical Safety (Harmonization Project Document, No. 11; https://www.who.int/ipcs/methods/harmonization/areas/hazard_assessment/en/).

Judson RS, Houck KA, Kavlock RJ, Knudsen TB, Martin MT, Mortensen HM et al. (2010). In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project. Environ Health Perspect. 118(4):485–92. doi:10.1289/ehp.0901392.

Kang SH, Kodell RL, Chen JJ (2000). Incorporating model uncertainties along with data uncertainties in microbial risk assessment. Regul Toxicol Pharmacol. 32(1):68–72. doi:10.1006/rtph.2000.1404.

Kavlock RJ, Allen BC, Faustman EM, Kimmel CA (1995). Dose–response assessments for developmental toxicity. IV. Benchmark doses for fetal weight changes. Fundam Appl Toxicol. 26(2):211–22. doi:10.1006/faat.1995.1092.

Li HH, Chen R, Hyduke DR, Williams A, Frotschl R, Ellinger-Ziegelbauer H et al. (2017). Development and validation of a high-throughput transcriptomic biomarker to address 21st century genetic toxicology needs. Proc Natl Acad Sci U S A. 114(51):E10881–E10889. doi:10.1073/pnas.1714109114.

Maurer JM, Schellekens RC, Van Rieke HM, Wanke C, Iordanov V, Stellaard F et al. (2015). Gastrointestinal pH and transit time profiling in healthy volunteers using the IntelliCap system confirms ileo-colonic release of ColoPulse tablets. PLoS One. 10(7):e0129076. doi:10.1371/journal.pone.0129076.

Mbah AK, Hamisu I, Naik E, Salihu HM (2014). Estimating benchmark exposure for air particulate matter using latent class models. Risk Anal. 34(11):2053–62. doi:10.1111/risa.12256.

Meek ME, Renwick A, Ohanian E, Dourson M, Lake B, Naumann BD et al. (2002). Guidelines for application of chemical-specific adjustment factors in dose/concentration–response assessment. Toxicology. 181–182:115–20. doi:10.1016/S0300-483X(02)00265-2.

Meek ME, Boobis A, Cote I, Dellarco V, Fotakis G, Munn S et al. (2014). 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. doi:10.1002/jat.2949.

Molenberghs G, Ryan LM (1999). An exponential family model for clustered multivariate binary data. Environmetrics. 10(3):279–300. doi:10.1002/(SICI)1099-095X(199905/06)10:3<279::AID-ENV352>3.0.CO;2-X.

Nair AB, Jacob S (2016). A simple practice guide for dose conversion between animals and human. J Basic Clin Pharm. 7(2):27–31. doi:10.4103/0976-0105.177703.

OECD (2010). Guidance for the derivation of an acute reference dose. Paris: Organisation for Economic Co-operation and Development (Series on Testing and Assessment, No. 124; ENV/JM/MONO(2010)15; http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2010)15&doclanguage=en).

OECD (2012). Statistical and dose response analysis, including benchmark dose and linear extrapolation, NOAELs and NOELs, LOAELs and LOELs. In: Guidance document 116 on the conduct and design of chronic toxicity and carcinogenicity studies, supporting Test Guidelines 451, 452 and 453, second edition. Paris: Organisation for Economic Co-operation and Development; 114–142 (Series on Testing and Assessment, No. 116; https://www.oecd-ilibrary.org/docserver/9789264221475-en.pdf?expires=1596994146&id=id&accname=guest&checksum=0FFC91C7D79025B7C43F075DD2264F6D).

Roca-Saavedra P, Mendez-Vilabrille V, Miranda JM, Nebot C, Cardelle-Cobas A, Franco CM et al. (2018). Food additives, contaminants and other minor

components: effects on human gut microbiota – a review. J Physiol Biochem. 74:69–83. doi:10.1007/s13105-017-0564-2.

Sakia RM (1992). The Box-Cox transformation technique: a review. Statistician. 41(2):169–78. doi:10.2307/2348250.

Sand S, Portier CJ, Krewski D (2011). A signal-to-noise crossover dose as the point of departure for health risk assessment. Environ Health Perspect. 119(12):1766–74. doi:10.1289/ehp.1003327.

Sand S, Parham F, Portier CJ, Tice RR, Krewski D (2017). Comparison of points of departure for health risk assessment based on high-throughput screening data. Environ Health Perspect. 125(4):623–33. doi:10.1289/EHP 408.

Shao K, Gift JS (2014). Model uncertainty and Bayesian model averaged benchmark dose estimation for continuous data. Risk Anal. 34(1):101–20. doi:10.1111/risa.12078.

Shao K, Shapiro AJ (2018). A web-based system for Bayesian benchmark dose estimation. Environ Health Perspect. 126(1):017002-1–017002-14. doi:10.1289/EHP1289.

Slob W (1999). Thresholds in toxicology and risk assessment. Int J Toxicol. 18(4):259–68. doi:10.1080/109158199225413.

Slob W (2002). Dose–response modeling of continuous endpoints. Toxicol Sci. 66(2):298–312. doi:10.1093/toxsci/66.2.298.

Slob W (2017). A general theory of effect size, and its consequences for defining the benchmark response (BMR) for continuous endpoints. Crit Rev Toxicol. 47(4):342–51. doi:10.1080/10408444.2016.1241756.

Slob W, Setzer RW (2014). Shape and steepness of toxicological dose–response relationships of continuous endpoints. Crit Rev Toxicol. 44(3):270–97. doi:10.3109/10408444.2013.853726.

Solecki R, Davies L, Dellarco V, Dewhurst I, Van Raaij M, Tritscher A (2005). Guidance on setting of acute reference dose (ARfD) for pesticides. Food Chem Toxicol. 43(11):1569–93. doi:10.1016/j.fct.2005.04.005.

Solecki R, Moeller T, Herrmann M, Stein B (2010). A retrospective analysis of acute reference doses for pesticides evaluated in the European Union. Crit Rev Toxicol. 40(1):24–34. doi:10.3109/10408440903436467.

Thomas RS, Bahadori T, Buckle TJ, Cowden J, Deisenroth C, Dionisio KL et al. (2019). The next generation blueprint of computational toxicology at the U.S. Environmental Protection Agency. Toxicol Sci. 169(2):317–32. doi:10.1093/ toxsci/kfz058.

Tice RR, Austin CP, Kavlock RJ, Bucher JR (2013). Improving the human hazard characterization of chemicals: a Tox21 update. Environ Health Perspect. 121(7):756–65. doi:10.1289/ehp.1205784.

Tukey JW, Ciminera JL, Heyse JF (1985). Testing the statistical certainty of a response to increasing doses of a drug. Biometrics. 41(1):295–301. doi: 10.2307/2530666.

USEPA (2011). Recommended use of body weight$^{3/4}$ as the default method in derivation of the oral reference dose. Washington (DC): United States Environmental Protection Agency (EPA/100/R11/0001 Final; https://www.epa.gov/sites/production/files/2013-09/documents/recommended-use-of-bw34.pdf).

USEPA (2012). Benchmark dose technical guidance. Washington (DC): United States Environmental Protection Agency (EPA/100/R-12/001; https://www.epa.gov/sites/production/files/2015-01/documents/benchmark_dose_guidance.pdf).

USEPA (2014). Guidance for applying quantitative data to develop data-derived extrapolation factors for interspecies and intraspecies extrapolation. Washington (DC): United States Environmental Protection Agency (EPA/100/R-14/002F; https://archive.epa.gov/raf/web/pdf/ddef-final.pdf).

USEPA (2020). Benchmark Dose Software (BMDS). Version 3.2. User guide. Washington (DC): United States Environmental Protection Agency (EPA/600/R-20/216; https://www.epa.gov/sites/production/files/2020-09/documents/bmds_3.2_user_guide.pdf).

VICH (2019). Studies to evaluate the safety of residues of veterinary drugs in human food: general approach to establish a microbiological ADI. Brussels: International Cooperation on Harmonisation of Technical Requirements for Registration of Veterinary Medicinal Products (VICH GL36 (R2); https://vichsec.org/en/guidelines/pharmaceuticals/pharma-safety/antimicrobial-safety.html).

