MCRA Documentation

Release 9.0

Biometris, Wageningen University and Research

Jan 08, 2020

CONTENTS

1 About the toolbox **3**
1.1 Data and calculation model . . . 3
1.1.1 Modular design . . . 3
1.1.2 Nominal run and uncertainty analysis . . . 4
1.1.3 Retain & Refine and tiered approaches . . . 4
1.1.4 Uncertainty . . . 4
Uncertainty due to limited sampled data . . . 5
Uncertainty due to missing data . . . 6
Uncertainty due to modelling approach . . . 6
1.2 Data repository . . . 6
1.2.1 Repository access levels . . . 6
1.2.2 Linking remote data repositories . . . 7
1.3 Workspaces and actions . . . 9
1.3.1 Workspace browser . . . 9
1.3.2 Workspace overview page . . . 10
1.3.3 Action page . . . 10
Scoping: entity selection . . . 11

2 Modules **17**
2.1 Primary entity modules . . . 17
2.1.1 Effects . . . 18
Effects data formats . . . 18
Effects settings . . . 19
Effects as data . . . 20
2.1.2 Foods . . . 20
Foods data formats . . . 20
Foods as data . . . 23
Food coding systems . . . 23
FoodEx2 . . . 27
Food hierarchies . . . 30
2.1.3 Populations . . . 31
Populations data formats . . . 31
Populations settings . . . 32
Populations as data . . . 32
2.1.4 Responses . . . 32
Responses data formats . . . 33
Responses settings . . . 33
Responses as data . . . 34
2.1.5 Substances . . . 34
Substances data formats . . . 34
Substances settings . . . 35
Substances as data . . . 35
2.1.6 Test systems . . . 35
Test systems data formats . . . 36

Test systems as data . . . 37
2.2 Consumption modules . . . 37
2.2.1 Consumptions . . . 37
Consumptions data formats . . . 37
Consumptions settings . . . 41
Consumptions uncertainty . . . 42
Consumptions as data . . . 42
2.2.2 Food recipes . . . 42
Food recipes data formats . . . 42
Food recipes as data . . . 43
2.2.3 Market shares . . . 43
Market shares data formats . . . 43
Market shares as data . . . 44
Market shares and brand loyalty . . . 44
2.3 Occurrence modules . . . 45
2.3.1 Concentration limits . . . 45
Concentration limits data formats . . . 45
Concentration limits as data . . . 46
2.3.2 Concentration models . . . 46
Concentration models calculation . . . 46
Concentration models settings . . . 51
Concentration models tiers . . . 53
Input tiers . . . 55
Input tiers . . . 56
Concentration models uncertainty . . . 56
Calculation of concentration models . . . 56
2.3.3 Concentrations . . . 56
Concentrations data formats . . . 56
Concentrations calculation . . . 63
Concentrations settings . . . 66
Concentrations tiers . . . 67
Concentrations uncertainty . . . 68
Concentrations as data . . . 68
2.3.4 Focal food concentrations . . . 68
Focal food concentrations data formats . . . 68
Focal food concentrations as data . . . 68
2.3.5 Food extrapolations . . . 69
Food extrapolations data formats . . . 69
Food extrapolations as data . . . 69
2.3.6 Foods as measured . . . 70
Foods as measured calculation . . . 70
Foods as measured settings . . . 70
Calculation of foods as measured . . . 70
2.3.7 Occurrence patterns . . . 70
Occurrence patterns data formats . . . 71
Occurrence patterns calculation . . . 72
Occurrence patterns settings . . . 72
Occurrence patterns tiers . . . 73
Input tiers . . . 74
Input tiers . . . 74
Occurrence patterns as data . . . 74
Calculation of occurrence patterns . . . 74
2.3.8 Processing factors . . . 74
Processing factors data formats . . . 74
Processing factors settings . . . 79
Processing factors uncertainty . . . 79
Processing factors as data . . . 79
2.3.9 Substance authorisations . . . 79

Substance authorisations data formats 80
Substance authorisations as data 80
2.3.10 Substance conversions 80
Substance conversions data formats 80
Substance conversions as data 82
2.3.11 Total diet study sample compositions 82
Total diet study sample compositions data formats 83
Total diet study sample compositions as data 83
2.3.12 Unit variability factors 83
Unit variability factors data formats 83
Unit variability factors as data 84
2.4 Exposure modules 84
2.4.1 Consumptions per food as measured 85
Consumptions by food-as-measured calculation 85
Consumptions by food as measured 85
Calculation of consumptions by food as measured 85
2.4.2 Dietary exposures 85
Dietary exposures calculation 86
Dietary exposures settings 98
Dietary exposures tiers 100
Input tiers 102
Input tiers 102
Input tiers 103
Input tiers 104
Calculation of dietary exposures 104
2.4.3 Dietary exposures with screening 104
Dietary exposures with screening calculation 104
Dietary exposures with screening settings 105
Calculation of dietary exposures with screening 105
2.4.4 Exposures 106
Exposures calculation 106
Exposures settings 116
Calculation of exposures 118
2.4.5 Exposure mixtures 119
Exposure mixtures calculation 119
Exposure mixtures settings 122
Calculation of exposure mixtures 128
2.4.6 Food conversions 128
Food conversions calculation 128
Food conversion settings 129
Calculation of food conversions 131
2.4.7 Human monitoring analysis 131
Human monitoring analysis calculation 131
Human monitoring analysis settings 132
Calculation of human monitoring analysis 133
2.4.8 Human monitoring data 134
Human monitoring data data formats 134
Human monitoring data settings 140
Human monitoring data as data 140
2.4.9 Non-dietary exposures 141
Non-dietary exposures data formats 141
Non-dietary exposures settings 146
Non-dietary exposures uncertainty 146
Non-dietary exposures as data 146
2.5 Hazard modules 146
2.5.1 Active substances 147
Active substances data formats 147
Active substances calculation 149

Active substances settings . . . 150
Active substances as data . . . 150
Calculation of active substances . . . 150
2.5.2 AOP networks . . . 151
AOP networks data formats . . . 151
AOP networks settings . . . 153
AOP networks as data . . . 153
2.5.3 Dose response data . . . 153
Dose response data data formats . . . 153
Dose response data settings . . . 159
Dose response data as data . . . 159
2.5.4 Dose response models . . . 160
Dose response models data formats . . . 160
Dose response models calculation . . . 163
Dose response models as data . . . 163
Calculation of dose response models . . . 163
2.5.5 Effect representations . . . 164
Effect representations data formats . . . 164
Effect representations as data . . . 164
2.5.6 Hazard characterisations . . . 165
Hazard characterisations calculation . . . 165
Hazard characterisations settings . . . 171
Calculation of hazard characterisations . . . 172
2.5.7 Inter-species conversions . . . 172
Inter-species conversions data formats . . . 172
Inter-species conversions settings . . . 173
Inter-species conversions as data . . . 174
2.5.8 Intra species factors . . . 174
Intra-species factors data formats . . . 174
Intra species factors settings . . . 175
Intra species factors as data . . . 175
2.5.9 Points of departure . . . 175
Points of departure data formats . . . 176
Points of departure settings . . . 178
Points of departure as data . . . 178
2.5.10 Relative potency factors . . . 179
Relative potency factors data formats . . . 179
Relative potency factors calculation . . . 180
Relative potency factors settings . . . 180
Relative potency factors as data . . . 180
Calculation of relative potency factors . . . 181
2.6 In-silico modules . . . 181
2.6.1 Molecular docking models . . . 181
Molecular docking models data formats . . . 181
Molecular docking models as data . . . 183
2.6.2 QSAR membership models . . . 183
QSAR membership models data formats . . . 183
QSAR membership models as data . . . 185
2.7 Kinetic modules . . . 185
2.7.1 Kinetic models . . . 185
Kinetic models data formats . . . 186
Kinetic models settings . . . 187
Kinetic models as data . . . 188
Available kinetic models . . . 188
2.8 Risk modules . . . 201
2.8.1 Risks . . . 201
Risks calculation . . . 201
Risks settings . . . 201

Calculation of risks 202

3 Examples **219**
3.1 Cumulative dietary exposure assessment 219
3.1.1 Introduction 219
3.1.2 Preparation 219
3.1.3 Example 1 219
3.1.4 Example 2 222
3.1.5 Example 3 222

4 Appendices **223**
4.1 Munro collection 223
4.2 Unit definitions 223
4.2.1 Benchmark response types 223
4.2.2 Body weight units 224
4.2.3 Concentration units 225
4.2.4 Consumption units 225
4.2.5 Dose response model types 225
4.2.6 Dose units 226
4.2.7 Exposure route types 227
4.2.8 Exposure units 228
4.2.9 Point of departure types 228
4.2.10 Response types 229
4.2.11 Test system types 229
4.3 Transformations 229
4.3.1 Box Cox power transformation 229
4.4 Gauss-Hermite 230
4.4.1 Gauss-Hermite integration 230
4.4.2 One-dimensional Gauss-Hermite integration 230
4.4.3 Two-dimensional Gauss-Hermite integration 230
4.4.4 Maximum likelihood for the LNN model with two-dimensional Gauss-Hermite integration 231
4.5 Concentration models 231
4.5.1 Mixture zero spike and censored lognormal 231
4.5.2 Censored lognormal 231
4.5.3 Mixture non-detect spike and truncated lognormal 232
4.5.4 Mixture non-detect spike and lognormal 232
4.6 Chronic exposure assessment, daily consumed foods 232
4.6.1 Model based usual intake 232
4.6.2 Model based usual intake on the transformed scale 232
4.6.3 Model based using a logarithmic transformation 233
4.6.4 Model based using a power transformation 233
4.6.5 Model assisted usual intake on the transformed scale 233
4.6.6 Model assisted using a logarithmic transformation 234
4.6.7 Model assisted using a power transformation 235
4.7 Chronic exposure assessment, episodically consumed foods 235
4.7.1 Beta-Binomial model for frequencies (BBN) 236
4.7.2 Model based frequencies for usual intake 236
4.7.3 Model assisted frequencies for usual intake 237
4.7.4 Logistic-Normal model for frequencies (LNN0) 237
4.7.5 Model based frequencies for usual intake 238
4.7.6 Model assisted frequencies for usual intake 238
4.7.7 Logistic-Normal model for frequencies correlated with amounts (LNN) 238
4.7.8 Model based usual intake. 238
4.7.9 Model assisted usual intake. 239
4.8 Unit variability 239
4.8.1 Beta distribution 239
4.8.2 Lognormal distribution 240
4.8.3 Bernoulli distribution 241

4.9 Screening calculation for large Cumulative Assessment Groups 241
4.9.1 Statistical model for the screening step (acute exposure) 241
4.9.2 Statistical model for the screening step (chronic exposure) 243

5 Colophon **245**
5.1 Contributors to MCRA 245

Bibliography **247**

Reference and user manual for MCRA 9.

CHAPTER

ONE

ABOUT THE TOOLBOX

Humans are exposed to a mixture of multiple chemicals via food intake, inhalation and dermal contact. The risk to health that may result from this depends on the effects of different chemicals in the mixture and how they combine.

MCRA 9 is the model and data toolbox developed in the EuroMix project (http://www.euromixproject.eu). It implements methods for exposure, hazard and risk assessment, following guidelines from a.o. the Joint Research Centre (JRC) of the European Commission and the European Food Safety Authority (EFSA). The toolbox should provide computational tools for future risk management decisions on the safety of chemicals in mixtures to be taken by the European Commission and the Codex Alimentarius.

MCRA 9 is a collection of data and models. The system consists of modules that are arranged in eight categories according to a *modular design*. See *Modules overview*.

Each module represents a certain type of data, which can be computed from data provided by other (sub)modules, or the data may be obtained from a dataset selected from the *data repository*. Likewise, each module may be of interest by its own merit, or may just be required as a sub-part of larger calculations. The modular design of the toolbox reveals a network of data and models, and shows how data of types and from various sources can be combined in overarching modules. The most overarching module is *health impact estimates*. The toolbox allows the user to start in any of the modules in the modular design for performing calculations.

For each module, an *action* can be created to configure and run the module. For data modules, such as the concentrations module, such an action comprises specifying the dataset, specifying the scope (i.e., foods of interest, substances of insterest, etc.), and perhaps specifying specific selections or model settings for data manipulations (e.g., imputation of water concentrations in the concentrations module). For calculation modules, when calculating the data of the module based on other data, configuration of an action comprises specification of the model settings and selection of the calculation inputs, which is data provided by other (sub-)modules. When running an action in the toolbox, the module produces output of its associated data type (which can be used as input for other modules), and a report will be generated of the selected data, the selection and model settings, and the module and all intermediate (i.e., sub-modules) results.

1.1 Data and calculation model

1.1.1 Modular design

The modular design distinguishes between three types of modules: primary entity modules, data modules, and calculation modules. For an overview see under *Modules*

- The primary entity modules are data modules determining the scope of the assessments in the toolbox. That is, in each assessment, the scope specifies the *foods*, *substances*, *effects*, *populations*, *responses*, and/or *test systems* that are of interest.
- The data modules give summaries of the available data which depend on (some of) the primary entities. For example *consumptions* data.
- The calculation modules perform calculations on input data to produce data on another type, as specified by the module name. E.g. the *dietary-exposures* calculation module calculates dietary exposures from consumption and occurrence data. Some calculation modules can also act as a data module, in which case the data are directly specified rather than calculated. An example of this is the *relative potency factors* module.

1.1.2 Nominal run and uncertainty analysis

The toolbox distinguishes between two types of simulation runs; the nominal run and the uncertainty analysis loop. The nominal run represents a single simulation run in which the aim is to compute the most likely, unbiased estimates for the model at hands. E.g., when computing dietary exposure distributions, the nominal run computea single exposure distribution, using nominal values for all uncertain values. In the uncertainty analysis, on the other hand, the simulation run is repeated a number of times, each time with a different uncertainty scenario obtained using bootstrapping, parametric resampling, and/or re-calculation of uncertain values, yielding uncertainty distributions and confidence intervals for the nominal estimates.

Making the distinction between the nominal run and the uncertainty loops has the practical advantage that it allows the user to setup and evaluate complex simulations first using only the nominal runs to quickly obtain a picture of the results and identify possible errors in the data or in the model settings before running the more time-consuming uncertainty analysis loop.

1.1.3 Retain & Refine and tiered approaches

A basic idea of Retain & Refine is that entities (e.g., substances) can be handled in different ways (more or less refined) while still being considered together in the same risk assessment (retain). We refer to such different ways as tiers.

In the modular design, a tier is defined here as a specific set of settings for a module or a group of modules. Tiers can differ in many respects, and there is no single dimension to rank tiers as low vs. high. In risk assessment, typical tiers contrast deterministic to probabilistic approaches, conservative to realistic approaches, approaches using restricted data to approaches using more extensive data, and approaches using different degrees of model complexity. For each of the modules of the toolbox, as many tiers are implemented as considered useful for the practice of risk assessment.

Each calculation in the modular design may involve multiple, nested, calculations of sub-modules. A risk (or health impact) assessment builds on an exposure assessment and a hazard assessment, the exposure assessment builds on a dietary and a non-dietary exposure assessment, the dietary exposure assessment builds on a consumption assessment and an occurrence assessment, etc. Tiers can be defined at each node of the assessment network. An example consists of the tiers 'IESTI', 'EFSA basic optimistic' and 'EFSA basic pessimistic' which are defined at the level of a dietary exposure assessment, but include the settings for the corresponding tiers at the level of the concentration model calculator.

Each calculator has as a main output entities that can be specified to have different tiers (tiered entities). For example, in a hazard assessment, some substances may be assessed using a tier 'Hazard Dose from dose-response data', other compounds may be assessed using a tier 'TTCx100' or 'sample from general NOAEL distribution x100' (which only requires knowledge of the Cramer class of the compound). As another example, in dietary exposure assessment some food-compound combinations may be recognised as risk drivers for which a more complex approach (e.g. probabilistic modelling) is required, whereas a simpler approach (e.g. deterministic modelling) may be sufficient for all other food-compound combinations. So in this case the tiered entity is 'food-compound'. A typical risk assessment will start at a tier that is simple to perform for all tiered entities (potential risk drivers). Note that, based on data availability and ease of application, the initial assessment can already include more complex elements, such as probabilistic modelling. If the initial calculations produce risk estimates that do not exclude concern, refinement of the modelling for the perceived risk drivers is useful for checking whether this concern is real.

1.1.4 Uncertainty

Uncertainties may arise in different forms in many of the models and data of the toolbox. One may encounter uncertainty in the data values (e.g., uncertain NOAELs, uncertain RPFs, or uncertain processing factors), uncertainty due to limited data (e.g., a limited number of food samples), uncertainty due to a lack of data (e.g., missing concentration data for some foods/substances or missing processing factors), and uncertainty of the models, (e.g., due to a lack of detail). In many situations it is desirable to analyse how the model outcomes vary for the different scenarios that uncertainties give rise to. For this, the toolbox offers:

1) for many types of data, the possibility to provide data including quantifications of uncertainty for many types of data,

2) imputation methods for filling in missing data in various types of models, and
3) a generic uncertainty analysis method that providing uncertainty estimates of the modelling results for many of the modules, which are based on bootstrapping, parametric resampling, and/or re-calculation on all sub-modules for which this is possible.

Uncertainty due to limited sampled data

For some type of data, e.g., processing factors, it may be that in some cases it is possible to not only provide nominal estimates of the data values, but also to provide quantified estimates of the uncertainties of these values. In other cases, it may happen that quantifications of the uncertainties of these estimates are not available. In the toolbox, the aim is to provide the possibility to work with both quantified and unquantified uncertainties. That is, include quantified uncertainties in a quantitive uncertainty analysis when available, or to ignore their absence and only use the nominal estimates, perhaps in combination with an offline qualitative uncertainty analysis.

Uncertainties of the data values may be expressed in different forms, and it depends on the type of data which forms are available, suitable, and implemented in the toolbox. For some data values, uncertainty may be quantified by means of parametric distribution parameters (e.g., processing factor uncertainties, or kinetic model instance parameter uncertainties). Alternatively, uncertainty values may be provided in the form of an empirical set of uncertainty values (e.g., relative potency factor uncertainties, or points of departure uncertainties).

Whenever data include quantified uncertainties, and the data module to which they belong is included as a sub-module of a calculation module. These uncertainties may be chosen to be included in an uncertainty analysis of the main module, and if this is so, the data values are resampled in each uncertainty analysis cycle based on the uncertainty quantifications.

The basic acute exposure distribution is estimated in a Monte Carlo simulation by combining dietary consumption records (person-days) with sampled residue values. The resulting distribution represents a combination of variability in consumption within the population and between residues in a food lot. Percentiles may be used for further quantification e.g. the median or 99th percentile. Due to the limited size of the underlying data, these outcomes are uncertain. Confidence (or uncertainty) intervals reflect the uncertainty of these estimates, where MCRA uses bootstrap methodology and/or, depending on the available data, parametric methods to estimate the uncertainty.

Empirical method, resampling

The empirical bootstrap is an approach to estimate the accuracy of an outcome. In its most simple, non-parametric form, the bootstrap algorithm resamples a dataset of n observations to obtain a *bootstrap sample* or *resampled set* of again n observations (sampling with replacement, that is: each observation has a probability of $1/n$ to be selected at any position in the new resampled set). By repeating this process B times, one can obtain B resampled sets, which may be considered as alternative data sets that might have been obtained during sampling from the population of interest. Any statistic that can be calculated from the original dataset (e.g. the median, the standard deviation, the 99th percentile, etc.) can also be calculated from each of the B resampled sets. This generates a *uncertainty distribution* for the statistic under consideration. The uncertainty distribution characterises the uncertainty of the inference due to the sampling uncertainty of the original dataset: it shows which statistics could have been obtained if random sampling from the population would have generated another sample than the one actually observed [18], [19].

Parametric methods

Instead of bootstrapping the observed data, inference about parameters is based on parametric methods. For processing, where factors are specified through a nominal and/or upper value this is the natural choice. For concentration data, where the lognormal model is used to represent less conservative scenario's (EFSA, 2012) [3], the parametric bootstrap may be an alternative, especially when data are limited and the empirical bootstrap fails.

According to Cochran's theorem, sample variance $\hat{\sigma}_y^2$ follows a scaled chi-square distribution. In the parametric bootstrap for the lognormal distribution, the sample variance $\hat{\sigma}_y^2$ is replaced by a random draw from a chi-square distribution with $n_1 - 1$ degrees of freedom; the sample mean $\hat{\mu}_y$ is replaced by a random draw from a normal distribution with parameters $\hat{\mu}_y$ and $\hat{\sigma}_y^{*2}/n_1$, giving a new set of parameters $\hat{\mu}_y$ and $\hat{\sigma}_y^{*2}$. This is repeated B times.

For the truncated lognormal and censored lognormal, large sample maximum likelihood theory is used to derive new parameters $\hat{\mu}_y$ and $\hat{\sigma}_y^{*2}$. This is repeated B times.

The binomial fraction of non-detects for the mixture lognormal and mixture truncated distribution is sampled using the beta distribution with uniform priors $a = b = 1$ (with the *beta* distribution as the empirical Bayes estimator for the binomial distribution). This is repeated B times.

Uncertainty due to missing data

In some cases, it may be that data as only available for specific (primary)entities and missing for others. E.g., points of departure (such as NOAELs or BMDs) may only be available for some of the substances of interest.

Uncertainty due to modelling approach

There is also uncertainty of model outcomes that may arise by conducting different modelling approaches or applying alternative modelling assumptions.

Note: TODO

1.2 Data repository

Figure 1.1 shows a screenshow of the toolbox data repository browser. The data repository provides users with the functionality upload and organise their own datasets and share these with other users. The data sources available in the data repository can be used directly as data sources for *modelling actions*. Each user has a private repository for which the user is free to upload data files and organise these files into folders and sub-folders at own liking. Besides this, the user may be granted access to one or more shared repositories, which are repositories that are shared, maintained, and used by multiple users. The datasets of these shared repositories may be used by the users in their own calculations.

The central panel of the repository browser shows the data sources and sub-folders of the currently opened folder/repository. The top bar of the repository browser shows the path of the currently opened repository, buttons to collapse/expand the repository folder tree-view sidebar on the left ≡ and the info-sidebar on the right, and a button to open the action menu ⋮. The tree-view sidebar shows the hierarchical structure of the repositories and sub-repositories to which the user has access. The info-panel shows the details of the currently selected data source or folder. If the currently selected item is a data source, then the info panel shows the types of data available in the data source and the different data source versions of the data source. If the currently selected item is a folder, then the info panel shows info about the owner of the repository, the *access level* of the user, and info about the other users and user groups that have access to this repository.

When a user has read-write access (or higher), new data source files can be uploaded by pressing the add button + on the bottom right and selecting the *upload new file(s)* item. Likewise, a new sub-repository can be created by pressing the same add button and selecting the *create new folder* item. A third option is to create an external Proast link, which can be seen as a data source repository folder in which the data sources link to datasets (outputs) available on Proast web.

1.2.1 Repository access levels

Shares and access rights can be granted on the level of repositories and sub-folders. Data sources inherit the access rights of the repository/folder in which these are located. The following access rights are available:

- **visible:** the user can only see that the repository exists, but cannot see its contents, except for sub-folders that may also visible to the user.
- **use:** the user is only allowed to use the data sources in this repository, but is **not** allowed to download the original data of the data sources of the repository.

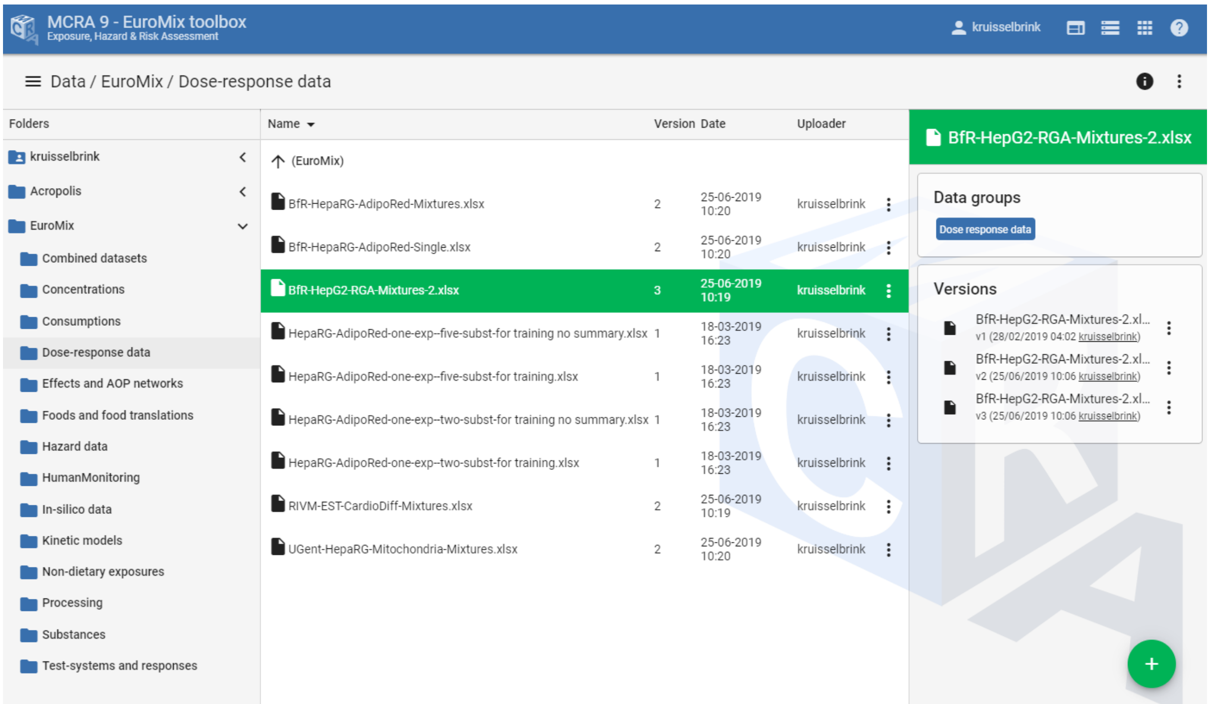


Figure 1.1: Screenshot of the toolbox data repository browser.

- **read:** the user can use data sources in this repository **and** is allowed to download the original data files of the data sources of the repository.
- **read/write:** the user can use and download data sources in this repository and is allowed to add/remove files and folders to/from this repository.
- **admin:** the is considered as an administrator of this repository and has full control over it, including the rights to add/remove files and folders to/from this repository and to add/remove user and group shares.
- **owner:** the user is considered to be the owner of this repository and therefore has full control over it.

When a user is administrator or owner of a repository/folder, then he or she is allowed to add/remove user and group access using the *edit shares dialog* (Figure 1.2) that can be opened by pressing the *edit shares* button.

1.2.2 Linking remote data repositories

Besides the normal data sources and folders of the data repository, the toolbox also offers the possibility to link to external data repositories. These are remote websites that are not part of the toolbox, but which contain data sources that can be used for calculations. Currently, there is only one remote source that can be linked as external repository in the toolbox, which is PROASTweb (https://proastweb.rivm.nl/). PROASTweb users have the possibility to directly link the outputs of their PROAST analyses (i.e., dose response models) as an external repository in the toolbox.

Figure 1.3 shows an example of how PROAST outputs of a PROASTweb user are linked in an external repository in the toolbox. Data sources of remote repositories have to be explicitly imported in the toolbox before they can be used in analyses. Initialy, all data sources in a remote repository have a status of not-imported. When pressing the import button, the toolbox will attempt to import the data source and once that is finished, the data source is ready to be used in analyses.

A new PROAST remote repository link can be created by pressing the add button + on the bottom right and selecting the *Create Proast link* option. This will open a dialog (Figure 1.4) asking for the local name of the external repository/folder, the PROASTweb username of the user of which the outputs should be linked, and the PROASTweb access key of the user, which is required as authentication token to access the analyses of the specified user.

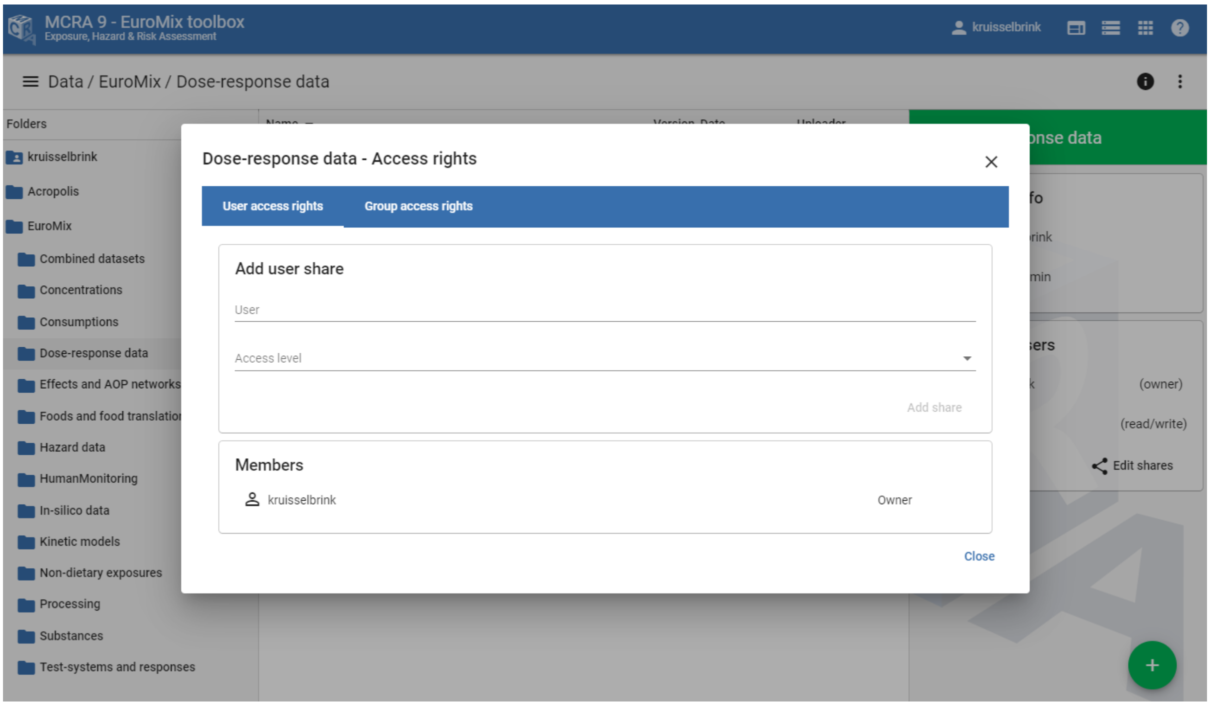


Figure 1.2: Screenshot of edit-shares dialog of the toolbox data repository browser in which user and group access rights can be added and removed by repository owners and administrators.

