PROBABILITY HAZARD ASSESSMENT

RESEARCH PROJEKT F2437: Derivation of occupational exposure limits for airborne chemicals – Comparison of methods and protection levels

prepared on behalf of:

Federal Institute for Occupational Safety and Health (BAuA)
Dortmund

Forschungs- und Beratungsinstitut Gefahrstoffe GmbH
Klarastraße 63 • 79106 Freiburg
Germany

Prepared by:
Dr. Klaus Schneider
Dr. Eva Kaiser
Dr. Marco Dilger

With support by
Dr. Werner Wosniok, Bremen

Freiburg, October 2019
## Content

Summary .......................................................................................................................... 4

Abbreviations .................................................................................................................. 5

1 Introduction .................................................................................................................. 8

2 Historical development and existing proposals .......................................................... 11
   2.1 Historical perspective ............................................................................................. 11
   2.2 The WHO-IPCS document .................................................................................... 13
      2.2.1 Definition of “Target human dose” depending on data type............................. 13
      2.2.2 Non-probabilistic approach .............................................................................. 14
      2.2.3 Approximate probabilistic approach (APROBA) ............................................. 15
      2.2.4 Full probabilistic approach .............................................................................. 15
      2.2.5 Uncertainty distributions ................................................................................. 15

3 Tools for risk assessment ............................................................................................. 19
   3.1 APROBA and related web tools .............................................................................. 19
   3.2 EFSA’s web-based tool .......................................................................................... 22
   3.3 Other tools for full probabilistic assessments ......................................................... 26

4 Analysis and discussion of existing approaches .......................................................... 27
   4.1 Principal approaches .............................................................................................. 27
   4.2 Preconditions, pros and cons ................................................................................ 27
   4.3 Uncertainty versus variability, secondary uncertainty .......................................... 29
   4.4 Types of dose-response data ................................................................................ 30
   4.5 Practicality ............................................................................................................ 32

5 Description of risks above the OEL .......................................................................... 34

6 Conclusions .................................................................................................................. 35

7 Annex 1: Examples ..................................................................................................... 36
   7.1 Quantal data set ..................................................................................................... 36
      7.1.1 Input data – dose response data ................................................................. 36
      7.1.2 Input data – uncertainty distributions ......................................................... 37
      7.1.3 Modelling results .......................................................................................... 38
      7.1.4 Summary ....................................................................................................... 39
      7.1.5 Comparison and discussion ......................................................................... 40
   7.2 Continuous dataset ............................................................................................... 41
      7.2.1 Input data – dose response data ................................................................. 41
Summary

In probabilistic approaches the input data to the following equation are given as distributions, expressing uncertainty (in the POD and the assessment factors) and variability (in the human population) of the input parameters:

\[ GV = \frac{POD}{AF_1 \times AF_2 \times AF_3 + \ldots} \]

with: GV = guidance value, POD = point of departure, AF = assessment factor.

These distributions are combined using probabilistic methods (Monte Carlo analysis), resulting in a distribution of GV.

This approach requires to decide on the critical effect size or benchmark rate (BMR), in order to determine the POD, if a benchmark dose is used, on the percentage of the target population to be covered by GV and on the probability of achieving the defined protection level.

The distribution of GV then allows describing uncertainty and variability of the output, to better characterise the protection level achieved and to estimate the size and likeliness of health effects at higher concentrations.

With regard to the practical implementation of probabilistic approaches in risk assessment two major recent developments are described in this report:

- The APROBA tool developed in the frame of the WHO/IPCS project on “Evaluating and Expressing Uncertainty in Hazard Characterization” (WHO, 2014).

This EXCEL®-based tool allows rapid approximations to full Monte Carlo analyses by using lognormal input distributions for all parameters.

- The Monte Carlo tool developed by EFSA

This full Monte Carlo analysis tool is currently under development at EFSA and allows not only Monte Carlo analyses, but also distribution fitting and use of various kinds of distributions.

The principles, pros and cons of the various approaches, the input data used by APROBA and the implications of different types of dose-response data are discussed in this report. With two simple examples the workability of the tools are demonstrated and results are compared.

In view of these new developments use of probabilistic approaches to hazard assessment are simplified and their use for

- method development and discussion of (combination of) deterministic factors
- comparison with standard assessments using deterministic factors
- refined assessments of complex cases

is encouraged.
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AEL</td>
<td>Acceptable Exposure Levels</td>
</tr>
<tr>
<td>AAEL</td>
<td>Acute Acceptable Exposure Levels</td>
</tr>
<tr>
<td>AF</td>
<td>Assessment factor</td>
</tr>
<tr>
<td>AGS</td>
<td>Ausschuss für Gefahrstoffe</td>
</tr>
<tr>
<td>AIC</td>
<td>Akaike information criterion</td>
</tr>
<tr>
<td>ANSES</td>
<td>Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail</td>
</tr>
<tr>
<td>AOEL</td>
<td>Acceptable Operator Exposure Levels</td>
</tr>
<tr>
<td>AAOEL</td>
<td>Acute Acceptable Operator Exposure Levels</td>
</tr>
<tr>
<td>APROBA</td>
<td>Approximate probabilistic analysis</td>
</tr>
<tr>
<td>BAuA</td>
<td>Bundesanstalt für Arbeitsschutz und Arbeitsmedizin</td>
</tr>
<tr>
<td>BBMD</td>
<td>Bayesian Benchmark Dose</td>
</tr>
<tr>
<td>BMD</td>
<td>Benchmark dose</td>
</tr>
<tr>
<td>BMDL</td>
<td>Benchmark dose lower bound</td>
</tr>
<tr>
<td>BMDU</td>
<td>Benchmark dose upper bound</td>
</tr>
<tr>
<td>BMR</td>
<td>Benchmark response</td>
</tr>
<tr>
<td>BMDS</td>
<td>Benchmark dose software</td>
</tr>
<tr>
<td>BOELV</td>
<td>Binding occupational exposure level values</td>
</tr>
<tr>
<td>BPR</td>
<td>Biocidal products regulation</td>
</tr>
<tr>
<td>BS</td>
<td>Bootstrapping</td>
</tr>
<tr>
<td>CDS</td>
<td>Cumulative distribution function</td>
</tr>
<tr>
<td>CES</td>
<td>Critical effect size</td>
</tr>
<tr>
<td>CSAF</td>
<td>Chemical-specific adjustment factors</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>DFG</td>
<td>Deutsche Forschungsgesellschaft</td>
</tr>
<tr>
<td>DMEL</td>
<td>Derived minimal effect level</td>
</tr>
<tr>
<td>DNEL</td>
<td>Derived no effect level</td>
</tr>
<tr>
<td>ECETOC</td>
<td>European Centre for Ecotoxicology and Toxicology of Chemicals</td>
</tr>
<tr>
<td>ECHA</td>
<td>European Chemicals Agency</td>
</tr>
<tr>
<td>ED10</td>
<td>Effective dose 10% (dose corresponding to a 10% increase in an adverse effect, relative to the control response)</td>
</tr>
<tr>
<td>EFSA</td>
<td>European Food Safety Authority</td>
</tr>
<tr>
<td>GM</td>
<td>Geometric mean</td>
</tr>
<tr>
<td>GSD</td>
<td>Geometric standard deviation</td>
</tr>
<tr>
<td>GV</td>
<td>Guidance value</td>
</tr>
<tr>
<td>IPCS</td>
<td>WHO's International Programme on Chemical Safety</td>
</tr>
<tr>
<td>IRIS</td>
<td>Integrated Risk Information System</td>
</tr>
<tr>
<td>LOAEC</td>
<td>Lowest observed adverse effect concentration</td>
</tr>
<tr>
<td>LOAEL</td>
<td>Lowest observed adverse effect level</td>
</tr>
<tr>
<td>MAK</td>
<td>Maximale Arbeitsplatzkonzentration</td>
</tr>
<tr>
<td>MC</td>
<td>Monte Carlo</td>
</tr>
<tr>
<td>MCMC</td>
<td>Markov Chain Monte Carlo</td>
</tr>
<tr>
<td>MCRA</td>
<td>Monte Carlo Risk Assessment</td>
</tr>
<tr>
<td>MPPD</td>
<td>Multiple path particle dosimetry (model)</td>
</tr>
<tr>
<td>NAEC</td>
<td>No adverse effect concentration</td>
</tr>
<tr>
<td>NAEL</td>
<td>No adverse effect level</td>
</tr>
<tr>
<td>NOAEC</td>
<td>No observed adverse effect concentration</td>
</tr>
<tr>
<td>NOAEL</td>
<td>No observed adverse effect level</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>OEL</td>
<td>Occupational exposure limit</td>
</tr>
<tr>
<td>PBPK</td>
<td>Physiology-based pharmacokinetic (model)</td>
</tr>
<tr>
<td>PDF</td>
<td>Probability density function</td>
</tr>
<tr>
<td>POD</td>
<td>Point of departure</td>
</tr>
<tr>
<td>PPP</td>
<td>Plant protection products</td>
</tr>
<tr>
<td>PROAST</td>
<td>Dose-response modelling software by RIVM</td>
</tr>
<tr>
<td>QSAR</td>
<td>Quantitative structure activity relationship</td>
</tr>
<tr>
<td>RAC</td>
<td>Committee for Risk Assessment</td>
</tr>
<tr>
<td>REACH</td>
<td>Registration, Evaluation, Authorisation and Restriction of Chemicals,</td>
</tr>
<tr>
<td>RfD</td>
<td>Reference dose</td>
</tr>
<tr>
<td>RIVM</td>
<td>Dutch National Institute for Public Health and the Environment</td>
</tr>
<tr>
<td>SC</td>
<td>EFSA’s Scientific Committee</td>
</tr>
<tr>
<td>SCOEL</td>
<td>Scientific Committee on Occupational Exposure Limits</td>
</tr>
<tr>
<td>STEL</td>
<td>Short-term exposure limit</td>
</tr>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>TD</td>
<td>Toxicodynamics</td>
</tr>
<tr>
<td>TK</td>
<td>Toxicokinetics</td>
</tr>
<tr>
<td>TRGS</td>
<td>Technische Regeln für Gefahrstoffe</td>
</tr>
<tr>
<td>US EPA</td>
<td>Environmental Protection Agency in the US</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organisation</td>
</tr>
</tbody>
</table>
1 Introduction

Health-based guidance values such as OELs are typically derived by dividing a point of departure (POD) by assessment factors (including allometric scaling factors) accounting for uncertainty and variability. In addition, adjustments, e.g. for differences in exposure conditions, might be considered.