Wheeler MW, Bailer AJ (2007). Properties of model-averaged BMDLs: a study of model averaging in dichotomous response risk estimation. Risk Anal. 27(3):659–70. doi:10.1111/j.1539-6924.2007.00920.x.

Wheeler MW, Shao K, Bailer AJ (2015). Quantile benchmark dose estimation for continuous endpoints. Environmetrics. 26(5):363–72. doi:10.1002/env.2342.

Wheeler MW, Bailer AJ, Cole T, Park RM, Shao K (2017). Bayesian quantile impairment threshold benchmark dose estimation for continuous endpoints. Risk Anal. 37(11):2107–18. doi:10.1111/risa.12762.

WHO (1967). Procedures for investigating intentional and unintentional food additives. Report of a WHO scientific group. Geneva: World Health Organization (WHO Technical Report Series, No. 348; https://apps.who.int/iris/bitstream/handle/10665/40637/WHO_TRS_348.pdf?sequence=1&isAllowed=y).

WHO (2012). Pesticide residues in food – 2011. Evaluations 2011, Part II – Toxicological. Geneva: World Health Organization, Joint FAO/WHO Meeting on Pesticide Residues (http://www.inchem.org/documents/jmpr/jmpmono/v2011pr01.pdf).

WHO (2013). Pesticide residues in food – 2012. Evaluations 2012, Part II – Toxicological. Geneva: World Health Organization, Joint FAO/WHO Meeting on Pesticide Residues (http://www.inchem.org/documents/jmpr/jmpmono/v2012pr01.pdf).

WHO (2015). Pesticide residues in food: guidance document for WHO monographers and reviewers. Geneva: World Health Organization, WHO Core Assessment Group on Pesticide Residues (https://www.who.int/foodsafety/publications/JMPR-guidance-document/en/).

WHO (2017). Guidance document for the establishment of acute reference dose (ARfD) for veterinary drug residues in food. Geneva: World Health Organization (https://www.who.int/foodsafety/chem/jecfa/Guidance-document-ARfD-2017.pdf?ua=1).

Yeh FS, Yu MC, Mo CC, Luo S, Tong MJ, Henderson BE (1989). Hepatitis B virus, aflatoxins, and hepatocellular carcinoma in southern Guangxi, China. Cancer Res. 49(9):2506–9.

Appendix 5.1: Dose–response modelling for risk assessment

A5.1.1 Statistical modelling

Experimental dose–response data are expected to exhibit random variation around the responses. This means that slight differences in the responses are expected if the same experiment is repeated several times at the same dose. Nevertheless, we expect the emergence of similar behaviour across experiments. We assume that this behaviour can be described by dose–response functions. Mathematical models are needed that capture the dose–response relationship as well as the random part of the response. This leads to statistical models. Statistical modelling usually includes the assumption of a statistical distribution about the errors or random components of the model (e.g. normal and binomial distributions for continuous and quantal responses, respectively). There are several functions that are typically used to describe dose–response relationships, such as the Hill, logistic and probit functions, among others. There is a consensus on most of the suitable functions for DRM (e.g. the Hill function), but not on all functions (e.g. the use of polynomials).

The properties of statistical model estimates (also known as asymptotic behaviour) are studied for large samples. This is done by computing expectations, variances and covariances for model-derived estimates (e.g. the BMD is one such model-derived estimate). Uncertainty is introduced into estimation because of the observed randomness in the data; this implies that the estimates are not deterministic, but would change slightly from data set to data set (e.g. the number of observations or the number of groups). The properties of this estimation are studied asymptotically by imagining that some aspect of the data increases without bound.

A5.1.2 Model components

The model represents the data-generating mechanism and is defined by its components. Each component in the model must be correctly specified to sufficiently describe the data-generating mechanism. Model selection, diagnosis and discrimination are important steps in the model-building process. In this process, the first important step is to formulate knowledge about the data-generating mechanism and to express the real observed phenomenon,

which describes the distribution of the response, as a mathematical model. In many cases, the model description takes the form of a model with additive error structure that has two components:

response = mean + error

where the mean term represents the deterministic component and the error term specifies a stochastic component. If an additive error model is not appropriate, then one may use a multiplicative model to describe the data-generating mechanism. Models for such data require a more elaborate formulation involving statistical distributions. However, for illustrative purposes, we consider only the additive formulation, noting that the descriptions could easily be extended to other model formulations describing more complex data-generating mechanisms.

From this perspective, the model components can be defined by:

- the *mean* term, describing the tendency of the response variable(s) to vary with the predictor/covariate(s) in a systematic fashion; and
- the *error* term, denoting the statistical variation around the mean term, which includes distributional assumptions of the response variable(s) that we are modelling (it could be considered univariate or multivariate), as well as assumptions related to its variability around the mean term. Embedded in the error term is the variability of the response around the mean structure, which could be a constant regardless of the values of the covariate(s) or could also be a function of them.

The distributional assumptions and each of the model components described above are essential when performing inference using a selected model.

A5.1.2.1 Describing the mean

As described in the main text of Chapter 5, one can assume a family of models to estimate the mean, μ, of the response as a function of dose. In what follows, μ(dose) defines the dose–response curve. Table A5.1 gives five functions that can be considered when modelling dose–response data and can be used in the general family defined in the main text. For many of the simpler models, $d = 1$ in Table A5.1 (see also Table A5.2 below). These parameters are

included for a consistency across all models. The unknown parameters are represented in the vector $\Theta = (a, b, d, \theta)$.

Table A5.1. Five models that are included in the general family of models[a]

Functional family	Functional form
Gamma model function	$F(b,\text{dose}^d,\theta) = \frac{\int_0^{b\cdot\text{dose}^d} x^{\theta-1}\cdot e^{-x}dx}{\Gamma(\theta)}$
Lomax model function	$F(b,\text{dose}^d,\theta) = 1 - \left(\frac{b}{b + \text{dose}^d}\right)^{\theta}$
Logistic model function	$F(a,b,\text{dose}^d) = \frac{1}{e^{-a-b\cdot\text{dose}^d}}$
Probit model function	$F(a,b,\text{dose}^d) = \Phi(a + b\cdot\text{dose}^d)$
Monotonic polynomial functions[b]	$F(a,b,\text{dose}^d) = a + b\cdot\text{dose}^d$

[a] Here $\Gamma(\theta)$ is the gamma function and $\Phi(x)$ is the standard normal cumulative distribution function.

[b] Any monotonic polynomial function can be thought of as being in this suite.

These functions are by no means exhaustive, but represent plausible smooth shapes that are useful dose–response curves. These functions can generate all the model forms found in standard software packages, such as PROAST and BMDS. For example, noting that $\Gamma(\theta) = \int_0^{\infty} x^{\theta-1}e^{-x}dx$ and $\Gamma(\theta) = (\theta - 1)!$ for integral values θ, using the gamma family one can derive the exponential family of models for continuous data and the Weibull model for quantal outcomes by setting $\theta = 1$. That is,

$$F(b,\text{dose}^d,\theta) = \frac{\int_0^{b\cdot\text{dose}^d} x^{\theta-1}\cdot e^{-x}dx}{\Gamma(\theta)}$$

$$= 1 - \exp\left(b\cdot\text{dose}^d\right)$$

Using the general family of models $\mu(\text{dose}) = a \cdot \{1 + (c - 1) F(b,\text{dose}^d)\}$, one arrives at the Exponential 5 model by direct

substitution of the above expression and the Weibull model by setting $c = 1/a$. Other common model forms are given in Table A5.2.