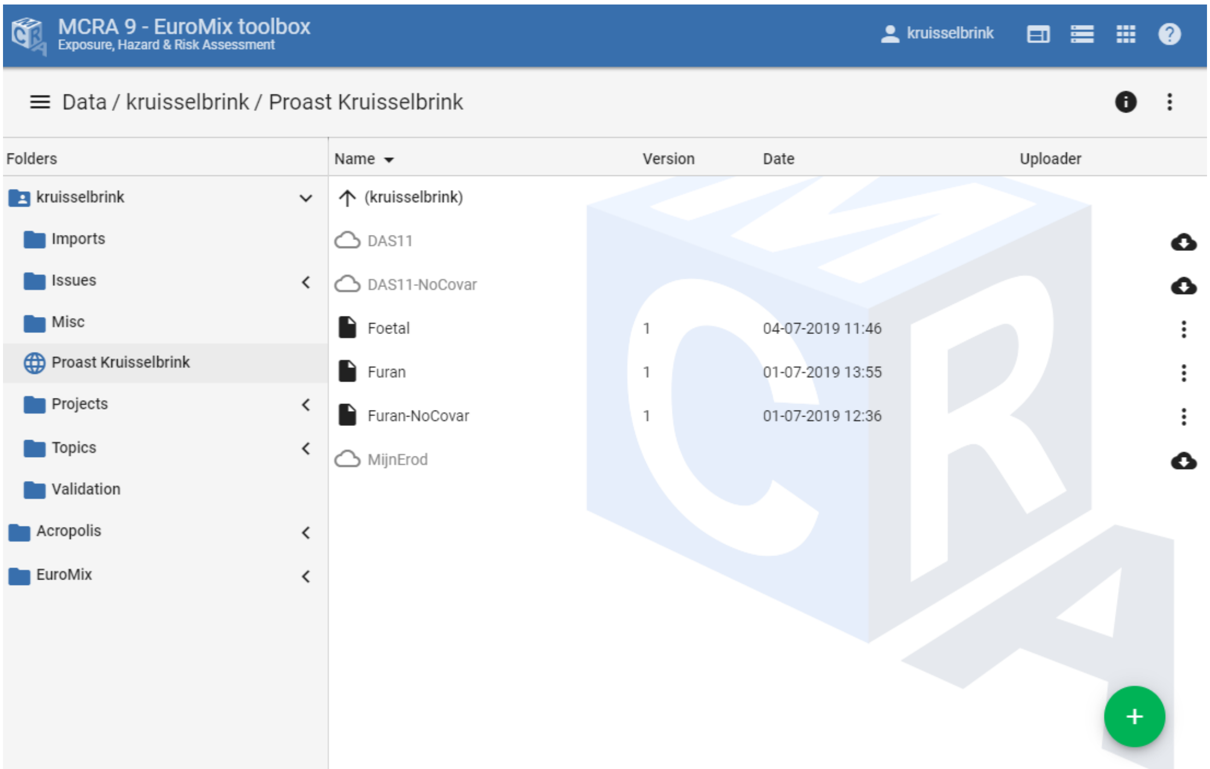


Figure 1.3: Screenshot of a remote (PROASTweb) repository in the toolbox data repository browser.

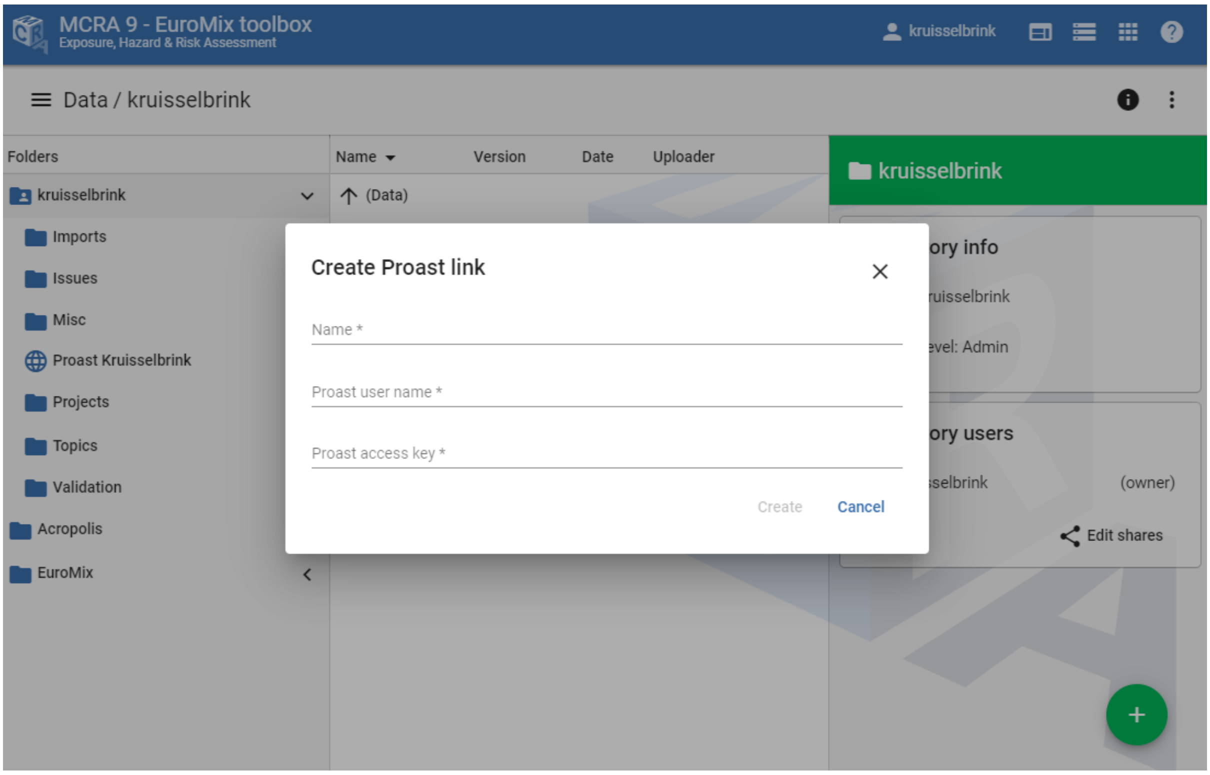


Figure 1.4: Screenshot of the dialog for creating a new PROASTweb remote repository link.

1.3 Workspaces and actions

User work is organized in workspaces. A workspace is a collection of work items that are logically grouped together. A workspace has a name, description and, optionally, a number of tags. Workspaces may be shared with other users. Users are the owners of their own workspace folders and possible subfolders.

Actions are configurations of the modules of the modular design. Each action is of a certain action type, which specifies the particular module for which this action is a configuration. An action can be available in two forms: 1) a data selection action and 2) a calculation action. A data selection action comprises the selection of already available data of that action type and specification of (subset) selections on that data. A calculation action is an action in which the data of that action is calculated based on relevant input and specific calculator settings. Within a workspace, multiple actions can be created.

When running an action, a task is spawned that produces output. Output is available in the form of reports or in the form of data that can be used as input in other actions. Actions have multiple outputs when settings are changed. Output reports are presented as screen reports or print reports. Output reports are composed of one or multiple sections.

1.3.1 Workspace browser

Figure 1.5 shows a screenshot of the workspace browser. In the workspace browser, users can scroll through their workspaces and select the workspace which they want to work with. Detailed information about the currently selected item in the browser is shown in the info panel, which can be expanded/collapsed using the info button on the right of the toolbar. The *filter text box* can be used to quickly find/filter workspaces by name or tag. A workspace can be opened by clicking on the workspace name or selecting the *open workspace* option of the *action menu* of the workspace. Opening a workspace will navigate you to the *workspace overview page*.

A new workspace can be added by pressing the add button + on the bottom right of the screen. A workspaces can be deleted by opening the *action menu* of the workspace item in the browser and selecting the delete option.

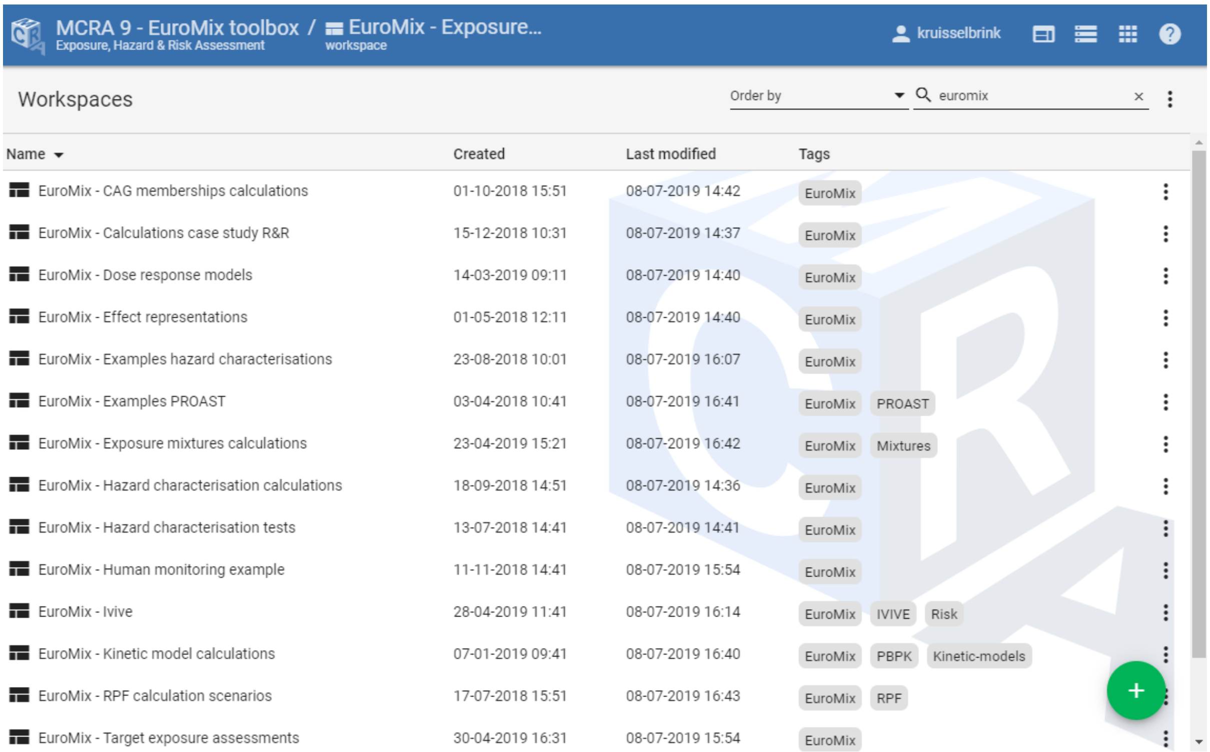


Figure 1.5: Screenshot of the workspace browser.

1.3.2 Workspace overview page

Figure 1.6 shows a screenshot of the workspace overview page. The workspace overview page gives an overview of the actions, data, tasks, and results of a workspace, which are shown as four tabs at the top of the page. The actions tab shows all actions of the workspace, and from this tab, actions can be opened. The data tab shows all data sources used in this workspace. I.e., all data sources that are used by the actions of the workspace. The results tab shows all tasks and results of simulation jobs that have been submitted by the actions of the workspace. The properties tab shows the general information of the workspace (i.e., name, descriptions, and tags) and allows for changing this information.

In the actions tab, all actions of the workspace are listed. The list of actions can be filtered by action type or by filter text using the controls on the toolbar. An action can be opened by clicking on the action name or by selecting the *open action* option of the action menu of the selected action item. Opening a workspace will navigate you to the action details pages. A new action can be added to the workspace by pressing the *add button* + at the bottom right of the page.

1.3.3 Action page

When opening an action, the user is directed to the main panel of the action, which is the panel associated with the action type of the action. In the main action page and subaction pages, an action can be configured, simulation jobs can be started, and the results can be evaluated. An example is shown in Figure 1.7. This panel shows the following sections:

- **Scope:** Links to the scope-panels in which the scope entities of the action can be set (e.g., foods or substances).
- **Inputs:** Links are shown for panels in which the calculation inputs or selection inputs can be set (e.g., concentration models that are inputs for computing dietary exposures).
- **Data source:** If the action is a data action, then a form is shown in which the data source should be specified (e.g., selection of the concentration data source in a concentrations action).

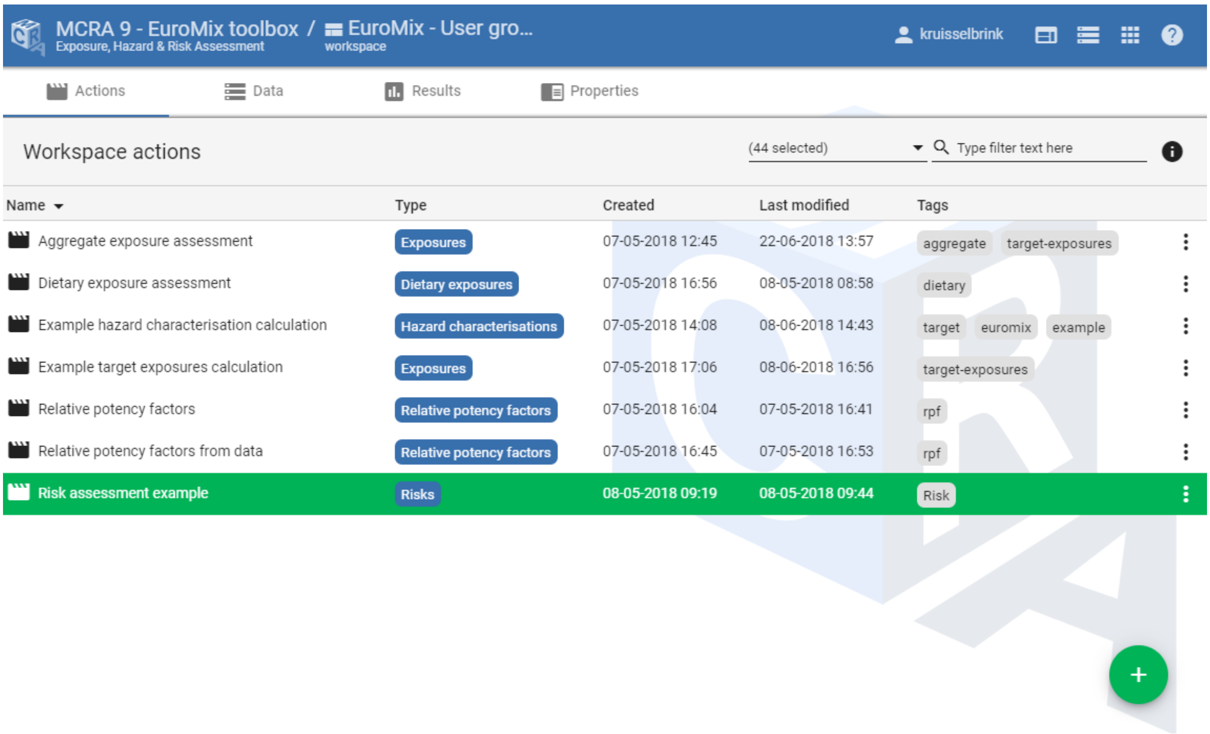


Figure 1.6: Screenshot of the workspace overview page.

- **Settings:** A form is shown in which the calculation and/or selection settings of the action can be set/changed (e.g., specify the exposure type, chronic/acute, of an exposure assessment).

For all modules of the toolbox, panels have the same structure. Within the panel, data sources and settings for the current action can be specified. In addition, the scope and input sub-module links that are relevant are shown. In this way, the user can follow the structure of the modular design to select the data and settings required for running the action. Besides the panels associated with a (sub)action, there is also a summary panel in which the main settings and data of the action are summarized, an output settings panel in which more general output settings can be specified, an uncertainty settings panel in which the uncertainty analysis options can be specified, and a results panel that shows the running tasks and the results of the action. An alternative form of navigating from action to sub-action is provided by the navigation menu in the left sidebar that can be expanded/collapsed by clicking on the menu button on the top left in the Action bar. In this menu, all necessary modules for the complete actions are shown in one list, allowing for a linear way of navigation.

An action is valid and ready to run if all scopes and inputs are valid and all required data and settings are configured. For each subaction, the check symbol ✓ indicates that it has been confidured correctly and is ready to run. In case the subaction has a warning symbol ⚠, some user action is required. When the main action is ready to run, then a simulation job for the action can be started by clicking the run button ▶ on the top right in the grey action bar. Optionally, subactions can be started by clicking the run button ▶ on the top right in the green (sub)action bar. Clicking the run button will send the simulation task of this (sub)action to the job-scheduler, and the progress of the task is shown in the results panel. When the task is completed, the output will be available in the form of a screen report and the report and downloads will be available to download the report as pdf, and the data in csv file format.

Scoping: entity selection

Selection of the primary entities so far as relevant for an action lies at the basis of each action of the toolbox. Also for other entities selection may be useful. Scoping of the action is defining the members for its primary entities, and sometimes also for other entities.

As an example, Figure 1.8 shows a screenshot of the substances module panel. At the top, the data source file can be selected containing the dataset of primary entity definitions. In the selection card, a selection can be made of the entities in the dataset that are relevant for the current action. Note that if no explicit selection is made, the scope is

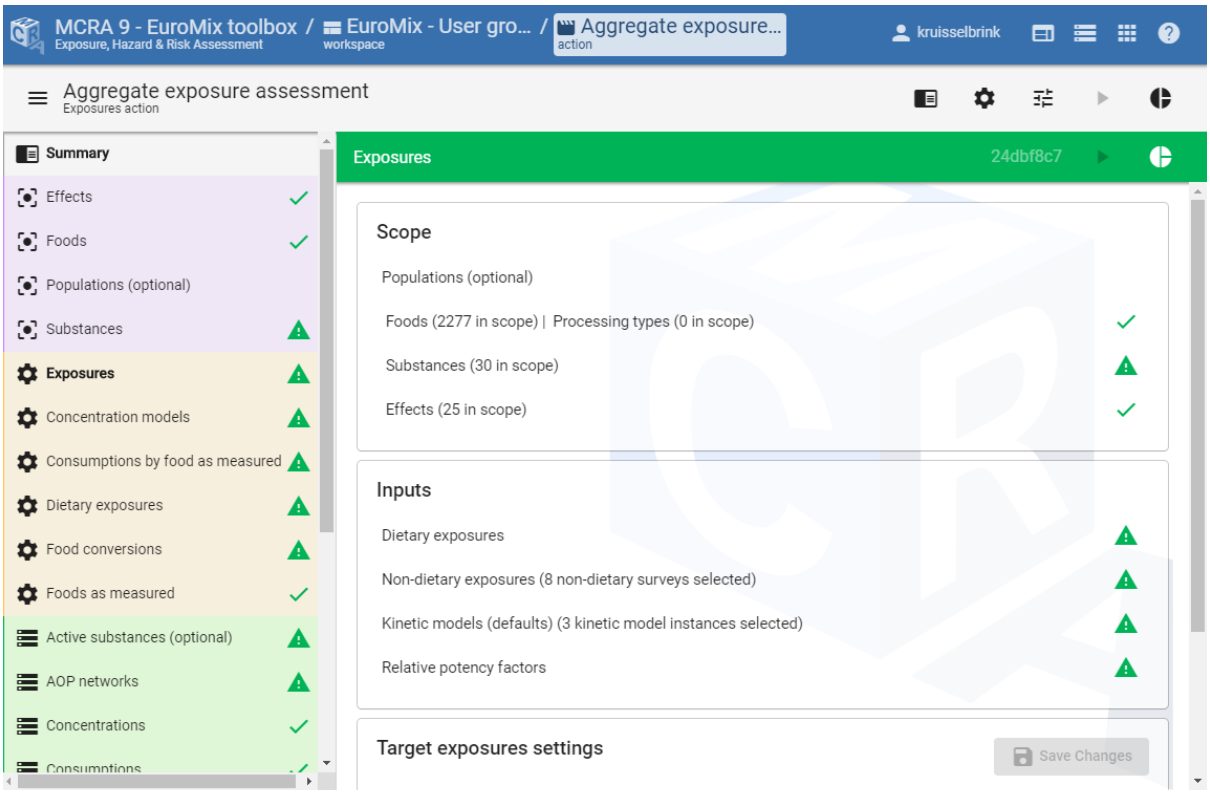


Figure 1.7: Screenshot of the main page of an action.

set to all entities by default. In the settings form, additional (selection) settings are shown, e.g., selection of the index substance which may be relevant for the chosen action type. In this way, the scope of the action can be specified by selection of the primary entities.

The panels for the data modules have a similar structure and selection is essentially the same. The only difference is that data actions always have a scope. I.e., data modules always relate to one or more primary entities.

Implicit versus explicit scoping

MCRA distinguishes between implicit and explicit selection of entities (scoping). By default, the selection is defined implicitly as 'all entities' found in all data linked to the action. For instance, the substance scope will be all substance codes found, not only in the substances data source, but also in the other data sources that link to substances (e.g., the substance codes of the concentration data or the points of departure data). Given an implicit selection, also explicit selections can be made in the specific module panel of this data type (e.g., by selecting the substances in the substances panel). Once made explicit, selections are no longer automatically expanded when new data sources are linked to the action.

For example, the substances scope shown in Figure 1.8 is defined explicitly, having three substances in the scope, and excluding 1626 substances that are present in the provided substances data source and/or other linked data sources. By pressing the *clear filter* button, the explicit scope can be cleared to become implicit again, effectively yielding a scope of all substances in all linked data sources, in this case 1626 + 3 substances.

Comparing new data to set scopes

When linking a data source to an action, MCRA checks whether the data links well to the current scope (selected entities) of the action and reports the results. For instance, when linking substance concentration data to an action which has already an implicit or explicit substance scope, it should be checked if and how the substance codes used in the concentration data match with the current substances scope. But the check is also performed when linking a

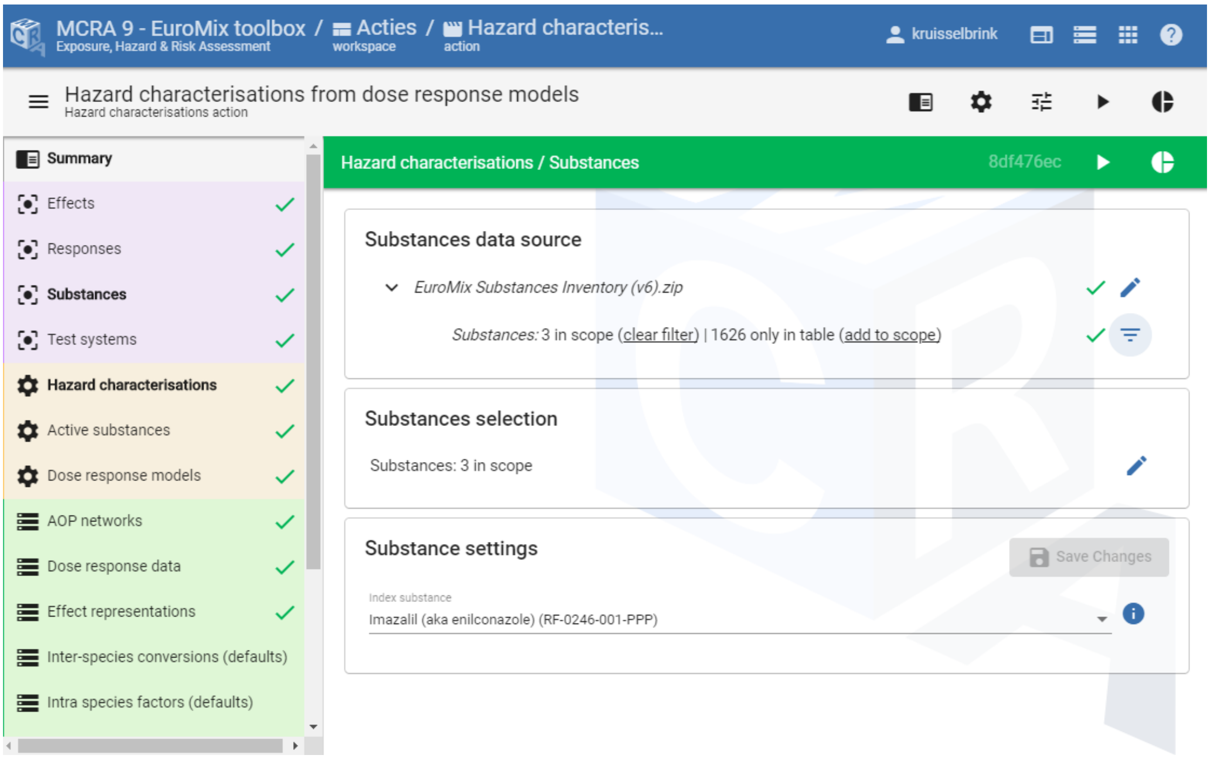


Figure 1.8: Screenshot of the substances module panel as an example of a primary entity module panel.

primary entity substances data source to an action wich has already a set substances scope due to already selected other data sources.

When linking a data table from a new data source to an action which has already a defined scope for one of the entities in the table, there are three possible states for entity codes:

- codes included in both the scope and the data source
- codes included in the scope, but not present in the data source
- codes included in the data source, but not present in the scope

The first case represents a successful link, which requires no further action. For the second and third type of mismatch,it depends on the type of data link whether this is considered a serious problem (red flag ⚠) or merely a point of attention (green flag ⚠). For instance, in the case of concentration data, these may be assumed to be zero for certain substances, and therefore MCRA allows missing concentration data for part of the substances in the scope: therefore a green warning symbol will be shown. The data sources may also contain codes that are not in the scope (e.g., substance concentrations for other substances as well), and it may be desirable to extend the scope with these substances. Also this situation is flagged with a green warning symbol.

Figure 1.9 shows an example of a point of departure action. The substances scope has already been defined by other data in the action (in this case points of departure data), and subsequently a substances data source is selected. Here, there are 140 substances in the current scope (explicitly defined). However, 132 of these 140 substances are not present in the substances data source (*not in table*). Hence, we are missing the definitions of these substances. This is considered a critical linking issue that should be solved by updating the substances data source to include these substances, therefore a red warning symbol is shown. On the other hand, the substances data source also contains 3 substances that are not part of the current scope (*only in table*), this situation on iots own would only lead to a green warning symbol, but is overruled in this case by the red warning symbol.

Another example is shown in Figure 1.10 where a points of departure data source is selected, dependent on the primary entities effects and substances. In the case of effects, we observe no linking errors, hence the data source matches perfectly with the effects scope. For substances, we see that there are 7 substances that are in the selected data source but not in the substances scope (*new*) and for 3 substances in the scope we don't have a point of departure (*not in table*). The former is fine, but we may want to extend the scope with these 7 substances (*add to scope*) and the

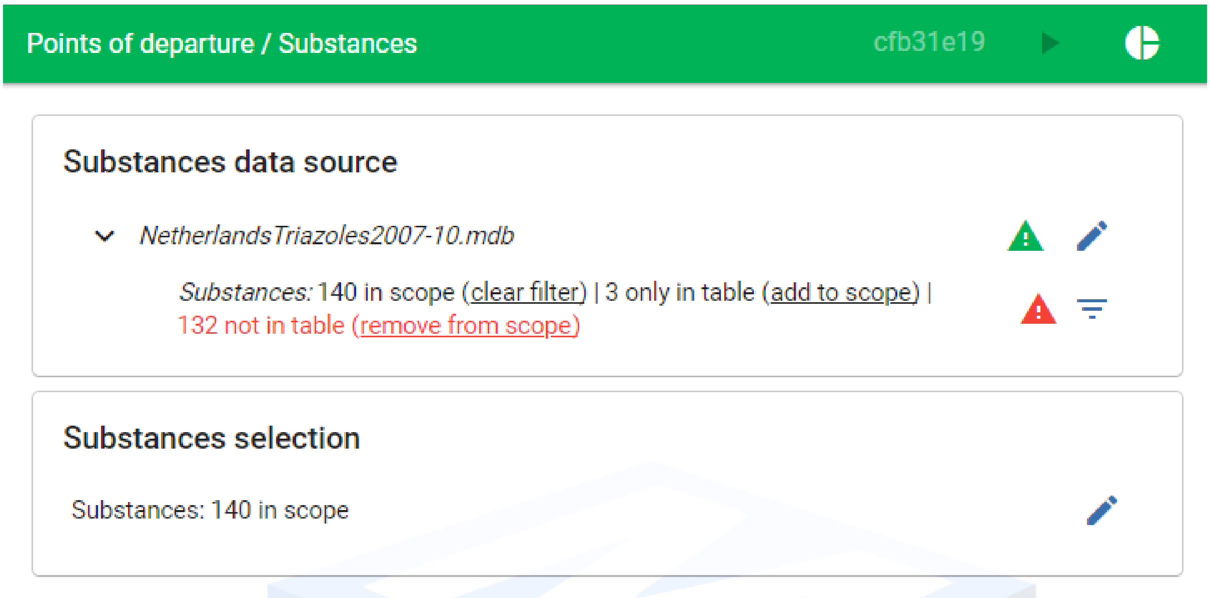


Figure 1.9: Screenshot of checking substances data in a substances data source against an already set substances scope.

latter should not have to be a problem, but it is a point of attention. Also, one may want to choose to remove these 3 substances from the scope (*remove from scope*).