Uncertainty

Uncertainty in this context means uncertainty of any input data due to e.g. inaccurate measurements or unknown representability of entity of tested chemicals for the target substance. Uncertainty can be reduced by improving input data.

Variability

Variability means the variation of e.g. internal doses and responses in the target population. This is a real property of the target population and cannot be reduced.

In the so-called deterministic approach, a single definite value is used for each input variable. If input data are properly chosen, the resulting value is thought to provide sufficient protection against harmful effects of the evaluated chemical. But such point estimates have some significant disadvantages:

- They do not come with a specification of the level of protection achieved
- They do not inform (quantitatively) about remaining uncertainties
- If for the applied assessment factors conservative estimates are used, multiplication of these conservative factors might lead to an overly conservative guidance value; on the other hand, using e.g. average values for each assessment factor (e.g. a factor sufficient to cover 50% of chemicals in an empirical dataset), might lead to a guidance value not sufficiently protective in the majority of cases.

Since several years so-called probabilistic approaches for deriving health-based guidance values are proposed (Mekel and Fehr, 2014). These approaches are characterised as follows:

- Input variables are not deterministic values but distributions (e.g. the distribution of possible values for extrapolating from subchronic to chronic study duration; these distributions are typically obtained from empirical datasets)
- These input distributions describe variability and/or uncertainty in the input data
As no discrete mathematical solutions are available for combining distributions of various forms, they need to be combined using probabilistic methods (Monte-Carlo analysis) (Aral, 2010). This approach is graphically presented in Figure 1-1.

\[ GV = \frac{POD}{AF \times AF_2 \times AF_3 \ldots} \]

Figure 1-1: Combination of input distributions (displayed as probability density functions) of assessment factors (AF) to probabilistically derive a guidance value (GV)

The input parameters as well as the output guidance value are distributions. (Note that distributions can either be depicted as density functions, chosen here, with an integral of 1, or as cumulative distribution functions, with values between 0 and 1, see 8.2.2).

The distribution obtained for the guidance value allows a better description and interpretation of uncertainties associated with the value – on costs of a mathematical more complex approach. Further, in order to apply a probabilistic approach several regulatory decisions are required:

- on the critical effect size or benchmark rate (BMR), in order to determine the point of departure (which is ideally a benchmark dose)
- on the percentage of the target population to be covered by the value
- on the probability of achieving the above defined protection level.

This will be further analysed and discussed in this report.

In order to facilitate the understanding for readers new to probabilistic methods in Annex 2 the differences between empirical and parametric distributions are discussed as well as the presentation of distributions either as
• probability density function or
• cumulative distribution function.

Figure 1-2: Probability density versus cumulative distribution function (with 10\textsuperscript{th} and 90\textsuperscript{th} percentiles as blue lines)

Further, some recurrent terms such as Monte Carlo simulation, bootstrapping, Bayesian methods, Markov Chain and Latin-Hypercube sampling are explained in Annex 2, section 8.2. But it is important to note that the tools presented here are able to be applied with a limited understanding of the mathematical background of the tools.
2 Historical development and existing proposals

2.1 Historical perspective

First approaches to probabilistic hazard assessment were published by authors in the USA (Baird et al., 1996; Baird et al., 2001; Price et al., 1997; Swartout et al., 1998). These approaches in the first instance tried to describe traditionally used assessment factors in the form of distributions. Hattis et al. described an approach for a probabilistic hazard assessment using empirical data for establishing distributions (Hattis et al., 2002). Weight was laid on the evaluation and use of the Hattis database on inter-individual differences in susceptibility. The results were compared with (then) existing RfD values from the IRIS database. This approach was further extended, again with using data from the Hattis database to describe both the toxicokinetic and toxicodynamic part of inter-individual variability (Hattis and Lynch, 2007).

Scientists from the Dutch RIVM were the first to derive distributions from empirical data on the size of assessment factors (Slob and Pieters, 1998; Vermeire et al., 1999). In several publications this approach was further developed and applied to example substances (Bosgra et al., 2005; van der Voet and Slob, 2007; van der Voet et al., 2009). The distribution for intraspecies extrapolation in this work is based on theoretical considerations, not on an empirical database. The concept was further expanded to propose an integrated approach for probabilistic modelling of both hazard and exposure assessment (Slob et al., 2014; van der Voet et al., 2009). This integrated approach was also used for assessing carcinogens. For this work the MCRA software (https://mcra.rivm.nl/), which was developed for exposure assessment in the food safety area, was extended to cover hazard aspects. The MCRA software is not publicly available.

Hattis and colleagues, but also others emphasize that probabilistic methods would allow to harmonize approaches for threshold and non-threshold (carcinogenic) substances (Hattis et al., 2002; Slob et al., 2014).

In a research project for BAuA, FoBiG developed a probabilistic framework for deriving OELs. Based on empirical data, distributions for assessment factors were derived by statistical parameter fitting. These were combined with distributions for benchmark doses in Monte Carlo analyses (Schneider et al., 2004; Schneider et al., 2006). With several examples the authors also showed how to use substance-specific data to derive and use distributions for chemical-specific adjustment factors (CSAF) in the probabilistic assessment.

A few years ago the US National Research Council requested the US Environmental Protection Agency (US EPA) to develop probabilistic methods for deriving guidance values (NRC, 2009). This fuelled the discussion on suitable approaches and pros and cons of probabilistic assessment procedures (Cooke, 2010; Crump et al., 2010; Goble and Hattis, 2010; Simon et al., 2016).
In 2015 Chiu of the US EPA and Wout Slob of RIVM proposed a probabilistic framework (Chiu and Slob, 2015), which is also the core of a WHO report, developed in the frame of the IPCS: *International Programme on Chemical Safety* (IPCS) - Harmonization Project (WHO, 2014). The approach presented there, which also includes an EXCEL®-based easy to use approximation to probabilistic modelling called APROBA, will be subject of a more detailed analysis in the following chapter. APROBA was further extended (APROBA-Plus) to cover also probabilistic exposure modelling (Bokkers et al., 2017).

The APROBA tool (for more information see chapter 3.1) was implemented as a web-based platform by Chiu and colleagues at two sites (Chiu et al., 2018):

- at https://wchiu.shinyapps.io/APROBAweb/
- and as part of the Bayesian Benchmark Dose (BBMD) Analysis tool platform by Shao and Shapiro (2018) (at https://benchmarkdose.org/).

The BBMD platform was discussed with its dose-response modelling functionalities in our respective report on dose-response modelling. The probabilistic assessment tool added by Chiu and colleagues refers to the work by Chiu and Slob (Chiu et al., 2018; Chiu and Slob, 2015) and is based on the approach proposed in the WHO/IPCS report. At both platforms all distributions are assumed to be lognormal as it is done within the APROBA tool. There are differences between the two platforms. For example, no distinction is made between continuous and quantal endpoints in BBMD, whereas APROBA web seems to be an exact implementation of the functionalities of the EXCEL® tool, but with different forms for providing input and for presenting results. No further guidance is available for the BBMD implemented form. The APROBA web and the EXCEL® tool are further discussed in chapter 3.

Oldenkamp et al. provided two example probabilistic evaluations for the pharmaceuticals ciprofloxacin and methotrexate, based on the WHO/IPCS framework (Oldenkamp et al., 2016). Continuous endpoints were modelled for both substances. The main differences to the IPCS approach is that inter-individual variability is modelled in two steps, describing first the differences between the median individual in the general population and the median individual in the (more susceptible) subpopulation and in a second step the variability within the subpopulation. Secondary uncertainty due to limitations of the dataset on inter-individual variability is included in the model. The authors also included substance-specific data for informing the input distributions and compared results with the default distributions as proposed by WHO/IPCS (WHO, 2014). As no substance-specific data on inter-individual differences in toxicodynamic could be found, data from the literature (Renwick and Lazarus, 1998) were used to derive a distribution for this step.

The authors modelled the distribution for the average individual as well as the one for covering 99% of the whole population. The difference between the medians of these two distributions was interpreted as a measure for inter-individual variability, whereas the width of the distribution of the 99% distribution was taken as a measure for the uncertainty of the health-based value (the latter being much larger than the former).
Similar results were obtained when using substance-specific distributions or default distributions as proposed by WHO/IPCS (WHO, 2014).

Chiu et al. tested the feasibility of the WHO/IPCS approach and its implications by performing (approximated) probabilistic assessments for a large set of substances and comparing it with RfD values of the US EPA (Chiu et al., 2018). In this comparison no benchmark doses were available, a fact that contributed much to the overall uncertainty. The values for 1% incidence at a 95% confidence level were found to be within one order of magnitude of the RfDs.

2.2 The WHO-IPCS document

2.2.1 Definition of “Target human dose” depending on data type

The WHO/IPCS Harmonization Project Document 11 is titled “Guidance document on evaluating and expressing uncertainty in hazard characterization” (WHO, 2014). It focuses on addressing uncertainty and variability in health-based guidance values by probabilistic methods. The following terms are important to understand the approach in this document:

“target human dose”, \( HD_M \)

where

- HD is the reference or health-based guidance value resulting from the risk characterisation
- \( I \) is the fraction of the population experiencing an effect (at dose HD)
- of magnitude (or severity) \( M \) or greater (for the critical effect considered).

In addition to \( I \) and \( M \), the coverage (also called confidence or probability of effects) needs to be defined. Typically, a 95% probability is used to describe the outcome of the assessment, which means that there is a 5% probability that effects are more severe at HD than described by \( I \) and \( M \).

\( I \) is the fraction of the general population not covered by HD. Typically, 1% is proposed in this tool, but a range from 0.1% to 5% is mentioned.

\( M \) is the critical effect size (or benchmark response, BMR) associated with the point of departure used for the assessment. \( M \) is set during dose-response modelling. For defining \( M \), a distinction is made between different types of effect data:

- **continuous data**

\( M \) characterises the effect size at the boundary to adversity (and is equal to the BMR set when applying dose-response modelling)

Example for defining a HD for continuous data:
HD_{05}^{01} (for critical effect “reduction in red blood cell count”): the human dose at which 1% of the population shows a decrease in red blood cell counts of 5% or greater.