Table A5.2. Common choices for the model parameters θ, *d* and *a* for DRM

Model	Distribution used
Continuous response	
Exponential 5 model	Gamma with θ = 1
Hill model	Lomax with θ = 1
Quantal response (setting *c* = 1/*a*)	
Log-logistic	Lomax with θ = 1
Log-probit	Probit with dose = log(dose) *d* = 1
Gamma	Gamma with *d* = 1
Logistic	Logistic with *d* = 1
Probit	Probit with *d* = 1 *a* = 1

A5.1.2.2 Describing the variability

Distributional assumptions describe the nature of the scatter – that is, variability of the response – that may be observed in an experiment. For example, a normal distribution assumes symmetric scatter around the mean and has relatively light tails when compared with other distributions. Alternatively, the log-normal distribution is an asymmetric distribution with heavy right tails. The former distribution is applicable for many analyses where the variability is symmetric around the mean, whereas the latter is useful in situations where observations exhibit strong right skew. Once the distributional assumptions are made and the form of the mean function is chosen, one can then fit the model to the given data.

A5.1.3 Statistical model fitting

For our purposes, statistical model fitting refers to estimating a BMD from the data. There is a vast literature on statistical model fitting for any purpose, and the review below is by no means comprehensive. For textbook treatments on frequentist maximum

likelihood estimation, see Millar (2011); for Bayesian inference, see Gelman et al. (2013). In our case, we restrict our discussion to DRM for BMD estimation using a dose–response function. As shorthand, we summarize all unknowns for a specific dose–response function using the vector Θ.

A5.1.3.1 Frequentist

Frequentist parametric methodologies assume that all information relevant for inference is encapsulated through a statistical distribution and the dose–response function; they further assume that no information, other than the observed data, will be included in formal inference. For example, when studying the effects of lead on cognitive function, one may assume that intelligence quotient (IQ) is distributed normally, which is the only distributional assumption used in the analysis. In this case, the mean and variance of the distribution specify all information about the scatter of IQs in a population, and one would be interested in estimating the dose–response relationship and standard deviation that best describe the observed data, which can be used for inference.

Assume that one observes the data vector *Y*. To perform frequentist inference, one constructs a likelihood, L(Θ, *Y*), which is the product of the observed data distribution with vector of the parameters Θ and finds the values of the mean and standard deviation that are most probable given the data. That is, one finds the value $\widehat{\Theta}$ that maximizes L(Θ, *Y*), which is known as the maximum likelihood estimate. Continuing the IQ example with the assumption of normality, assume that one observes three individuals with IQs = 95, 110 and 102; then

$$L(\theta = \{\mu,\sigma\}, Y = \{95,110,102\}) =$$

$$\frac{1}{(2\pi\sigma^2)^{3/2}} \exp(-{}^{1}/_{2\sigma^2} [\{95-\mu\}^2 + \{110-\mu\}^2 + \{102-\mu\}^2])$$

which is the product of three normal distributions with observations at 95, 110 and 102. In this example, the maximum likelihood estimation of μ is the sample mean, and σ is proportional to the sample standard deviation. Some methodologies, such as least squared estimation, can be framed as maximum likelihood estimation – that is, inference for least squares models is equivalent to

maximizing a normal likelihood – and we discuss only maximum likelihood estimation in what follows.

A5.1.3.2 Bayesian

Like frequentist inference, Bayesian inference encapsulates information into the analysis through the likelihood function $L(\Theta, Y)$, but assumes additional information on Θ through a prior probability distribution, $p(\Theta)$, on this parameter. This information may take the form of empirical information from previous studies, expert opinion on plausible values of Θ that are elicited before the analysis, as well as flat uniform priors (these are sometimes referred to as non-informative priors, depending on the limits of integration). To illustrate this point, continuing the IQ example above, it is reasonable to assume that μ should be near the population average of 100; however, as there may be other effects due to lead, one places a priori a variance of 25 on this value, which is reasonably diffuse given that the population average is 100. Assuming an approximate normal distribution over μ, one encapsulates the prior information using this distribution.

In Bayesian analysis, inference takes place using Bayes' rule, which is a conditional probability statement relating the observed data and the prior information to posterior inference. That is, the posterior probability of Θ given the data vector Y is expressed as

$$P(\Theta|Y) = \frac{L(\Theta,Y)P(\Theta)}{\int L(\Theta,Y)P(\Theta)d\Theta}$$

As the posterior is often not analytically tractable, inference on this quantity is done through approximation, which is traditionally done using Markov Chain Monte Carlo, but can be done using simpler approximations, such as maximum a posteriori estimation. The latter relies on optimization and asymptotic approximations to the posterior.

The choice of the prior impacts the posterior distribution; however, inference between the Bayesian approach and the frequentist approach is often qualitatively similar for large samples.

Bayesian analyses require prior assumptions on the dose–response function in the form of prior distributions on the parameters. Such priors can be constructed using historical data or could be based on expert knowledge. One can also employ priors on the parameters

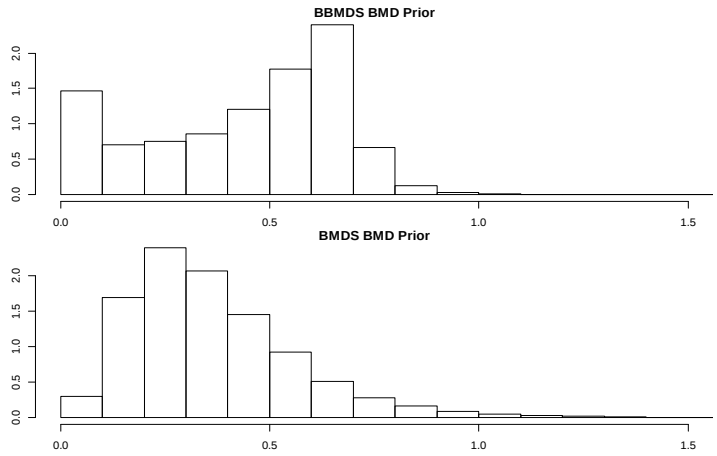
that will lead to no or the lowest bias on the final outcome of the value of the BMD. Such priors are commonly called objective or uninformative priors. Note that objective and uninformative are terms that are often used synonymously in the literature. As priors may influence inference, they should be carefully considered. Specifically, for BMD estimation, the choice of the prior distributions on the model parameters Θ results in a second prior distribution on the BMD. So, one is mainly concerned about the prior distribution of the BMD when choosing the prior distributions on the parameters Θ, because the prior distribution of the BMD depends on the choices one makes for the prior distributions of the model parameters. The prior distribution of the BMD is the quantity of interest. This is especially true when the model is a mathematical model and has no biological basis (i.e. the model is empirical). Consequently, if using priors outside of the default of either PROAST or BMDS, some analysis should be conducted on the priors' effect on the BMD.

As an example, one can compare the priors used in BMDS version 3.1 against the priors used in the Bayesian Benchmark Dose System (BBMDS; Shao & Shapiro, 2018) for the Weibull model. Here, an initial assumption may be to use a flat uniform prior, such as a prior that assumes that all values of the parameter are equally like a flat uniform prior, which is the case of the BBMDS as of this writing; however, this may imply a prior on the BMD that is not realistic. To illustrate this, Fig. A5.1 gives the BBMDS and BMDS priors on the BMD for a 10% response (BMD_{10}) as a percentage of the maximum dose tested (MD). In this figure, for the BBMDS, the default flat uniform prior is used for all parameters, and *d*, the steepness term above, is given a flat uniform prior over the interval (0,15), which represents an option in the BBMDS and the unconstrained default position of Chapter 5. The choice of a flat uniform prior on the parameters results in a highly biased prior distribution for the BMD, as is shown in the upper graph of Fig. A5.1. It displays the probability density function/probability mass function of the prior distribution on the BMD that results from the choice of flat prior distributions from the parameters.