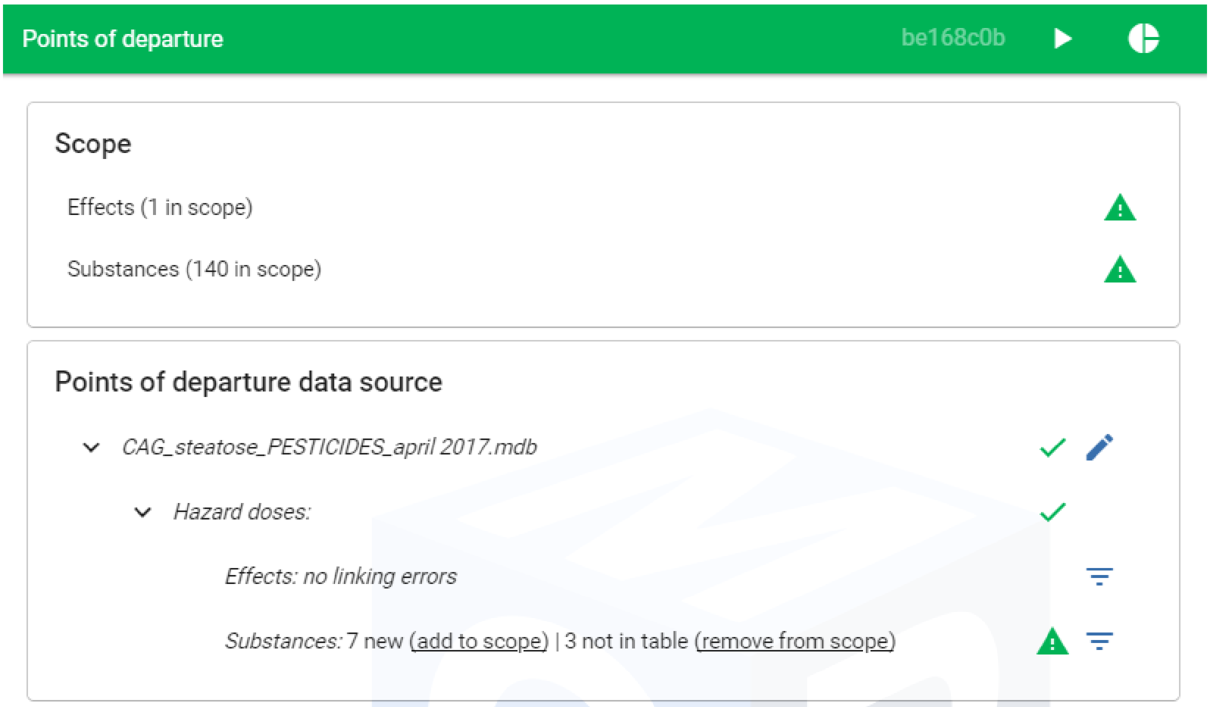


Figure 1.10: Screenshot of checking substances data in a POD data source against an already set substances scope.

CHAPTER

TWO

MODULES

MCRA is a modular system. The diagram of Figure 2.1 shows the modules and their relations. Each module is associated with its own type of data, and is linked to one or more other modules. Note that not all details can be fully shown in the scheme, for details consult the table below, which specifies all relations between the modules in MCRA.

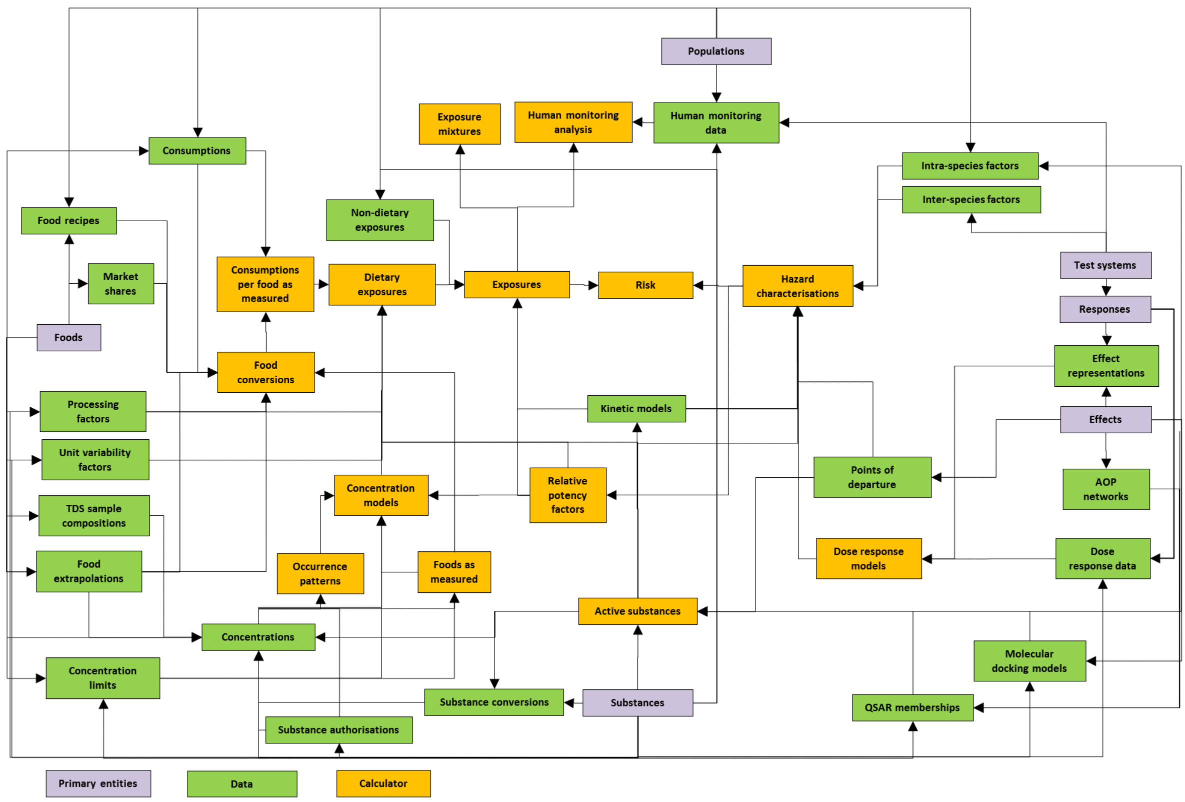


Figure 2.1: Diagram of the modular design of MCRA.

2.1 Primary entity modules

The MCRA modular system is based on six primary entities, defining who (*Populations*) is to be protected against what impact (*Effects*) caused by what agent (*Substances*) originating from where (*Foods*), with an indication how the effects are quantified (*Responses* in *Test systems*).

2.1.1 Effects

Effects are biological or toxicological consequences for human health, that may result from chemical exposure and are the focus of hazard or risk assessment.

Output of this module is used by: *Concentration models Dietary exposures with screening Dietary exposures Exposure mixtures QSAR membership models Molecular docking models Active substances Relative potency factors Hazard characterisations Points of departure Effect representations Inter-species conversions Intra species factors AOP networks Risks*

Effects data formats

Effects

Effects are primary entities of the data model. Health effects are defined as (critical) changes relative to a treatment or exposure.

Effects

Effects are uniquely identified by a code (idEffect). Optionally, a name and description can be added. Health effects are commonly distinguished in two types, acute and chronic. Further properties may be specified, e.g. in relation to decision schemes such as the use of thresholds of toxicological concern (TTCs).

Table 2.1: Table definition for Effects.

Name	Type	Description	Aliases	Required
idEffect	AlphaNumeric(50)	Unique identification code of the effect.	idEffect, EffectId, CodeEffect, Id, Code, KeyEvent, idKeyEvent	Yes
CodeSystem	AlphaNumeric(100)	Identifier of the coding system of the effect code.	CodeSystem	No
Name	AlphaNumeric(100)	Name of the effect.	Name	No
Description	AlphaNumeric(200)	Additional description or label of the effect.	Description	No
Biological-Organisation	AlphaNumeric(100)	Biological organisation of the effect: Molecular, Cellular, Tissue, Organ, Individual. This is in line with AOP wiki terminology and can be used for grouping.	Biological-Organisation	No
KeyEvent-Process	AlphaNumeric(100)	Description of AOP Key event component process. E.g., receptor signalling.	Process	No
KeyEvent-Object	AlphaNumeric(100)	Description of AOP Key event component object. E.g., PPAR-alpha.	Object	No
KeyEvent-Action	AlphaNumeric(100)	Description of AOP Key event component action. E.g., decreased.	Action	No
KeyEventOrgan	AlphaNumeric(100)	Description of AOP Key event organ. E.g., liver.	Organ	No
KeyEventCell	AlphaNumeric(100)	Description of AOP Key event organ. E.g., hepatocyte.	Cell	No
AOPwikiKE	AlphaNumeric(200)	Key event ID number in AOP wiki https://aopwiki.org/events Several ID possible Some effects might not be in the wiki, and this field will be empty.	AOPWikiIds, AOPwikiKE	No
Reference	AlphaNumeric(200)	External reference(s) to sources containing more information about the AOP key event. E.g., the AOP wiki, and the associated AOP wiki Ids.	References	No

Table aliases: Effects, Effect, KeyEvents, KeyEvent, RawEffects.

Effects settings

Selection settings

Table 2.2: Selection settings for module Effects.

Name	Description
Focal effect	The main (health) effect of interest.

Effects as data

Effect definitions are provided as lists/catalogues of effect definitions.

- *Effects data formats*

2.1.2 Foods

Foods are uniquely defined sources of dietary exposure to chemical substances. Foods may refer to 1) foods-as-eaten: foods as coded in food consumption data (e.g. pizza); 2) foods-as-measured: foods as coded in concentration data (e.g. wheat); 3) any other type of food (e.g. ingredients, e.g. flour).

Output of this module is used by: *Consumptions Market shares Food recipes Concentrations Processing factors Unit variability factors Occurrence patterns Substance authorisations Concentration limits Concentration models Foods as measured Total diet study sample compositions Food extrapolations Food conversions Consumptions by food as measured Dietary exposures with screening Dietary exposures Exposures Exposure mixtures*

Foods data formats

Foods

Foods are of interest in (dietary) consumption assessments and the sources of exposure within expore assessments. The foods table is the main table of the food definitions. Relevant food related data, such as processing types, additional properties (e.g., unit weight, large portion consumption amounts, and brand loyalty), facets, and hierarchies, can be described in the food properties, food hierarchies, and faces and facet descriptors tables.

Foods

Each food is identified by a unique code (idFood) in a code system of choice, a name, and a description. In the EuroMix data collection, FoodEx1 codes are used for both foods in consumption surveys (foods-as-eaten) and for raw agricultural commodities (foods-as-measured). Example: 'A.19.01.002.002' is pizza and pizza-like pies, cheese, and vegetables and 'A.01.02.001' is wheat grain. Food codes can have a hierarchical structure (as in the FoodEx1 and FoodEx2 coding systems), using '.' or '$' as separator between adjacent hierarchical levels, e.g. 'A.05' is fruits and fruit products, 'A.05.01' is citrus fruits, and 'A.05.01.001' is grapefruit (citrus paradisi). Additional forms of foods, such as foods in processed form, can be specified via food facets according to the FoodEx2 system of EFSA.

Table 2.3: Table definition for Foods.

Name	Type	Description	Aliases	Required
idFood	AlphaNumeric(50)	The unique identification code of the food.	idFood, Code, FoodId, FoodCode, Food, Id	Yes
Name	AlphaNumeric(100)	The name of the food.	Name, FoodName, Name1, FoodName1	No
Description	AlphaNumeric(200)	Food description.	Description	No

Table aliases: Foods, Food, RawFoods.

Food properties

Additional food properties, such as unit weight and portion sizes can be attached using the food properties table.

Table 2.4: Table definition for FoodProperties.

Name	Type	Description	Aliases	Required
idFood	AlphaNumeric(50)	The code of the food to which the property is attached. The provided food code should match with a code of the foods table.	idFood, FoodId, Food, FoodCode, Code	Yes
UnitWeight	Numeric	The nominal weight of a unit (dependent of the unit of consumption in g (default) or kg). Unit weights are relevant for foods-as-measured in the context of unit variability modelling.	UnitWeight	No
BrandLoyalty	Numeric	A parameter used in brand loyalty modelling, where 0 (default) specifies no brand loyalty (on each eating occasion a random selection of the next lower level in the hierarchy of food codes), and 1 specifies absolute brand loyalty (on subsequent eating occasions the same selection of the next lower level in the hierarchy of food codes).	BrandLoyalty	No
LargePortion	Numeric	Population (1 - 97 years): weight of a large portion (dependent of the unit of consumption in g (default) or kg). Used in deterministic modelling of exposure as in the IESTI equation.	LargePortion, LargePortion-Population, LargePortion-General-Population	No
LargePortion-Babies	Numeric	Babies (8 - 20 months): weight of a large portion (dependent of the unit of consumption in g (default) or kg). Used in deterministic modelling of exposure as in the IESTI equation.	LargePortion-Babies	No
LargePortion-Children	Numeric	Children (2 - 6 years) weight of a large portion (dependent of the unit of consumption in g (default) or kg). Used in deterministic modelling of exposure as in the IESTI equation.	LargePortion-Children	No

Table aliases: FoodProperties, FoodProperty, RawFoodProperties.

Food hierarchies

Food items are commonly categorised in hierarchies, e.g. oranges and mandarins are citrus fruits. For example FoodEx is a food description and food classification (FDFC) system consisting of a large number of individual food items aggregated into food groups and broader food categories in a hierarchical structure of parent-child relationships.

Table 2.5: Table definition for FoodHierarchies.

Name	Type	Description	Aliases	Required
idFood	AlphaNumeric(50)	Food node.	idFood, FoodId, Food, Code	Yes
idParent	AlphaNumeric(50)	Parent node of the food.	idParent, ParentId, Parent, ParentCode	Yes

Table aliases: FoodHierarchies, FoodHierarchy, FoodsHierarchy, RawFoodHierarchies.

Facets

Fode codes can be linked to facets, as e.g. in FoodEx.

Table 2.6: Table definition for Facets.

Name	Type	Description	Aliases	Required
idFacet	AlphaNumeric(5)	The food code of the food to which the facet is attached.	idFacet, Code, Id, FacetCode, FacetId	Yes
Name	AlphaNumeric(200)	Facet name	Name, FacetName	Yes

Table aliases: Facets, Facet, FoodFacets, FoodFacet, RawFacets.

Facet descriptors

Table 2.7: Table definition for FacetDescriptors.

Name	Type	Description	Aliases	Required
idFacet-Descriptor	AlphaNumeric(5)	The identification code of the facet descriptor.	idFacet-Descriptor, Code, Id, FacetCode, FacetId	Yes
Name	AlphaNumeric(200)	The name of the facet descriptor.	Name, Facet-DescriptorName	Yes

Table aliases: FacetDescriptors, FacetDescriptor, FoodFacetDescriptors, FoodFacetDescriptor, RawFacetDescriptors.

Processing types

Table 2.8: Table definition for ProcessingTypes.

Name	Type	Description	Aliases	Required
idProcessing-Type	AlphaNumeric(50)	The unique identification code of the processing type.	idProcessing-Type, ProcessingType-Id, ProcType, Id	Yes
Name	AlphaNumeric(100)	The processing name.	ProcName, Name	No
Description	AlphaNumeric(200)	The processing type description.	Description	No
Distribution-Type	AlphaNumeric	The distribution type. Simulated processing factors are restricted to the interval (0,1) using a logistic-normal distribution (= 1) or simulated processing factors are restricted to positive values using a log-normal distribution (= 2).	Distribution-Type, DistType	Yes
Bulking-Blending	AlphaNumeric(10)	For types of processing applied on large batches e.g. juicing, sauce/puree. No bulking/blending = 0, bulking blending = 1.	Bulking-Blending, BulkBlending	Yes

Table aliases: ProcessingTypes, ProcessingType, RawProcessingTypes.

Foods as data

Food definitions are provided as lists/catalogues of food definitions, optionally with encompassing processing type definitions, facet definitions, hiearchy definitions, and additional food property information.

- *Foods data formats*

Food coding systems

MCRA is intended to retain complete transparence of the results of risk assessment in terms of the foods that were actually consumed (foods-as-eaten). In many cases measurements of substances have not been made on the **food-as-eaten**, e.g. pizza, but on a raw agricultural commodity (RAC), e.g. tomato, onion etc. The food on which the concentration measurements have been made is termed the **food-as-measured**. MCRA implements a recursive search algorithm to link foods-as-eaten to foods-as-measured. This means that there can be intermediate steps, e.g. if unpeeled *apple* and *grapes* are the foods-as-measured, the food-as-eaten *apple pie* contains *peeled apple* and *raisins*, *peeled apple* is linked to unpeeled *apple*, and *raisins* are dried *grapes*.

Food classification: FoodEx1

Food code definition

In MCRA, a food code is a string consisting of symbols. Some special symbols (. $ - ∗) are reserved for special use (see below), and can not be used freely in own codes.

Codes can be hierarchical. Any code can be followed by $ or . plus a subtype code. This can be repeated any number of times, e.g. ABC$D, or A.B.C.D.

Codes can specify food processing. Any code can be followed by - plus a processing code. Only one level of processing code is allowed (e.g. FP0226-2). Subtype codes should precede processing codes (e.g. NL005$123$456-2). The asterisk (∗) serves as a wildcard for the preceding code: the processing information is valid for all codes that start with the code preceding the ∗.

Food codes in consumption surveys

Any coding system for foods-as-eaten can be used in MCRA. For example, in Europe EFSA develops a Food Classification and Description System for exposure assessment named FoodEx 2 ([1], [2]), featuring a hierarchical system of a core list of foods, an extended list, and domain-specific hierarchies.

Food codes in concentration data

Any coding system for foods-as-measured can be used in MCRA.

Food processing

Concentrations of substances in foods may change when foods are processed. Examples of processing types are peeling (e.g. of apples), cooking (e.g. of spinach), drying (e.g. of grapes), juicing (e.g. of oranges). In MCRA a processing factor can be specified for any food. Processing factors specify the ratio of concentrations in the processed and unprocessed food. The food code of the processed food will be converted to the food code of the unprocessed food. In simulations the concentration in the unprocessed food will then be multiplied by the processing factor. Special attention is needed if food processing also changes the weight of the food. Traditionally, processing factors combine the effects of chemical alteration and weight change, so the weight change should not be double-counted. The processing correction factor is introduced to correct processing factors that combine both effects, e.g. when 100g *raisins* (dried grapes) are translated to 300g *grape* (food-as-measured) and the processing factor for drying combines both effects, the processing correction factor is 3.

Recipes and food translation

Recipes specify the composition of composite foods, e.g. *pizza*, in terms of relevant ingredients, e.g. 100g pizza contains 10g *tomato*, 5g *cheese* and 50g *flour*. Recipes are also used to specify weight changes, e.g. to obtain 100g *raisins* (dried grapes) 300g of the food-as-measured *grape* is needed.

A special use of recipes and food translation is found in Total Diet Studies. Here, the composition of a Total Diet Study food is specified, e.g. TDS-food *FruitMix* is composed of *apple*, *orange* and *pear* with a default translation proportion of 100%. So in MCRA, the food-as-eaten *apple* is converted to *FruitMix* (100%) and *FruitMix* is considered as the food-as-measured (TDS-food). A conversion from *apple-pie* (food-as-eaten) to *FruitMix* (food-as-measured) is based on a recipe for apple-pie and a TDS composition for FruitMix.

Another use of converting foods (as-eaten or as an intermediate step), is through the specification of so-called food extrapolations (read across translations), e.g. for *pineapple* no measurements are found but by specifying that *pineapple* is converted to *FruitMix* (with a default proportion of 100%), the TDS sample concentration value of *FruitMix* will be used for *pineapple* (as-eaten or as ingredient).

Market shares and brand loyalty

Sometimes measurements of substances in food are available at a more detailed food coding level than consumption data. For example, measurements may have been made for specific brands of a food whereas the consumption survey did not record the brand. MCRA allows to specify market share data for subtypes of a food (e.g. A$1, A$2, A$3 are three brands of food A), and to calculate acute exposure based on such *market shares*.

Supertypes

Sometimes measurements of substances on food are available at a less detailed food coding level than consumption data. MCRA allows to use the concentration data of a supertype for all underlying food codes. However, this is not the default, and an explicit permission should be given to allow this feature.

Maximum Residue Levels

Maximum residue levels are the upper legal levels of a concentration for substance residues in a food, e.g. pesticide, or feed based on good agricultural practices and to ensure the lowest possible consumer exposure.

MCRA food code conversion algorithm

Food codes are linked using a *7-step procedure*.

Food classification: FoodEx2

'The collection and evaluation of data on levels of chemical occurrence or presence of biological agents in food and feed are important tasks of EFSA. By combining the data with information on food consumption allows for detailed intake and exposure estimates crucial to any food and feed safety risk assessment or nutrient adequacy analysis. The EU Member States provide an increasing volume of data to EFSA and other European bodies. To provide a common link to all the diverse food and feed databases, a system for the unique and universal identification and characterisation of food and feed items is essential. EFSA has developed a preliminary standardised food classification and description system called FoodEx2 (version 2 of the EFSA Food Classification and Description System [FCDC] for exposure assessment). The system consists of descriptions of a large number of individual food items aggregated into food groups and broader food categories in a hierarchical parent-child relationship. Central to the system is a common 'core list' of food items or generic food descriptions that represent the minimum level of detail needed for intake or exposure assessments. More detailed terms may exist in addition to the core list and these are identified as the 'extended list'. A parent-child relationship exists between a core list food item and its related extended list food items. The terms of the core and extended list may be aggregated in different ways according to the needs of the different food safety domains. In the present version four hierarchies are proposed: three domain-specific and a general purpose one. Facets are used to add further detail to the information provided by the food list term. Facets are collections of additional terms describing properties and aspects of foods from various perspectives'. For more information visit: http://www.efsa.europa.eu/en/datex/datexfoodclass.htm.

For MCRA, having a different set of food codes is in itself not a problem. That is, for MCRA, it does not matter how foods are coded, as long as they can be linked to consumptions and concentrations within an exposure assessment. What makes FoodEx2 different from other food coding systems is that it provides additional food hierarchies, food facets, and a combined food/facet coding system. Below follows a brief summary of these main features of the FoodEx 2 coding system from the perspective of exposure assessment using MCRA.

Foods and food hierarchies

FoodEx 2 contains different food hierarchy definitions and allows for creation of custom food hierarchy definitions. These hierarchies could, for exposure assessment, allow to assess intake or consumption data based on the groups defined by these hierarchies.

Table 2.9: Food hierarchy export from FOODEX 2.0 Browser version 0.1.3

Code	Level	Name	ParentCode	Scopenotes
A000J	1	Grains and grain-based products	ROOT	The category covers all …
A000K	2	Cereals and similar	A000J	…
A000l	3	Cereal and cereal-like grains	A000K	…
A000M	4	Amaranth grain	A000L	…
A000N	5	Buckwheat grain	A000L	…
A000P	6	Barley grain	A000L	…
…	…	…	…	…

Facets and facet descriptors

FoodEx 2 allows to provide supplementary details on specific aspects of foods by means of so-called facets and facet descriptors. Facets are collections of terms defining specific characteristics of food from particular points of view and facet descriptors describe specific characteristics foods. For example, *processing technology* is a facet, and *baking* is a facet descriptor belonging to this facet. Currently, 26 facets are defined, containing in total 2172 descriptors (EFSA 2011b) [2]. Facets are also defined in a hierarchical system. For instance, *cooking in fat (A07GR)* and *baking (A07GX)* are sub-items of the descriptor *cooking and similar thermal preparation processes (A0BA1)*. Facets are coded as small strings that consist of a facet code and a facet descriptor code separated by a '.'-character. For example, the facet code *F28.A07GX* holds

1. the facet code *F28*, which is the facet code for *process technology*, and
2. *A07GX* , which is the descriptor code for *baking*.

Table 2.10: Part of the FoodEx 2 facet descriptor codes of the source facet (F01).

Code	Level	Name	ParentCode	Scopenotes
A04SF	1	Animals	ROOT	…
A056H	2	Mammals (food source animal)	A04SF	…
A056Z	3	Farmed / non-game mammals (food source animal)	A056H	…
A057A	4	African buffalo (food source animal)	A056Z	…
A057B	4	American buffalo (food source animal)	A056Z	…
A057C	4	Buffalo (food source animal)	A056Z	…
A057D	4	Cape buffalo (food source animal)	A056Z	…
A057E	4	Cattle (food source animal)	A056Z	…
…	…	…	…	…

Implicit facets

Implicit facets are facets of a product that are already implied by the food product itself. Consider, for example, *potato boiled (A011P)*, where *boiling (A011P)* is an implicit facet, because boiling is already implied by the product. According to EFSA [1] 'inclusion of implicit facets in the string recorded for each food database record is not encouraged' and it is suggested to identify and record the implicit facet descriptors in a separate table.

Foods as facets

Foods and facet descriptors share the same unique alphanumerical coding system; in some cases, like *characterising ingredient or sweetening agent* food list elements may be used as facet descriptors.

The FoodEx 2 coding system

In the coding system, facets can be added to the primary food codes to provide supplementary detailed information of particular data records. The structure of the FoodEx 2 codes is:

idFood#idFacet.idFacetDescriptor$idFacet.idFacetDescriptor$. . . .

The code starts with the primary FoodEx2 food code. Then, when there are supplementary facets, the food code is followed by a '**#**'-character and the facets string. The facets string is constructed as a concatenation of the individual facets strings, separated by means of the '**$**' character. As an example, consider the string *A011P#F28.A07GL$F28.A07KQ* which is composed of:

- Food: *A011P - Potato boiled*
- Facet 1: *F28.A07GL - Process technology - Boiling*
- Facet 2: *F28.A07KQ - Process technology - Freezing*

FoodEx2

For MCRA, FoodEx 2 introduces the following points of attention:

- Reading and dealing with FoodEx 2 coded data sets
- Reading and dealing with food facets
- Reading and exploiting food hierarchy data

Reading and dealing with FoodEx 2 codes

All data entities that contain foods data are potentially affected by the introduction of FoodEx 2. In MCRA, the following data tables are adapted to allow for input of full FoodEx 2 food codes:

- Foods
- Consumptions
- Concentrations

For these tables, the food code is allowed to be the complete FoodEx 2 food code and automatically recognized as such. As an example, Table 2.11 shows how the FoodEx 2 coded consumptions should be provided to the system. On important note: the maximum field length of the food code is 50. This means that there is a maximum of five facets that can be specified for a food.

Table 2.11: Integrated coding of the facets in the consumed foods field of food consumptions. Implementation.

Individual	DayOfSurvey	Food	Amount	FoodSurvey
14233701	1	A011R# F28.A07GX	153.43	FS01
18843004	1	A011R# F28.A07GX	125.23	FS01
34025701	1	A011R# F28.A07GX	153.60	FS01
14720005	2	A011R# F28.A07GX	105.00	FS01
49174010	1	A011R# F28.A07GX	140.00	FS01
62794010	1	A011R# F28.A07GX	67.00	FS01
61392002	1	A011P# F28.A07GL$F28.A07KQ	104.72	FS01
61281231	1	A011P# F28.A07GL$F28.A07KQ	109.72	FS01

Reading and dealing with facets data

Within MCRA, the following facets related aspects are accounted for:

- Reading facets data
- Dealing with facets
- Facets in concentration data
- Facets in food conversion
- Using facets as processing factors
- Using hierarchy data in the output

Reading facets data

To incorporate input of facets data in MCRA, two tables Facets and FacetDescriptors are introduced as optional tables of the Foods data group. The *table for Facets* and *table for FacetDescriptors*.

Within MCRA, the facets of FoodEx 2 coded foods, consumptions, and concentrations are automatically linked to the provided facets and facet descriptors. Also, the facet descriptor names are added automatically to the foods containing these facets.

Dealing with facets

The introduction of food facets allows for much more detailed specifications of consumption and concentration data. However, it introduces the problem of deciding on which level of detail the exposure assessment should be performed. That is, should concentration models be generated on the level of foods-without-facets or on the level of foods-with-facets? E.g., should the concentrations of *clementine peeled (A01CE#F28.A07LC)* and *clementine unprocessed (A01CE#F28.A0C0S)* be modelled separately or should one model be constructed for *clementine (A01CE)*? Treating all clementine's as equal may yield over-simplified conversions, whereas treating all separately may lead to many concentration models based on only few measurements. In MCRA, no implicit grouping of concentrations of equal foods with different facets is applied. If concentrations are provided for both *clementine peeled (A01CE#F28.A07LC)* and *clementine unprocessed (A01CE#F28.A0C0S)*, then these are modelled separately. Another question is whether the order of the facets is relevant or not. E.g., is *A0BYV#F02.A06GF$F03.A06HY* the same as *A0BYV#F03.A06HY$F02.A06GF*? Regarding this matter, MCRA considers the facet order to be important. I.e., *A0BYV#F02.A06GF$F03.A06HY* is not the same as *A0BYV#F03.A06HY$F02.A06GF*.