- **quantal deterministic data**

A distinction is made in the report between quantal data with an underlying continuous effect (e.g. histopathological effects in the liver, following (continuous) physiological changes) versus real stochastic effects (e.g. tumours or malformations). This will be discussed in more details further below.

Example for defining a HD for quantal deterministic data:

HD_{05}^{05} (for critical effect liver lesions): the human dose at which 5% of the population shows liver lesions.

- **quantal stochastic data**

Example for defining a HD for quantal stochastic data:

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

All these HD values are also a function of the probability (coverage) chosen: If a definite percentile of the probability function for HD_{M} is chosen, then HD_{M} would assume a single value (which, of course, would be different if a 90th or a 95th percentile is chosen).

The WHO/IPCS document describes three different approaches.

### 2.2.2 Non-probabilistic approach

In this approach, the lower and upper bounds for each hazard characterization aspect are combined by multiplication. Lower and upper bounds are typically chosen as 5th and 95th percentiles of uncertainty distributions, meaning, e.g. that the BMDL is divided by the 95th percentiles of the distributions for interspecies, intraspecies and time extrapolation. Not surprisingly, this approach leads to wide ranges, spanning several orders of magnitude, between lower and upper bound estimates of the hazard value. This approach is of limited practical value.

This non-probabilistic approach is included – for comparison – in the APROBA EXCEL® spreadsheet.
2.2.3 Approximate probabilistic approach (APROBA)

In the approximate probabilistic approach all uncertainties are assumed to be lognormally distributed and are described by independent lognormal probability distributions. With this assumption, algorithms can be implemented in an EXCEL® spreadsheet and can be solved numerically without Monte Carlo simulations.

The APROBA EXCEL® spreadsheet contains default distributions (see chapter 2.2.5), which can be changed by the user (see chapter 3.1 for more information).

2.2.4 Full probabilistic approach

With this approach, uncertainty distributions of any form are combined probabilistically. As the type of distributions is not restricted to lognormal distributions, a mathematical exact solution is not possible. Calculations need to be done using Monte Carlo simulations.

A full probabilistic assessment is more flexible, can use any kind of distribution and can include additional aspects of uncertainty. But it can be run with the same distributions as used in APROBA and – according to the report - would give similar results.

For a full probabilistic approach there are no ready-to-use tools described in the report.

2.2.5 Uncertainty distributions

2.2.5.1 Point of departure

Although the BMDL is given clear preference, also NOAEL and LOAEL are foreseen as possible points of departure (POD).

When dose-response modelling is performed, upper and lower bounds of the BMD are obtained (BMDU and BMDL, resp.). A distribution of the BMD expressing the uncertainty of the BMD can be obtained either

- using the BMDL and BMDU and assuming a lognormal distribution (this approach is used in APROBA)
- approximating a distribution by the bootstrap method or by a Bayesian approach.

Note that the POD and consequently also the resulting $\text{HD}_m$ is different for the different data types with regard to the effect size chosen:

Continuous data: $\text{BMDL}_M$ with $M$ being the BMR chosen for a specific continuous endpoint is the POD. As the data points used for dose-response modelling are the group averages, it can...
be assumed that each data point and also the BMDL stands for an average response rate of 50% within each group with regard to the chosen M.

Quantal deterministic data:
Although this does not become clear from the WHO/IPCS report and the description of the APROBA tool, it is proposed here to use the BMDL50 as the point of departure (personal communication Wout Slob, 21 May 2019). Implications will further be discussed in chapter 4.

Quantal stochastic data:
For stochastic endpoints the BMDL10 is proposed as POD, in agreement with current practice.

Also for using NOAEL and LOAEL values, approximate uncertainty distributions are derived and implemented in APROBA. These uncertainty distributions are wider than typical BMD distributions.

2.2.5.2 Exposure duration

The following distributions are proposed for exposure duration extrapolation:

Subchronic to chronic extrapolation, based on the evaluation of Bokkers and Slob (2005):

Distribution with GM = 2, with P95/P50 = 4 [(P05, P95) = (0.5, 8)]

Subacute to chronic extrapolation, based on various evaluations of NOAEL ratios in the literature up to the year 2011:

Distribution with GM = 5, with P95/P50 = 8 [(P05, P95) = (0.625, 40)].

The database on exposure duration extrapolation will be discussed in detail in a separate report of this project.

2.2.5.3 Interspecies extrapolation

Interspecies extrapolation is performed in two steps:

- adjustment of the dose for differences in body size between test animal and humans by allometric scaling

- accounting for potential and unknown (chemical-specific) differences between species in toxicokinetics and/or toxicodynamics.

Based on evaluations in the literature on quantitative species differences the following distributions were proposed:
For step 1:

An allometric scaling exponent of 0.7 is used; the distribution is established based on the 95% confidence interval of the exponent of 0.66 – 0.74.

For step 2:

TK/TD uncertainty after accounting for body size differences, based on evaluations in the literature on quantitative species differences the following distributions:

Distribution with GM = 1, P95/P50 = 3 [(P05, P95) = (1/3, 3)]

Again, the underlying data on interspecies extrapolation will be discussed in a separate report.

For inhalation exposure the report assumes as a default that in the central tendency there are no differences between animals and humans with regard to deposition of particles or doses of gases (i.e. median of ratios between animals and humans is 1). A low uncertainty with a 95th percentile to median ratio of the lognormal distribution of 2 is assumed, but no data are presented to support these assumptions. But this assumption is not used in the APROBA tool, where the same distribution (see above) for oral and inhalation studies is used.

2.2.5.4 Inter-individual variability

The distributions for inter-individual variability are derived from literature evaluations, mainly from Hattis and co-workers (Hattis et al., 2002; Hattis and Lynch, 2007) and Renwick and Dorne (Dorne et al., 2005; Renwick and Lazarus, 1998).

The derived distributions cover the whole population (no distinction is made, e.g., between adults and children, as the differences were not statistically significantly different). Separate analysis and distributions are presented for toxicokinetics and – dynamics, but – assuming independency – the distributions can also be combined to one. A generic lognormal distribution results, which is characterised by \( \log(\text{GSD}_H) \), which represents the distribution of the spreads of groups of individuals with differing susceptibility found in the literature. The \( \log(\text{GSD}_H) \) is characterised by the following values: median: 0.324; 95th percentile 0.697. How geometric standard deviations are used to describe variability in human populations will be discussed in more details in another report (on intraspecies extrapolation).

To exemplify the meaning of this distribution, it can be translated into assessment factors (ratio between the equipotent doses in average and susceptible humans):

\[
AF_{\text{intra-i}} = \text{Factor covering } (1 - I) \text{ of the population} = \text{GSD}_H^{z_{1-I}},
\]

For \( I = 5\% \), 1\% and 0.1\%, the corresponding values for \( z_{1-I} \) are 1.6449, 2.3263 and 3.0902.
Table 2-1: Factors for intraspecies extrapolation

<table>
<thead>
<tr>
<th>chosen incidence</th>
<th>5th perc</th>
<th>50th perc</th>
<th>95th perc</th>
</tr>
</thead>
<tbody>
<tr>
<td>5%</td>
<td>1.8</td>
<td>3.41</td>
<td>14</td>
</tr>
<tr>
<td>1%</td>
<td>2.2</td>
<td>5.67</td>
<td>42</td>
</tr>
<tr>
<td>0.1%</td>
<td>2.9</td>
<td>14.23</td>
<td>143</td>
</tr>
</tbody>
</table>

Interpretation for a chosen incidence level of 1%: a factor of 5.67 is required to cover 99% of the population with a probability of 50%. At the 95% probability level a factor of 42 is required to cover 99% of the population.

With this concept of having individual distributions for covering certain percentiles of the target human population the WHO/IPCS report follows the methodological approach as developed in the BAuA project (Schneider et al., 2004; Schneider et al., 2006).

2.2.5.5 Route-to-route extrapolation

The uncertainties of route-to-route extrapolation are discussed, but no default distribution is proposed.

2.2.5.6 Other uncertainties

The APROBA tool as well as the full probabilistic model are open for adding additional uncertainty distributions (in case of APROBA up to 3, all in the form of lognormal distributions). Examples are uncertainties due to differences in exposure patterns (e.g. 4 hour daily inhalation exposure instead of 8 or 24 hours) or the uncertainty introduced by an incomplete database.

2.2.5.7 Secondary uncertainties

Secondary uncertainties are uncertainties associated with the used distributions themselves. For example, there is uncertainty in the data used for deriving the ratios between studies of different exposure duration, e.g. the NOAELs used for calculating NOAEL ratios. With regard to inter-individual differences in susceptibility, the individuals studied and/or the chemicals under study might not be representative for the target population and/or the chemical under assessment.

Secondary uncertainties are considered in some probabilistic models (Oldenkamp et al., 2016). They cannot be assessed in APROBA. They are generally considered to be substantially lower than primary uncertainties.
3 Tools for risk assessment

3.1 APROBA and related web tools

The APROBA (“Approximate Probabilistic Analysis) tool was developed in the frame of the WHO/IPCS “Guidance document on evaluating and expressing uncertainty in hazard characterization” (WHO, 2014). In its original form it is EXCEL®-based, but two web-based forms have been made available by Chiu and colleagues (Chiu et al., 2018) at:

- [https://wchiu.shinyapps.io/APROBAweb/](https://wchiu.shinyapps.io/APROBAweb/)
- and as part of the Bayesian Benchmark Dose (BBMD) Analysis tool platform by Shao and Shapiro (2018) (at [https://benchmarkdose.org/](https://benchmarkdose.org/)).

The EXCEL®-form and the APROBAweb-version are discussed in more details. The BBMD version does not have the full functionality and no additional features.