Many toxicology experiments test doses at geometric titrations of the MD (e.g. MD/2, MD/4). Fig. A5.1 highlights why uniform priors may unintentionally bias the analysis to arbitrarily high or low doses. The upper graph shows that the priors for the BBMDS place a high prior probability on doses greater than the MD/2 as well as very low doses, which are often well outside of the doses tested. In fact,

approximately 75% of the prior probability for the BMD_{10} is placed on values outside of the range of 10–50% of the MD. In contrast, the BMDS version 3.1 software priors imply a prior distribution of the BMD (shown in Fig. A5.1 lower graph) that places approximately 76% of the prior probability within this range (10–50% of the MD), which is consistent with a large body of literature linking the observed POD to titrations of the MD (e.g. see Krewski et al., 1993). In fact, if one were to make the priors even more diffuse (i.e. raise the upper limit from 15 to a larger number), the problem becomes exacerbated. As this value gets arbitrarily large, the prior probability for the BMD_{10} concentrates on the MD. Note that biases due to priors typically occur only when there are limited data. For large samples, the data overwhelm the priors for both the BBMDS and BMDS version 3.1.

Fig. A5.1. Induced default prior for the BMD_{10} for the BBMDS (top) and the default settings for BMDS version 3.1 (bottom) as a fraction of the MD based upon the actual priors defined for Θ[a]



[a] For example, the parameters *a*, *b* and *d* place equal prior probability on all outcomes (within a defined range) for the parameters. This results in an implied density over the BMD shown in the top pane. This prior density is plotted as a relative frequency of occurrence against the percentage of MD and does not result in an equal probability of BMD values.

A5.1.3.3 Model averaging

Once the model components have been defined, a model selection phase is undertaken, and several models can be fitted to the

data at hand. Common strategies involve comparing the fitted model to a full model to determine if the model fits the data, comparing the AIC (Akaike, 1973) to the other fitted models, examining the visual fit of the model and looking at goodness-of-fit statistics such as the Pearson χ^2 statistic to determine the appropriateness of the "best model" to use in a risk assessment (EFSA, 2017). It is known that this process might suffer from model selection bias and that any single estimate from an individual model ignores model uncertainty; this is true with any model selection strategy. For example, choosing the model with the best *P*-value, or AIC, which has been the default procedure for risk assessment, will lead to model uncertainty that is not encapsulated in the final estimate.

In BMD estimation for risk assessment, the recommended method is model averaging (Kass & Raftery, 1995), and the previous methods (e.g. selecting the best model) should be used only when model averaging is not available. Hoeting et al. (1999) wrote a tutorial on Bayesian model averaging and showed how the uncertainty in model selection leads to overconfident inferences when one model is chosen.

Several frequentist approaches for model averaging have been presented in the literature. Buckland, Burnham & Augustin (1997) proposed model averaging to deal with model selection bias in the case of regression models. Hjort & Claeskens (2003) studied these frequentist approaches, giving several asymptotic results on such estimators.

The model-averaging approach assumes a model from the general model family (i.e. the *F*() defined above) and considers the model *m*, being one model out of the total of *R* models, described in the *F*s above (see Table A5.1). The AIC is calculated for each model, and the difference between the AIC for each model and the minimum AIC over all possible models is computed:

$$\Delta_m = \mathrm{AIC}_m - \mathrm{AIC}_{\min}$$

To calculate the new average model over *R* models, the weights are constructed using the following formula for weight w_m:

$$w_m = \frac{e^{-\frac{\Delta_m}{2}}}{\sum_{r=1}^{R} e^{-\frac{\Delta_r}{2}}}$$

Burnham & Anderson (2004) argued that where a model-averaged estimator can be used, it often has reduced bias and better precision compared with the best model. For Bayesian models, weights are computed similarly, but can be interpreted as the posterior probability that the given model is the correct model. For more information on Bayesian model averaging, see Hoeting et al. (1999).

Currently, the BMDS version 3.1 system implements Bayesian model averaging as described in Hoeting et al. (1999). PROAST implements frequentist model averaging using the AIC as the basis for the weighting scheme. Although both approaches may give different weights, final inference on the BMD is typically qualitatively the same. In practice, either approach can be used, and the focus on the analysis should be the quantities of interest, which are the BMD and BMDL.

The frequentist and Bayesian model-averaging approaches result in averages of probability distributions for the BMD. This is distinguished from alternatives that involve taking averages of the BMD values themselves under different models. Averages taken over BMD or BMDL values will be more sensitive to high than to low model results, and such approaches may be at risk of overestimating the true BMD. For example, the PROAST frequentist model averaging, which uses methods similar to those of Wheeler & Bailer (2007), uses averages of distributions rather than averages of BMD values.

A5.1.4 Dose–response modelling and weight of the evidence

Dose–response models are quantitative expressions of relationships in pharmacology and toxicology that are thought to encapsulate causation (e.g. the effect of an exposure to a hazard). However, even when it is expressed mathematically, the validity of the expression of causality ultimately depends on a judgement that is not mathematical (Illari & Russo, 2015). In the fields of medicine and physiology, perhaps the best evidence of this comes from the fact that when Hill (1965) gave his widely known lecture on causality before a group of statisticians, he used no mathematical equations whatsoever.

Weight-of-evidence approaches have been used for dose–response modelling by both JECFA (e.g. for lead: WHO, 2000, 2011) and JMPR, as well as elsewhere (e.g. Morgan, Henrion & Morris, 1980; Evans et al., 1994). Using weight of the evidence to address

dose–response model uncertainties is largely the same as when Bayesian methods are used. There is still a need to identify a finite set of alternative models or hypotheses, and the models are still either directly fit to data or designed to be consistent with the empirical record. Furthermore, both approaches utilize expert opinion, and at the end of the process, probabilities are assigned to each alternative model so that they all add up to 1. There are, however, important differences:

1) Bayesian methodology uses expert opinion prior to curve fitting and then "updates" the probabilities initially assigned by the experts as part of the curve-fitting process to yield the final model probabilities. In contrast, a weight-of-evidence approach does not assign model probabilities until after curve fitting has taken place; experts may use information about how well each model describes the data, but they also use other theoretical and experiential criteria as well. Because the Bayesian approach alters expert opinion after it is expressed, it has the potential of yielding final model probabilities that contradict what experts believe.

2) Because it is amenable to automation, Bayesian methodology is far more reproducible than a methodology that depends solely on expert opinion. Model probabilities assigned by experts may vary among experts or even for a single expert over time. That fact perhaps makes the Bayesian methodology preferable when a standardized approach is desirable and there is no strongly held expert opinion.

3) Because it is thought of as a mathematical exercise, calculating Bayesian probabilities requires the use of analytically tractable models. If this is not possible, consulting expert opinion may be the only option.

Although the Hill criteria have been used for other purposes (e.g. Suter & Cormier, 2011), a formal process of the same ilk as the Hill criteria (Hill, 1965) is not typically implemented for quantitative dose–response modelling. A process for weighing evidence could temper differences of opinion among experts regarding dose–response model form without eliminating expert opinion altogether. Assigning probabilities by committee would also help. In place of the "associations" that concerned Hill (1965), one or more numerical goodness-of-fit measures could be used to argue for or against

specific models. The Hill criterion that is directly relevant to dose–response modelling is the requirement for a "biological gradient". Quite simply, a dose–response model ought to look like what a dose–response relationship is supposed to look like. That criterion could perhaps be subdivided into theoretical and experiential components. As an example of the former, a dose–response relationship cannot be supralinear; this is argued based on the fact that when the dose approaches zero, the supralinear dose–response relationship will violate the generally accepted biochemical law of mass action (Tallarida, Laskin & Jacob, 1976). An experiential argument would reflect the experience of toxicologists with other analogous dose–response relationships.

A5.1.5 Examples

A5.1.5.1 Quantal data

In the case of quantal data, PROAST and BMDS are most alike with respect to the suite of models applied to such data and default statistical assumptions (e.g. binomial distributions). However, although both software platforms implement model averaging for quantal end-points, the methods they use differ: PROAST uses a frequentist approach to model averaging, whereas BMDS uses Bayesian model averaging with informative priors.

Consider the data set for hepatocellular hyperplasia given in Table A5.3.