Facets in food conversion

For conversion of foods-as-eaten to foods-as-measured, MCRA considers foods with different facet strings as different foods. I.e., there is no implicit conversion of foods-with-facets to foods-without-facets and also the order of the facets is important. However, as it is realistic to convert food-with-facets to the base food without facets, an additional (explicit) conversion step remove-all-facets is added that converts foods with facets to the base foods. I.e., the action is "remove all". There is no conversion step for "stripping off one facet at a time". The reason for this is that there is no good way of deciding which facet to strip off first. This new conversion step is somewhat equivalent to the already existing default processing conversion step (step 6), and is therefore implemented as step 6b of the conversion algorithm. Particular rules followed by this step:

- Conversion of food-with-facets to food-without-facets.

Using facets that reveal processing data

Facets containing processing information, such as *part-consumed-analysed (F20)* and *processing technology (F28)* could be integrated with processing data. As an example, consider *clementine peeled (A01CE#F28.A07LC)*. This could be linked to *clementine (A01CE)*, with processing type *removal of external layer (A07LC)*. Linking to processing data could be achieved by entering processing data using the facet codes. As an alternative to the current processing factor tables, a facet-based processing factors table is defined for processing facets. That is, the codes for food processed and unprocessed are implicitly defined for FoodEx 2.

Table 2.12: Example of a MCRA processing factors table using FoodEx 2 foods and facets codes.

FacetCode	Compound	FoodCode	ProcNom	ProcUpp	Proc-NomUnc-Upp	Proc-UppUnc-Upp
A07LC	CompoundX	A01CE	0.5	0.6	0.05	0.06
F28.A07GV	CompoundX	A0BY	0.2	0.1	0.03	0.04

Note that in the example, the facet code could be specified as the full facet code, or just the code of the facet descriptor. As a more elaborate example consider

French fries from cut potato (A0BYV#F02.A06GF$F03.A06HY$F04.A00ZT$F28.A07GR)

For this food code, the substring of the processing facet is extracted from the list of facets.

- *A0BYV#F02.A06GF$F03.A06HY$F28.A07GR$F04.A00ZT* with processing facet link A07GR
- *A0BYV#F02.A06GF$F03.A06HY$F04.A00ZT*

In MCRA, a table FacetProcessingFactors is introduced that allows for specification of processing factors by means of facets. This table has the following structure:

Table 2.13: Table FacetDescriptors of the Food data group.

Column name	Key	Required	Type	Size	Description
idProcessingType	Yes	Yes	String	5	The facet code of this processing factor definition. May be specified as full facet code, i.e., facet code plus facet descriptor code, or as the facet descriptor code.
idFood	Yes	Yes	String	200	The food code
idCompound	Yes	No	String	50	The compound for which this processing factor is defined.
Nominal	No	Yes	Double		Nominal value (best estimate of 50th percentile) of processing factor (defines median processing factor)
Upper	No	Yes	Double		Upper value (estimate of 95th percentile or “worst case” estimate) of processing factor due to variability
NominalUncertaintyUpper	No	Yes	Double		Upper 95th percentile of nominal value (Nominal) due to uncertainty. A standard deviation for uncertainty of the nominal value (Nominal) is derived using the nominal value (Nominal) and upper 95th percentile (NominalUncertaintyUpper)
UpperUncertaintyUpper	No	Yes	Double		Upper 95th percentile of upper value (Upper) due to uncertainty. From the nominal value (Nominal), upper value (Upper) and the specified uncertainties of these values (NominalUncertaintyUpper and UpperUncertaintyUpper, respectively) the degrees of freedom of a chi-square distribution describing the uncertainty of the standard

The integration with the food conversion algorithm is as follows: Conversion step 2 (*processing*) is extended with a step 2c (*processing facet*) that attempts to match facets of a food code to processing data provided in the processing facets table. The following important rules are followed:

- Processing factors can be defined for base-food-code/facet-code combinations and translate as food-with-processing-facet to food-without-processing-facet.
- If multiple processing facets are present in the food-as-eaten code, then the last processing facet is used first for conversion.
- Facet processing factors can be specified using the full facet code (i.e., facet-code plus facet-descriptor-code) or just the facet descriptor code. If both are specified for the same food, the full facet code is used.
- Facet processing factors can be defined compound-specific, and non-compound-specific. Processing factors that are defined compound-specific always precede non-compound specific processing factors.
- Processing factors defined by a food-processed/food-unprocessed combination precede processing factors defined through facets.

Weight reduction factors for processing factors defined for facets should be included in the food translation table and should match exactly.

Food hierarchies

Reading and dealing with food hierarchy data

Within MCRA, the following hierarchy related aspects are accounted for:

- Reading food hierarchy data
- Using hierarchical data for conversion of foods
- Using hierarchy data in the output

Reading food hierarchy data

A new data group named *Foods* is added. In this group, a new *table for FoodHierarchies* is used for input of food hierarchies. This table contains food hierarchy node-definition records that reflect a hierarchical structure. For foods that are not in this list as idFood, it is implicitly assumed that these foods are root items.

Note: It is common practice to describe hierarchies using tree structures. Here, the elements of the tree are named *nodes*, the lines connecting the nodes are named *branches*, and nodes without children are *leaf nodes/end-nodes*. This terminology is also used throughout the remainder of this document.

Using food hierarchies for food conversion

The introduction of the hierarchy structure allows for integration with step 4 and step 5 of the food conversion algorithm; the *subtype* and *supertype* linking steps. That is, when no concentration data is found for a certain product, the concentration data of a (according to the hierarchy) related product could be used. In MCRA, the *supertype* conversion step also contains a *hierarchy-supertype* step based on the food hierarchy.

Supertype link (step 5):

a) **Supertype**: Try to find supertypes base on '$'-coded strings, e.g., 'xxx$yyy' is converted to 'xxx'

b) **Hierarchy-supertype**: try to find the supertype of the current food based on the food hierarchy (i.e., convert the current food to its parent).

Note 1: the *supertype* conversion step is optional and should be specified in the conversion settings panel.

Note 2: the *hierarchy-supertype* step only applies for foods-without-facets. The reason for this is that for the conversion, the base type of a food-with-facets can be considered as a better conversion candidate than the parent food with the same facets.

Using hierarchy data in the output

Food hierarchy information could be used in presentation of various tables of the output of MCRA. That is, in the tables in which foods data is presented, these records could be grouped based on the hierarchy and/or a tree-like display can be built for the presentation of this data. Tables that are candidate for being extended are, for example, the input data tables foods-as-eaten/foods-as-measured and the exposure by food-as-eaten/food-as-measured output tables.

Summarizing over the food hierarchy is many cases not a straightforward task. Consider, for instance, the statistic *number of consumption days* given the artificial hierarchy of *Citrus Fruits* containing two child-nodes *Mandarin* and *King Mandarin*: the number of consumption of *Citrus Fruits* is not "just" the sum of the consumption day of *Mandarin* and *King Mandarin*. A difficulty for summarizing based on a hierarchy arises when a node contains both data and child-nodes with data. E.g., concentrations are defined on the level of *Citrus Fruits* and on the level of *Mandarin*. In this case, the hierarchy view should ideally summarize for both *Citrus Fruits* as data record and *Citrus Fruits* as summary node. An additional complication is the status of facet-coded foods within the hierarchy. In a hierarchical view, foods-with-facets should ideally be added to their base-foods for visualization.

In MCRA, an alternative view (treetable) is added that can display hierarchical data. This alternative view is used to present a hierarchical view based on the foods hierarchy for the consumption input summary tables food as eaten and food as measured. The data summary methods for these tables are updated such that the data is also summarized per hierarchy-node.

Food name	Food code	Mean consumption (g)	Mean consumption days (g)	Consumption days	Percentage consumption days	Total weights consumption days	Percentage total weights consumption days
⊟ Fruit and fruit products	A01BS	167	200	5	83.3 %	5.0	83.3 %
⊞ Fresh fruit	A04RK	167	200	5	83.3 %	5.0	83.3 %
⊟ Starchy roots or tubers and products thereof, sugar plants	A00ZR	100	600	1	16.7 %	1.0	16.7 %
⊟ Starchy root and tuber products	A011B	66.7	400	1	16.7 %	1.0	16.7 %
⊟ Processed root and tuber products	A04MJ	66.7	400	1	16.7 %	1.0	16.7 %
⊟ Potato boiled	A011P	66.7	400	1	16.7 %	1.0	16.7 %
Potato boiled Tuber (as part-nature)	A011P#F02.A067V	16.7	100	1	16.7 %	1.0	16.7 %
Potato boiled Tuber (as part-nature), Potatoes, Boiling	A011P#F02.A067V$F27.A00ZT$F28.A07GL	16.7	100	1	16.7 %	1.0	16.7 %
Potato boiled Tuber (as part-nature), Potatoes, Boiling	A011P#F02.A067V$F28.A07GL$F27.A00ZT	16.7	100	1	16.7 %	1.0	16.7 %
Potato boiled Tuber (as part-nature), Potatoes, Boiling, Baking	A011P#F02.A067V$F27.A00ZT$F28.A07GL$F28.A07GX	16.7	100	1	16.7 %	1.0	16.7 %
⊟ Starchy roots and tubers	A00ZS	33.3	200	1	16.7 %	1.0	16.7 %
⊟ Tubers	A04MC	33.3	200	1	16.7 %	1.0	16.7 %
⊟ Potatoes	A00ZT	33.3	200	1	16.7 %	1.0	16.7 %
Potatoes Potatoes (food source plant), Tuber (as part-nature)	A00ZT#F01.A05KG$F02.A067V	16.7	100	1	16.7 %	1.0	16.7 %
Potatoes Potatoes (food source plant), Tuber (as part-nature), Baking	A00ZT#F01.A05KG$F02.A067V$F28.A07GX	16.7	100	1	16.7 %	1.0	16.7 %

Figure 2.2: Hierarchy view for the foods as eaten input summary table.

If a node contains both data and a child record, then this node is split-up in two nodes: a summary node that summarizes the data of the node and all of its child nodes, and a data record with the string "(unspecified)" added as a child of this summary node. See Figure 2.2 for an example (*Citrus Fruits* versus *Citrus Fruits (unspecified)*). In MCRA, foods-with-facets are added as child nodes of the foods-without-facets.

2.1.3 Populations

Populations are groups of human individuals that are the scope of exposure or risk assessments. Optional descriptors of populations are location (e.g. a country), time period (start date, end date), age range and gender. Example: the French population in 2005-2007 of women of child-bearing age (18-45 yr).

Output of this module is used by: *Consumptions Consumptions by food as measured Dietary exposures Non-dietary exposures Exposures Human monitoring analysis Risks*

Populations data formats

Populations

Populations are primary entities of the data model.

Populations

Populations are used to select dietary, nondietary and human monitoring surveys. Optionally, a name and description can be added. Population can be restricted to a certain time period. AgeMin, AgeMax and Gender are optional properties of a population.

Table 2.14: Table definition for Populations.

Name	Type	Description	Aliases	Required
idPopulation	AlphaNumeric(50)	Unique identification code of the population.	IdPopulation, PopulationId, Code, Id	Yes
Name	AlphaNumeric(100)	The name of the population.	Name, PopulationName	No
Description	AlphaNumeric(200)		Description	No
Location	AlphaNumeric(50)	Location.		No
StartDate	DateTime	Starting date	StartDate	No
EndDate	DateTime	End date	EndDate	No
AgeMin	Integer	Minimum age	AgeMinimum	No
AgeMax	Integer	Maximum age	AgeMaximum	No
Gender	AlphaNumeric(50)	All levels of gender	Sex	No

Table aliases: Populations, Population, RawPopulations.

Populations settings

Selection settings

Table 2.15: Selection settings for module Populations.

Name	Description
Population	Specifies which population is selected.

Populations as data

- *Populations data formats*

2.1.4 Responses

Responses are measurable entities in test systems. Responses are used to represent effects (see effect representations) and their measured values are collected in dose response data.

This module has as primary entities: *Test systems*

Output of this module is used by: *Dose response models Dose response data Effect representations*

Responses data formats

Responses

A response is a measurable endpoint on in a test system. E.g., in a rat test system a response may be the percentage of fatty hepatocytes observed after 90 days. Responses are defined in the responses table.

Responses

Each response is identified by a unique code (idResponse) in a code system of choice, a name, and a description. Also, each response should be linked to a test system (idTestSystem) on which the response is measured. Responses can be of various types (ResponseType), e.g., ContinuousMultiplicative (= non-negative real values using a ratio scale), ContinuousAdditive (= real values using an interval scale), Ordinal, Quantal, or Binary. For continuous variables, the response unit (ResponseUnit) is also relevant. Additionally, also a reference to the test method guideline, e.g., standaridised assay kit may also be specified (GuidelineMethod).

Table 2.16: Table definition for Responses.

Name	Type	Description	Aliases	Required
idResponse	AlphaNumeric(50)	Unique identification code of the response. In the EuroMix data collection, a EuroMix coding system has been set up in which the id of the test system prefixes the id of the response. E.g., 'HepaRG-PCR-PPARA', 'RatWEC-PCR-CYP26a1' and 'MouseDevelopmental-FacialPrimordia-malformed-E9'.	idResponse, ResponseId, Response, Id	Yes
CodeSystem	AlphaNumeric(100)	Identifier of the coding system of the response code.	CodeSystem	No
Name	AlphaNumeric(100)	Name of the response.	Name	No
Description	AlphaNumeric(200)	Additional description or label of the response.	Description	No
idTestSystem	AlphaNumeric(50)	Unique identification code of the test system.	idTestSystem, idSystem, SystemId, TestSystem	Yes
Guideline-Method	AlphaNumeric(200)	Reference to the test method guideline, e.g., standaridised assay kit.	Guideline-Method	No
ResponseType	*ResponseTypes*	The data type of the response measurements (e.g., continuous multiplicative, ordinal, categorical).	ResponseType	Yes
ResponseUnit	AlphaNumeric(100)	If the response type is Continuous, then this should be the unit of the response, e.g., kg.	ResponseUnit	No

Table aliases: Responses, Response, RawResponses.

Responses settings

Selection settings

Table 2.17: Selection settings for module Responses.

Name	Description
Response(s)	The response(s) of interest.

Responses as data

A response is a measurable endpoint defined in a test system. It has a unit and a measurement type (e.g., continuous non-negative, quantal).

- *Responses data formats*

2.1.5 Substances

Substances are chemical entities that can refer to: 1) active substances such as investigated in toxicology; 2) measured substances such as defined in specific analytical methods. MCRA assessments can have one or more substances as the scope. When more than one substance is specified, there is an option to perform a cumulative assessment. In that case one of the substances has to be indicated as the index/reference substance, and results will be expressed in equivalents of the index substance.

Output of this module is used by: *Concentrations Processing factors Unit variability factors Occurrence patterns Substance authorisations Substance conversions Concentration limits Concentration models Foods as measured Food conversions Consumptions by food as measured Dietary exposures with screening Dietary exposures Non-dietary exposures Exposures Exposure mixtures Human monitoring data Human monitoring analysis QSAR membership models Molecular docking models Kinetic models Active substances Relative potency factors Hazard characterisations Points of departure Dose response models Dose response data Inter-species conversions Intra species factors Risks*

Substances data formats

Substances

Substances are primary entities of the data model. Substance intakes are of main interest in exposure assessments and the effect of intake on human health is of interest in risk assessments. In the substances table, the substance entities and other relevant substance properties that are relevant for the assessment at hand should be defined.

Substances

Each substance should have a unique identification code (idSubstance), and optionally, a name and description may be used for a more detailed description of the entity. Additional properties, such as the molecular mass (MolecularMass) and Cramer class (CramerClass) may also be specified. Example: Captan (idSubstance RF-0061-001-PPP) has MolecularMass 300.5922 and CramerClass 3.

Table 2.18: Table definition for Compounds.

Name	Type	Description	Aliases	Required
idSubstance	AlphaNumeric(50)	The unique identification code of the substance. This code may be from an existing coding system, such as CAS-codes or Param codes of EFSA, or it may be a used-defined code.	idSubstance, SubstanceId, Substance, Code, Id, idCompound, CompoundId, Compound	Yes
Name	AlphaNumeric(100)	The substance name.	Name, Compound-Name, SubstanceName, PesticideName	No
Description	AlphaNumeric(200)	Substance description.	Description	No
ARFD	Numeric	The acute reference dose of the critical effect. Note that this is always specified in mg/kg bw/day (exposure).	ARFD	No
ADI	Numeric	The acceptable daily intake. Note that this is always specified in mg/kg bw/person (exposure).	ADI	No
SF	Numeric	The safety factor belonging to the ADI/ARFD.	SF	No
CramerClass	Integer	The Cramer class of the substance.	CramerClass	No
MolecularMass	Numeric	The molecular (molar) mass.	MolecularMass, Mass, MolarMass, Molecular-Weight, MolarWeight	No

Table aliases: Compounds, Compound, Substances, Substance, RawCompounds.

Substances settings

Selection settings

Table 2.19: Selection settings for module Substances.

Name	Description
Index substance	The substance of interest or index substance.

Substances as data

- *Substances data formats*

2.1.6 Test systems

Test systems are biological or artificial systems used for assessing hazard in relation to chemical exposure from substances in varying doses. Test systems may refer to 1) in-vivo test systems (e.g. a rat 90-day study, a human biomonitoring study); 2) in-vitro test systems (e.g. HepaRG cells).

Output of this module is used by: *Responses Dose response models Dose response data*

Test systems data formats

Test Systems

Test systems are the biological systems (e.g., animals) or in-vitro systems on which responses related to health effects can be measured.

Test Systems

Each test system should have a unique identification code (idSystem), and (optionally) a name and a description. The test system's type (TestSystemType) indicates the type whether the test system is an in-vivo test system (in which case it is a model for external exposure) or any of a range of other, in-vitro, options (cell-line, etc., which all will be interpreted as models for internal exposure). Additionally, if applicable, the organ (e.g., liver) of the test system and the route of exposure (RouteExposure) for in-vivo test systems (oral, dermal or inhalation) may be specified.

Table 2.20: Table definition for TestSystems.

Name	Type	Description	Aliases	Required
idTestSystem	AlphaNumeric(50)	Unique identification code of the test system.	idTestSystem, idSystem, Id, Code	Yes
CodeSystem	AlphaNumeric(50)	Identifier of the code system of the test systems.	CodeSystem	No
Name	AlphaNumeric(100)	Name of the test system.	Name	No
Description	AlphaNumeric(200)	Additional description or label of the test system.	Description	No
TestSystem-Type	*TestSystemTypes*	The type of the test system, i.e., in-vivo, cell-line, etc.	TestSystem-Type, SystemType	No
Organ	AlphaNumeric(100)	If applicable, the organ that the cells originate from associated with the in vitro test-system.	Organ	No
Species	AlphaNumeric(100)	If applicable, the species associated with the test-system.	Species	No
Strain	AlphaNumeric(100)	If applicable, the strain of the species associated with the test-system.	Strain	No
RouteExposure	*ExposureRouteTypes*	If applicable, the route of exposure associated with the in vivo test-system, oral, dermal, inhalation, s.c., i.v.	ExposureRoute-Type, ExposureRoute, RouteExposure	No
Guideline-Method	AlphaNumeric(200)	Reference to test guideline.	GuidelineStudy	No
Reference	AlphaNumeric(200)	External reference(s) to other sources containing more information about the test system. E.g., publications, website, documents.	Reference	No

Table aliases: TestSystems, TestSystem, Systems, System, RawTestSystems.

Test systems as data

- *Test systems data formats*

2.2 Consumption modules

Consumption modules specify the *consumptions* of *foods* by surveyed individuals in *populations*. Foods can be related to each other using *food recipes*.

2.2.1 Consumptions

Consumptions data are the amounts of foods consumed on specific days by individuals in a food consumption survey. For acute exposure assessments, the interest is in a population of person-days, so one day per individual may be sufficient. For chronic exposure assessments, the interest is in a population of person, so preferably two or more days per individual are needed.

This module has as primary entities: *Populations Foods*

Output of this module is used by: *Food conversions Consumptions by food as measured*

Consumptions data formats

Consumption data is often collected in 24-hour dietary recall studies and contains the food consumptions and consumption amounts for a number of individuals on a number of days. For each of the individuals, the bodyweight should be specified, and optionally also age, sex, and other properties may be recorded. If applicable, sampling weights may also be specified that can be used to correct the sample of individuals in the survey to a more representative sample of the targetted population. The consumption amounts are usually expressed in grams, but may also be expressed in alternative units of plates, cups, or spoons. Optionally, the uncertainty of food consumption quantifications can be specified, see [42].

Consumptions

Consumption surveys are described using three tables: FoodSurveys, Individuals, and Consumptions. Individuals are linked to food surveys using the survey code (idFoodSurvey), and consumptions are linked to individuals using the individual codes (idIndividual). The food codes used to identify the consumed foods should match with the codes provided by the foods entity definitions.

Food consumption surveys

The records of the food consumption surveys table contain the ids, names, descriptions, and other relevant metadata of consumption surveys.

Table 2.21: Table definition for FoodSurveys.

Name	Type	Description	Aliases	Required
idSurvey	AlphaNumeric(50)	Unique identification code of the food consumption survey.	idSurvey, idFoodSurvey, Survey, FoodSurvey, SurveyId, FoodSurveyId, Name, Code, Id	Yes
Description	AlphaNumeric(200)	Description of the food consumption survey.	Description	No
Location	AlphaNumeric(50)	The location or country where survey is held. It is recommended to use ISO Alpha-2 country codes.	Location, Country	No
BodyWeight-Unit	*BodyWeightUnits*	The unit of bodyweight of the individuals of the survey: kg (default) or g.	BodyWeight-Unit, UnitBody-Weight, WeightIn	No
AgeUnit	AgeUnit	The unit of age, i.e., year or month.	UnitAge, agein, AgeUnit	No
Consumption-Unit	*ConsumptionUnits*	The unit of the use/consumption amounts of the consumptions of the survey: g (default) or kg or CustomUnit (see table food consumption quantifications table).	AmountUnit, UnitAmount, AmountUnit, Consumption-Unit	No
StartDate	DateTime	The start date of the survey.	StartDate	No
EndDate	DateTime	The end date of the survey.	EndDate	No
NumberOf-SurveyDays	Integer	The number of days each individual participated in the survey.	NumberOf-SurveyDays, NDaysInSurvey	Yes
idPopulation	AlphaNumeric(50)	Unique identification code of the population.	IdPopulation, PopulationId	No

Table aliases: FoodSurvey, FoodSurveys, Survey, Surveys, RawFoodSurveys.

Individuals

The individuals of a survey are recorded in the individuals table.

Table 2.22: Table definition for Individuals.

Name	Type	Description	Aliases	Required
idIndividual	AlphaNumeric(50)	Unique identification code of the individual.	idIndividual, IndividualId, Individual, Id	Yes
idFoodSurvey	AlphaNumeric(50)	The identification code / short name of survey.	idSurvey, idFoodSurvey, Survey, FoodSurvey, SurveyId, FoodSurveyId, SurveyCode	Yes
BodyWeight	Numeric	The body weight of the individual.	BodyWeight, Weight	Yes
Sampling-Weight	Numeric	The sampling weight for an individual (default = 1).	SamplingWeight	No
NumberOf-SurveyDays	Integer	The number of days the individual participated in the survey.	NumberOf-SurveyDays, NumberOfDays-InSurvey, DaysInSurvey, NDaysInSurvey	No
Age	Numeric	The age of the individual.	Age	No
Gender	AlphaNumeric(50)	The gender of the individual. It is recommended to use the codes Male/Female for coding the gender.	Gender	No
Other inidividual properties		Other individual properties can be added just like the fields age and gender. These properties are automatically parsed as co-factors or co-variables.		No

Table aliases: Individuals, Individual, RawIndividuals.

Individual properties

Individual properties, additional columns that can also be specified as additional columns in the Individuals table

Table 2.23: Table definition for IndividualProperties.

Name	Type	Description	Aliases	Required
Name	AlphaNumeric(50)	The name of the property.	Id	Yes

Table aliases: IndividualProperties, IndividualProperty, RawIndividualProperties.

Individual property values

Individual property values, additional columns that can also be specified as additional columns in the Individuals table

Table 2.24: Table definition for IndividualPropertyValues.

Name	Type	Description	Aliases	Required
idIndividual	AlphaNumeric(50)	The identification number of the Individual.	Id	Yes
PropertyName	AlphaNumeric(50)	The name of the property.	Name	Yes
TextValue	AlphaNumeric(50)	The value of the property as text value.		No
DoubleValue	Numeric	The value of the property as number.		No

Table aliases: IndividualPropertyValues, IndividualPropertyValue, RawIndividualPropertyValues.

Consumptions

The individual consumptions are recorded in the consumptions table.

Table 2.25: Table definition for Consumptions.

Name	Type	Description	Aliases	Required
idIndividual	AlphaNumeric(50)	The unique identification code of the consumer (individual).	idIndividual, IndividualId, Individual	Yes
idFood	AlphaNumeric(50)	The food code (food as eaten code).	idFood, Food, FoodId, FoodConsumed, FoodAsEaten	Yes
idUnit	AlphaNumeric(50)	Identification code of the unit in which the food is consumed (e.g. plate, cup, spoon).	idUnit, Unit, UnitId	No
idDay	AlphaNumeric(50)	Identification code of the day of consumption, sequential number	idDay, DayId, Day, DayOfSurvey	Yes
idMeal	AlphaNumeric(50)	Identification code of the meal (eating occasion within a day).	idMeal, MealId, Meal	No
Amount	Numeric	The consumed portion of food in g (default) or kg or quantity of a plate, cup, spoon. Days without consumptions are not recorded.	Amount, Amount-Consumed	Yes
DateConsumed	DateTime	The date of the consumption.	DateConsumed, Consumption-Date	No

Table aliases: FoodConsumptions, FoodConsumption, Consumptions, Consumption, RawFoodConsumptions.

Food consumption quantificiations

Food consumption quantifications record information about food consumption quantities that are associated with unit-consumptions of foods.

Table 2.26: Table definition for FoodConsumptionQuantifications.

Name	Type	Description	Aliases	Required
idFood	AlphaNumeric(50)	The food code of the quantification.	idFood, FoodId, Food	Yes
idUnit	AlphaNumeric(50)	The code of the unit of consumption. E.g spoon, plate, cup. Units may depend on food.	idUnit, UnitId, Unit	Yes
UnitWeight	Numeric	The unit weight/portion size of the food, specified in grams.	UnitWeight	Yes
UnitWeight-Uncertainty	Numeric	The uncertainty in unit weight/portion size (%).	UnitWeight-Uncertainty, UnitWeight%	No
Amount-Uncertainty	Numeric	The uncertainty in amount consumed (%). The label 'general' specifies a default value for the uncertainty when specific information for combinations of food and unit in food consumptions table is not available.	Amount-Uncertainty, Amount%	No

Table aliases: FoodConsumptionQuantifications, FoodConsumptionQuantification, RawFoodConsumptionQuantifications.

Consumptions settings

Selection settings

Table 2.27: Selection settings for module Consumptions.

Name	Description
Food survey	The food consumption representative for the population of interest.
Restrict population to consumers or consumer days only	Specifies whether the population should be restricted to the individuals (chronic) or individual days (acute) that have non-zero consumption.
Restrict population to consumers or consumer days with consumptions of specific foods	Specifies whether the population should be restricted to the individuals (chronic) or individual days (acute) consuming any of the foods of the specified subset.
Selected foods-as-eaten	Set of consumed foods that are of particular interest for restricting the consumers / consumption days.
Consumption subset: restrict to consumptions of specific foods	If checked, then the consumptions are restricted to those of the specified food-as-eaten subset.
Selected foods-as-eaten	Set of consumed foods that are of particular interest.
Ignore sampling weights	If checked, individual sampling weights are not used (sampling weight = 1). If unchecked, the specified sampling weights are used.

Uncertainty settings

Table 2.28: Uncertainty settings for module Consumptions.

Name	Description
Resample individuals	Individual data are resampled from the original database using the bootstrap methodology (Efron 1979, Efron & Tibshirani 1993).
Resample portion sizes	Specifies whether portion sizes should be resampled based on food consumption quantification data, see (Souverein et al. 2011).

Consumptions uncertainty

In MCRA, in an acute exposure assessments, individual consumption day data are resampled, thus preserving the multivariate consumption patterns and associated weights and/or other individual characteristics. In MCRA we resample the set of individuals x number of survey days. We think that this implementation better reflects the notion of acute exposure which is expressed as the normalized intake per day. For chronic exposure assessments the resampling algorithm remained unchanged and the set of individuals (with corresponding days) is resampled.

Consumptions as data

Consumptions data are the amounts of foods consumed on specific days by individuals in a food consumption survey.

- *Consumptions data formats*

2.2.2 Food recipes

Food recipes data specify the composition of specific foods (typically: foods-as-eaten) in terms of other foods (intermediate foods or foods-as-measured) by specifying proportions in the form of a percentage.

This module has as primary entities: *Foods*

Output of this module is used by: *Food conversions*

Food recipes data formats

Food recipes

Recipe data to specify the ingredients of foods. Food recipes can be used to describe the ingredients of a composite food (e.g., of apple pie), or to specify the amount of a primary ingredient needed to obtain 100g of the food (e.g., grapes to raisins). Recipe is commonly used recursively (e.g., apple pie contains apple and flour, flour contains wheat).

Recipes

Table 2.29: Table definition for FoodTranslations.

Name	Type	Description	Aliases	Required
idFromFood	AlphaNumeric(50)	The code of the composite food (from-code), i.e., the code of the food for which the ingredient(s) are specified.	idFromFood, FromFoodId, FromFood, FoodFrom, Food	Yes
idToFood	AlphaNumeric(50)	The code of the ingredient food (to-code).	idToFood, ToFoodId, ToFood, FoodTo, Ingredient	Yes
Proportion	Numeric	Proportion of each ingredient in the food (%).	Proportion, Proportion%	Yes
idPopulation	AlphaNumeric(50)	Unique identification code of the population.	IdPopulation, PopulationId	No

Table aliases: FoodTranslations, FoodTranslation, FoodCompositions, FoodComposition, RawFoodTranslations.