The EXCEL® version has the following main features:

- APROBA is specifically designed for probabilistic hazard assessment
- as an EXCEL-based tool it works on the simplification that all distributions are lognormally distributed; this allows an analytical solution of the algorithms
- three types of dose-response data are discerned and dealt with in slightly different form (see chapter 2.2):
  - continuous data
  - quantal deterministic data
  - quantal stochastic data.
- BMDs, LOAELs or NOAEL can be used as input data for the point of departure
  - the BMD is assumed to be lognormally distributed and the BMDU and BMDL are required to define the distribution
  - default uncertainty distributions for NOAELs and LOAELs are implemented.
- There are default distributions implemented for
  - uncertainties associated with allometric scaling
  - interspecies extrapolation
  - exposure duration extrapolation
  - inter-individual variability
- The tool allows to compare with a deterministic assessment
- A sensitivity analysis in tabular form allows to assess the contribution of individual factors to the overall uncertainty
- Results are presented graphically and in tables.

In principal, the web-based version contains all these features as well. Differences between the two versions are listed in the following table.
Table 3-1: Comparison of main features of the EXCEL® and the web-based form of APROBA

<table>
<thead>
<tr>
<th></th>
<th>APROBA EXCEL®</th>
<th>APROBAweb</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>input data for BMD distribution</td>
<td>BMDL and BMDU</td>
<td>BMDL and BMD</td>
<td></td>
</tr>
<tr>
<td>default values for body weight</td>
<td>e.g. rat: 0.4</td>
<td>e.g. rat: 0.3045</td>
<td>different body weight defaults are used in the two versions, but can be changed</td>
</tr>
<tr>
<td>distributions for exposure duration, interspecies extrapolation</td>
<td>input data required as lower and upper bound (5th and 95th percentile) values</td>
<td>input data required as central value (50th percentile) and ratio 95/50 percentile</td>
<td>Example: sub-chronic to chronic defaults: xls: 0.5 / 8 web: 2 /4</td>
</tr>
<tr>
<td>distribution for inter-individual variability</td>
<td>input data required as lower and upper bound (5th and 95th percentile) values</td>
<td>input data required as GSDH (50th perc.) and log GSDH</td>
<td>for interpretation of GSDH see 2.2.5.4</td>
</tr>
<tr>
<td>graphical presentation</td>
<td>yes, plot of dose versus incidence, at various probability levels</td>
<td>yes, plot of dose versus incidence, at various probability levels (with normal and lognormal y scale); in addition PDF* of HDMI</td>
<td>extended graphical presentation in web tool</td>
</tr>
<tr>
<td>sensitivity analysis of uncertainties</td>
<td>in tabular form</td>
<td>in graphical form</td>
<td></td>
</tr>
<tr>
<td>export and report generation</td>
<td>no, but excel file can be saved</td>
<td>only export as csv</td>
<td></td>
</tr>
</tbody>
</table>

*Probability density function

The following figure shows the graphical presentation of results in the APROBA EXCEL® tool. On the x axis the dose is given (on a log scale). The y axis indicates the incidence or the percentage of the population covered (incidence 1% means 99% of
the population is covered). The coloured lines give the relationships between dose and incidence for selected probabilities (from left to right): 99%, 95%, 90%, 10%, 5%, 1%). For example, 99% probability here means that with a 99% percent probability the dose leading to the incidence on the y axis is not higher than the dose given on the x axis.

If incidence and probability are determined, then a fixed dose is obtained: $HD_M$ (blue square). The black vertical line indicates where a deterministic value would be located, using default assessment factors. The red vertical line is shown, if an exposure estimate is entered.

![Figure 3-1: Presentation of results in the APROBA EXCEL® tool](image)

The presentation of results in the web-based APROBA tool is very similar (see Figure 3-2). The incidence scale can be depicted either normal or on a logarithmic scale.
In addition, the web tool provides a probability density function of the resulting HDMI (with fixed incidence), describing the uncertainty of HDMI (note that the PDF takes the shape of a normal distribution, as the x axis (dose) is given in a log scale).

3.2 EFSA’s web-based tool

On behalf of EFSA a publicly accessible tool for Monte-Carlo simulations was developed respectively is under development (Seynaeve and Verbeke, 2017). The tool is accessible (after simple registration) at EFSA’s Statistical Models website\(^1\). Main features of the tool are:

- It is designed to be used for risk assessments (exposure, hazard, any kind), but it is not populated with default distributions.
- It allows to use predefined parametric distributions (with their parameter values), but it is also possible to enter distributions via a range of percentile values (various import functions are available for use of existing data; however, guidance needs to be developed for full usage).
- Algorithms (model equations) can be entered by an easy-to-use formula editor.
- Two modes are available for Monte Carlo simulations: simple random sampling and Latin hypercube sampling.
- There is a sophisticated presentation of results, in tabular and graphical form.
- Sensitivity analysis is performed as part of the routine and presented graphically.
- Input data and assessments can be stored and output reports can be exported in PDF or word format.

The tool was initially developed for specific problems in food safety, but is designed to be applied for any kind of question to be solved with a probabilistic approach. During its development it was compared and validated against @risk (see chapter 3.3), both with regard to model fitting and Monte Carlo analysis. So far, only a technical report describing the main features of the tool is available (Seynaeve and Verbeke, 2017). But a user manual is in preparation and is expected to become available until end of 2019 (José Cortinas Abrahantes, EFSA, personal communication, 10 Sept 2019).

The following figures illustrate some of the main features of the tool.

**Creation of distributions of input variables:**
Parametric distributions of input variables can be determined by entering determinants like GM and GSD (for a lognormal distribution), but also, as shown in the figure below, by determining the type of distribution (e.g. lognormal) and entering percentile values determining the shape of the distribution.

![Figure 3-4: Creation of a lognormal input distribution by providing percentiles](image-url)
Functionalities are planned to allow fits to empirical data, but these are not yet fully implemented.

**Editor for model equations**

![Editor for model equations](image)

Figure 3-5: Editor for model equations

An easy-to-use formula editor is available in order to connect input and output variables by a defined algorithm.

**Monte-Carlo simulation**

Simulation conditions can be set in the respective module and the simulation can be followed in real-time.
Results presentation

The resulting distribution is given in tabular form (last row in table), together with all input distributions, as well as in graphical form as probability density function (PDF, left) and cumulative distribution function (CDF, right).

Figure 3-7: Presentation of results in the EFSA tool
Input distributions are displayed in a similar way. Results can be exported as pdf, docx, or csv file.

### 3.3 Other tools for full probabilistic assessments

Various software implementations exist to perform Monte Carlo Simulations. Commercially available add-in based tools are e.g. @risk, Crystal Ball, Risk Solver or Model Risk. These tools can be used for fitting models to empirical distributions and provide a large variety of uncertainty distributions as potential data input for Monte Carlo simulations. A Bayesian approach of sample generation (Markov chain Monte Carlo) is implemented in non-commercially and freely available software as WinBUGS, JAGS or STAN (Chiu and Slob, 2015) or Nimble.

It should be noted that these software tools were not developed for the specific use in toxicology. Main application areas, for example of @risk are economics or insurance business. With the availability of the EFSA tool described there exist an easy to use alternative allowing a full probabilistic assessment.
4 Analysis and discussion of existing approaches

4.1 Principal approaches

The common principle in all probabilistic approaches published so far is to replace the input parameters of the algorithm

\[ HD = \text{POD} / (AF_1 \times AF_2 \times AF_3 \times \ldots) \]

by distributions characterising uncertainty and/or variability in these parameters.

The POD typically used in probabilistic approaches is the distribution to the benchmark dose (BMD), although APROBA also allows using NOAELs or LOAELs (with predefined uncertainty distributions).

In early publications distributions were merely reflections of the default assessment factors used (Baird et al., 1996; Baird et al., 2001; Price et al., 1997; Swartout et al., 1998). Authors from RIVM then proposed data-derived distributions for exposure duration and interspecies extrapolation (Slob and Pieters, 1998; Vermeire et al., 1999). Hattis and colleagues (Hattis et al., 2002) and later Schneider et al. (Schneider et al., 2004; Schneider et al., 2006) proposed empirical distributions also for intraspecies extrapolation. In principal, all systems allow to add additional uncertainty distributions and to replace empirical default distributions by substance-specific distributions for certain extrapolation steps.

All approaches published since 2001 use allometric scaling for accounting for differences in body size in addition to empirically derived distributions for various uncertainties. It can be concluded that there is a high agreement in the principal construction of probabilistic models for hazard assessment.

Slight methodological differences in how Monte Carlo simulations are performed (e.g. whether random or Latin Hypercube sampling is used) are not expected to lead to relevant differences.

With APROBA a tool is now available, which is based on the assumption that all distributions are lognormally distributed, which allows easy analytical solutions, which can be implemented in EXCEL®. The new EFSA tool, although still under development and without adequate guidance yet, is a ready-to-use tool for full probabilistic assessments. Practicality and adequacy of both tools is investigated with the two examples analysed in chapter 6.

4.2 Preconditions, pros and cons

Probabilistic methods are not yet often used for OEL setting. The model by Schneider et al. was specifically developed for the workplace, but is not used on a regular basis
(Schneider et al., 2006). ANSES is experimenting with a probabilistic approach, in which distributions for assessment factors are numerically set to represent default values used previously (Vernez et al., 2018). ANSES combines the probabilistic approach with expert decisions and no rule is mentioned by Vernez et al. for the coverage of the exposed population. For the two examples presented, OELs were around the 9th percentile of the resulting probability density distribution.

In order to perform a probabilistic assessment to derive a definite OEL value, principal regulatory decisions need to be taken:

- on the level of adversity by determining a benchmark response (BMR)
- on the percentage of the whole population, which should be covered by the OEL
- on the coverage or confidence of the assessment by determining the probability level (e.g. 95% percentile of the probability distribution).

If the first two parameters are set, then a distribution is obtained, which describes the probability at a given dose of not exceeding the chosen effect size and incidence. If all three parameters are set, a single dose value is obtained at which with at a selected confidence level (e.g. 5th percentile of the distribution for 95% confidence) the determined critical effect size (BMR) is not exceeded at the chosen incidence (e.g. in 99% of the target population).

Although assessors would prefer to derive “definitely” safe levels, it is obvious that uncertainties and probabilities are associated also with deterministically derived OELs. The main difference is that with deterministic OELs these uncertainties are hidden. Probabilistic methods provide more information on the characteristics of the outcome of the assessment. By explicitly stating the percentile of the population covered, the misinterpretation that the OEL is safe for everybody in a population of individuals with diverse susceptibilities is avoided (Hattis et al., 2002). On the other hand it can be argued that the complex presentation might imply a (too) high precision (Goble and Hattis, 2010).