Table A5.3. Example of a quantal data set

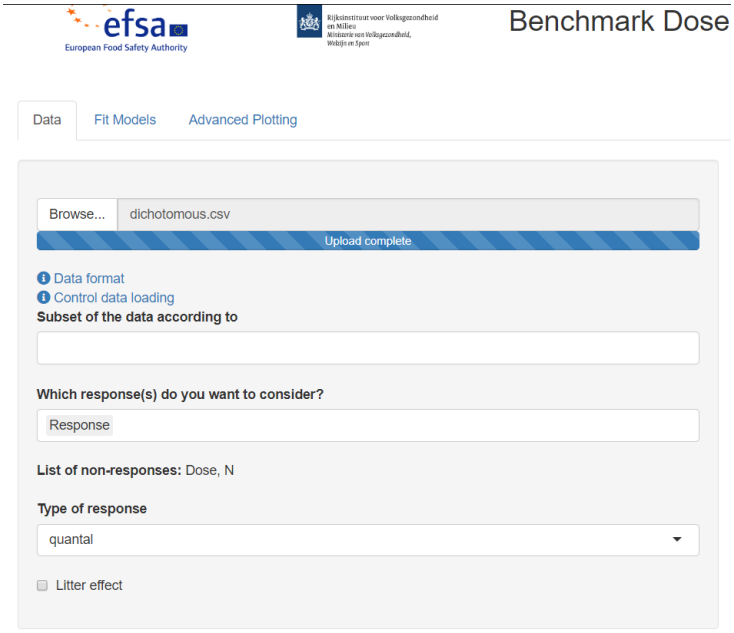
Dose (mg/kg body weight per day)	Number of animals with hepatocellular hyperplasia	Number of animals per dose group	Sex
0	1	50	Male
100	2	50	Male
200	5	50	Male
400	30	50	Male
800	48	50	Male

Using PROAST (desktop version, https://www.rivm.nl/en/proast; PROASTweb, https://proastweb.rivm.nl/; or EFSA Open Analytics version, https://shiny-efsa.openanalytics.eu/app/bmd),

model averaging utilizes information from all fitted models to derive the 95% lower confidence limit on the BMD (the BMDL).

In this example, EFSA's PROAST web tool is used. This requires importing data as a .csv file. After loading the data in the Data tab, users can choose the end-point of response and the type of data for modelling (Fig. A5.2), a quantal response in this case. Using the Fit Models tab, additional modelling details (e.g. Max. difference in AIC for acceptance, Value for the BMR, BMD confidence level) can be set (Fig. A5.3). The result of fitting all available quantal models is presented in Table A5.4 and Fig. A5.4, together with the statistical details and the visual fits.

Fig. A5.2. Data tab in EFSA PROAST web tool



When using frequentist model averaging for quantal data, available in PROAST, a bootstrap approach with multiple iterations is used. Users can specify the number of bootstrap iterations on the Fit Models tab (200 iterations were selected by default for this example). PROAST's model averaging averages the results of the models using weights based on the individual model AIC values (Table A5.5). In this case, the log-logistic, log-probit, gamma

Fig. A5.3. Fit Models tab in EFSA PROAST web tool

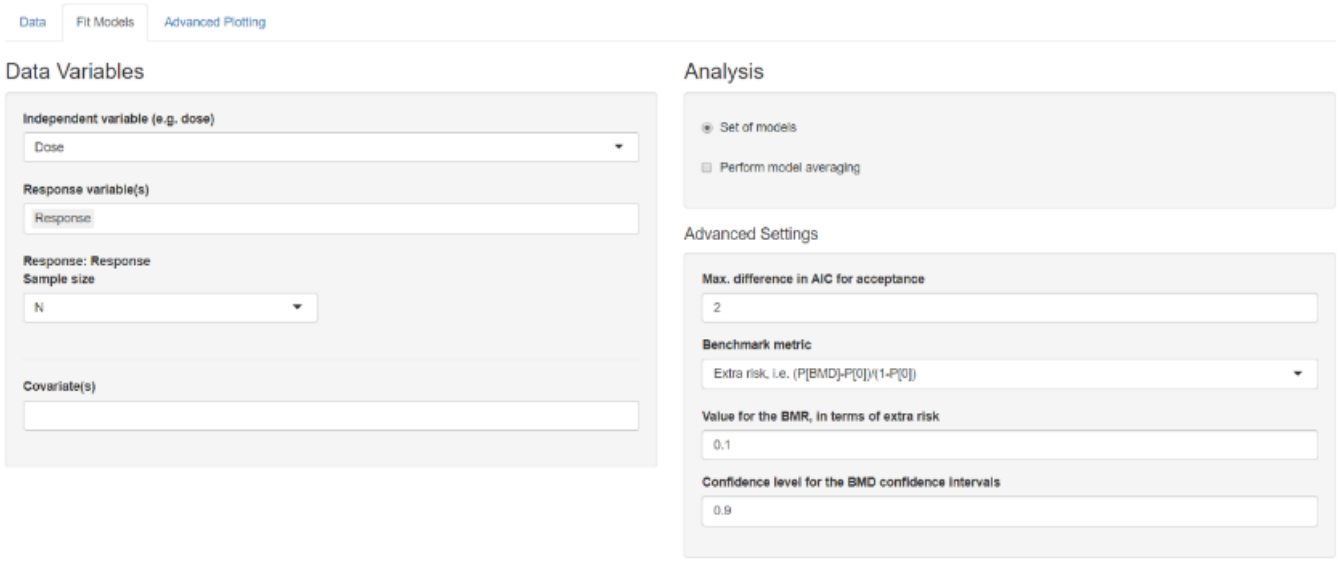


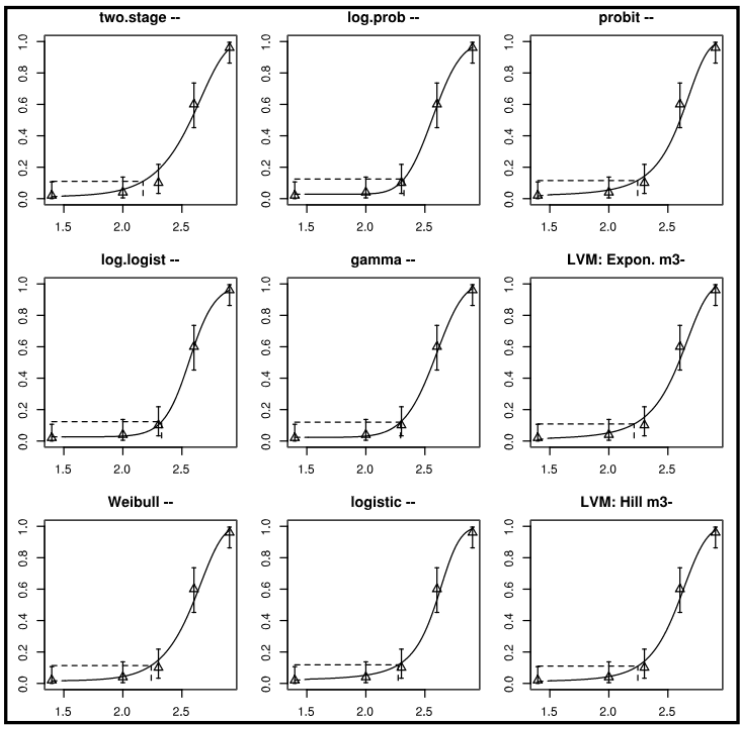
Table A5.4. PROAST quantal model fits to hepatocellular hyperplasia data set

model	No.par	loglik	AIC	accepted	BMDL	BMDU	BMD	conv
null	1	-160.91	323.82		NA	NA	NA	NA
full	5	-71.6	153.2		NA	NA	NA	NA
two.stage	3	-73.63	153.26	no	NA	NA	148	yes
log.logist	3	-71.7	149.4	yes	169	261	213	yes
Weibull	3	-72.91	151.82	no	NA	NA	174	yes
log.prob	3	-71.79	149.58	yes	169	253	210	yes
gamma	3	-72.05	150.1	yes	153	243	197	yes
logistic	2	-73.08	150.16	yes	156	221	188	yes
probit	2	-73.77	151.54	no	NA	NA	175	yes
LVM: Expon. m3-	3	-73.43	152.86	no	NA	NA	164	yes
LVM: Hill m3-	3	-72.68	151.36	yes	138	215	176	yes

and logistic models and a latent variable Hill model with three parameters (LVM-Hill3) consume approximately 80% of the weight used for calculating the model-averaged BMD as well as the BMDL and BMDU.