Food recipes as data

Food recipes are provided as data in the form of simple composition tables.

- *Food recipes data formats*

2.2.3 Market shares

Market shares data specify for a given food, percentages of more specific foods (subfoods, e.g. brands) representing their share in a market. Market shares are used when consumption data are available at a more generalised level than concentration data.

This module has as primary entities: *Foods*

Output of this module is used by: *Food conversions*

Market shares data formats

MarketShares

Describes the shares (proportions) in a market.

Market shares

Market shares main table.

Table 2.30: Table definition for MarketShares.

Name	Type	Description	Aliases	Required
idFood	AlphaNumeric(50)	The subtype of the food.	idFood, FoodId, Food, FoodType	Yes
Percentage	Numeric	Market share of each subtype (%)	Percentage, Marketshare-Percentage, MarketShare, MarketShare-Percentage, MarketShare%	Yes

Table aliases: MarketShares, MarketShare, FoodMarketShares, FoodMarketShare, RawMarketShares.

Market shares as data

Market shares are provided as data in the form of percentages.

- *Market shares data formats*

Market shares and brand loyalty

Sometimes measurements of substances in food are available at a more detailed food coding level than consumption data. For example, measurements may have been made for specific brands of a food whereas the consumption survey did not record the brand. MCRA allows to specify market share data for subtypes of a food (e.g. A$1, A$2, A$3 are three brands of food A), and to calculate acute exposure based on such market shares.

For chronic assessments **brand loyalty** should be specified according to a simple Dirichlet model [22]. Technically, the Dirichlet model for brand choice needs nbrand parameters α_i (which should be positive real numbers). The average brand choice probability for each brand is

$$\alpha_i/S$$

where

$$S = \sum \alpha_i$$

By definition, the market shares m_i should be proportional to the brand choice probabilities, and thus to the parameters α_i. Thus means that S, the sum of the alphas, is the only additional parameter that should be specified, and indeed this is the parameter that determines brand loyalty. $S = 0$ corresponds to absolute brand loyalty, and brand loyalty decreases with increasing S. We define $L = (1 + S)^{-1}$ as an interpretable brand loyalty parameter, where now $L = 0$ and $L = 1$ correspond to the situations of no brand loyalty and absolute brand loyalty, respectively. Given empirical or parametric distributions of consumption and concentration values, the algorithm for chronic exposure assessment now operates as follows:

1. Simulate consumptions for a large number n of individuals.
2. Simulate n selection probabilities from the Dirichlet distribution
3. For each individual, simulate d brand choices from a multinomial distribution using the individual specific selection probabilities from step 2.
4. For all individuals and days simulate values from the appropriate concentration distribution.
5. Multiply consumption with concentration to obtain exposure.

2.3 Occurrence modules

The basic occurrence data are *concentrations* for *substances* in *foods*, sometimes specified separately for a focal food as *focal food concentrations*. In some cases *concentration limits* are used as a stand-in when data are missing.

Concentration data are recalculated (if needed) as *active substance concentrations* in *foods-as-measured*. If substance concentrations are not specified directly for the *active substances*, then they are converted using *substance conversions* and/or specified authorised *occurrence patterns*. The composition of mixed samples in total diet studies is described in *total diet study sample compositions*. *Food extrapolation rules* specify if insufficient data for a food can be suppleted with data from another food. From these basic data the list of *foods-as-measured* is derived.

Active substance concentrations in *foods-as-measured* are modelled in *concentration models*, optionally allowing for *occurrence pattern models*. In addition, *processing factors* and *unit variability factors* can be provided for further use in *dietary exposure assessment*.

2.3.1 Concentration limits

Concentration limits specify (legal) limit values for substance concentrations on foods and are sometimes used as conservative values for concentration data. In the framework of pesticides the legal Maximum Residue Limit (MRL) is the best known example.

This module has as primary entities: *Foods Substances*

Output of this module is used by: *Concentrations Concentration models Foods as measured*

Concentration limits data formats

The concentration limits table describes limit values (e.g., MRLs) for specific food/substance combinations. This data may be used, for instance, for the food/substance combinations for which no concentration data is available. The food codes (idFood) and substance codes (idSubstance) should match the codes of the foods and substances table respectively.

Concentration limits

Concentration limits are concentration limit values for specific food and substance combinations originatin from regulations (e.g., MRLs). This data may be used, for instance, for the food/substance combinations for which no concentration data is available.

Concentration limits

The food codes (idFood) and substance codes (idSubstance) should match the codes of the foods and substances table respectively.

Table 2.31: Table definition for MaximumResidueLimits.

Name	Type	Description	Aliases	Required
idFood	AlphaNumeric(50)	Code of the food of this residue limit definition.	idFood, FoodId, Food	Yes
idSubstance	AlphaNumeric(50)	Code of the substance of this residue limit definition.	idSubstance, SubstanceId, SubstanceCode, Substance, idCompound, CompoundId, Compound-Code, Compound	Yes
Limit	Numeric	Residue limit value.	Limit, Maximum-ResidueLimit, Maximum-ResidueLimits, MRL	Yes
StartDate	DateTime	Start date of the period during which the limit applies.	StartDate	No
EndDate	DateTime	End date of the period during which the limit applies.	EndDate	No
Concentration-Unit	*ConcentrationUnits*	The unit of the limit value (default mg/kg).	Concentration-Unit, Unit	No

Table aliases: ResidueLimits, ResidueLimit, MaximumResidueLimits, MaximumResidueLimit, MRLs, MRL, RawMaximumResidueLimits.

Concentration limits as data

- *Concentration limits data formats*

2.3.2 Concentration models

Concentration models are distributional models of substance concentrations on foods. They describe both the substance presence (yes/no, with no representing an absolute zero concentration) and the substance concentrations. Concentration models are specified per food/substance combination.

This module has as primary entities: *Foods Substances Effects*

Output of this module is used by: *Dietary exposures with screening Dietary exposures*

Concentration models calculation

There are a number of *concentration model types* are available. A basic distinction is between using the empirical concentration data (empirical model), fitting a statistical model to the concentration data (parametric model), or to construct a model from (conservative) limit values. Settings relevant for some of these model types as well as other settings are described under *concentration model settings*.

Concentration model types

Empirical model

Data points are sampled at random from the available set. Non-detects are handled by imputation. If occurrence patterns are used, a proportion p_0/p_{ND} of non-detects is set as 0. See also *appendix*.

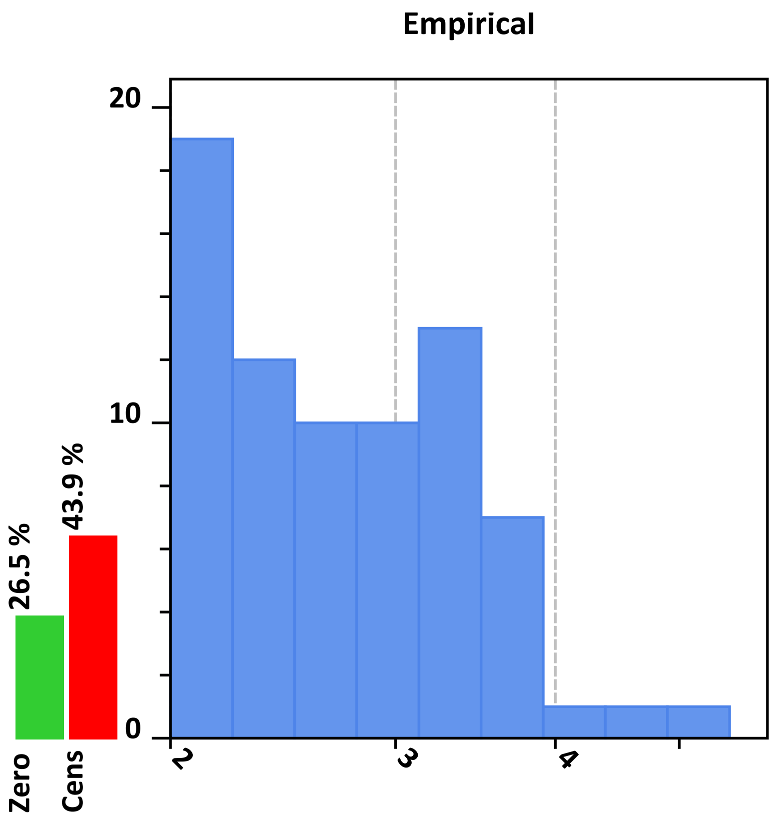


Figure 2.3: Empirical distribution

Non-detect spike lognormal model

A binomial model is used to estimate the proportion p of positive values (detects). This is just the proportion observed in the data (unless agricultural use data have been used to set a proportion of true zeroes). A lognormal model is fitted to the positive data. This provides estimates of μ and σ, which are the mean and standard deviation of the natural logarithm of the concentration. Simulated concentrations are a non-detect with probability $p_{ND} = 1 - p$ or a value sampled from the fitted lognormal distribution with probability p. Non-detects are handled by imputation. If occurrence patterns are used, a proportion p_0/p_{ND} of non-detects is set as 0. Minimum requirements: at least two positive concentration values. See also *appendix*.

Non-Detect-Spike Truncated lognormal model

A binomial model is used to estimate the proportion p of positive values (detects). This is just the proportion observed in the data (unless agricultural use data have been used to set a proportion of true zeroes in which case p is calculated on the remaining proportion). A truncated lognormal model, with LOR as the truncation limit, is fitted to the positive data, leading to estimates of μ and σ, which are the mean and standard deviation of the natural logarithm of the concentration. Simulated concentrations are a non-detect with probability $p_{ND} = 1 - p$ or a value sampled from the fitted lognormal distribution with probability p. Non-detects are handled by imputation. If occurrence patterns are used, a proportion p_0/p_{ND} of non-detects is set as 0. Minimum requirements: at least two positive concentration values, all non-detects must have one LOR value. See also *appendix*.

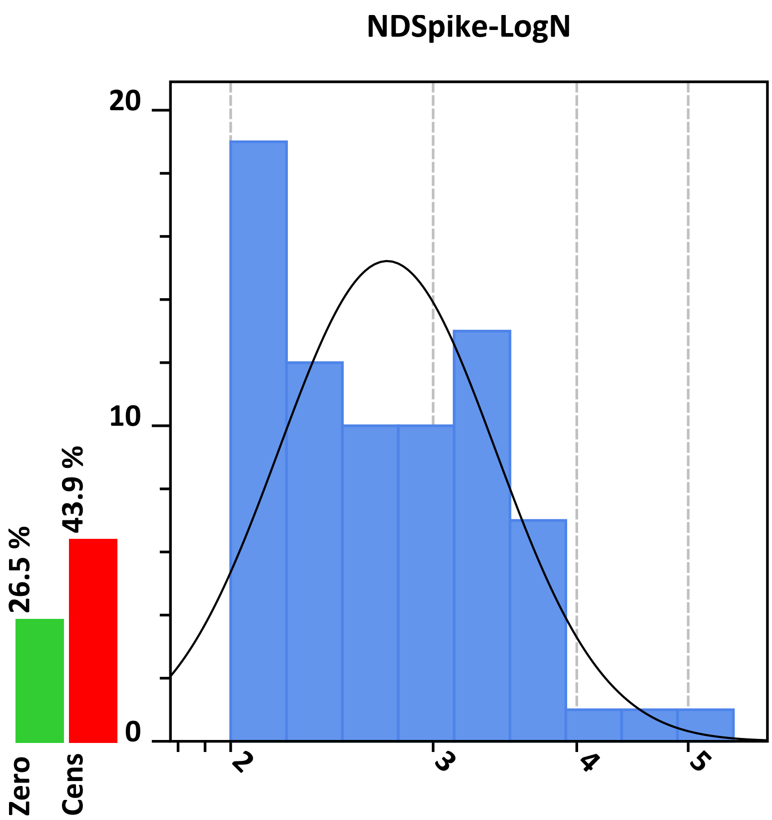


Figure 2.4: Nondetect Spike Lognormal distribution

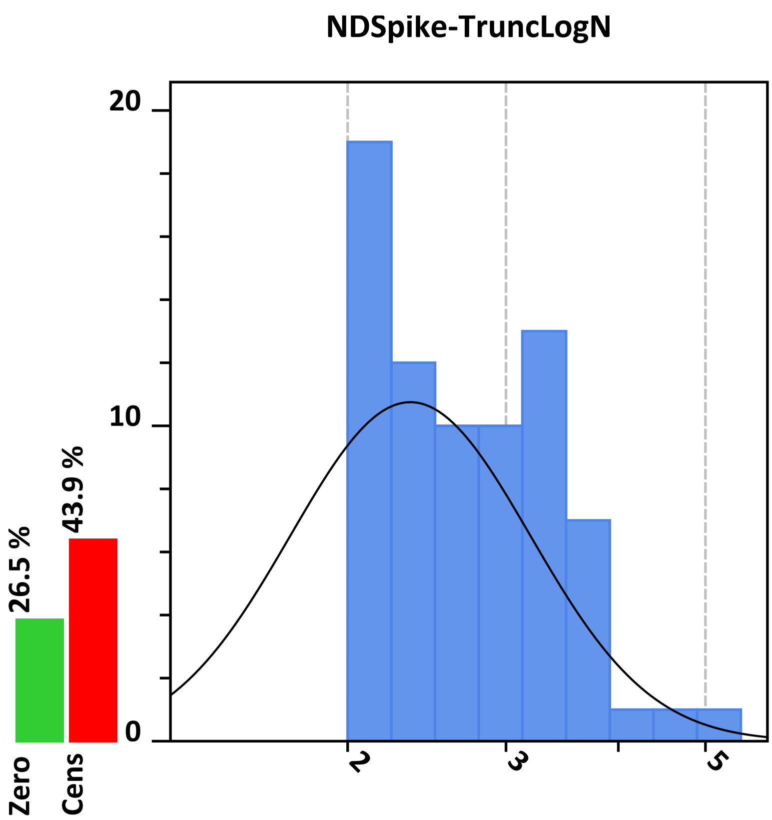


Figure 2.5: Nondetect Spike Truncated Lognormal distribution

Censored Lognormal model

A censored lognormal model, with LOR as the censoring limit, is fitted to the data, both positives and non-detects. This provides estimates of μ and σ, which are the mean and standard deviation of the natural logarithm of the concentration. If agricultural use data are being used, then a proportion p_0/p_{ND} of non-detects will be excluded, where p_0 will be lowered to p_{ND} if it would be higher. Simulated concentrations are sampled from the fitted lognormal distribution. If agricultural use data have been used, simulated concentrations are 0 with probability p_0 or are sampled from the fitted lognormal distribution with probability $1 - p_0$. Minimum requirements: at least one positive concentration value. See also *appendix*.

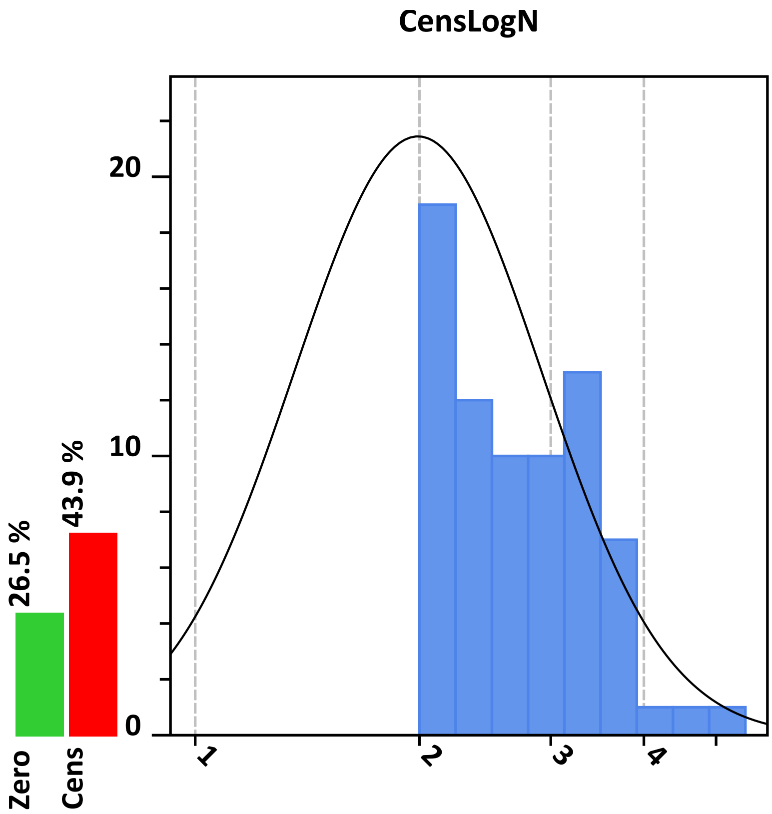


Figure 2.6: Censored Lognormal distribution

Zero-spike censored lognormal model

A mixture distribution of a spike of true zeroes and a censored lognormal model, with LOR as the censoring limit, is fitted to the data (non-detects and positives. This provides estimates of p_0, which is the proportion of true zeroes, and of μ and σ, which are the mean and standard deviation of the natural logarithm of the concentration. Simulated concentrations are 0 with probability p_0 and are sampled from the fitted lognormal distribution with probability $1-p_0$. Minimum requirements: at least one positive concentration value, no agricultural use data for the food-compound combination (which directly specify p_0, therefore it should not be estimated from the data). See also *appendix*.

Non-detect spike MRL model

This model simply takes values specified in an input table as Maximum Residue Limit (MRL) to be used for the proportion of positive values in the concentration dataset, and can be used to force the use of a pessimistic value.

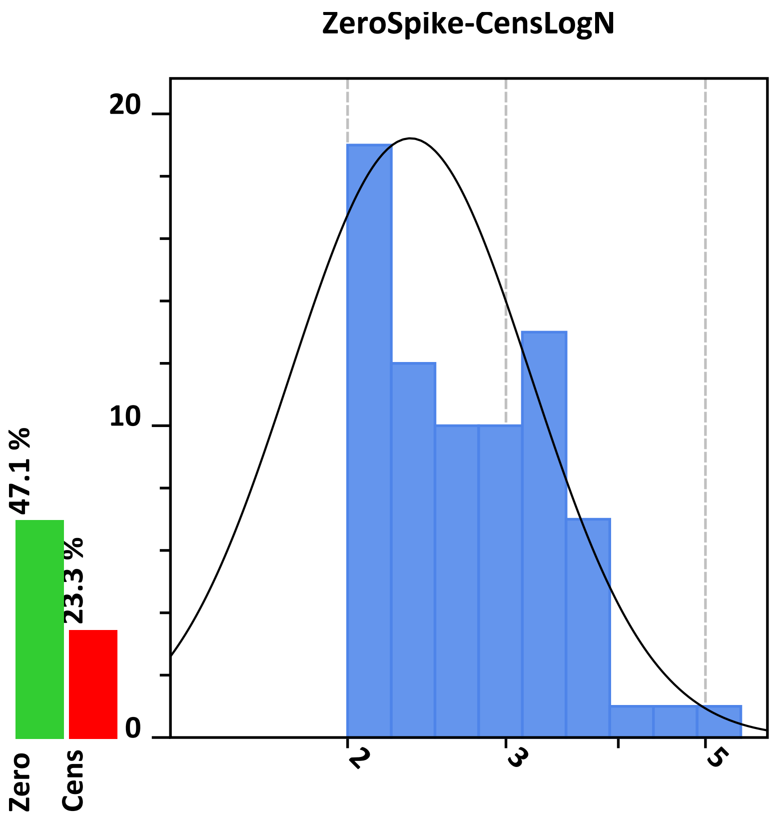


Figure 2.7: Zero Spike Censored Lognormal distribution

Summary statistics model

For this model, no individual measurements on raw agricultural commodities are needed. The final estimates of μ and σ are simply provided or pooled or estimated using e.g. a coefficient of variation. Specific use of this model is found in Total Diet Studies. In general, each TDS food sample is prepared only once, yielding one measurement for a TDS food sample. The variability of the underlying distribution is unknown. However, a rough guess can be made using the e.g. coefficient of variation of the subsamples (in general raw agricultural commodities) that compose the TDS food sample. The estimated standard deviation is calculated as a pooled estimate using the coefficient of variation and the count of each subsample in the TDS food.

Imputation

A complication in concentration modelling occurs if results are reported as being below a limit. Different names may be used for such a limit, e.g. limit of detection or limit of quantification. For the purpose of exposure assessment it is only relevant whether results are reported as a positive value or as a non-detect, therefore we refer to any limit as the **Limit Of Reporting** (LOR), and any result reported as '<LOR' is termed a **nondetect**. The value of LOR should always be known for the particular analytical method used.

Non-detects are a very common phenomenon for some classes of substances like pesticides. Non-detects can be handled by replacing them with a given value (**imputation**), or by incorporating them in a parametric model. In the imputation approach, non-detects (values reported less than LOR) can be replaced in simulations by any value between 0 and LOR * *constant*.

Imputation may be also dependend on the authorisation status of a substance i.c. whether the use of a substance on a agricultural crop is allowed or not.

In Figure 2.8 to Figure 2.11, the various scenarios are displayed. Two substances, Fenamidine and Hexythiazox are indicated with a brown box, these substances are authorized.

No imputation

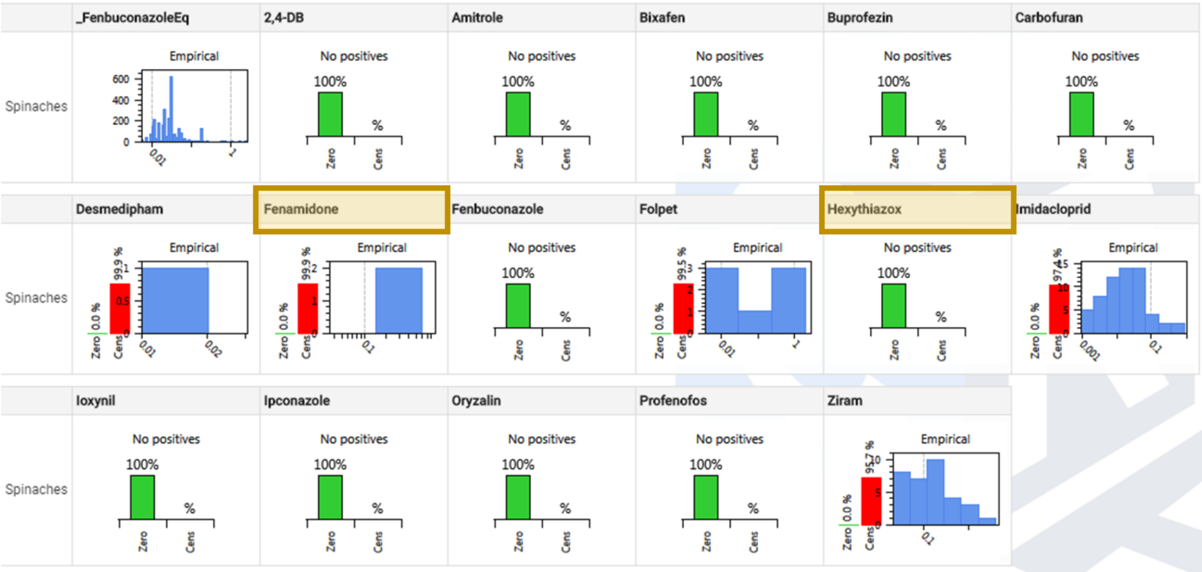


Figure 2.8: Tier 1: Nondetects are not replaced. For Fenamidine and Hexythiazox (brown boxes) authorized use is assumed.

Impute all nondetects

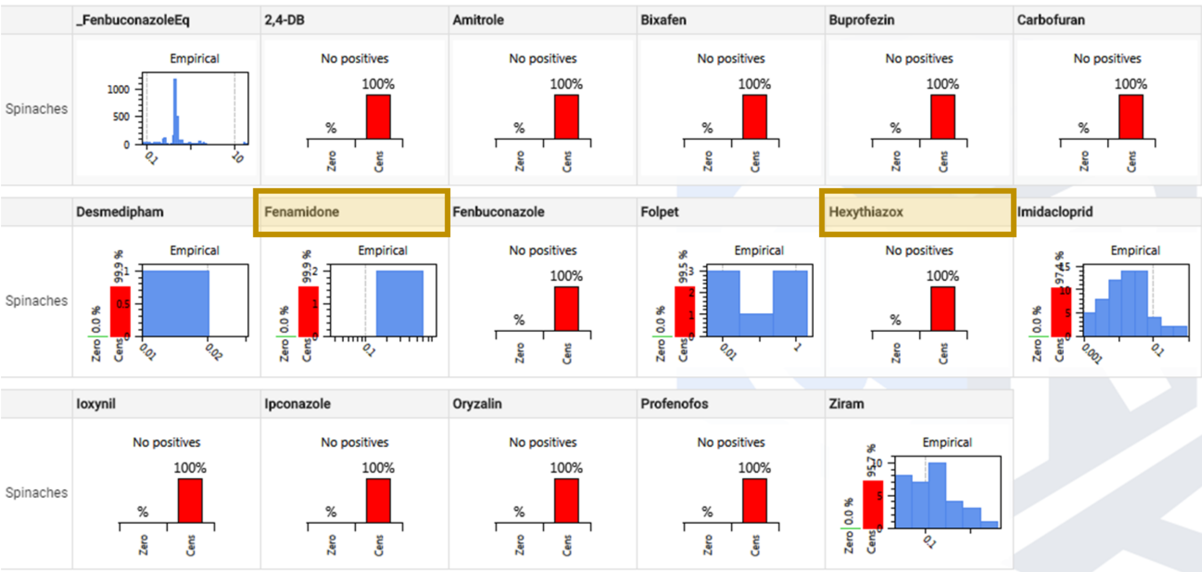


Figure 2.9: All nondetects are replaced by a constant factor x LOR. For Fenamidine and Hexythiazox (brown boxes) authorized use is assumed.

Impute nondetects based on authorized uses

No imputation except for authorized uses

Concentration models settings

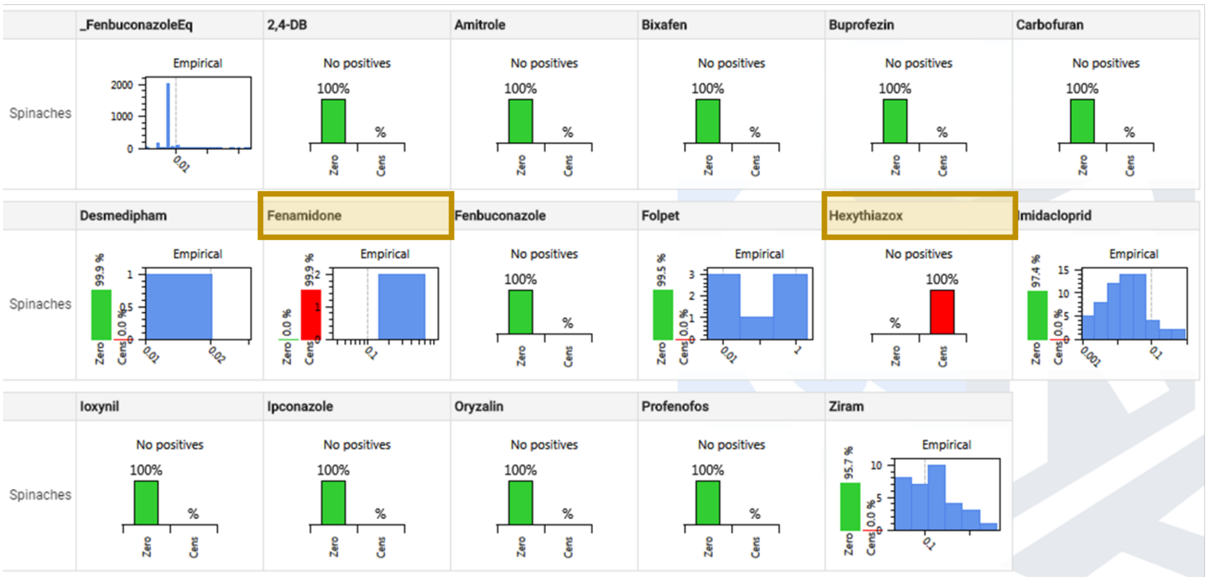


Figure 2.10: Nondetects are replaced by a constant factor x LOR for authorized uses. For Fenamidine and Hexythiazox (brown boxes) authorized use is assumed.

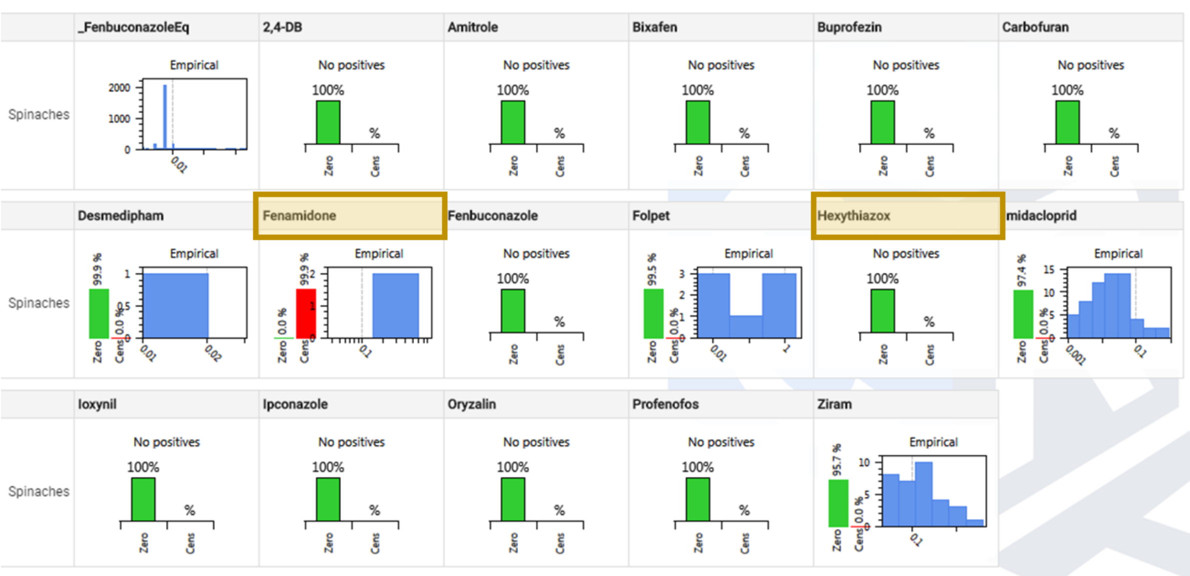


Figure 2.11: Tier 1: Nondetects are not replaced except for authorized uses (replaced by a constant factor x LOR). For Fenamidine and Hexythiazox (brown boxes) authorized use is assumed.