Advantages of probabilistic methods:

- It requires responsible persons or bodies to make conscious choices on the aim of the assessment and the protection objectives of the OEL
- These choices are made transparent
- Uncertainty and variability are quantified
- Risk management receives information on the probability of adverse effects in the chosen part of the population in a three-dimensional matrix of dose, percent of population and probability of effects
- This matrix allows for conclusions on effects at dose levels above the OEL (see next chapter)
- These methods allow for a sensitivity analysis, which helps to focus resources on the most relevant sources of uncertainty.

Disadvantages:

- Probabilistic modelling requires a principal understanding of the approach and the interpretation of the results by risk managers, stakeholders etc.
- Its complexity might cause acceptance problems, requires education and may lead to a reduction in transparency (if not understood)
- It is more resource-intensive
- Ready-to-use tools (APROBA, EFSA tool) have been developed only recently; they need proper testing and introduction and use experiences need to be gathered in a larger community.

4.3 Uncertainty versus variability, secondary uncertainty

As already explained above, the inherent uncertainty of the BMD and values for interspecies extrapolation, exposure duration extrapolation and others is described and quantified by their distributions:

- The BMD distribution describes the uncertainty and variability observed in an experimental study under the given conditions, i.e. the uncertainty about the observed BMD being the “real” BMD.

- The distribution for interspecies extrapolation quantifies the uncertainty of not knowing whether for the assessed substance the differences in toxicokinetics and –dynamics between humans and experimental animals is larger or smaller than the central tendency of the empirical database of chemicals.

- The same holds true for exposure duration extrapolation: the distributions describe the ratios (chronic to shorter-term studies) found for a large set of substances; the uncertainty stems from not knowing which factor is the correct one for the substance in question.

Further uncertainty distributions can be added where required (e.g. to describe uncertainties in route-to-route extrapolation, uncertainties due to an incomplete database, or uncertainty introduced by applying read-across concepts). Also adjustments of the POD, for example for differences in the exposure scenarios of the animal experiment compared to the workplace situation, may contain uncertainties, which might be approximated by distributions.

The intraspecies distribution describes the variability in susceptibility between individuals in the target group, but also contains aspects of uncertainty: the distributions are derived from individual datasets on susceptibility differences between individuals of small groups (e.g. volunteers of a toxicokinetic study of a pharmaceutical) from a certain range of chemicals (many of them pharmaceuticals). Such datasets might underestimate the variability in the target population because the groups of volunteers (younger, healthy adults) and/or pharmaceuticals might not be representative for the more variable entity of the workforce and chemical substances used at the workplace (Hattis et al., 2002).

Differentiation between uncertainty and variability (within the worker population) might be simply achieved approximately by omitting/adding the intraspecies distribution, although (as discussed above) this distribution contains some aspects of uncertainty as well.
Some approaches (e.g. Oldenkamp et al., 2016) also include aspects of secondary uncertainty. Secondary uncertainties result from lack of knowledge on whether the chosen distributions are actually the right ones. For example, inter-study variation is included in the BMD distribution to a limited extent and repeating the experimental study several times might lead to a different BMD distribution. As mentioned above, the datasets used for describing inter-individual variability might not be representative with regard to the human individuals participating in those studies and the chemical substances used. This leads to secondary uncertainty not included in the model when using the distribution derived in the first place.

4.4 Types of dose-response data

Dose-response modelling typically distinguishes between continuous and quantal data (see separate report on “Dose-response modelling”, Part 2). The WHO/IPCS report further divides data presented in quantal form into quantal deterministic and quantal stochastic data (WHO, 2014), with consequences for the handling in the APROBA tool and for the interpretation of the resulting “target human dose”, $\text{HD}^t_M$.

The interpretation of $\text{HD}^t_M$ in the case of continuous data is that a part of the target population, which is defined by the incidence $I$ (e.g. 1% of the target population) can experience an effect defined by $M$ or higher. For example, if 5% increase in thyroid weight is chosen as $M$, then at or above dose $\text{HD}^t_M$ (with $I=1\%$ and the chosen probability) 1% of the population will suffer from a $\geq 5\%$ increase in thyroid weight. The POD used here, the $\text{BMD05-thyroid weight}$ represents the dose, at which the average group response in the experiment was 5% increase in thyroid weight.

For quantal data according to the WHO/IPCS document the assessor needs to decide whether the quantal effect is a “real” stochastic effect or whether there is an underlying (unknown or unreported) continuous effect. For the former type of effects, a POD of 10% extra risk is proposed. In contrast, for the latter the $\text{BMDL50}$ is proposed by the authors as a POD (personal communication with Wout Slob, 19 May 2019 and Weihsueh Chiu, 30 May 2019), because

- the difference between the BMDL50 and the BMDL10 reflects only the variability in the experimental animals, which is not of interest
- the underlying continuous effect, if measured, would be provided as group averages; hence, the quantification would not be based on the most sensitive animal
- the BMDL50 would better correlate with the effect level related to the critical continuous effect.

The relationship between the data types are shown in the following figure. This theoretical dataset consists of the individual responses of 5 animals in each of 4 dose groups (including control). The effect measured is weight of organ $X$, measured in g. In order to transform these continuous data into quantal data, a cut-off needs to be
defined. Here, 10 g organ weight is chosen: any individual animal showing a weight of organ X below 10 g is considered affected.

Table 4-1: Theoretical dataset – organ weights of animals in 4 dose groups

<table>
<thead>
<tr>
<th>Dose (mg/kg bw/day)</th>
<th>Organ weight of individual animals (g)</th>
<th>Arithmetic mean (g)</th>
<th>Incidence (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>17 11 14 15 13</td>
<td>14.0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>14 8 11 16 12</td>
<td>12.2</td>
<td>20</td>
</tr>
<tr>
<td>30</td>
<td>9 11 12 14 6</td>
<td>10.4</td>
<td>40</td>
</tr>
<tr>
<td>100</td>
<td>5 7 11 8 6</td>
<td>7.4</td>
<td>80</td>
</tr>
</tbody>
</table>

Figure 4-1: Dose-response data for continuous effect data and the transformed quantal data (bold line: adversity criterion: 10 g organ weight)

This example shows that indeed transformation of a set of continuous data into quantal data leads to a lower BMD(L), as the quantal BMD is based on the most sensitive animal(s): the BMD_{10} can be expected at approx. 5 mg/kg bw/day, whereas the BMD for the critical effect 10 g organ weight is close to 30 mg/kg bw/day.

As inter-individual variability in the human population is addressed separately, it can be argued that the variability between animals does not need to be considered. The BMD(L)_{50} is expected to be much closer to the BMD(L)_{critical continuous effect} than the BMD(L)_{10}. This problem was also already addressed in the BAuA research project (Schneider et al., 2004).

In contrast to these quantal deterministic data with underlying continuous effects, the authors assume that there can be real stochastic effects (i.e. effects, which, upon
repetition realise not with certainty, but only with a certain probability; examples:
tumour formation, malformations), for which not different susceptibilities of individual
animals causes the variability, but the random stochastic event. In that case they
propose to use the $\text{BMDL}_{10}$ as POD.

The practical consequences of this differentiation need further discussion. Whereas it
is likely that many or most quantal effects fall under the quantal deterministic category
(i.e. it is likely that there are some measurable physiological or pathogenic changes
occurring as the basis of the quantal effect), it is difficult to decide upon this in individual
cases. Also, deviating from the standard procedure of using the $\text{BMD(L)}_{10}$ for quantal
effects in dose-response modelling certainly would need a broader discussion and consensus in the risk assessor community. Further, examples are required to
demonstrate the practical and numerical consequences.

In the current situation assessments with quantal data, if carried out probabilistically,
should use the $\text{BMDL}_{10}$ as POD to allow comparing it with the traditional deterministic
approach. But wherever feasible, in addition, $\text{BMDL}_{50}$ should be used as POD to gather
experience and to discuss and compare the approaches.

4.5 Practicality

With APROBA (EXCEL® or web-form) and the EFSA web-based tool there are now
easy to use (approximate in the case of APROBA) probabilistic tools available, which
allow for a fast and easy application. The EFSA tool is under development and proper
guidance documentation is still required. This is expected to become available in the
near future. Nevertheless, in order for probabilistic assessments to be properly
performed, a good understanding of principles, objectives and rules by the assessor is
required.

APROBA is restricted to using predefined lognormal distributions for describing
assessment factors. In contrast, the EFSA tool allows fitting models to empirical
datasets or to use other parametric distributions and therefore is broader in its
applicability. However, no predefined default distributions are given. Therefore, more
expertise is required for the EFSA tool in defining the distributions. A detailed guidance
document (expected to come soon) is still lacking.

Both tools come with an easy to use sensitivity analysis, which indicates which input
distribution contributes most to overall uncertainty. This is very helpful asset for every
assessment. (But note that differences were observed in the sensitivity results
obtained with the examples in chapter 6, which need further analysis.)

With these available tools the next stage in probabilistic modelling can be achieved.
This would consist in probabilistic assessments performed in parallel to deterministic
ones for comparison and discussing pros and cons and possibilities for improvements
of the approach. Nevertheless, it is expected that the use of probabilistic tools will
increase only slowly. A main reason definitely is the higher complexity of probabilistic
approaches. Risk assessors need to be familiar with the tools, its advantages and
disadvantages. Probabilistic approaches might be especially useful under specific circumstances. Several applicability cases can be discerned for a substance, for which an OEL needs to be derived:

- Case 1: Uncertainty is assumed to be high, but exposures are orders of magnitude below the deterministic OEL
- Case 2: Uncertainty is low (e.g. OEL is based on qualitatively good human data), and exposure is in the range or above the OEL
- Case 3: Uncertainty is high and exposure is in the range or above the OEL.