From the full set of bootstrap BMD estimates, BMDL and BMDU values are obtained as the 5th and 95th percentile values of the BMD posterior distribution. The averaged BMD interval (BMDL–BMDU) is calculated as 152–243 mg/kg body weight per day. Given that the model-averaging approach implemented in PROAST uses random bootstrap iterations, this result (i.e. BMD interval) may differ slightly when different users analyse the same data set. Increasing the number of iterations should produce results that are closer to unity, but small (perhaps inconsequential) differences will always persist. However, increasing the number of iterations will also increase the analysis time. Therefore, it is important that users of this method determine, prior to the analysis, the most efficient trade-off between number of iterations (more precise estimation of the BMD interval) and analysis time. Another option is to run the analysis with different numbers of iterations as a sensitivity analysis to objectively report the effect that this choice has on the final results.

Fig. A5.4. PROAST quantal visual model fits to hepatocellular hyperplasia data set[a]



[a] Given are the modelled dose–response curves for the nine applied models (solid line) and the BMR/BMD (dotted line). Here, all models are considered for the analysis. Models showing poor fit (e.g. *P*-values < 0.1) are effectively removed by having low weights.

Table A5.5. Model weights based on AIC used in PROAST model-averaging analysis of hepatocellular hyperplasia

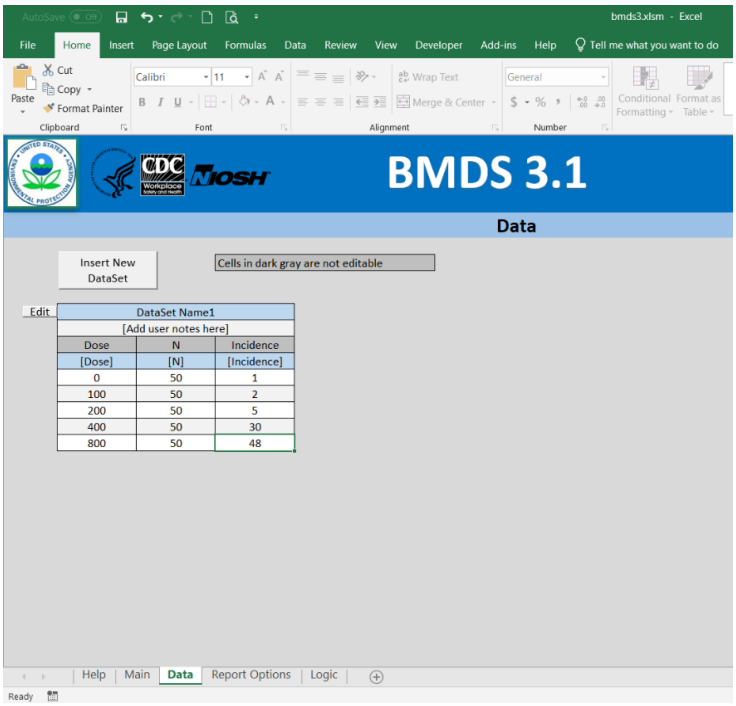
Weights for Model Averaging

two.stage	log.logist	Weibull	log.prob	gamma	logistic	probit	EXP	HILL
0.0313	0.2155	0.0642	0.1969	0.1518	0.1473	0.0739	0.0382	0.0809

When performing BMD analyses using the USEPA's BMDS (available at https://www.epa.gov/bmds), users can utilize a model-averaging approach for quantal end-points. The version of BMDS used for this analysis (version 3.1) is available as a desktop Excel

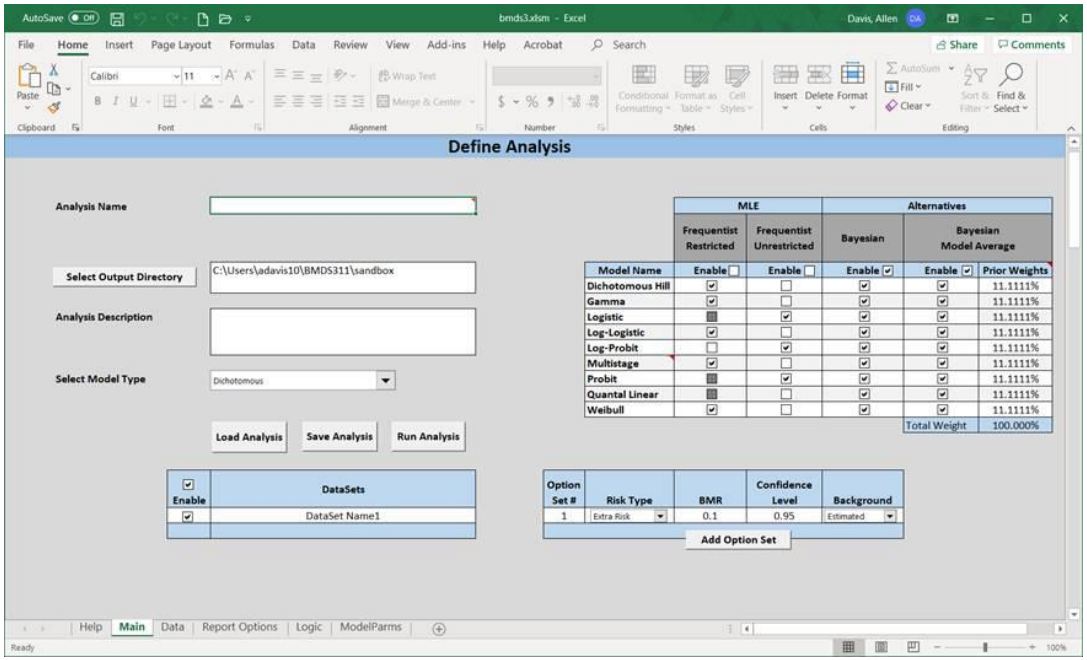
workbook. Users can easily enter the data they are modelling on the Data tab (Fig. A5.5) and then parameterize their analysis to their needs on the Main tab (Fig. A5.6).

Fig. A5.5. Data entry in BMDS version 3.1



When using BMDS's Bayesian model-averaging approach to model the hepatocellular hyperplasia data in Table A5.3, the BMD (BMDL–BMDU) is 184 mg/kg body weight per day (143–233 mg/kg body weight per day). Remember that the Bayesian method used in BMDS uses prior information on the distribution of the BMD and that prior information along with the Laplace approximation determines the posterior distribution of the BMD (see main text of Chapter 5 and Hoeting et al., 1999). Hence, although the model-averaging results are very similar between PROAST and BMDS, even though the two software platforms use different averaging approaches, the weights of the frequentist approach in PROAST are quite different from the posterior probabilities estimated in BMDS. Users can generally expect results to be similar between the two software platforms when

Fig. A5.6. Analysis setup in BMDS version 3.1



modelling data-rich dose–response data sets (i.e. high number of animals per dose group, large number of dose groups); results will most likely differ the most in information-poor data sets (i.e. few dose groups, low number of animals per dose group). Table A5.6 lists the modelling fits of the individual Bayesian quantal models along with the posterior probabilities.