Calculation settings

Table 2.32: Calculation settings for module Concentration models.

Name	Description
Concentration model tier (Note: set tier separately in sub-modules)	Custom model, or set according to EFSA Guidance 2012. Note: you may need to set the tier separately in sub-modules.
Default concentration model	The concentration model type that will be used as default for all food/substance combinations. If this model type cannot be fitted, e.g., due to a lack of data, a simpler model will be chosen automatically as a fall-back.
Include MRL fallback model	Use the MRL as fallback model in case the occurrence data is insufficient for other concentration modelling options.
Restrict LOR imputation to authorised uses	Specifies whether imputation of factor x LOR should be limited to authorised uses only.
Non-detects replacement	How to replace non-detects (when not co-modelled, as in censored models).
Factor f (f x LOR)	Replace non-detects by Limit Of Reporting (LOR) times this factor. Constant (f), e.g. 0.5.
MRL Factor (f x MRL)	Use f x MRL as concentration estimate of the MRL models.
Sample based	Include co-occurrence of substances in samples in simulations. If checked, substance residue concentrations are sampled using the correlations between values on the same sample. If unchecked, any correlation between substances is ignored, substance residue concentrations are sampled ignoring the correlations between values on the same sample.
Imputation of missing values	If checked, in procedure of EFSA Guidance 2012, Appendix 1, impute missing values using substance based concentration models. If unchecked, missing values are not imputed (set to 0).
Correlate imputed values with sample potency	If checked, in procedure of EFSA Guidance 2012, Appendix 1, correlate high imputed values with high cumulative potency samples. If unchecked, random imputation.
Use occurrence patterns for imputation	Use of AU data is relevant for imputation of non-detects in the concentration data. Non-detects can be imputed with zero when an agricultural use is un-authorised (all non-detects) or when the AU percentage is less than 100% (part of the non-detects). If checked, AU data is expected and will be used for the imputation of non-detects. If unchecked, 100% potential presence is assumed for all substances on all foods.

Uncertainty settings

Table 2.33: Uncertainty settings for module Concentration models.

Name	Description
Parametric uncertainty	For resample concentrations: specifies whether the uncertainty assessment is based on a parametric approach.

Concentration models tiers

In addition to the possibility for users to work with their own choices for all settings, MCRA implements four tiers from two documents:

- The optimistic and pessimistic basic assessments from the *EFSA 2012 Guidance on the Use of Probabilistic Methodology for Modelling Dietary Exposure to Pesticide Residues* [3].

- Tier 1 and 2 from the *European Commission working document SANTE-2015-10216 rev. 7 (2018)* on risk management aspects related to the assessment of cumulative exposure [37].

Overview

Table 2.34: Tier overview for module Concentration models.

Name	EFSA 2012 Optimistic	EFSA 2012 Pessimistic	EC 2018 Tier 1	EC 2018 Tier 2
Default concentration model	Empirical	NonDetectSpikeLogNormal	Empirical	Empirical
Include MRL fallback model	false	true	false	false
Restrict LOR imputation to authorised uses		false	false	false
Non-detects replacement	ReplaceByZero	ReplaceByLOR	ReplaceByLOR	ReplaceByLOR
Factor f (f x LOR)		1	0.5	0.5
MRL Factor (f x MRL)		1		
Sample based	true	true	true	true
Imputation of missing values	false	true	true	true
Correlate imputed values with sample potency	false	true	true	false
Use occurrence patterns for imputation			true	true
Parametric uncertainty	false	true	false	false

The sections below describe the settings specified by each tier in detail.

EFSA 2012 Optimistic

Use the optimistic model settings according to the EFSA Guidance 2012. Non-detects and missing values are replaced by zero.

Table 2.35: Tier definition for EFSA 2012 Optimistic.

Name	Setting
Default concentration model	Empirical
Include MRL fallback model	false
Non-detects replacement	ReplaceByZero
Sample based	true
Imputation of missing values	false
Correlate imputed values with sample potency	false
Parametric uncertainty	false

EFSA 2012 Pessimistic

Use the pessimistic model settings according to the EFSA Guidance 2012. A nondetect spike lognormal model is fitted to the positive residue values. Non-detects are replaced by the LOR. When the number of positives is smaller than 2, the maximum residue limit (if available) is used instead. Missing values are imputed.

Table 2.36: Tier definition for EFSA 2012 Pessimistic.

Name	Setting
Default concentration model	NonDetectSpikeLogNormal
Include MRL fallback model	true
Restrict LOR imputation to authorised uses	false
Non-detects replacement	ReplaceByLOR
Factor f (f x LOR)	1
MRL Factor (f x MRL)	1
Sample based	true
Imputation of missing values	true
Correlate imputed values with sample potency	true
Parametric uncertainty	true

EC 2018 Tier 1

Table 2.37: Tier definition for EC 2018 Tier 1.

Name	Setting
Default concentration model	Empirical
Include MRL fallback model	false
Restrict LOR imputation to authorised uses	false
Non-detects replacement	ReplaceByLOR
Factor f (f x LOR)	0.5
Sample based	true
Imputation of missing values	true
Correlate imputed values with sample potency	true
Use occurrence patterns for imputation	true
Parametric uncertainty	false

Input tiers

Table 2.38: Input tiers for EC 2018 Tier 1.

Module	Input tier
Occurrence patterns	*EC 2018 Tier 1*
Concentrations	*EC 2018 Tier 1*

EC 2018 Tier 2

Table 2.39: Tier definition for EC 2018 Tier 2.

Name	Setting
Default concentration model	Empirical
Include MRL fallback model	false
Restrict LOR imputation to authorised uses	false
Non-detects replacement	ReplaceByLOR
Factor f (f x LOR)	0.5
Sample based	true
Imputation of missing values	true
Correlate imputed values with sample potency	false
Use occurrence patterns for imputation	true
Parametric uncertainty	false

Input tiers

Table 2.40: Input tiers for EC 2018 Tier 2.

Module	Input tier
Occurrence patterns	*EC 2018 Tier 2*
Concentrations	*EC 2018 Tier 2*

Concentration models uncertainty

When using empirical distributions, concentration model uncertainty is covered by the the inputs. I.e., concentration models can be recomputed from resampled/bootstrapped concentration data. This happens for both the univariate concentration models, being recomputed from the bootstrapped residue collections for each food and substance, and also for the samples of the sample-based approach that are re-generated from the bootstrapped samples (including the necessarry steps of missing value imputation and imputation of non-detects).

When parametric uncertainty is prefered over emperical bootstrapping, the parameters of the univariate concentration models fitted as a parametric distributions can be resampled parametrically.

Let x denote a random variable from the specified distribution. The log transformed variable $y = ln(x)$ is normally distributed with mean μ_y and variance σ_y. The maximum likelihood estimates are $\hat{\mu}_y$ and $\hat{\sigma}_y$. In each bootstrap sample, values are drawn from a normal distribution where the maximum likelihood estimates are replaced by ($\hat{\mu}^*_y$, $\hat{\sigma}^*_y$).

Calculation of concentration models

Concentration models can be computed from concentration data.

- *Concentration models calculation*

Inputs used: *Concentrations Concentration limits Foods as measured Occurrence patterns Relative potency factors*

Settings used

- *Calculation Settings*

2.3.3 Concentrations

Concentrations data are analytical measurements of chemical substances occurring in food samples. In their simplest form, concentration data can just be used as provided by datasets. Optionally, concentrations data can be manipulated for active substances, extrapolated to other foods, and/or default values can be added for water.

This module has as primary entities: *Foods Substances*

Output of this module is used by: *Occurrence patterns Concentration models Foods as measured*

Concentrations data formats

Three schemes for data are implemented:

1. MCRA scheme: relational tables that can hold all information about Food samples (e.g. sampling date and location), Analytical methods, Analytical method properties for substances (e.g. LOR), Analysis samples (e.g. analysis date) and Concentrations;
2. SSD scheme: data according to the EFSA Standard Sample Description (SSD) guideline; SSD data are converted automatically to the MCRA scheme;
3. Tabulated data scheme: simplified data format, where samples and analytical methods are not explicitly specified.

Concentration data

In this group all tables are collected that store information related to concentration or concentration related entities.

Sample-based concentration data

This sub-group contains five tables to specify food samples, analytical methods, their properties for given substances, analyses and concentrations.

Analytical methods

The analytical methods used for analyzing the samples are recorded in the analytical methods table. Each analytical method should have a unique identification code (idAnalyticalMethod). The description field may be used for a more detailed description of the analytical method. The records of this table should be linked to one or more analytical-method-substance records, which record the substances that are measured by this method (and their limits of reporting).

Table 2.41: Table definition for AnalyticalMethods.

Name	Type	Description	Aliases	Required
idAnalytical-Method	AlphaNumeric(50)	The code for the method of analysis.	idAnalytical-Method, Analytical-MethodId, Analytical-MethodName, Id	Yes
Description	AlphaNumeric(200)	Additional description of method of analysis.	Description	No

Table aliases: AnalyticalMethod, AnalyticalMethods, RawAnalyticalMethods.

Analytical method properties for substances

Table 2.42: Table definition for AnalyticalMethodCompounds.

Name	Type	Description	Aliases	Required
idAnalytical-Method	AlphaNumeric(50)	The code of method of analysis.	idAnalytical-Method, Analytical-MethodName, Analytical-MethodId	Yes
idSubstance	AlphaNumeric(50)	The substance code.	idSubstance, SubstanceId, Substance, idCompound, CompoundId, Compound	Yes
LOR	Numeric	The limit of reporting (LOR). In MCRA, LOR just means the limit below which no quantitative result has been reported. Depending on a laboratory's format of reporting, LOR may be a limit of detection (LOD), a limit of quantification (LOQ) or another limit.	LOR	Yes
Concentration-Unit	*ConcentrationUnits*	The unit of the concentrations/LORs reported by the analytical method for this substance (default mg/kg).	Concentration-Unit, Units, Unit	No

Table aliases: AnalyticalMethodSubstances, AnalyticalMethodSubstance, AnalyticalMethodCompounds, AnalyticalMethodCompound, RawAnalyticalMethodCompounds.

Food samples

Food sample for analysis of concentrations. May be characterised by location and/or date of sampling. A sample can be analysed multiple times, the results per analysis are stored as analysis samples.

Table 2.43: Table definition for FoodSamples.

Name	Type	Description	Aliases	Required
idFoodSample	AlphaNumeric(50)	The identification number of the food sample.	idFoodSample, idSample, SampleId, Id	Yes
idFood	AlphaNumeric(50)	The food code.	idFood, FoodId, Food, FoodCode	Yes
Location	AlphaNumeric(50)	The location or country code, sampling location.	Location, Location-Sampling, Sampling-Location, Country	No
DateSampling	DateTime	The date of sampling.	DateSampling, SamplingDate	No

Table aliases: FoodSamples, FoodSample, Samples, Sample, PrimarySample, PrimarySamples, RawFoodSamples.

Sample properties

Food sample properties, additional columns that can also be specified as additional columns in the food samples table

Table 2.44: Table definition for SampleProperties.

Name	Type	Description	Aliases	Required
Name	AlphaNumeric(50)	The name of the property.	Id	Yes

Table aliases: SampleProperties, SampleProperty, RawSampleProperties.

Sample property values

Food sample property values, additional columns that can also be specified as additional columns in the food samples table

Table 2.45: Table definition for SamplePropertyValues.

Name	Type	Description	Aliases	Required
idSample	AlphaNumeric(50)	The identification number of the food sample.	Id	Yes
PropertyName	AlphaNumeric(50)	The name of the property.	Name	Yes
TextValue	AlphaNumeric(50)	The value of the property as text value.		No
DoubleValue	Numeric	The value of the property as number.		No

Table aliases: SamplePropertyValues, SamplePropertyValue, RawSamplePropertyValues.

Analysis samples

An analysis sample specifies the analysis of a sample by an analytical method. A sample can be analysed multiple times, the results per analysis are stored as analysis samples.

Table 2.46: Table definition for AnalysisSamples.

Name	Type	Description	Aliases	Required
idAnalysis-Sample	AlphaNumeric(50)	The identification number of the analysed sample.	idAnalysis-Sample, AnalysisSample-Id, id	Yes
idFoodSample	AlphaNumeric(50)	The identification number of the food sample.	idFoodSample, idSample, SampleId, Sample	Yes
idAnalytical-Method	AlphaNumeric(50)	The code of method of analysis.	idAnalytical-Method, Analytical-MethodId	Yes
DateAnalysis	DateTime	The date of the analysis.	DateAnalysis, AnalysisDate, Date	No

Table aliases: AnalysisSamples, AnalysisSample, SampleAnalysis, SampleAnalyses, RawAnalysisSamples.

Sample concentrations

The positive concentration values for substances from analysis in the unit specified in table AnalysisSamples. Non-detects (i.e. results 'less than LOR') are not included, their existence can be inferred from the tables AnalysisSamples and AnalyticalMethodSubstances, and the LOR itself from the table AnalyticalMethods.

Table 2.47: Table definition for ConcentrationsPerSample.

Name	Type	Description	Aliases	Required
idAnalysis-Sample	AlphaNumeric(50)	The identification number of the analysed sample.	idAnalysis-Sample, AnalysisSample-Id	Yes
idSubstance	AlphaNumeric(50)	The substance code.	idSubstance, SubstanceId, Substance, idCompound, CompoundId, Compound	Yes
Concentration	Numeric	The measured concentration.	Concentration	Yes

Table aliases: ConcentrationsPerSample, ConcentrationPerSample, RawConcentrationsPerSample.

Tabulated concentration data

Tabulated concentration data provide a simplified concentration data format, where samples and analytical methods are not explicitly specified and analysis results can be tabulated for repeats of the same outcome. This is a convenient data format for single-substance analyses, but it should be noted that it is not possible to use this data in sample-based methods of multiple substances, because it does not record co-occurrence information of substances in samples.

Tabulated concentrations

In the tabulated concentration data table, each sample has a unique identifier and contains a concentration value for a food/substance combination. Non-detects (i.e. results 'less than LOR') are specified as negative values, i.e. 'less than LOR' should be specified as minus the LOR value.

Table 2.48: Table definition for ConcentrationTabulated.

Name	Type	Description	Aliases	Required
GUID	AlphaNumeric(50)	Unique identifier of the analysis sample of this tabulated concentration record.	idAnalysis-Sample, SampleId, SampleCode, Code, Id	No
idSubstance	AlphaNumeric(50)	The code of the substance of this concentration value.	idSubstance, SubstanceId, Substance, idCompound, CompoundId, Compound	Yes
idFood	AlphaNumeric(50)	The food code.	idFood, FoodId, FoodMeasured, Food	Yes
DateSampling	AlphaNumeric(10)	The date of sampling.	DateSampling	No
SamplingType	AlphaNumeric(50)	The type of sampling (monitoring).	SamplingType	No
Location	AlphaNumeric(50)	The location or country of sampling.	Location, Country	No
NumberOf-Samples	Integer	The count of the number of times the specified concentration or limit of reporting (LOR) occurs.	NumberOf-Samples	Yes
Concentration	Numeric	The concentration or LOR. LORs are specified using a minus (-) sign.	Concentration, Value	Yes
Concentration-Unit	*ConcentrationUnits*	The unit of the specified concentrations/LORs (default mg/kg).	Concentration-Unit, Unit	No

Table aliases: ConcentrationTabulated, ConcentrationValues, TabulatedConcentrations, TabulatedConcentration, RawTabulatedConcentrations.

EFSA SSD concentration data

MCRA provides an option to upload concentration data that is formatted according to the EFSA Standard Sample Description (SSD) guideline. SSD formatted concentrations data is converted to the internal, relational data format of MCRA.

SSD concentrations

MCRA uses the concept of samples analysed by analytical methods, where the analytical method is characterised by the substances analysed and the LORs for these substances. However, the SSD data do not provide information on the analytical methods at this level of detail. Therefore, from the provided SSD records, analytical methods are reconstructed and samples are linked to these analytical methods. All SSD records with the same labSampCode and labSubSampCode are considered to be from the same sample. All SSD samples that have records for the same substances, with the same LOQ/LOD values and resUnit are considered to originate from the same reconstructed analytical method. If both LOQ and LOD are provided, LOQ is used as LOR of the reconstructed analytical method. It is highly recommended to supply LOQ/LOD values, even for positive measurement, because this reduces the number of reconstructed analytical methods.

Table 2.49: Table definition for ConcentrationsSSD.

Name	Type	Description	Aliases	Required
labSampCode	AlphaNumeric(30)	Code of the laboratory sample. MCRA will use the combination of labSampCode and labSubSampCode as unique code for a sample.	labSampCode	Yes
labSubSamp-Code	AlphaNumeric(4)	Code of the laboratory sub-sample. MCRA will use the combination of labSampCode and labSubSampCode as unique code for a sample.	labSubSamp-Code	No
sampCountry	AlphaNumeric(2)	Two-letter code to identify the country of sampling.	sampCountry	Yes
prodCode	AlphaNumeric(20)	Code identifying the food as measured. Should be equal to a code idFood in the Foods table.	prodCode	Yes
sampY	Integer(4)	Year of sampling.	sampY	Yes
sampM	Integer(2)	Month of sampling.	sampM	No
sampD	Integer(2)	Day of sampling.	sampD	No
analysisY	Integer(4)	Year of analysis.	analysisY	Yes
analysisM	Integer(2)	Month of analysis.	analysisM	No
analysisD	Integer(2)	Day of analysis.	analysisD	No
paramCode	AlphaNumeric(20)	Code identifying the substance.	paramCode	Yes
resUnit	AlphaNumeric(5)	Unit of residue measurement.	resUnit	Yes
resLOD	Numeric	Residue Limit Of Detection. Required if resType is LOD. MCRA will use resLOD as LOR if resLOQ is not provided.	resLOD	No
resLOQ	Numeric	Residue Limit Of Quantification. Required if resType is LOQ MCRA will use resLOQ as LOR if provided.	resLOQ	No
resVal	Numeric	Required if resType is VAL.	resVal	No
resType	AlphaNumeric(3)	Type of residue data. Should be VAL, LOQ or LOD.	resType	Yes

Table aliases: ConcentrationsSSD, SSDConcentrations.

Concentration distributions

Substance concentrations on foods specified in the form of summary statistics.

Table 2.50: Table definition for ConcentrationDistributions.

Name	Type	Description	Aliases	Required
idFood	AlphaNumeric(50)	Food code, the raw agricultural commodity.	idFood	Yes
idSubstance	AlphaNumeric(50)	The code of the substance.	idSubstance, SubstanceId, SubstanceCode, Substance, idCompound, CompoundId, Compound-Code, Compound	Yes
Mean	Numeric	The mean of (monitoring) samples, on the original scale (in mg/kg).	Mean	Yes
CV	Numeric	Coefficient of variation, for samples of the size of the TDS pooled amount.	CV	No
Percentile	Numeric	The percentile at the point specified by the percentage.	Percentile	No
Percentage	Numeric	The percentage that belongs to the given the percentile, e.g., 95 (in mg/kg).	Percentage	No
Limit	Numeric	The specified norm value or limit value (in mg/kg).	Limit	No
Concentration-Unit	*ConcentrationUnits*	The unit of the limit value (default mg/kg).	Concentration-Unit, Unit	No

Table aliases: ConcentrationDistributions, ConcentrationDistribution, RawConcentrationDistributions.

Concentrations calculation

Substance conversions data may be used to *convert* concentration data at the level of measured substances to concentration data at the level of potentially active substances. These rules may be applicable, e.g., when a measured substance represents multiple substances and measurements of this measured substance should be converted into measurement values for these substances. This conversion may depend on *substance authorisations* which provides information on the likelihood of certain translations to occur and one may need *points of departure* or *relative potency factors* when the substance conversion should select the most toxic candidate in case a measured substance translates to multiple active substances.

In some cases, it may be that for a certain food/substance combination, there are few measurements in the concentration data. In this case, *extrapolation of concentration data* may be desired. If this is the case, *food extrapolation rules* may be provided to specify per food, the alternative foods from which extrapolation is allowed. The extrapolation of concentrations will then be performed within this module and the results are included in the resulting active substance concentrations data. *Substance authorisations* and/or *concentration limits* may be used to further restrict the to-food/from-food combinations per substance for which extrapolation is possible.

Concentration data for water are often not available in the *concentrations* data, but it may be desirable to include water concentrations in the assessments. For this, *imputation* of low-tier, deterministic estimates of water concentrations of the most toxic substances may be used to include (typically conservative) estimates in the calculations.

Substance conversion

When concentration data at the level of measured substances have to be converted to concentration data at the level of *active substances*, then *substance conversions* should be specified to provide the rules for deriving active or inactive substance concentrations from measured substance concentrations. This section will first describe the basic substance conversion, and following this the way in which *substance authorisations* can be used to refine these calculations.

For each measured substance in the concentration data, there may be zero or more conversion rules (records in the data source), each linking to an active or inactive substance. Note that the substance conversion procedure has been implemented for two cases:

1. Measured substances link to one or more exclusive substances which are possible translations (i.e. the measured concentration is assumed to be the sum of concentrations for all linked substances, but it is assumed that only one substance is present in the sample, therefore the measured substance is considered to be one of the linked substances);
2. Measured substances link to one or more exclusive substances plus one (non-exclusive) substance that is a metabolite of the others. The metabolite can occur together with any of the exclusive substances.

It is assumed that either all conversion rules linked to a measured substance are marked as exclusive (case 1), or precisely one rule is marked as exclusive and the other rules are marked as not exclusive (case 2). If this is not the case for any set of rules linked to a measured substance, then this is regarded as erroneous data.

For each measured substance concentration measurement on a food sample (positive or non-detect), the active substance concentration allocation is done using the following procedure:

1. If there is no conversion rule available for the measured substance linked to any active substance of interest, then no conversion is required. The measured substance is ignored in the active substance concentrations set, unless it is an active substance itself.
2. If there are conversion rules available for the measured substance, then concentrations (positive or non-detect) are converted from the measured substance to one or more linked substances as specified by the conversion rules. Two substance allocation methods are available, one of which should be selected:

Tier 1) Most potent: For each measured substance, the linked substances are restricted to the active substances of interest. The concentration of the measured substance is assigned to the most potent active substance in this set. Potency is specified by the *relative potency factors*. All other candidate active substances are assigned a zero concentration. I.e., the measured substance concentration is allocated for 100% to the most potent substance specified by the conversion rules and for this allocation, the concentration or LOR is multiplied by the molecular weight correction factor.

Tier 2) Random: One of the conversion rules is drawn randomly (with equal probability), including the rules of both active and other substances. This drawn rule is used as follows to generate active substance concentrations:

- **If the drawn conversion rule is marked as exclusive**, the concentration or LOR is allocated to the linked substance.
- **If the drawn conversion rule is marked as not exclusive**, a proportion *p*, specified by the drawn conversion rule, of the concentration or LOR is allocated to the linked substance. The remaining proportion *(1-p)* is allocated to one other substance, which is the substance that is linked to the measured substance in a conversion rule marked as exclusive (in this case it is assumed that precisely one record per measured substance is marked as exclusive).

In Tier 2 all assigned concentrations are multiplied by the molecular weight correction factor. All unselected candidate substances are assigned a zero concentration. After allocation, the resulting set of substance concentrations is restricted to contain only concentrations for the active substances of interest.

(Included for research purposes) Nominal: The substances specified through the conversion rules are allocated a nominal value based on all possible conversion rules. This may be regarded as the nominal or average allocation value of the random sampling method.

- **All conversion rules are marked as exclusive:** The measured substance concentration is divided over all *n* active substances specified with equal proportions *1/n*, accounting for the molecular weight correction factor for all substances.

- **Precisely one conversion rule is marked exclusive and n conversion rules are marked as not exclusive:** The measured substance concentration is divided over all active substances specified, with a proportion *1/2 + 1/n* for the substance belonging to the exclusive conversion rule, and equal proportions *1/n* for the other substances, accounting for the molecular weight correction factor for all substances.

Use of substance authorisations in substance conversion

When *substance authorisations* are available, then these can be used to exclude conversions of measured substances to unauthorised substances on a given food. The information is used as follows in the substance conversion procedures:

> **Option 1) Most potent:** The set of candidate active substances from which the most potent active substance is to be drawn is reduced to only the substances with authorised uses. However, if none of the candidate active substances is authorised, then the most potent of the unauthorised substances is selected for active substance allocation.
>
> **Option 2) Random:** The set of conversion rules from which to draw is reduced to the rules linking to authorised substances or the non-exclusive substance (thus allowing the selection of a possibly unauthorised metabolite of an authorised substance). If none of the conversion rules links to an authorised substance, then one rule is drawn from the full set of all (unauthorised) conversion rules.
>
> **Nominal:** The set of conversion rules is reduced in the same way as in Tier 2. Nominal calculation is performed on the resulting set of conversion rules.

Food extrapolation

If the *food extrapolation* setting has been checked, extrapolation of concentrations is performed for all food/*active substance* combinations for which:

1. the number of measurements in the analytical scope is smaller than a given threshold for extrapolation (default 10), and
2. there is an *extrapolation rule* allowing extrapolation of concentrations from one or more other foods (the from-food(s)) to the given food (the to-food), and
3. (optional criterion:) the substance is associated with *authorised use* for both foods, and
4. (optional criterion:) *concentration limits (e.g. MRLs)* on the from-food and to-food exist and are equal. Note: if the **active substance** is not a **measured substance**, then the MRL check has to be made per measurement at the level of the measured substance which provided the concentrations assigned to the active substance.

Food extrapolation is performed by one of the following procedures: 1) Substance-specific imputation of missing values by extrapolated measurements, or 2) Extrapolation of complete samples for multiple substances.

1. Substance-specific imputation of missing values by extrapolated measurements

The missing values in the active substance concentrations of the tofood are imputed in a random order by active substance concentrations (positive, nondetect or zero) from a randomised list obtained from the fromfood(s). By matching the randomised lists, each fromfood measurement is assigned at most once, so after extrapolation there may still be missing values left, or not all measurements of the from-food(s) may have been used for extrapolation.

Note: In this method, it is assumed that the to-food has a sufficient number of samples. No extrapolation is applied for foods with no samples at all, and data gaps will also remain for foods with fewer than *n* samples, because no new samples are added.

Note: the resulting occurrence patterns will be random with respect to the extrapolated substances, i.e., observed occurrence patterns for the from-food are not extrapolated to the to-food.

2. Extrapolation of complete samples for multiple substances

(not yet implemented)

All samples of the from-food(s), i.e., complete samples with data for all active substances, are copied as samples for the to-food and added to the existing to-food samples. For example, extrapolate all apple sample records to the available pear sample records. However, measurements for substances that do not fulfill the (optional) criteria 3 and 4 above are non-valid extrapolations and are replaced by missing values. The status of the extrapolated samples is stored to distinguish between extrapolated and non-extrapolated sample records. Note that this method maintains correlations in the occurrence patterns and postpones imputation of MVs until the concentration models step.

Water imputation

If water has been selected as an additional source of exposure, but concentration data is missing, then, fixed concentration values can be assigned to water for the five most toxic *active substances*, with the toxicity ranking being based on the *relative potency factors*. For all other substances, zero concentrations are imputed. The default imputation value is 0.05 µg/L, but this value can be chosen as a setting. If specified, *substance authorisations* may be used to restrict to the set of active substances for which water concentrations are imputed to only those for which concentrations may be expected from authorised use.

Concentrations settings

Selection settings

Table 2.51: Selection settings for module Concentrations.

Name	Description
Concentrations tier	Specifies the concentration data should be treated according to a pre-defined tier or custom.
Use substance conversions	If true, concentrations are modelled in terms of active substances (using substance conversion).
Substance conversion method	Allocation method for assigning active substance concentrations from measured substance concentrations based on substance translations.
Retain all allocated substances after active substance allocation	If true, all allocated substances kept after substance conversion, if false, the concentration data is restricted to the active substances of the assessment group.
Account for substance authorisations in substance conversions	Account for substance authorisations when allocating measured substances to active substance using substrance conversions.
Use extrapolation rules	Use extrapolation rules.
Threshold for extrapolation	Threshold for extrapolation.
Restrict extrapolations to equal MRLs	Restrict extrapolations to equal MRLs.
Restrict extrapolations to authorised uses	Only extrapolate if substance use is authorised.
Impute water concentrations	Impute constant concentration values for the five most toxic substances on the selected (water) commodity.
Water commodity	The commodity for which constant concentration values should be added.
Water concentration value (µg/kg)	Constant concentration value that should be used for water (in µg/kg).
Restrict water imputation to authorised uses	Restrict water imputation to authorised uses.

Uncertainty settings

Table 2.52: Uncertainty settings for module Concentrations.

Name	Description
Resample concentrations	Specifies whether concentrations are resampled by empirical bootstrap or using a parametric uncertainty model.