In the first case there is not much need for refining the OEL by probabilistic methods, as it would not have practical consequences in a situation considered to be safe. In the second case sufficient knowledge is available to conclude on adequate risk management measures. Case 3 is the one, which would benefit from a probabilistic approach:

- the sensitivity analysis would indicate major sources of uncertainty (which might be reduced by additional efforts before investing into costly risk management measures)
- the derived probabilistic OEL would indicate the confidence the assessor can have when using the value
- the assessment would inform about the likeliness of adverse effects at the determined exposure levels (see following chapter).
5 Description of risks above the OEL

The derivation of OELs aims at identifying a concentration, below which adverse effects are unlikely to occur (see separate report on the “Comparison of methods” for details and differences in the definition of existing approaches). But for certain situations and regulatory problems it is also important to have information on the likelihood and incidence/severity of effects above the OEL. Examples for such problems are

- the OEL is exceeded in a certain real situation and adequate risk management measures need to be defined
- an impact assessment is performed to identify the consequences of setting an OEL at a certain level.

An example for the latter is the setting of binding OELs by the European Commission for carcinogens under the Carcinogens and Mutagens Directive (Directive 2004/37/EC). Although this is typically performed for non-threshold carcinogens, recently an impact assessment was carried out for substances with a presumed threshold (nickel compounds, benzene, acrylonitrile).

The dose-response modelling allows to predict the incidence (for quantal data) or the severity (in terms of the effect size in the case of continuous data) of effects above the benchmark dose, these predictions apply strictly only to the test system and test conditions applied to gain the dose-response data, typically the experimental setting of the toxicity study used (see separate report on “Dose-response modelling”). Only where adequate long-term human data are used, a direct conclusion may be derived for the situation in humans at higher doses.

The result of a probabilistic assessment is a three-dimensional matrix with the dimensions dose (or concentration), incidence and probability. The consequences of increasing the dose for the expected incidence of effects at a given probability can directly be derived from the obtained distribution. Especially, the graphical presentation in APROBA allows to rapidly conclude on the increase in incidence with increasing dose (see examples in chapter 6):

For Example 1 (renal tubule hyperplasia induced by 3-MPMD) the following result was obtained:

\[ \text{HDMI} = 0.0011 \text{ mg/kg bw/day} \]

M: extra risk of 10 percent for renal hyperplasia
I: 1% of the population
which indicates that at this dose – with a 95% confidence – 1% of the population will have an extra risk of 10% for renal tubule hyperplasia (which can be roughly approximated to a 0.1% extra risk for that effect in the population)

At a 10fold dose at the same confidence level of 95% the incidence level would be approx. 25% for a 10% extra risk.
6 Conclusions

Two recent developments in the area of probabilistic hazard assessment are described in this report:

The APROBA tool developed in the frame of the WHO/IPCS project on “Evaluating and Expressing Uncertainty in Hazard Characterization” (WHO, 2014) is an EXCEL®-based tool allowing approximations to full Monte Carlo analyses by using lognormal input distributions for all parameters.

The Monte Carlo tool developed by EFSA, currently under development at EFSA, allows for full Monte Carlo analyses, including distribution fitting and use of various kinds of distributions.

In view of these new developments use of probabilistic approaches to hazard assessment are simplified and their use for

- method development and discussion of (combination of) deterministic factors
- comparison with standard assessments using deterministic factors
- refined assessments of complex cases

is encouraged.
7  Annex 1: Examples

Both examples are based on oral studies. In order to keep these examples simple, no conversion to air concentrations is performed. As at this stage of the project no conclusions on adequate distributions for assessment factors can be drawn, the distributions as proposed by WHO in the APROBA tool are used. Although these distributions are meant for assessing risks in the general population, they are useful for demonstrating how the tools work and results can be interpreted. But note that using these distributions by no means does imply that they are recommended for deriving OELs.

7.1  Quantal data set

7.1.1  Input data – dose response data

Example dataset for quantal data:
- Substance: 3-monochloropropane-1,2-diol
- Study type: chronic toxicity study in rats
- Effect: renal tubule hyperplasia in male rats
- Source: Cho et al. (2008)

Dose-response data used for analysis:

<table>
<thead>
<tr>
<th>dose (mg/kg bw/day)</th>
<th>effect # affected animals</th>
<th># animals in group</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>1.97</td>
<td>11</td>
<td>50</td>
</tr>
<tr>
<td>8.27</td>
<td>21</td>
<td>50</td>
</tr>
<tr>
<td>29.50</td>
<td>36</td>
<td>50</td>
</tr>
</tbody>
</table>

Dose-response modelling with the EFSA Benchmark dose modelling tool,
- with BMR = 10% extra risk compared to the controls
- with model averaging

yielded the following result:

<table>
<thead>
<tr>
<th>BMDL</th>
<th>BMDU</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.19 mg/kg bw/day</td>
<td>1.88 mg/kg bw/day</td>
</tr>
</tbody>
</table>
7.1.2 Input data – uncertainty distributions

Allometric Scaling, based on rat weight 0.4 kg and human weight 70 kg:
- Median of allometric exponent: 0.7
- Standard deviation of estimate of exponent using in allometric scaling by body weight (based on 95% CI of 0.66 - 0.74)
- results in lognormal distribution of allometric scaling factor for rats (body weight 400 g) versus humans (70 kg):
  - 5 percentile 3.83
  - 95 percentile 5.97

Remaining interspecies variability:
- Median estimate of remaining chemical-specific TK and TD uncertainty after allometric scaling: 1
- Geometric standard deviation (GSD) of chemical-specific TK and TD uncertainty after allometric scaling: 1.95
- results in distribution with:
  - 5 percentile 0.33
  - 95 percentile 3

Intraspecies variability (humans):
- Median estimate of the Log(GSDh) for human variability: 0.324
- GSDu of the Log(GSDh) for human variability = (P95/P50)^^(1/1.645): 1.59
- results in (approximated) distribution with:
  - 50 percentile 9.69
  - 95 percentile 41.88
7.1.3 Modelling results

7.1.3.1 APROBA EXCEL®

<table>
<thead>
<tr>
<th>Probabilistic guidance value =</th>
<th>0.0011 mg/kg bw/day</th>
</tr>
</thead>
<tbody>
<tr>
<td>= Approximate probabilistic HD_{M1} at specified % confidence</td>
<td>= Estimate of dose (mg/kg body weight per day) at which, with 95% confidence of the population will have renal hyperplasia of magnitude $\geq$ 10%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>UNCERTAINTY ANALYSES</th>
<th>% contribution to overall uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASPECT</td>
<td></td>
</tr>
<tr>
<td>PoD</td>
<td>28%</td>
</tr>
<tr>
<td>Allometric scaling</td>
<td>1%</td>
</tr>
<tr>
<td>Interspecies TK/TD</td>
<td>26%</td>
</tr>
<tr>
<td>Intraspecies</td>
<td>46%</td>
</tr>
<tr>
<td>Greatest contributor to overall uncertainty</td>
<td>Intraspecies</td>
</tr>
</tbody>
</table>
7.1.3.2 EFSA tool

7.1.4 Summary

<table>
<thead>
<tr>
<th>variable</th>
<th>mean</th>
<th>sd</th>
<th>1%</th>
<th>5%</th>
<th>50%</th>
<th>90%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMD</td>
<td>0.7606</td>
<td>0.6049</td>
<td>0.1188</td>
<td>0.1904</td>
<td>0.5931</td>
<td>1.441</td>
<td>1.876</td>
<td>3.049</td>
</tr>
<tr>
<td>AS</td>
<td>4.741</td>
<td>0.5997</td>
<td>3.514</td>
<td>3.829</td>
<td>4.699</td>
<td>5.533</td>
<td>5.79</td>
<td>6.359</td>
</tr>
<tr>
<td>InterAF</td>
<td>1.246</td>
<td>0.9456</td>
<td>0.2131</td>
<td>0.3322</td>
<td>0.9915</td>
<td>2.361</td>
<td>2.996</td>
<td>4.639</td>
</tr>
<tr>
<td>IntraAF</td>
<td>14.31</td>
<td>15.72</td>
<td>1.212</td>
<td>2.195</td>
<td>9.712</td>
<td>30.14</td>
<td>41.48</td>
<td>77.93</td>
</tr>
<tr>
<td>GV</td>
<td>0.03121</td>
<td>0.06231</td>
<td>0.0006273</td>
<td><strong>0.00154</strong></td>
<td>0.01311</td>
<td>0.07234</td>
<td>0.1133</td>
<td>0.2874</td>
</tr>
</tbody>
</table>

The 5\textsuperscript{th} percentile of the distribution of the guidance value is 0.0015 mg/kg bw/day.
7.1.5 Comparison and discussion

The full probabilistic assessment carried out with the EFSA tool leads to a value with a 95% probability (5th percentile of the distribution) of 0.00154 mg/kg bw/day, compared to APROBA, which results in a value of 0.0011.

POD / APROBA HD\textsuperscript{M} = factor 173

POD / EFSA GV = factor 123.

It is reasonable that the full probabilistic model is somewhat less conservative compared to the approximated model (WHO, 2014). The results are reasonably close.

Note that the absolute figures and the absolute distance between POD and the obtained values are not meaningful, as long no distributions adequate for worker assessments are used.