Table A5.6. BMDS Bayesian model fits to the hepatocellular hyperplasia data

Model Results

Benchmark Dose	
BMD	184.2822259
BMDL	143.1280017
BMDU	233.2425117

MA - Individual Models				
Model	Posterior Probability	BMD	BMDL	BMDU
Dichotomous Hill	0.05584784	201.803422	160.223651	246.6868
Gamma	0.020543394	162.416935	125.799036	201.4622
Logistic	0.136751704	177.7816415	148.355794	209.9621
Log-Logistic	0.027577479	191.5786147	151.476657	234.4709
Log-Probit	0.297903926	206.1234236	166.62941	247.3338
Multistage	0.004860768	115.7768011	91.3348973	137.9889
Probit	0.301050677	170.8027363	141.737914	203.3391
Quantal Linear	6.45776E-08	55.90221286	46.2311715	68.33326
Weibull	0.155464147	171.6930985	133.266699	214.0427

A5.1.5.2 Continuous data

In the case of continuous data, PROAST and BMDS differ substantially in several aspects: default options for modelling, assumed distributions and recommended choice of the BMRs. Additionally, only PROAST offers continuous model averaging (as of May 2020). As with quantal data, PROAST uses a frequentist approach to model averaging, whereas BMDS plans to implement a Bayesian model averaging with informative priors. As model averaging is not available at present for BMDS, only results obtained in PROAST are presented here.

Consider the data set for body weight given in Table A5.7. Loading and parameterizing the BMD analysis is the same as for quantal data. The resulting model fits are shown in Table A5.8. Note that PROAST assumes by default a log-normal distribution for responses – that is, constant variance of the log-transformed response data over the dose groups (including controls). The relative deviation definition of risk was used in this example, with a BMR of 5%.

Table A5.7. Example of a continuous end-point

Dose (mg/kg body weight per day)	Body weight, group mean (g)	SD	Number of animals per dose group
0	43.85	2.69	37
0.1	43.51	2.86	35
0.5	40.04	3.00	43
1.1	35.09	2.56	42

SD: standard deviation

Table A5.8. PROAST continuous model fits to hepatocellular hyperplasia data set

Fitted Models

model	converged	loglik	npar	AIC
full model	yes	198.36	5	-386.72
null model	yes	119.03	2	-234.06
Expon. m3-	yes	198.25	4	-388.5
Expon. m5-	yes	198.36	5	-386.72
Hill m3-	yes	198.25	4	-388.5
Hill m5-	yes	198.36	5	-386.72
Inv.Expon. m3-	yes	198.35	4	-388.7
Inv.Expon. m5-	yes	198.36	5	-386.72
LN m3-	yes	198.33	4	-388.66
LN m5-	yes	198.36	5	-386.72

Although PROAST implements continuous model averaging, given that all the continuous models are nested within four families (exponential, Hill, inverse exponential and log-normal), the representative model for averaging (either model 3 or model 5) must first be selected from each family based on the lowest AIC (see Table A5.8). In this example, model 3 from all four families is selected.

Using continuous model averaging and 200 bootstrap iterations (as described above in section A5.1.5.1) and selecting model 3 from each nested continuous model family give similar results of about 25% (0.24–0.26) model-averaging weights for the four models (Table A5.9).

Table A5.9. Model weights based on AIC used in PROAST model-averaging analysis of body weight

Weights for Model Averaging

EXP	HILL	INVEXP	LOGN
0.2388	0.2388	0.2639	0.2586

The final BMDL and BMDU values for this model-averaging analysis of body weights are 0.231 mg/kg body weight per day and 0.397 mg/kg body weight per day, respectively.

A5.1.5.3 Epidemiological data example

To give an example of human BMD modelling for continuous outcome, the association between serum concentration of perfluorooctanoic acid (PFOA) and total cholesterol is modelled. This association was observed in a cross-sectional study of 46 294 participants from the PFOA cohort summarized in Table A5.10 (for more information on this cohort and study, see Steenland et al., 2009). Subjects in this cohort were exposed to relatively high levels of PFOA through contaminated drinking-water. The reasons for selecting this example are that 1) there is a relatively clear dose–response relationship; 2) the results were presented as summary statistics that could be extracted from the publication; and 3) the number of participants was large enough to allow for relatively precise quantification of the BMD and its confidence interval (BMDL–BMDU).

This same study was used as the basis for deriving an HBGV for PFOA in the 2018 EFSA Opinion (EFSA, 2018a). For modelling, median PFOA serum concentrations, mean serum cholesterol concentrations and 95% confidence intervals around the mean were extracted by digitizing the results from Fig. 1 in the paper by Steenland et al. (2009). As the results were presented by dividing

PFOA into deciles, the number of subjects can be assumed to be about 4629 (without any meaningful loss of precision).

Table A5.10. The cross-sectional association between median serum PFOA and mean serum cholesterol concentrations as extracted from the study by Steenland et al. (2009)

Deciles	Median PFOA concentration (ng/mL)	Mean cholesterol concentration (mg/dL)	Number of subjects	SD
1	5.5	199	4629	60
2	9.6	202	4629	60
3	13.5	204	4629	60
4	18.2	205	4629	60
5	24.1	206	4629	60
6	33.5	206	4629	60
7	48.3	208	4629	60
8	70.9	207	4629	60
9	117	209	4629	60
10	344	210	4629	60

PFOA: perfluorooctanoic acid; SD: standard deviation

Concerning the selection of this example, it should be noted that the causality of the association between PFOA and total cholesterol is subject to some uncertainty (EFSA, 2018b). The selection of this example should by no means reflect any judgement of whether this association may be causal or coincidental.

A5.1.5.3.1 Selection of BMR

Despite a clear dose–response relationship, the mean increase in total cholesterol is small (~5%), at least compared with effect sizes often observed in controlled laboratory animal experiments. In contrast, it is important to keep in mind that variability in response among free-living humans is usually much larger than variability in response among homogeneous experimental animals. In Steenland et al. (2009), the mean increase in cholesterol was approximately 8% when comparing subjects with normal weight compared with those who were overweight or obese. The maximum mean increase in total cholesterol across age categories (18–80+ years) in this study was

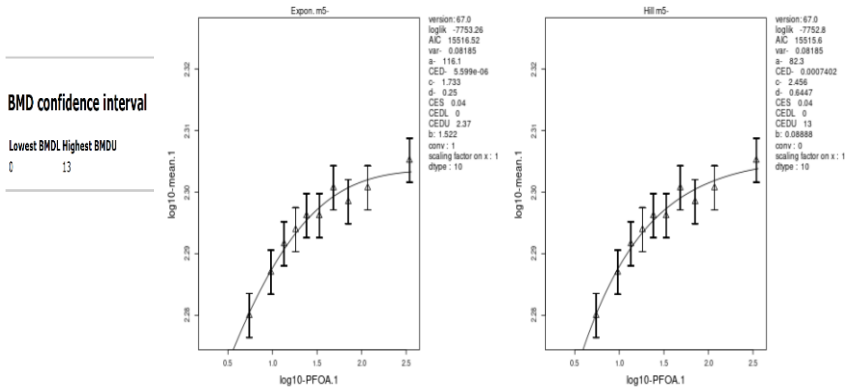
approximately 16%. With age and body mass index being two of the strongest determinants of total cholesterol, a BMR of 4% may be justified.

A5.1.5.3.2 "Zero dose" as point of comparison

When the data are modelled with dose divided into deciles, BMD estimates are computed by extrapolating beyond the observed (or reported) dose range down to "zero dose", which may be unrealistic in this case – that is, PFOA is environmentally persistent, and there is no true zero dose. As the dose–response curve monotonically increases from the lowest decile (5.5 ng/mL), the resulting BMD/ BMDL may be suspect. That is, if individual data had been available, the curve would be expected to level off and stabilize at concentrations below 5.5 ng/mL. Similar situations may also occur in controlled toxicological experiments in cases where the dose–response curve is steep at the lower end of the dose range and well-defined zero dose does not exist (e.g. for environmental contaminants or due to cross-contamination).

To analyse these data, PROAST version 67.0 (online version available on the RIVM homepage; https://proastweb.rivm.nl/) is used to fit the data. The reported BMDU and BMDL are derived on the basis of the five-parameter exponential model and the five-parameter Hill model, as illustrated in Fig. A5.7.

Fig. A5.7. Model output from PROAST version 67.0, online RIVM version



For these data, there is significant uncertainty in the shape of the dose–response curve below 5.5 ng/mL, which is caused by a mixture of the model chosen and the fact that the data do not contain enough information to estimate background response (e.g. 0 ng/mL); consequently, the choice of the given BMR may not be appropriate.