Concentrations tiers

In addition to the possibility for users to work with their own choices for all settings, MCRA implements Tier 1 and 2 from the European Commission working document SANTE-2015-10216 rev. 7 (2018) on risk management aspects related to the assessment of cumulative exposure.

Overview

Table 2.53: Tier overview for module Concentrations.

Name	EC 2018 Tier 1	EC 2018 Tier 2
Substance conversion method	UseMost-Toxic	DrawRan-dom
Retain all allocated substances after active substance allocation	true	true
Account for substance authorisations in substance conversions	false	true
Use extrapolation rules	true	true
Threshold for extrapolation	10	10
Restrict extrapolations to equal MRLs	true	true
Restrict extrapolations to authorised uses	true	true
Impute water concentrations	true	true
Water concentration value (µg/kg)	0.1	0.05
Restrict water imputation to authorised uses	false	false

EC 2018 Tier 1

Table 2.54: Tier definition for EC 2018 Tier 1.

Name	Setting
Substance conversion method	UseMostToxic
Retain all allocated substances after active substance allocation	true
Account for substance authorisations in substance conversions	false
Use extrapolation rules	true
Threshold for extrapolation	10
Restrict extrapolations to equal MRLs	true
Restrict extrapolations to authorised uses	true
Impute water concentrations	true
Water concentration value (µg/kg)	0.1
Restrict water imputation to authorised uses	false

EC 2018 Tier 2

Table 2.55: Tier definition for EC 2018 Tier 2.

Name	Setting
Substance conversion method	DrawRandom
Retain all allocated substances after active substance allocation	true
Account for substance authorisations in substance conversions	true
Use extrapolation rules	true
Threshold for extrapolation	10
Restrict extrapolations to equal MRLs	true
Restrict extrapolations to authorised uses	true
Impute water concentrations	true
Water concentration value (µg/kg)	0.05
Restrict water imputation to authorised uses	false

Concentrations uncertainty

Uncertainty due to a limited number of samples can be accounted for by resampling/bootstrapping. Resampling is done on a sample-based basis preserving co-occurrence of substance residue values on the same sample for multiple-substance analyses.

Concentrations as data

Concentration data can be entered using the internal, relational data format or using the EFSA SSD format. Depending on the settings, the entered concentration data can be pre-processed for conversion to active substances, extrapolation to other foods, and/or default values can be added for water.

- *Concentrations data formats*
- *Concentrations calculation*

Inputs used: *Focal food concentrations Food extrapolations Substance conversions Relative potency factors Substance authorisations Active substances Concentration limits*

2.3.4 Focal food concentrations

In some cases the attention in an assessment is on a specific food (focal food), against the background of other foods. Focal food concentrations are separate concentration data for one or more focal food commodities, that will take the place of any other concentration data for the focal food in the ordinary concentration data.

Output of this module is used by: *Concentrations*

Focal food concentrations data formats

See *concentration data formats*.

Focal food concentrations as data

Focal food concentrations are concentration data and specified in the exact same manner. The difference is that this data will be used to replace part of the concentration data in order to combine specific concentration data with a background of ordinary concentration data.

- *Focal food concentrations data formats*

2.3.5 Food extrapolations

Food extrapolations data specify which foods (data rich foods) can be used to impute concentration data for other foods with insufficient data (data poor foods).

This module has as primary entities: *Foods*

Output of this module is used by: *Concentrations Food conversions*

Food extrapolations data formats

Food extrapolations

Food extrapolations (or read-across food translations) can be used to specify whether data (e.g, occurrence data) on a food for which this is missing (a data poor food) may be extrapolated from another food for which data is available (read-across food).

Food extrapolations

Food extrapolations are simply specified as combinations of two food codes. One code for the food for the data poor food, and one for the data rich food (or read-across food).

Table 2.56: Table definition for FoodExtrapolations.

Name	Type	Description	Aliases	Required
DataPoorFood	AlphaNumeric(50)	The code of the data poor food. I.e., the food for which missing data is allowed to be extrapolated.	IdFoodDataPoor, FoodDataPoor, idFromFood, FromFoodId, FromFood, FoodFrom, Food, IdFood	Yes
CodeDataRichFood	AlphaNumeric(50)	The code of the read-across food (or data rich food). I.e., the food from which data is used for extrapolation.	IdFoodDataRich, FoodDataRich, IdFoodReadAcross, FoodReadAcross, IdReadAcrossFood, ReadAcrossFood, idToFood, ToFoodId, ToFood, FoodTo	Yes

Table aliases: ReadAcrossFoodTranslations, ReadAcrossFoodTranslation, ReadAcrossTranslations, ReadAcrossTranslation, FoodExtrapolations, FoodExtrapolation, RawReadAcrossFoodTranslations.

Food extrapolations as data

Food extrapolations are specified as data in the form of simple tuples of data rich food and data poor food for which extrapolation is allowed/reasonable.

- *Food extrapolations data formats*

2.3.6 Foods as measured

Foods as measured are foods within the foods scope for which concentration data of substances are available (or expected).

This module has as primary entities: *Foods Substances*

Output of this module is used by: *Concentration models Food conversions*

Foods as measured calculation

Foods as measured, or modelled foods, are the foods within the foods scope for which concentration data of substances are available (or expected). Foods as measured are derived primarily from *concentration data*. That is, all foods for which food samples are available in the concentration data are considered to be foods-as-measured. In addition, this set may be extended when *concentration limits* such as MRLs are available (see *calculation settings*) and/or when *food extrapolation rules* are used. Foods for which such data is available are considered to be foods as measured. The set of foods as measured can also be restricted by omitting foods with only non-detect measurements (see *calculation settings*).

Foods as measured settings

Calculation settings

Table 2.57: Calculation settings for module Foods as measured.

Name	Description
Include foods with only non-detect measurements	Specifies whether foods with only non-detect measurements are part of the exposure assessment (default yes).
Include foods without occurence data but with specified maximum residue limits	Include foods without concentration data but for which for which concentration limits such as MRLs are defined (default: no).
Food-as-measured subset: restrict to specific foods-as-measured	If checked, then the assessment is restricted to the specified foods-as-measured.
Selected foods-as-measured	Set of measured foods that are of particular interest.

Calculation of foods as measured

Foods-as-measured are computed from occurrence of concentration data and (if specified) availability of maximum residue limits.

- *Foods as measured calculation*

Inputs used: *Concentrations Concentration limits*

Settings used

- *Calculation Settings*

2.3.7 Occurrence patterns

Occurrence patterns (OPs) are the combinations (or mixtures) of substances that occur together on foods and the frequencies of these mixtures occurring per food, expressed in percentages. In the context of pesticides, occurrence patterns can be associated with agricultural use percentages. Occurrence patterns are relevant to account for co-occurrence of active substances in exposed individuals. Occurrence patterns may be specified as data or modelled based on observed patterns of positive concentrations.

This module has as primary entities: *Foods Substances*

Output of this module is used by: *Concentration models*

Occurrence patterns data formats

Agricultural uses

Agricultural use percentages for plant protection products (PPPs) may be of use for concentration modelling, as they provide information about what substance mixtures can be expeceted to be present simultaneously on food samples. Especially for non-detect concentration measurements, this information may aid to determine whether the non-detect measurement originated from a true zero or may be a positive measurement below the limit of detection. Agricultural use percentages can be specified using the agricultural uses and agricultural use substances table. This data format expects agricultural use percentages to be specified for mixtures of substances. Each mixture has an id (idAgriculturalUse) and a list of substances that are part of this mixture (agricultural use substances). These agricultural uses are assumed to be exclusive (i.e., only one mixture or PPP is used per sample). Hence, the sum of the agricultural uses for one food should not exceed 100%.

Agricultural uses

The AgriculturalUses contains the definitions of the agricultural use mixtures, or PPPs and the specification of the percentage of the products treated with this mixture. Optionally also the time period of the use percentage may be specified.

Table 2.58: Table definition for AgriculturalUses.

Name	Type	Description	Aliases	Required
idAgricultural-Use	AlphaNumeric(50)	The unique identification code of the agricultural use group / plant protection product (PPP).	idAgricultural-Use, AgriculturalUse-Id, Id	Yes
idFood	AlphaNumeric(50)	The food code.	idFood, FoodId, Food	Yes
Location	AlphaNumeric(50)	The location or country code, agricultural use location.	Country, Location	No
StartDate	DateTime		StartDate	No
EndDate	DateTime		EndDate	No
Percentage-CropTreated	Numeric	The percentage agricultural use (%).	PercentageCrop-Treated, Percentage, PercCrop-Treated, PercentageUse	Yes

Table aliases: AgriculturalUses, AgriculturalUse, RawAgriculturalUses.

Agricultural use substances

The agricultural use substances table records the substances that are part of the agricultural use mixtures (PPPs).

Table 2.59: Table definition for AgriculturalUsesHasCompounds.

Name	Type	Description	Aliases	Required
idAgricultural-Use	AlphaNumeric(50)	The agricultural use code, normally a code for a combination of authorised substances.	idAgricultural-Use, AgriculturalUse-Id	Yes
idSubstance	AlphaNumeric(50)	The code of the substance.	idSubstance, SubstanceId, SubstanceCode, Substance, idCompound, CompoundId, Compound-Code, Compound	Yes

Table aliases: AgriculturalUseHasCompounds, AgriculturalUsesHasCompounds, AgriculturalUseCompounds, AgriculturalUseHasSubstances, AgriculturalUsesHasSubstances, AgriculturalUseSubstances, AgriculturalUseGroups, AgriculturalUseGroup, AgriculturalUses_has_Compounds, RawAgriculturalUses_has_Compounds.

Occurrence patterns calculation

Assumptions can be made for each food on the basis of findings in concentration data.

Tier 1: 0% occurrence is assumed for all substances with no positive concentrations at all; 100% occurrence is assumed for all substances with at least one positive concentration;

Tier 2: 0% occurrence is assumed for all substances with no positive concentrations at all; for substance-food combinations with at least one positive (finding), use findings patterns to implement a specific interpretation of Option 5 in the SANTE document, as described below.

Therefore in both tiers, substance-food combinations without any positive finding are handled in the optimistic way by assuming absolute zeroes for any non-detect observation.

If Tier 2 is selected, then for each of the modelled foods a tabulation is made of the observed frequencies of positives for all substance combinations (including the empty set), based on the *active substance concentrations*. For an OP consisting of just one substance, the basic frequency is the number of samples with a positive concentration divided by the number of samples where the substance has been measured (i.e., is not a MV). For an OP consisting of multiple substances, the basic frequency is the number of samples with all concentrations positive for the members divided by the number of samples where all members of the set have been measured.

After calculation of the basic frequencies for all occurrence patterns, these frequencies are rescaled such that the overall sum of frequencies is 100%. When *substance authorisations* are available, then patterns involving unauthorised substances are not rescaled and only those patterns for which all substances are authorised are rescaled such that the sum of all frequencies is 100%.

Note: the Tier 2 procedure is not what is literally written in the SANTE document, but is an interpretation agreed upon by EFSA and RIVM. An alternative model, not yet implemented, but perhaps more in line with the text of the SANTE document, would be to double the basic frequencies to modelled occurrence pattern frequencies. Only if the sum of all frequencies becomes larger than 100%, the set of frequencies would be normalised to 100% sum.

Occurrence patterns settings

Selection settings

Table 2.60: Selection settings for module Occurrence patterns.

Name	Description
Associate the unspecified percentage with no-occurrence for foods with at least one specified occurrence pattern	If checked, for foods with at least one specified occurrence pattern, unspecified occurrence patterns for the same food are assumed to be associated with no use. If unchecked, all substances are considered to be authorised (potentially present in samples). Note that this setting cannot be used for foods that have no specified AUs. These foods have 100% potential presence of all substances. To declare all AUs on such a food un-authorised, include an empty AU with percentage 100% in the AU data table (i.e., use an AU for this food, without specifying substances in the AU Substances table)
Apply occurrence pattern percentages	If checked, use the percentages of potential presence as specified by the occurrence patterns. If unchecked, 100% potential presence in samples is assumed for all substances identified by the occurrence patterns.
Scale up use percentages to 100%	Scale up use percentages to 100%.
Restrict use percentage up-scaling to authorised uses	Restrict use percentage up-scaling to authorised uses.

Uncertainty settings

Table 2.61: Uncertainty settings for module Occurrence patterns.

Name	Description
Recompute occurrence patterns	Specifies whether occurrence patterns should be recomputed in the uncertainty runs.

Occurrence patterns tiers

Overview

Table 2.62: Tier overview for module Occurrence patterns.

Name	EC 2018 Tier 1	EC 2018 Tier 2
Apply occurrence pattern percentages	false	true
Scale up use percentages to 100%		true
Restrict use percentage up-scaling to authorised uses		true

EC 2018 Tier 1

Table 2.63: Tier definition for EC 2018 Tier 1.

Name	Setting
Apply occurrence pattern percentages	false

Input tiers

Table 2.64: Input tiers for EC 2018 Tier 1.

Module	Input tier
Concentrations	*EC 2018 Tier 1*

EC 2018 Tier 2

Table 2.65: Tier definition for EC 2018 Tier 2.

Name	Setting
Apply occurrence pattern percentages	true
Scale up use percentages to 100%	true
Restrict use percentage up-scaling to authorised uses	true

Input tiers

Table 2.66: Input tiers for EC 2018 Tier 2.

Module	Input tier
Concentrations	*EC 2018 Tier 2*

Occurrence patterns as data

Occurrence patterns can be provided as data by specification of the occurrence mixtures and their associated occurrence/agricultural use percentages.

- *Occurrence patterns data formats*

Inputs used: *Concentrations Substance authorisations Active substances*

Calculation of occurrence patterns

Occurrence patterns can be computed from the observed patterns of positive concentrations in the concentration data.

- *Occurrence patterns calculation*

2.3.8 Processing factors

Processing factors are multiplication factors to derive the concentration in a processed food from the concentration in an unprocessed food and can be specified for identified processing types (e.g., cooking, washing, drying). Processing factors are primarily used in dietary exposure assessments to correct for the effect of processing on substance concentrations in dietary exposure calculations.

This module has as primary entities: *Foods Substances*

Output of this module is used by: *Food conversions Dietary exposures*

Processing factors data formats

Processing factors connect two food codes, one for the processed food and one for the unprocessed food. There are two schems to make this connection:

1) specify the two food codes and the processing type, or

2) use food facets, i.e. specify only the code of the unprocessed food and the processing type (facet), the code of the processed food is defined by the other two.

Processing factors

Processing factors are defined for triplets of processing type, food, and substance. The processing types are defined in the processing types table and the processing factors are defined in the processing factors table.

Processing factors

Processing factor records should be linked to processing types using the processing type code (idProcessingType) and for the foods and substances. The codes of the processing factor records should match the codes of the foods, substances, and processing type definitions.

Table 2.67: Table definition for ProcessingFactors.

Name	Type	Description	Aliases	Required
idProcessing-Type	AlphaNumeric(50)	The code of the processing type.	idProcessing-Type, ProcessingType-Id, ProcessingType, ProcType	Yes
idSubstance	AlphaNumeric(50)	The code of the substance.	idSubstance, SubstanceId, SubstanceCode, Substance, idCompound, CompoundId, Compound-Code, Compound	No
idFood-Processed	AlphaNumeric(50)	The code of the processed food (may contain a wildcard e.g. 'FP*'. The wildcard matches all characters preceded by the startcode 'FP').	idFood-Processed, FoodProcessed-Id, FoodProcessed	Yes
idFood-Unprocessed	AlphaNumeric(50)	The code of the unprocessed food (may contain a wildcard e.g. 'FP*').	idFood-Unprocessed, Food-UnprocessedId, idFood, FoodId, Food-Unprocessed	Yes
Nominal	Numeric	The nominal value (best estimate of 50th percentile) of processing factor (defines median processing factor).	Nominal, ProcNom	No
Upper	Numeric	The upper value (estimate of 95th percentile or "worst case" estimate) of processing factor due to variability.	Upper, ProcUpp	No
Nominal-Uncertainty-Upper	Numeric	The upper 95th percentile of nominal value (Nominal) due to uncertainty. A standard deviation for uncertainty of the nominal value (Nominal) is derived using the nominal value (Nominal) and upper 95th percentile (NominalUncertaintyUpper).	Nominal-Uncertainty-Upper, ProcNomUnc-Upp	No
Upper-Uncertainty-Upper	Numeric	The upper 95th percentile of upper value (Upper) due to uncertainty. From the nominal value (Nominal), upper value (Upper) and the specified uncertainties of these values (NominalUncertaintyUpper and UpperUncertaintyUpper, respectively) the degrees of freedom of a chi-square distribution describing the uncertainty of the standard deviation for variability is derived.	Upper-Uncertainty-Upper, ProcUppUnc-Upp	No

Table aliases: ProcessingFactors, ProcessingFactor, Processing, RawProcessingFactors.

Food facet processing factors

This table can be used to define processing factors for (FoodEx2) food/food-facet combinations.

Table 2.68: Table definition for FoodFacetProcessingFactors.

Name	Type	Description	Aliases	Required
idProcessing-Type	AlphaNumeric(50)	The code of the processing type.	idProcessing-Type, ProcessingType-Id, ProcessingType, ProcType, facet, idFacet, codeFacet	Yes
idFood	AlphaNumeric(50)	The food to which this facet should be linked.	idFood, FoodId, Food	Yes
idSubstance	AlphaNumeric(50)	The code of the substance.	idSubstance, SubstanceId, SubstanceCode, Substance, idCompound, CompoundId, Compound-Code, Compound	No
Nominal	Numeric	The nominal value (best estimate of 50th percentile) of processing factor (defines median processing factor).	Nominal, ProcNom	No
Upper	Numeric	The upper value (estimate of 95th percentile or "worst case" estimate) of processing factor due to variability.	Upper, ProcUpp	No
Nominal-Uncertainty-Upper	Numeric	The upper 95th percentile of nominal value (Nominal) due to uncertainty. A standard deviation for uncertainty of the nominal value (Nominal) is derived using the nominal value (Nominal) and upper 95th percentile (NominalUncertaintyUpper).	Nominal-Uncertainty-Upper, ProcNomUnc-Upp	No
Upper-Uncertainty-Upper	Numeric	The upper 95th percentile of upper value (Upper) due to uncertainty. From the nominal value (Nominal), upper value (Upper) and the specified uncertainties of these values (NominalUncertaintyUpper and UpperUncertaintyUpper, respectively) the degrees of freedom of a chi-square distribution describing the uncertainty of the standard deviation for variability is derived.	Upper-Uncertainty-Upper, ProcUppUnc-Upp	No

Table aliases: FoodFacetProcessingFactors, FoodFacetProcessingFactor, FacetProcessingFactors, FacetProcessingFactor, FacetProcessing, RawFoodFacetProcessingFactors.

Processing factors settings

Uncertainty settings

Table 2.69: Uncertainty settings for module Processing factors.

Name	Description
Resample processing factors	Specifies whether processing factors are resampled from a parametric uncertainty distribution.

Processing factors uncertainty

Processing effects are modelled either by a fixed processing factor, or by a lognormal or logistic-normal distribution (depending on the distribution type of the *processing type*). In case of a fixed factor, the uncertainty distribution is lognormal or logistic-normal with the same mean μ as the fixed value, and with a standard deviation σ_{unc} which is calculated from the specified central value μ (or nominal) and an estimate of the p95 of the uncertainty distribution (set *NominalUncertaintyUpper* in the *table for ProcessingFactors*).

The calculation is:

$$\sigma_{unc} = \frac{f(NominalUncertaintyUpper) - f(\mu)}{1.645}$$

with $f() = logit$ for the logistic-normal distribution (distribution type 1) and $f() = ln$ for the lognormal distribution (distribution type 2). Values lower than 0.01 or higher than 0.99 (distribution type 1 only) are replaced by default values (0.01 and 0.99); this is useful computationally to avoid problems. In each iteration of the uncertainty analysis a new value is drawn from this distribution to be used as a fixed factor in the Monte Carlo calculation. In case of distribution based processing factors (describing the variability of processing factors) two uncertainties can be specified. For σ_{unc}, specification and calculation is as before (set *NominalUncertaintyUpper* in the *table for ProcessingFactors*).

The uncertainty about the variability standard deviation

$$\sigma_{var} = \frac{f(Upper) - f(\mu)}{1.645}$$

can be specified by the *UpperUncertaintyUpper* value. This value is specified as the p95 upper limit on *Upper*. The specified value is used to derive in a iterative search the number of degrees of freedom *df* (van der Voet et al. 2009) [46]. In the uncertainty analysis, a modified chi-square distribution with *df* degrees of freedom is used to generate new values of σ_{var}. A very high value of *df* means litte uncertainty and σ_{var} will be almost equal in all iterations of the uncertainty analysis. A *df* close to 0 means a large uncertainty and very different values of σ_{var} will be obtained in the iterations of the uncertainty analysis. The p95 upper limit on Upper is set through parameter *UpperUncertaintyUpper*.

Processing factors as data

- *Processing factors data formats*

2.3.9 Substance authorisations

Substance authorisations specify which food/substance combinations are authorised for (agricultural) use. If substance authorisations are used, then only the food/substance combinations that are specified in the data are assumed to be authorised and all other combinations are assumed to be not authorised. This information may, for instance, be used to determine whether concentration measurements below the LOR could be assumed true zeros. I.e., if a food/substance combinations is assumed to be unauthorised, then the LOR may be assumed to be a zero.

This module has as primary entities: *Foods Substances*

Output of this module is used by: *Concentrations Occurrence patterns*

Substance authorisations data formats

Substance authorisations

Authorised uses data provides information about whether substance use is allowed for specified foods. For cumulative exposure assessments, this information can be used for imputation of non-detects/missing values.

Authorised uses

The authorised uses table

Table 2.70: Table definition for AuthorisedUses.

Name	Type	Description	Aliases	Required
idFood	AlphaNumeric(50)	The food code.	idFood, FoodId, Food	Yes
idSubstance	AlphaNumeric(50)	The food code.	idSubstance, Substance, SubstanceId	Yes
Reference	AlphaNumeric(200)	External reference(s) to sources containing more information about the effect (key event) relationships.	Reference, References	No

Table aliases: AuthorisedUses, AuthorisedUse, RawAuthorisedUses.

Substance authorisations as data

Substance authorisations are specified as data in the form of a list of authorised food/substance combinations, with combinations not on the list assicated with no authorised use.

- *Substance authorisations data formats*

2.3.10 Substance conversions

Substance conversions specify how measured substances are converted to active substances, which are the substances assumed to cause health effects. In the pesticide legislation such measured substances and the substance conversion rules are known as residue definitions.

This module has as primary entities: *Substances*

Output of this module is used by: *Concentrations*

Substance conversions data formats

Two types of substance conversions are implemented, with two subtypes for the first type:

> 1a) The measured substance is one or more of a set of possible substances (e.g. isomers or metabolites), and the toxicity of all substances in this set is assumed to be the same and is expressed in one active substance. Example: The measured substance Parathion-methyl(RD) is either Parathion-methyl or paraoxon-methyl, but both are expressed as the active substance Parathion-methyl.
>
> 1b) The measured substance is one or more of a set of possible substances (e.g. isomers or metabolites), and the toxicity of all substances in this set is assumed to relate with equal probability to one of a subset of active substances. Example: The measured substance Dithiocarbamates includes the active substances maneb, mancozeb, metiram, propineb, thiram and ziram, one of which will be assumed to be the active substance present with equal probability.

2) If n active substances all metabolise to the same active substance (the metabolite), it is assumed that all n+1 substances have equal probability of being the source of the measured concentration. The measured substance then is either one active substance (the metabolite) or a mixture of two active substances, one being the metabolite and the other one of the possible parent substances. Example: The measured substance Carbofuran(RD) is either the active substance Carbufuran or a mixture of Carbofuran and one of the possible active parent substances Benfuracarb or Carbosulfan.

Substance conversions

Substance conversions are described by a single substance conversions table.

Substance conversions

The records of the substance translations definitions table specify which active substances (idActiveSubstance) link to a measured substance (idMeasuredSubstance). Each record contains a conversion factor that specifies how a concentration of the measured substance translates to a concentration of the active substance, a flag that states whether the residue definition should be assumed to translate exclusively to one of its active substances, and a proportion. The proportion specifies the proportion of the samples that should translate to this specific active substance in case the translation is exclusive, otherwise it specifies the proportion of the concentration that is assumed to be attributed to the active substance.

Table 2.71: Table definition for ResidueDefinitions.

Name	Type	Description	Aliases	Required
idMeasured-Substance	AlphaNumeric(50)	Substance code of the measured substance.	idResidue-Definition, Residue-Definition, Measured-Substance	Yes
idActive-Substance	AlphaNumeric(50)	Substance code of the active substance.	idActive-Substance, idSubstance, Active-Substance, Substance	Yes
Conversion-Factor	Numeric	Specifies the (molecular weight) conversion factor to translate the concentration of the residue definition to a concentration of the active substance	Conversion-Factor	Yes
IsExclusive	Boolean	Specifies whether a measurement of the residue substance should be translated exclusively to this active substance, or if the residue definition represents/breaks down to a mixture of active substances.	IsExclusive	Yes
Proportion	Numeric	In case the definition is exclusive: the proportion of measurements of the residue definition that can be assumed to translate exclusively to a concentration of the active substance. In case the residue definition is not exclusive, the proportion of the concentration that is assumed to be attributed to the active substance.	Proportion	No

Table aliases: ResidueDefinitions, ResidueDefinition, RawResidueDefinitions.

Substance conversions as data

- *Substance conversions data formats*

Inputs used: *Active substances*

2.3.11 Total diet study sample compositions

Total diet study sample compositions specify the composition of mixed food samples, such as used in a total diet study (TDS), in terms of their constituting foods.

This module has as primary entities: *Foods*

Output of this module is used by: *Food conversions*

Total diet study sample compositions data formats

Total diet study data

Total diet studies (TDS) complement traditional monitoring and surveillance by providing a scientific basis for population dietary exposure to nutrients and non-nutrients including contaminants, and potential impact on public health.

TDS food sample compositions

The TDS food sample compositions table contains the descriptions of the TDS samples.

Table 2.72: Table definition for TDSFoodSampleCompositions.

Name	Type	Description	Aliases	Required
idTDSFood	AlphaNumeric(50)	The code of the TDS food.	idTDSFood	Yes
idFood	AlphaNumeric(50)	Sub-food of the TDS food.	idFood	Yes
PooledAmount	Numeric	Total weight (in g) or volume (in ml) of the food.	PooledAmount, Weight	Yes
Description	AlphaNumeric(200)	Additional description of the TDS sample (e.g. number of subsamples).	Description	No
Regionality	AlphaNumeric	Regionality information.	Regionality	No
Seasonality	AlphaNumeric	Seasonality information.	Seasonality	No

Table aliases: TDSFoodSampleCompositions, TDSFoodSampleComposition, CompositionTDSFoodSamples, CompositionTDSFoodSample, RawTDSFoodSampleCompositions.

Total diet study sample compositions as data

- *Total diet study sample compositions data formats*

2.3.12 Unit variability factors

Unit variability factors specify the variation in concentrations between single units of the same food, which have been put together in a mixture sample on which the concentration measurements have been made. Unit variability factors are used to account for the fact that concentration data often relate to composite samples, whereas an acute risk may result from single food units.

This module has as primary entities: *Foods Substances*

Output of this module is used by: *Dietary exposures*

Unit variability factors data formats

Unit variability factors

Unit variability factors specify the unit-to-unit variation of substance concentrations on foods. Unit variability factors are described by a single unit variability factors table.

Unit variability factors

Unit variability factors are defined for a food, and may possibly also be specified for a specific substance and/or processing type. The unit variability factors are linked to the foods by means of the food code (idFood). Unit variability factors can be specified as unit variability factors (P97.5/mean) or as coefficients of variation of a statistical distribution.

Table 2.73: Table definition for UnitVariabilityFactors.

Name	Type	Description	Aliases	Required
idFood	AlphaNumeric(50)	The food code.	idFood, FoodId, Food	Yes
idSubstance	AlphaNumeric(50)	The code of the substance.	idSubstance, SubstanceId, SubstanceCode, Substance, idCompound, CompoundId, Compound-Code, Compound	No
idProcessing-Type	AlphaNumeric(50)	The processing type code.	idProcessing-Type, ProcessingType-Id, ProcessingType, ProcType	No
Factor	Numeric	The variability factor.	Factor, VarFac, VariabilityFactor	No
UnitsIn-Composite-Sample	Numeric	The number of units in the composite sample.	UnitsIn-Composite-Sample, NoUnitComp	Yes
Coefficient	Numeric	The coefficient of variation.	Coefficient, Variability-Coefficient, CoefVar, VarCoef	No

Table aliases: UnitVariabilityFactors, UnitVariabilityFactor, VariabilityFactor, VariabilityFactors, VariabilityProcCompProd, UnitVariability, RawUnitVariabilityFactors.

Unit variability factors as data

- *Unit variability factors data formats*

2.4 Exposure modules

Exposures are, in the simplest applications, *dietary exposures*, which combine consumption and occurrence data, either for single or for multiple *substances* causing the same adverse *effect*. Links between the foods-as-eaten and the *foods-as-measured* are made using *food conversions*, and the consumptions are expressed as *consumptions per food as measured*. For large assessment groups, the use of *dietary exposures screening* may be used to reduce the complexity of the calculations and only focus calculations on the risk drivers.

In aggregate exposure assessments, *exposures* combine *dietary exposures* with *non-dietary exposures*, which have to be entered as pre-calculated data.

Human monitoring data can be compared to *exposures* using *human monitoring analysis*.

In cumulative assessments, important mixtures of *substances* can be identified using *exposure mixtures*.

2.4.1 Consumptions per food as measured

Consumptions by food as measured are consumptions of individuals expressed on the level of the foods for which concentration data are available (i.e., the foods-as-measured). These are calculated from consumptions of foods-as-eaten and food conversions that link the foods-as-eaten amounts to foods-as-measured amounts.