The sensitivity analysis shows differences: Intraspecies extrapolation is considered to contribute most to uncertainty in APROBA, whereas the BMD accounts for the highest uncertainty in the EFSA tool. The reason for this discrepancy is that different algorithms are used for the sensitivity analysis: in APROBA for each variable the spread of the distribution is calculated relative to the sum of spreads, irrespective of whether it is placed in the enumerator or denominator. In the EFSA tool the impact of each variable on the outcome parameter is calculated on a normal scale. This implies that variables in the enumerator (BMD) get a higher sensitivity score than those in the denominator (José Cortinas Abrahantes, EFSA, personal communication, 18 Sept 2019).
7.2 Continuous dataset

7.2.1 Input data – dose response data

Example dataset for quantal data:
- Substance: nalidixic acid
- Study type: chronic toxicity study in rats
- Effect: body weight changes in male and female rats
- Source: NTP (1989), NTP TR No. 368

Dose-response data used for analysis:

<table>
<thead>
<tr>
<th>concentration in food (ppm)</th>
<th>bw</th>
<th>SEM</th>
<th>n</th>
<th>sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>0</td>
<td>29.4</td>
<td>1.13</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>1000</td>
<td>28.2</td>
<td>1.18</td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>2000</td>
<td>28.7</td>
<td>1.06</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>4000</td>
<td>27.1</td>
<td>0.42</td>
<td>10</td>
</tr>
<tr>
<td>11</td>
<td>8000</td>
<td>24.8</td>
<td>0.84</td>
<td>10</td>
</tr>
<tr>
<td>12</td>
<td>16000</td>
<td>23.6</td>
<td>0.55</td>
<td>10</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>36.1</td>
<td>0.89</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>1000</td>
<td>35.0</td>
<td>0.64</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>2000</td>
<td>34.9</td>
<td>0.71</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>4000</td>
<td>33.6</td>
<td>0.41</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>8000</td>
<td>32.4</td>
<td>0.47</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>16000</td>
<td>31.4</td>
<td>0.71</td>
<td>10</td>
</tr>
</tbody>
</table>

Dose-response modelling with the EFSA Benchmark dose modelling tool,
- with BMR = 10% relative difference in final body weight compared to the controls
- with model averaging

yielded the following result:

<table>
<thead>
<tr>
<th>endpoint</th>
<th>subgroup</th>
<th>BMDL (ppm food)</th>
<th>BMDU (ppm food)</th>
</tr>
</thead>
<tbody>
<tr>
<td>bw</td>
<td>f</td>
<td>1800</td>
<td>10700</td>
</tr>
<tr>
<td>bw</td>
<td>m</td>
<td>5950</td>
<td>14900</td>
</tr>
</tbody>
</table>

For the more sensitive female rats these food concentrations can be converted to a dose (with defaults as given in ECHA (2012)):

1800 ppm, with a food factor of 50 g per kg bw per day: 90 mg/kg bw/day
10700 ppm, with a food factor of 50 g per kg bw per day: 535 mg/kg bw/day
Note: in the EFSA tool with 90 and 535 mg/kg bw/day as 5th and 95th percentile, no lognormal distribution could be fitted. Instead a Weibull function with a similar shape was used.

7.2.2 Input data – uncertainty distributions

See chapter 7.1.2., apart from:

Allometric Scaling, for female rat weight 0.35 kg and human weight 70 kg:

- Median of allometric exponent: 0.7
- Standard deviation of estimate of exponent using in allometric scaling by body weight (based on 95% CI of 0.66-0.74)
- results in normal distribution of allometric scaling factor for female rats (body weight 350 g) versus humans (70 kg):
  o 5 percentile 3.97
  o 95 percentile 6.06

Note: in the EFSA tool with these percentile values, no lognormal distribution could be fitted. Instead a normal function with a similar shape was used.
### 7.2.3 Modelling results

#### 7.2.3.1 APROBA

| Probabilistic guidance value | = Approximate probabilistic $\text{HD}_{M1}^*$ at specified % confidence |
| 0.597 | = Estimate of dose (mg/kg body weight per day) at which, with 95% confidence 1% of the population will have reduced body weight development of magnitude $\geq 10\%$ |

<table>
<thead>
<tr>
<th>ASPECT</th>
<th>% contribution to overall uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>PoD</td>
<td>19%</td>
</tr>
<tr>
<td>Interspecies scaling</td>
<td>1%</td>
</tr>
<tr>
<td>Interspecies TK/TD</td>
<td>29%</td>
</tr>
<tr>
<td>Intraspecies</td>
<td>51%</td>
</tr>
<tr>
<td>Greatest contributor to overall uncertainty</td>
<td>Intraspecies</td>
</tr>
</tbody>
</table>
7.2.3.2 EFSA tool

<table>
<thead>
<tr>
<th>variable</th>
<th>mean</th>
<th>sd</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
<th>50%</th>
<th>90%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMD</td>
<td>294.1</td>
<td>135.3</td>
<td>43.91</td>
<td>91.78</td>
<td>126.8</td>
<td>283.6</td>
<td>475.5</td>
<td>537.4</td>
<td>645.1</td>
</tr>
<tr>
<td>AS</td>
<td>5.016</td>
<td>0.6365</td>
<td>3.524</td>
<td>3.973</td>
<td>4.196</td>
<td>5.012</td>
<td>5.822</td>
<td>6.07</td>
<td>6.5</td>
</tr>
<tr>
<td>InterAF</td>
<td>1.243</td>
<td>0.9726</td>
<td>0.2108</td>
<td>0.3292</td>
<td>0.4178</td>
<td>0.9824</td>
<td>2.362</td>
<td>3.017</td>
<td>4.85</td>
</tr>
<tr>
<td>IntraAF</td>
<td>14.33</td>
<td>15.5</td>
<td>1.248</td>
<td>2.26</td>
<td>3.102</td>
<td>9.63</td>
<td>30.55</td>
<td>41.45</td>
<td>75.72</td>
</tr>
<tr>
<td>GV</td>
<td>11.27</td>
<td>18.45</td>
<td>0.2754</td>
<td>0.6649</td>
<td>1.096</td>
<td>5.587</td>
<td>26.16</td>
<td>39.64</td>
<td>87.83</td>
</tr>
</tbody>
</table>

Histogram of GV

Sensitivity
7.2.4 Comparison and discussion

The full probabilistic assessment carried out with the EFSA tool leads to a very similar value as APROBA: with 95% probability (5th percentile of the distribution) GV is 0.66 mg/kg bw/day, compared to the value of 0.60 mg/kg bw/day obtained with APROBA.

POD / APROBA HDM = factor 151

POD / EFSA GV = factor 136.

Note that the absolute figures and the absolute distance between POD and the obtained values are not meaningful, as long no distributions adequate for worker assessments are used.

These examples will be reassessed based upon agreed distributions for all extrapolation steps later in the project.

The sensitivity analysis again shows some differences (see above for explanations): intraspecies extrapolation is contributing most to uncertainty in both tools, but in the EFSA tool uncertainty is more even distributed between intraspecies extrapolation, POD and interspecies extrapolation.
8 **Annex 2: Explanations**

8.1 **Distributions**

8.1.1 **Empirical distribution (frequency distribution)**

If a parameter $x$ is measured many times, the more likely values will be measured more often. A graphical presentation is given in Figure 8-1. The frequency distribution of measured continuous parameters can be visualised as such a histogram (Fahrmeier et al., 2001). A histogram approximates the underlying theoretical probability density function by a step function. It is generated by splitting the range of measured values into intervals and by showing for each interval the number of values in this interval.\(^2\)

![Figure 8-1: Example for an empirical distribution (x axis unitless)](image)

8.1.2 **Model fitting**

Curve fitting may lead to a parametric model representing the empirical distribution with high accuracy. Such curve fitting in the case of the empirical dataset above might

\(^2\) In a normalised histogram the $x$ axis gives the proportion (= number of values in the interval / number of all values) of values in this interval as a rectangle over the interval which has an area equal to this proportion. The area is calculated as interval width $x$ rectangle height, and the rectangle height is chosen such that the sum over all rectangle areas is equal to 1. Usually, all intervals have the same width, though this is not a necessary requirement. The number of intervals together with the interval limits control the appearance of the histogram. For equidistant interval limits there are recommendations regarding the number of intervals to employ, which aim at providing a good representation of the distribution shape.
lead to a lognormal distribution by its expected value $\mu$ and its standard deviation $\sigma$, both expressed on the log scale. The parameter $\mu$ is estimated from data as the arithmetic mean of the logarithms of the data), and the parameter $\sigma$ is estimated by the standard deviation of the data logarithms (natural logarithms or logarithm to base 10 can be used).

$$f(x) = \frac{1}{\sigma x \sqrt{2\pi}} \exp\left(-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right)$$

Figure 8-2: Probability density of a log-normal distribution

8.1.3 Probability density function versus cumulative distribution function

The distribution above shows a probability density function (PDF). Its advantage is to show the shape of the distribution more clearly than the alternative presentation as cumulative distribution function (CDF) discussed below. The probability of getting values lying between two limits $x_1$ and $x_2$ is the area below the curve (integral) between $x_1$ and $x_2$. The 10th percentile is the point on the x axis, under which 10% of all values lie (and, correspondingly, 90% are larger than the 10th percentile). Blue lines in the figure below show the 10th and 90th percentile for the density above.

An alternative presentation of a distribution is the cumulative distribution function, which has values between 0 and 1. For each x value the y axis gives the probability that the parameter has a value of x or below. The 10th and 90th percentiles are the values on the x-axis with y values of 0.1 and 0.9, resp. (blue lines in figure below). The advantage of a CDF is that probabilities can be read directly from the graph.
8.2 Explanation of terms

8.2.1 Bayesian methods

Bayesian methods is the summary term for a certain philosophy in statistics. It starts from the assumption that there is certain prior information about the problem to solve, typically from earlier studies. The prior information is typically an assumption about the distribution of the quantity to analyse. Prior information and data are combined using Bayes' law of total probability and result in a posterior distribution, which is the central result of the analysis. As an example, highest posterior density regions are used as
equivalent for confidence intervals known from standard (frequentist) analysis (Gelman et al., 1998; Lambert, 2018).

The concept of Bayes analysis is old, but has for a long time come out of focus, because the computational effort was too large. Meanwhile, due to fast computers and some mathematical developments, Bayesian approaches have become feasible. The philosophical counterpart to Bayesian analysis is frequentist analysis, which does not use prior information, but derives its results only from the analysed data.

### 8.2.2 Monte-Carlo simulation

**Monte Carlo simulation** (also MC experiment or MC analysis) is a mathematical approach in statistics for deriving (among others) the distribution of an estimated quantity or of a test statistic. It is also used to analyse the properties of a sampling design, of an experimental design or of more general stochastic processes. It is typically applied, when it is difficult or impossible to derive this distribution analytically.

Finding the distribution of an output statistic that results from combining values from various input distributions by a mathematical formula is an example for such a problem.

The idea of MC simulation is to generate many fictive input data sets of the type under study. From each data set, the interesting statistic (estimated value or test statistic) is calculated. Doing this for each fictive data set generates the desired distribution of the statistic. Often this distribution itself is of interest. It can also be used to e.g. calculate a confidence interval for an estimated parameter or the p value for an observed test statistic.

The basic operation in an MC simulation is generating the fictive input data sets. The procedure can be illustrated by using Figure 1-1. There, four input variables (POD, AF1, AF2, AF3) are combined to give the output GV by the equation $GV = POD/(AF1 \times AF2 \times AF3)$. All input variables are random variables. The distribution of each input variable is completely known (i.e. including the numerical values of the distribution parameters like mean value or standard deviation). Each input may have its own distribution. Together, all input distributions induce a distribution of GV, which is the quantity of interest.