As PFOA is environmentally persistent and thus ubiquitous in the environment, there is no reason to assume that 0 ng/mL is a realistic dose. A more realistic alternative would be to use the lowest reported dose (5.5 ng/mL) as referent. This can be justified, as there is some indication from other studies that the dose–response curve starts to level off somewhere between 2 and 3 ng/mL (Eriksen et al., 2013; Skúladóttir et al., 2015). Thus, assuming 5.5 ng/mL as a referent may only modestly overestimate the BMD. To analyse the data in this manner and to avoid the extrapolation problem, we subtract the lowest dose (5.5 ng/mL) from all dose groups.

Using PROAST version 67.0 and the lowest reported exposure quantile (5.5 ng/mL) as the referent group, the BMDL–BMDU range is given as 13–108 ng/mL.

If one does not want to accept the underestimation of the BMD using zero dose as the point of comparison and the slight overestimation when using the lowest reported exposure decile as the point of comparison, the third option would be to restrict the dose–response curve somewhere below 5.5 ng/mL based on a priori information from other studies. As mentioned previously, there is some empirical information that the curve is levelling off around 2–3 ng/mL (Eriksen et al., 2013; Skúladóttir et al., 2015). Such restriction would give an estimate somewhere in between the results of the two examples shown in Fig. A5.8. In any case, the differences in the BMD estimates observed between BMDS and PROAST are somewhat larger than the uncertainty associated with extrapolation to zero or using the lowest decile as the point of comparison.

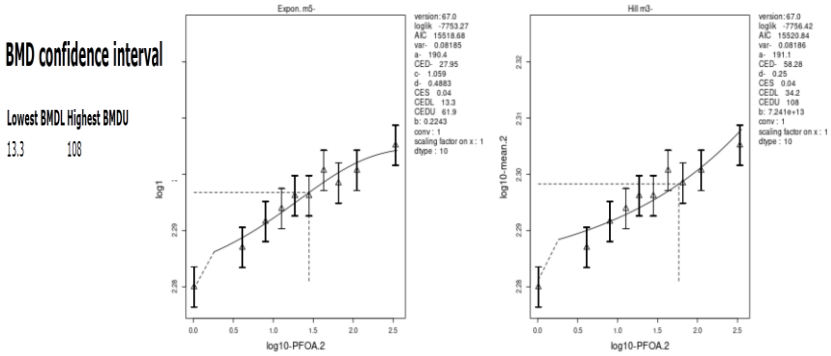
A5.1.6 References

Akaike H (1973). Information theory and the maximum likelihood principle. In: Petrov BN, Csäki F, editors. Proceedings of the 2nd International Symposium on Information Theory. Budapest: Akademiai Ki à do; 267–81.

Buckland ST, Burnham KP, Augustin NH (1997). Model selection: an integral part of inference. Biometrics. 53(2):603–18. doi:10.2307/2533961.

Burnham KP, Anderson DR (2004). Multimodel inference: understanding AIC and BIC in model selection. Sociol Methods Res. 33(2):261–304.

Fig. A5.8. Model output from PROAST version 67.0, online RIVM version



EFSA (2017). Update: guidance on the use of the benchmark dose approach in risk assessment. European Food Safety Authority Scientific Committee. EFSA J. 15(1):4658. doi:10.2903/j.efsa.2017.4658.

EFSA (2018a). Scientific Opinion on the risk to human health related to the presence of perfluorooctane sulfonic acid and perfluorooctanoic acid in food. European Food Safety Authority Panel on Contaminants in the Food Chain. EFSA J. 16(12):5194. doi:10.2903/j.efsa.2018.5194.

EFSA (2018b). Minutes of the expert meeting on perfluorooctane sulfonic acid and perfluorooctanoic acid in food assessment. Parma: European Food Safety Authority, Unit on Biological Hazards and Contaminants (EFSA/CONTAM/ 3503; https://www.efsa.europa.eu/sites/default/files/news/efsa-contam-3503.pdf).

Eriksen KT, Raaschou-Nielsen O, McLaughlin JK, Lipworth L, Tjønneland A, Overvad K et al. (2013). Association between plasma PFOA and PFOS levels and total cholesterol in a middle-aged Danish population. PLoS One. 8(2):e56969. doi:10.1371/journal.pone.0056969.

Evans JS, Graham JD, Gray GM, Sielken RL (1994). A distributional approach to characterizing low-dose cancer risk. Risk Anal. 14(1):25–34. doi:10.1111/ j.1539-6924.1994.tb00025.x.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013). Bayesian data analysis, third edition. Boca Raton (FL): CRC Press.

Hill AB (1965). The environment and disease: association or causation? Proc R Soc Med. 58(5):295–300. doi:10.1177/0141076814562718.

Hjort NL, Claeskens G (2003). Frequentist model average estimators. J Am Stat Assoc. 98(464):879–99. doi:10.1198/016214503000000828.

Hoeting JA, Madigan D, Raftery AE, Volinsky CT (1999). Bayesian model averaging: a tutorial. Stat Sci. 14(4):382–417. doi:10.1214/ss/1009212519.

Illari P, Russo F (2015). Chapter 6: Evidence and causality. In: Causality: philosophical theory meets scientific practice. Oxford: Oxford University Press; 46–59.

Kass RE, Raftery AE (1995). Bayes factors. J Am Stat Assoc. 90(430):773–95. doi:10.1080/01621459.1995.10476572.

Krewski D, Gaylor DW, Soms AP, Szyszkowicz M (1993). An overview of the report: Correlation between carcinogenic potency and the maximum tolerated dose: implications for risk assessment. Risk Anal. 13(4):383–98. doi:10.1111/j.1539-6924.1993.tb00738.x.

Millar RB (2011). Maximum likelihood estimation and inference: with examples in R, SAS and ADMB. Chichester: John Wiley & Sons.

Morgan MG, Henrion M, Morris SC (1980). Expert judgments for policy analysis. Upton (NY): Brookhaven National Laboratory.

Shao K, Shapiro AJ (2018). A web-based system for Bayesian benchmark dose estimation. Environ Health Perspect. 126(1):017002. doi:10.1289/EHP1289.

Skúladóttir M, Ramel A, Rytter D, Haug LS, Sabaredzovic A, Bech BH et al. (2015). Examining confounding by diet in the association between perfluoroalkyl acids and serum cholesterol in pregnancy. Environ Res. 143(Part A):33–8. doi:10.1016/j.envres.2015.09.001.

Steenland K, Tinker S, Frisbee S, Ducatman A, Vaccarino V (2009). Association of perfluorooctanoic acid and perfluorooctane sulfonate with serum lipids among adults living near a chemical plant. Am J Epidemiol. 170(10):1268–78. doi:10.1093/aje/kwp279.

Suter GW, Cormier SM (2011). Why and how to combine evidence in environmental assessments: weighing evidence and building cases. Sci Total Environ. 409(8):1406–17. doi:10.1016/j.scitotenv.2010.12.029.

Tallarida RJ, Laskin OL, Jacob LS (1976). Perturbation of drug receptor equilibrium in the presence of competitive blocking agents. J Theor Biol. 61(1):211–9. doi:10.1016/0022-5193(76)90115-6.

Wheeler M, Bailer AJ (2007). Properties of model-averaged BMDLs: a study of model averaging in dichotomous response risk estimation. Risk Anal. 27(3):659–70. doi:10.1111/j.1539-6924.2007.00920.x.

WHO (2000). Lead. In: Safety evaluation of certain food additives and contaminants. Geneva: World Health Organization (WHO Food Additives Series, No. 44; http://www.inchem.org/documents/jecfa/jecmono/v44jec12.htm).

WHO (2011). Lead. In: Safety evaluation of certain food additives and contaminants. Geneva: World Health Organization (WHO Food Additives Series, No. 64; https://apps.who.int/iris/bitstream/handle/10665/44521/9789241660648_eng.pdf?sequence=1).