For the MC solution, $n$ fictive datasets are generated. Each dataset consists of four numbers, one value for each of the four input variables. Each number is a computer generated random number from the relevant distribution. Such a random number is calculated by first drawing a number from a uniform distribution. All possible values of a uniform distribution have the same probability like (theoretically) the possible outcomes when playing roulette, hence the name “Monte Carlo” method. The uniform random number is then transformed to a random number having the required input distribution. The four transformed numbers are then combined by the above equation to obtain (one realisation of) the output quantity $GV$. Repeating the process of data generation and calculation of $GV$ $n$ times gives the desired distribution of $GV$. The larger $n$ is, the better is the knowledge of the $GV$ distribution.
This description of an MC simulation for independent inputs refers to the simplest form of a MC simulation. However, it already makes obvious that the MC simulation provides a result only for exact those input conditions that were used to generate the fictive data sets. Also the number $n$ of data sets is not obvious to determine. A large $n$ is desirable, but can generate large computing time. Therefore, methods have been developed to accelerate random number generation. Special action is also required, when input variables are not statistically independent from another, as assumed in the example.

MC analysis has the advantage of being conceptually simple and of being nearly universal in the sense that many problems can be solved that cannot be solved by other methods. MC analysis has the disadvantage of being (computer-) time-consuming, with some uncertainty whether the number $n$ of generated datasets is sufficient. Also, different from a formal solution, MC analysis gives no structural insights into the problem, but provides only a numerical solution under exactly the conditions that were used in the simulation. If a parameter of the input distribution is changed, the whole MC calculation must be done again, while the formal solution, if it exists, provides an answer for each input distribution parameter without new derivation (Gilks et al., 1996; Monteith et al., 2011; Vose, 1996).

8.2.3 Random sampling, Latin-Hypercube sampling, Markov Chain Monte Carlo sampling

There are several possibilities to perform drawings:

**Random sampling** means drawing values randomly from a distribution, without further conditions on the sampling process. The distribution is either given by a formula or by a density estimate or as the empirical distribution of a data set. The “random” component in this operation means that in the initial step of random number generation, described in the previous section, uniformly distributed numbers from the whole range $[0, 1]$ are used. As in practice the number of values is always finite, the distribution of the numbers actually drawn will not exactly be uniform. This causes transformed random numbers which do not have exactly the theoretically required form.

A random sample has the advantage that its statistical properties can easily be generalized to the underlying universe. However, random sampling requires large samples to ensure that generated numbers have the intended distribution. This holds especially if properties of an extreme part of the distribution are sought (e.g. the location of the 0.1% quantile). This has led to the development of other sampling schemes as described below.

**Stratified sampling** is an enhancement of random sampling, which subdivides the range of possible input values into disjoint subgroups (strata), then first selects a stratum to sample from and subsequently takes a random sample from the selected stratum. In this way, a good coverage of the universe can be achieved with smaller samples than under random sampling. This approach is also applicable for sampling multidimensional input variables.
Latin hypercube sampling is an enhancement of stratified sampling. It is particularly useful for sampling multidimensional input variables, when not all input variables are of same importance to the outcome. Latin hypercube sampling ensures that the full range of each input variable gets represented by the sample. If certain conditions hold for the relationship between input and outcome variables, then the variance of the output variable under Latin hypercube sampling is less than or equal to the variance under random sampling.

Markov Chain Monte Carlo (MCMC) is a concept to generate vectors of dependent random numbers. These are needed e.g. for an MC analysis, which deals with problems involving statistically dependent input random variables. In the example from chapter 8.2.2, such a dependency would occur if the probability of AF1 having a certain value would depend on the value of POD. In this case, random values for AF1 and POD can no more be obtained by independent drawing of these two values from the distributions of AF1 and POD, respectively. Instead, the pair (AF1, POD) must be drawn from the joint distribution of AF1 and POD. The joint distribution is not the product of the two single distributions, but must be obtained from corresponding two-dimensional data on the joint occurrence of AF1 and POD. A simple example of a two-dimensional distribution is that ln(AF1) and ln(POD) have a joint normal distribution, which is characterized by a vector of means $(\mu_{AF1}, \mu_{POD})$ and a $2 \times 2$ covariance matrix containing the variances $(\sigma_{AF1})^2$ and $(\sigma_{POD})^2$ and the covariance $\sigma_{AF1,POD}$, which quantifies the degree of dependency between both variables.

The MCMC method starts like a usual MC analysis, but then generates vectors of dependent input variables by feeding independent random numbers into a Markov chain. A Markov chain is a stochastic process in discrete space and over discrete time, where the state of the chain at time $t+1$ depends only on the state at time $t$, but not on earlier states. This property allows a simple computer generation of dependent random numbers (Gelman et al., 1998; Gilks et al., 1996).

8.2.4 Bootstrapping

Bootstrapping (BS) is a technique to obtain properties of a statistic derived from data by resampling new data sets from the same data from which the statistic was derived. It is closely related to MC simulation, because the desired properties result from computing the statistic from randomly selected data sets. However, BS samples its data sets only on basis of the real data, not from distributions that were selected from additional considerations. There are several versions of bootstrapping: empirical BS takes random samples from the real data set, parametric BS fits a parametric distribution to the real data sets and generates random numbers from the fitted distribution. Semi-parametric BS fits a nonparametric density to the data (e.g. by a kernel density estimate) and gets random numbers from this density (Davison and Hinkley, 2009; Efron and Tibshirani, 1994).
9 References

Uncertainty and variability analysis

Noncancer risk assessment: a probabilistic alternative to current practice
Human and Ecological Risk Assessment, 2, 79-102

Probabilistic noncancer risk estimation
Comments on Toxicology, 7, 541-574

A comparison of ratio distribution based on the NOAEL and the benchmark approach for subchronic- to-chronic extrapolation
Toxicological Sciences, 85, 1033-1040

APROBA-Plus: A probabilistic tool to evaluate and express uncertainty in hazard characterization and exposure assessment of substances
Food and Chemical Toxicology, 110, 408-417

Probabilistic risk characterization: an example with di(2-ethylhexyl)phthalate
Regulatory Toxicology and Pharmacology, 43, 104-113

A unified probabilistic framework for dose-response assessment of human health effects
Environmental Health Perspectives, 123, 1241-1254

Beyond the RfD: Broad application of a probabilistic approach to improve chemical dose-response assessments for noncancer effects
Environmental Health Perspectives, 126, 067009

Carcinogenicity study of 3-monochloropropane-1,2-diol in Sprague-Dawley rats
Food and Chemical Toxicology, 46, 3172-3177

Cooke, R. (2010)
Conundrums with uncertainty factors
Risk Analysis, 30, 330-339

Issues in using human variability distributions to estimate low-dose risk
Environmental Health Perspectives, 118, 387-393

Bootstrap methods and their application
Cambridge University Press, Cambridge

Human variability in xenobiotic metabolism and pathway-related uncertainty factors for chemical risk assessment: a review
_Food and Chemical Toxicology_, 43, 203-216

ECHA, European Chemicals Agency (2012)

An introduction to the bootstrap
Chapman & Hall, Boca Raton

Statistik. Der Weg zur Datenanalyse
Springer-Verlag, Berlin Heidelberg New York

Bayesian data analysis
Chapman & Hall, London

Markov chain Monte Carlo in practice
Chapman & Hall, London

Are conundrums with uncertainty factors an obstacle to developing probabilistic interpretations of noncancer risks from chemicals?
_Risk Analysis_, 30, 340-345; author reply 353

A Straw Man Proposal for a Quantitative Definition of the RfD

Empirically Observed Distributions of Pharmacokinetic and Pharmacodynamic Variability in Humans Implications for the Derivation of Single Point Component Uncertainty Factors Providing Equivalent Protection as Existing RfDs
In: Lipscomb, J.C.; Ohanian, E.V., Toxicokinetics in risk assessment, Informa Healthcare USA, Inc., 69-93

Lambert, B. (2018)
A student’s guide to Bayesian statistics
Mekel, O.; Fehr, R. (2014) Probabilistic methods in toxicology
In: Reichl, F.X.; Schwenk, M., Regulatory Toxicology, Springer Berlin Heidelberg, 253-259

Monteith, K.; Carroll, J.; Seppi, K.; Martinez, T. (2011) Turning Bayesian Model Averaging into Bayesian Model Combination
Proceedings of the International Joint Conference on Neural Networks IJCNN’11, 2657-2663

NRC, National Research Council (2009) Science and Decisions: Advancing Risk Assessment

Critical Reviews in Toxicology, 46, 261-278

Risk Analysis, 17, 427-437

Regulatory Toxicology and Pharmacology, 27, 3-20


Regulatory Toxicology and Pharmacology, 44, 172-181


Shao, K.; Shapiro, A.J. (2018) A Web-Based System for Bayesian Benchmark Dose Estimation
Environmental Health Perspectives, 126, 017002

Bayesian methods for uncertainty factor application for derivation of reference values
*Regulatory Toxicology and Pharmacology*, 80, 9-24

A probabilistic approach for deriving acceptable human intake limits and human health risks from toxicological studies: general framework
*Risk Analysis*, 18, 787-798

Slob, W.; Bakker, M.I.; te Biesebeek, J.D.; Bokkers, B.G.H. (2014)
Exploring the uncertainties in cancer risk assessment using the Integrated Probabilistic Risk Assessment (IPRA) approach
*Risk Analysis*, 34, 1401-1422

A probabilistic framework for the reference dose
*Risk Analysis*, 18, 271-282

Integration of probabilistic exposure assessment and probabilistic hazard characterization
*Risk Analysis*, 27, 351-371

A model for probabilistic health impact assessment of exposure to food chemicals
*Food and Chemical Toxicology*, 47, 2926-2940

Assessment factors for human health risk assessment: a discussion paper
*Critical Reviews in Toxicology*, 29, 439-490

Experiencing a probabilistic approach to clarify and disclose uncertainties when setting occupational exposure limits

Quantitative Risk Analysis: A guide to Monte Carlo simulation modelling
John Wiley & Sons, Chichester

WHO, World Health Organization (2014)