This module has as primary entities: *Populations Foods Substances*

Output of this module is used by: *Dietary exposures with screening Dietary exposures*

Consumptions by food-as-measured calculation

Consumptions by food as measured are calculated from *consumptions* of *foods-as-measured* and *food conversions* that link the foods-as-eaten amounts to foods-as-measured amounts. Given that the food conversion is already available, the procedure for computing the consumptions by food-as-measured is straightforward. For each consumption of each individual, a food-as-measured consumption record is created for each food-as-measured that is linked to the consumed foods through the food conversion, with the amount being the total consumption amount multiplied by the proportion indicated by the food conversion. Also, if the food conversion includes one or more processing steps, then these are recorded in the consumption per food as measured record.

Consumptions by food as measured

Calculation settings

Table 2.74: Calculation settings for module Consumptions by food as measured.

Name	Description
Restrict population to consumers or consumer days only (food-as-measured)	Specifies whether the population should be restricted to the individuals (chronic) or individual days (acute) with consumptions containing any of the foods-as-measured.
Risk type	The type of exposure considered in the assessment; acute (short term) or chronic (long-term).
Restrict population to consumers or consumer days with consumptions of specified foods-as-measured only	Specifies whether the population should be restricted to the individuals (chronic) or individual days (acute) with consumptions containing any of the specified food-as-measured subset.
Selected foods-as-measured	Set of consumed foods as measured that are of particular interest for restricting the consumers / consumption days.

Calculation of consumptions by food as measured

Consumptions by food as measured are calculated from consumptions of foods-as-eaten and food conversions that link the foods-as-eaten amounts to foods-as-measured amounts.

- *Consumptions by food as measured calculation*

Inputs used: *Consumptions Food conversions*

Settings used

- *Calculation Settings*

2.4.2 Dietary exposures

Dietary exposures are the amounts of substances, expressed per kg bodyweight or per individual, to which individuals in a population are exposed from their diet per day. Depending on the exposure type, dietary exposures can be short-term/acute exposures and then contain exposures for individual-days, or they can be long-term/chronic exposures, in which case they represent the average exposure per day over an unspecified longer time period.

This module has as primary entities: *Populations Foods Substances Effects*

Output of this module is used by: *Exposures*

Dietary exposures calculation

In probabilistic exposure assessment we consider a population of individuals. Exposure assessment with MCRA can address *acute exposure* or *chronic exposure*. Acute exposure is relevant when the short-term effect on individuals is relevant, chronic exposure when the long-term effects on the individuals matter. In MCRA short-term is operationalised as one day, so effectively acure exposure assessment is concerned with a population of person-days, whereas chronic exposure assessment is concerned with a population of persons.

The basic operation in exposure assessment is integrating consumptions and concentrations per food. With multiple foods, consumptions are typically correlated, therefore MCRA works with the multivariate distribution of a consumption vector, as represented by the consumption data of individuals in a consumption survey. In contrast, the distributions of concentration for each food are typically considered to be independent between foods. E.g., eating an apple with an accidentally high residue concentration does not predict that another food eaten on the same day will also have a high residue concentration. As a consequence of this assumption, concentrations of substances as modelled for each food independently.

For large assessment groups, the use of *dietary exposures screening* may be used to reduce the complexity of dietary exposures calculations and only focus calculations on the risk drivers. In this case, only detailed information is recorded for the risk drivers. With or without screening MCRA produces the same estimated cumulative exposure distribution summarized by percentiles and exceedance percentages, the same contributions of all substances and all foods-as-measured. After screening, contributions related to food-as-eaten are available for the risk drivers only.

Acute exposure assessment

In an acute exposure assessment, the short term exposure to a substance or group of substances is estimated. The interest is in the distribution of individual day exposures and derived statistics like the fraction of days that exceed an intake limit or point of departure (PoD). The PoD is calculated as the acute reference dose (ARfD) * safety factor (SF). The basic model for the exposure to a substance in an acute exposure assessment is:

$$y_{ij} = \frac{\sum_{k=1}^{p} x_{ijk} c_{ijk}}{bw_i}$$

where y_{ij} is the intake by individual i on day j (in microgram substance per kg body weight), x_{ijk} is the consumption by individual i on day j of food k (in g), c_{ijk} is the *(simulated) concentration* of that substance in food k eaten by individual i on day j (in mg/kg), and bw_i is the body weight of individual i (in kg). Finally, p is the number of foods accounted for in the model. Within parenthesis, the default unit definitions are assumed, but decimal multiples or submultiples of units are easily specified using the relevant tables.

In the exposure assessment, individual days enter the Monte Carlo sample using the inverse of the sampling weights wi when the number of MC iterations is > 0 (see *table for Individuals*, field *SamplingWeight*).

Modelling unit to unit variation

In the basic model for an acute exposure assessment, it is assumed that the concentration of the substance displays the variation of residues between units in the marketplace. In general, both monitoring data and controlled field trial data are obtained using composite samples and, as a result, some of the unit to unit variation is averaged out. The model for unit variability aims to adjust the composite sample mean such that sampled concentrations represent the originally unit to unit variation of the units in the compositie sample.

MCRA offers three distributions to sample from:

1. the *beta distribution*,

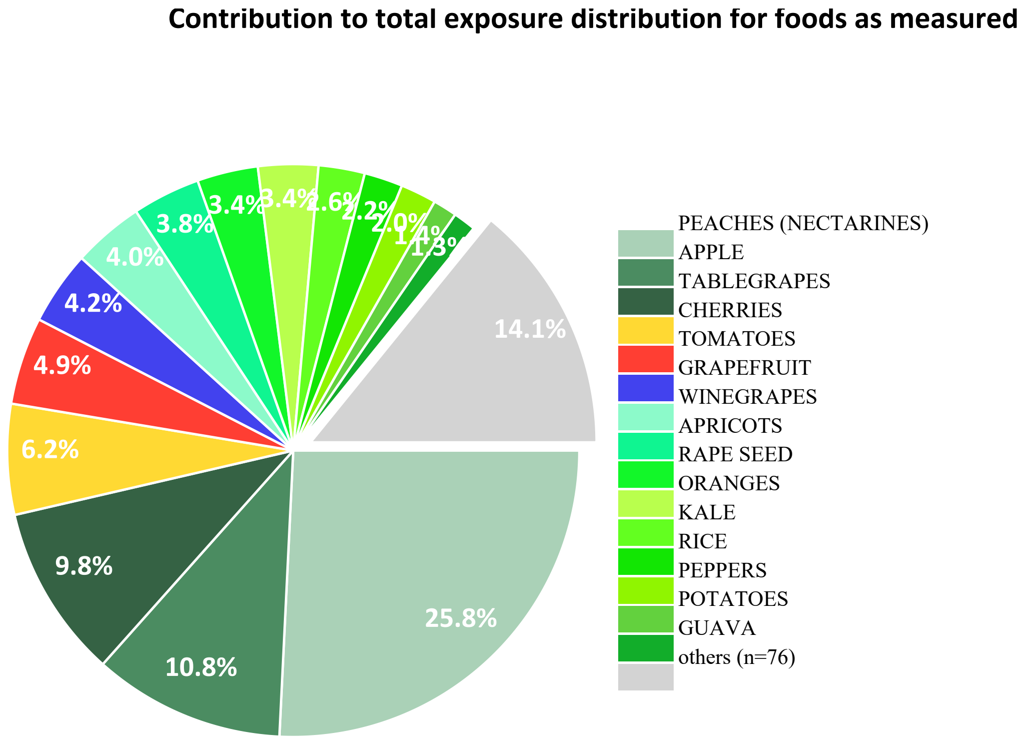


Figure 2.12: Example MCRA dietary exposure contributions foods as measured.

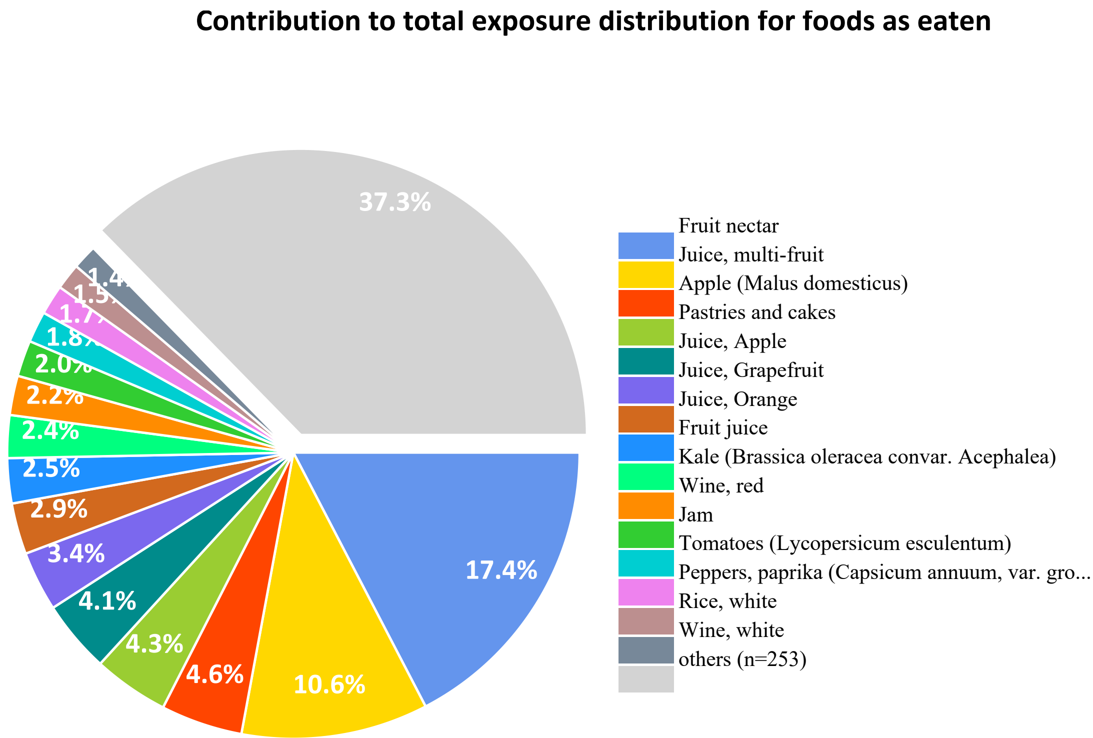


Figure 2.13: Example MCRA dietary exposure contributions foods as eaten

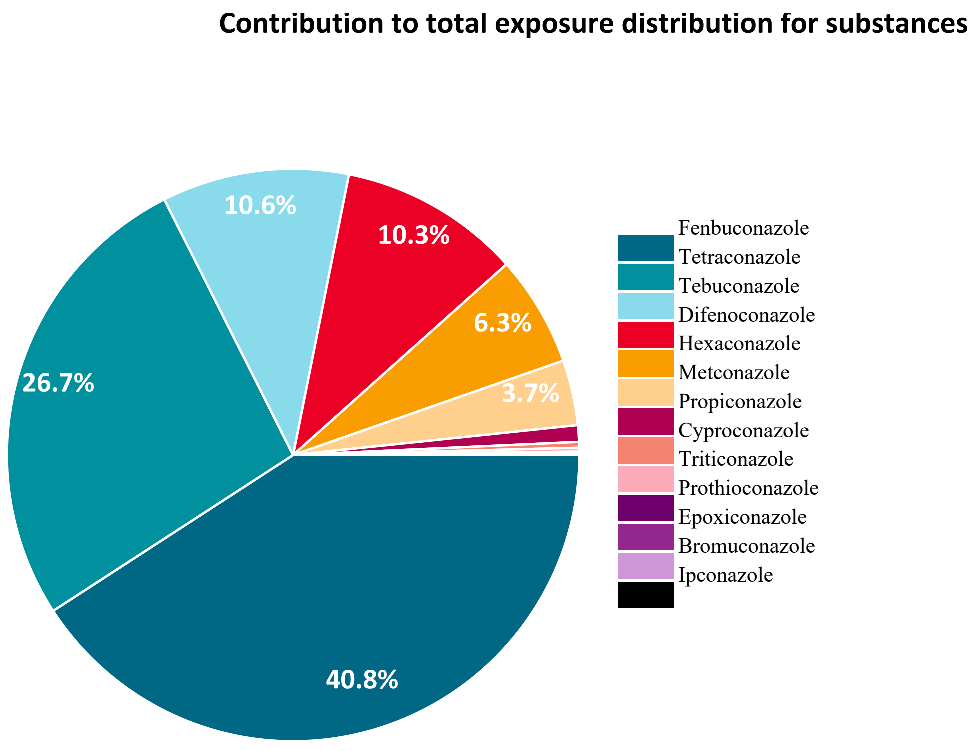


Figure 2.14: Example MCRA dietary exposure contributions substances

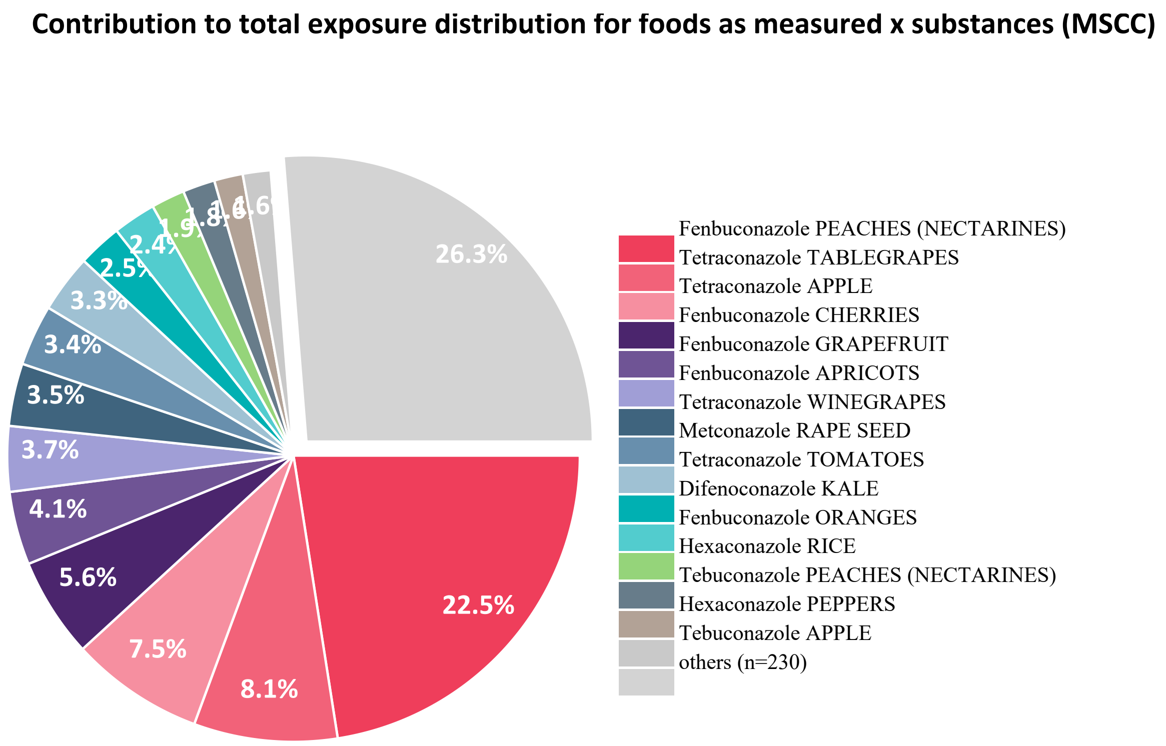


Figure 2.15: Example MCRA dietary exposure contributions foods as measured x substances

2. the *lognormal distribution*,
3. and the *bernoulli distribution*.

The beta distribution simulates values for a unit in the composite sample and requires knowledge of the number of units in a composite sample and of the variability between units. The lognormal distribution simulates values for a new unit in the batch and requires only knowledge of the variability between units. The bernoulli distribution is considered as a limiting case of the beta distribution when knowledge of the variability between units is lacking and only the number of units in the composite sample is known. For the beta and lognormal distribution, estimates of unit variability are realistic (no censoring at the value of the monitoring residue) or conservative (unit values are left-censored at the value of the monitoring residue). For the lognormal distribution, sampled concentrations have no upper limit whereas for the beta distribution, sampled concentration values for a unit are never higher than the monitoring residue * the number of units in the composite sample.

Variability between units is specified using a variability factor v (defined as 97.5th percentile divided by mean) or a coefficient of variation cv (standard deviation divided by mean). Following FAO/WHO recommendations, for small crops (unit weight < 25 g), a default variability factor v = 1 is used, for large crops (unit weight ≥ 25 g), a variability factor v = 5 is used. For foods which are processed in large batches, e.g. *juicing, marmalade/jam, sauce/puree, bulking/blending* a variability factor v = 1 is proposed.

Estimation of intake values using the concept of unit variability

A composite sample for food k is composed of nu_k units with nominal unit weight wu_k. The weight of a composite sample is $wm_k = nu_k \cdot wu_k$ with mean residue value cm_k.

- For each iteration i in the MC-simulation, obtain for each food k a simulated intake x_{ik}, and a simulated composite sample concentration cm_{ik}.
- Calculate the number of unit intakes nux_{ik} in x_{ik} (round upwards) and set weights w_{ikl} equal to unit weight wu_k, except for the last partial intake, which has weight $w_{ikl} = x_{ik} - (nux_{ik} - 1)wu_k$.
- For the beta or bernoulli distribution: draw nux_{ik} simulated values bc_{ikl} from a beta or bernoulli distribution. Calculate concentration values as $c_{ikl} = bc_{ikl} \cdot cm_{ik,max} = bc_{ikl} \cdot cm_{ik} \cdot nu_k = svf_{ikl} \cdot cm_{ik}$, where nu_k is the number of units in a composite sample of food k, and svf_{ikl} is the stochastic variability factor for this simulated unit, i.e. the ratio between simulated concentration c_{ikl} and the simulated composite sample concentration cm_{ik}. Sum to obtain the simulated concentration in the consumed portion:

$$c_{ik} = \sum_{l=1}^{nux_{ik}} w_{ikl}c_{ikl}/x_{ik}$$

- For the lognormal distribution: draw nux_{ik} simulated logconcentration values lc_{ikl} from a normal distribution with (optional) a biased mean $\mu = ln(cm_{ik})$ or (default) unbiased mean $\mu = ln(cm_{ik}) - 1/2\sigma^2$ and standard deviation σ. Calculate concentration values as

$$c_{ikl} = \exp(lc_{ikl}) = svf_{ikl} * cm_{ik}$$

where svf_{ikl} is the stochastic variability factor for this simulated unit, i.e. the ratio between simulated concentration c_{ikl} and the simulated composite sample concentration cm_{ik}. Back transform and sum to obtain the simulated concentration in the consumed portion:

$$c_{ik} = \sum_{l=1}^{nux_{ik}} w_{ikl}c_{ikl}/x_{ik}$$

For cumulative exposure assessments, a sensitivity analysis may be performed by specifying a full correlation between concentrations from different substances on the same unit. As a result, high (or low) concentrations from different substances occur together on the same unit. In MCRA, for each unit the random sequence is repeatedly used to generate concentration values for all substances.

Chronic exposure assessment

In a chronic exposure assessment, the main interest goes to the fraction of individuals with a usual intake per day higher than an intake limit or point of departure (PoD). The PoD is calculated as the acceptable daily intake (ADI) * safety factor (SF). Usual intake is defined as the long-run average of daily exposure to a substance or group of substances by an individual. Usually, for an individual, dietary recall data are available on 2 (or more) consecutive days. We assume an equal number of days for each individual, unless specified differently in *table for Individuals*.

For a chronic exposure assessment the available data are used to calculate exposures per person-day (daily intake):

$$y_{ij} = \frac{\sum_{k=1}^{p} x_{ijk} c_{ijk}}{bw_i}$$

where y_{ij}, x_{ijk} and bw_i are defined as before but now concentrations of the substance found in food k enter the model as the *estimated mean substance concentration value* c_k. Using the person-day intakes MCRA, provides a number of *intake models* to calculate the distribution of usual intake at the person level.

Chronic intake models

Using the person-day intakes MCRA uses one of the following models to calculate the distribution of usual intake at the person level:

1. The observed individual means observed individual means (OIM) model;
2. The logisticnormal-normal model, in a full version that includes the estimation of correlation between intake frequency and amount (LNN), and in a simpler version without this estimation (LNN0);
3. The betabinomial-normal (BBN) model;
4. The discrete/semi-parametric model known as the Iowa State University Foods (ISUF) model. For this model, an equal number of days per individual is assumed.

In modelling usual intake, two situations can be distinguished. Foods are consumed on a *daily basis* or foods are *episodically consumed*. For the logisticnormal-normal model and the betabinomial-normal model, the latter requires fitting of a two-part model,

1. a model for the frequency of consumption, and
2. a model for the intake amount on consumption days.

In the final step, both models are integrated in order to obtain the usual intake distribution. For daily consumed foods, fitting of the frequency of consumption is skipped and modelling resorts to fitting the model to daily intake amounts only. Note that the distinction between BNN, LNN and LNN0 disappears and modelling will give equivalent results.

Observed individual means (OIM)

The usual intake distribution for a population is estimated with the empirical distribution of individual means. Each mean is the average of all single-day intakes for an individual. The mean value for an individual still contains a considerable amount of within-individual variation. As a consequence, the distribution of within-individual means has larger variance than the true usual intake distribution and estimates using the OIM-method are biased, leading to a too high estimate of the fraction of the population with a usual intake above some standard. Despite its known tendency to over-estimate high-tail exposures, the OIM method is the method to be used in EFSA (2012) [3] basic assessments.

Model based and model assisted

Following Kipnis et al. [31], some of the models available in MCRA are extended to predict individual usual intakes. This model assisted approach has been added to BBN and LNN0 and may be a useful extension in evaluating the relationship between health outcomes and individual usual intakes of foods. In contrast, the estimation of the usual intake distribution in the general population is called the model based approach. Summarizing, we get Table 2.75:

Table 2.75: Model based and assisted approach available for chronic exposure models

Model based approach	Model assisted approach
	observed individual means (OIM)
betabinomial-normal (BBN)	betabinomial-normal (BBN)
logisticnormal-normal without correlation (LNN0)	logisticnormal-normal without correlation (LNN0)
logisticnormal-normal with correlation (LNN)	
Iowa State University Foods (ISUF)	

The model assisted approach builds on the proposal of Kipnis et al. [31], but is modified to ensure that the population mean and variance are better represented. The method is based on shrinkage of the observed individual means (modified BLUP estimates) and shrinkage of the observed intake frequencies. The model-assisted usual intake distribution applies to the population for which the consumption data are representative, and automatically integrates over any covariates present in the model. Model-assisted intakes are not yet available for LNN, and when a covariable is modelled by a spline function of degree higher than 1. In case of a model with covariates the usual intake is presented in graphs and tables as a function of the covariates (conditional usual intake distributions).

Betabinomial-Normal model (BBN)

The *Betabinomial-Normal (BBN)* model for chronic risk assessment is described in [16], including its near-dentity to the STEM-II model presented in [40].

Logisticnormal-Normal model (LNN with and without correlation)

An alternative to the betabinomial modelling of intake frequencies in BBN model is modelling these frequencies by a logistic normal distribution. In notation, for probability p:

$$\text{logit}(p) = \log(p/1 - p) = \mu - i + \underline{c}_i$$

where μ_i represents the person specific fixed effect model and $\underline{c}_i$ represent person specific random effects with estimated variance component $\sigma^2_{between}$. This model is referred to as the *LogisticNormal-Normal (LNN0)* model. The full *LNN model* model includes the estimation of a correlation between intake frequency and intake amount. This is similar to the NCI model described in Tooze et al. [43]. A simple and computationally less demanding version of the LNN method which does not estimate the correlation between frequency and amount is termed LNN0, where the '0' indicates the absence of correlation. The models are fitted by maximum likelihood, employing *Gauss-Hermite integration*.

For chronic models amounts are usually transformed before the statistical model is fit. The power transformation, given by y^p, has been replaced by the equivalent Box-Cox transformation. The Box-Cox transformation is a linear function of the power transformation, given by $(y^p - 1)/p$, and has a better numerical stability. Gauss-Hermite integration is used for back-transformation (see also *Box Cox power transformation*).

Discrete/semi-parametric model (ISUF)

Nusser et al. [35] described how to assess chronic risks for data sets with positive intakes (a small fraction of zero intakes was allowed, but then replaced by a small positive value). The modeling allowed for heterogeneity of variance, e.g. the concept that some people are more variable than others with respect to their consumption habits. However, a disadvantage of the method was the restricted use to contaminated foods which were consumed on an almost daily basis, e.g. dioxin in fish, meat or diary products. The estimation of usual intake from data sets with a substantial amount of zero intakes became feasible by modeling separately zero intake on part or all of the days via the estimation of intake probabilities as detailed in Nusser et al. [36] and Dodd [17]. In MCRA, a discrete/semi-parametric model is implemented allowing for zero intake and heterogeneity of variance following the basic ideas of Nusser et al. and Dodd ([35], [36], [17]). This implementation of the ISUF model for chronic risk assessment is fully described in de Boer et al. [16].

Model-Then-Add

The traditional approach can be termed the Add-Then-Model approach, because adding over foods precedes the statistical modelling of usual exposure. MCRA offers, as an advanced option, an alternative approach termed Model-Then-Add (van der Voet et al. 2014). In this approach the statistical model is applied to subsets of the diet (single foods or food groups), and then the resulting usual exposure distributions are added to obtain an overall usual exposure distribution. The advantage of such an approach is that separate foods or food groups may show a better fit to the normal distribution model as assumed in all common models for usual exposure (including MCRA's betabinomial-normal (BBN) model and logisticnormal-normal model models). That this principle can work in practice was shown in previous work (de Boer et al. 2009 [16], Slob et al. 2010 [41], Goedhart et al. 2012) [21], and a simulation model was developed and implemented in MCRA 7.1 to show how multimodal distributions can arise from adding unimodal distributions of foods that are not always consumed (Slob et al. 2010 [41], de Boer and van der Voet 2011, [15]). For specific cases involving separate modelling of dietary supplements and the rest of the diet, proposals have been made (Verkaik-Kloosterman et al. 2011) [47]. However, a practical approach to apply the Model-Then-Add approach to general cases of usual exposure estimation was still missing. Therefore a module in MCRA was developed to implement such an approach based on a visual inspection of a preliminary estimate of the usual exposure distribution using the Observed Individual Means (OIM) method.

The Model step

At this stage of development the division of foods into a number of food groups is performed in an interactive process, where the MCRA user is presented with a visual display (see example in Figure 2.16) which shows:

1. The OIM distribution represented as a histogram, where each bar shows the frequency of exposures (summed over foods) of individuals in a certain exposure interval; each bar is subdivided according to the contributions of the individual foods contributing to those exposures (left panel Figure 2.16).
2. The contributions graph, where each of the bars in the OIM histogram is expanded to 100%. This graph allows a better view of the lower bars in the OIM histogram.

The visual display identifies the nine foods that contribute most to the total exposure; the remaining foods are grouped in a rest category to avoid identification problems because of too many colours (right panel Figure 2.16).

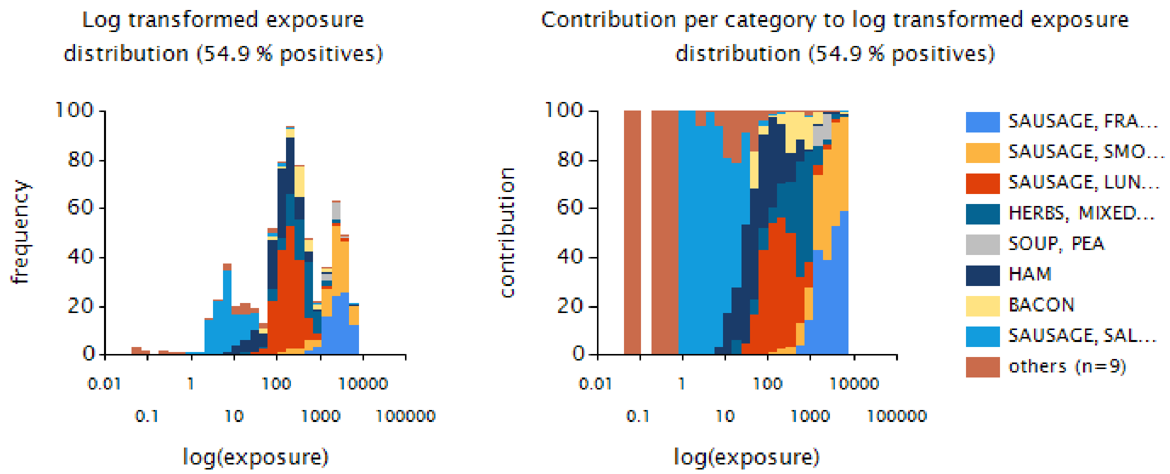


Figure 2.16: Left panel: OIM usual exposure distribution to smoke flavours via the different foods (excluding the zero exposures) in young children; right panel: Contribution of foods to exposures within each bar of the OIM distribution histogram.

The user has now the possibility to select one or more foods and to split these from the main exposure histogram. A separate graph shows the OIM distribution for the split-off food or food group. The graphs for the main group (now called the rest group) are adapted to show the OIM distribution and the contributions for the remaining foods only (see Figure 2.17 upper two panels). This splitting-off can be repeated several times for other foods or food groups. In this way the user can try to obtain foods or food groups that show unimodal OIM distributions. If the result is not what is intended, a food or food group can be added again to the rest group. Per split-off food or food group the usual exposure can be modelled using either BBN or LNN, with a logarithmic or power transformation. The rest group will always be modelled as OIM. It is possible that the rest group is empty, when the total exposure via the different split-off foods and /or food groups is modelled with BBN or LNN.

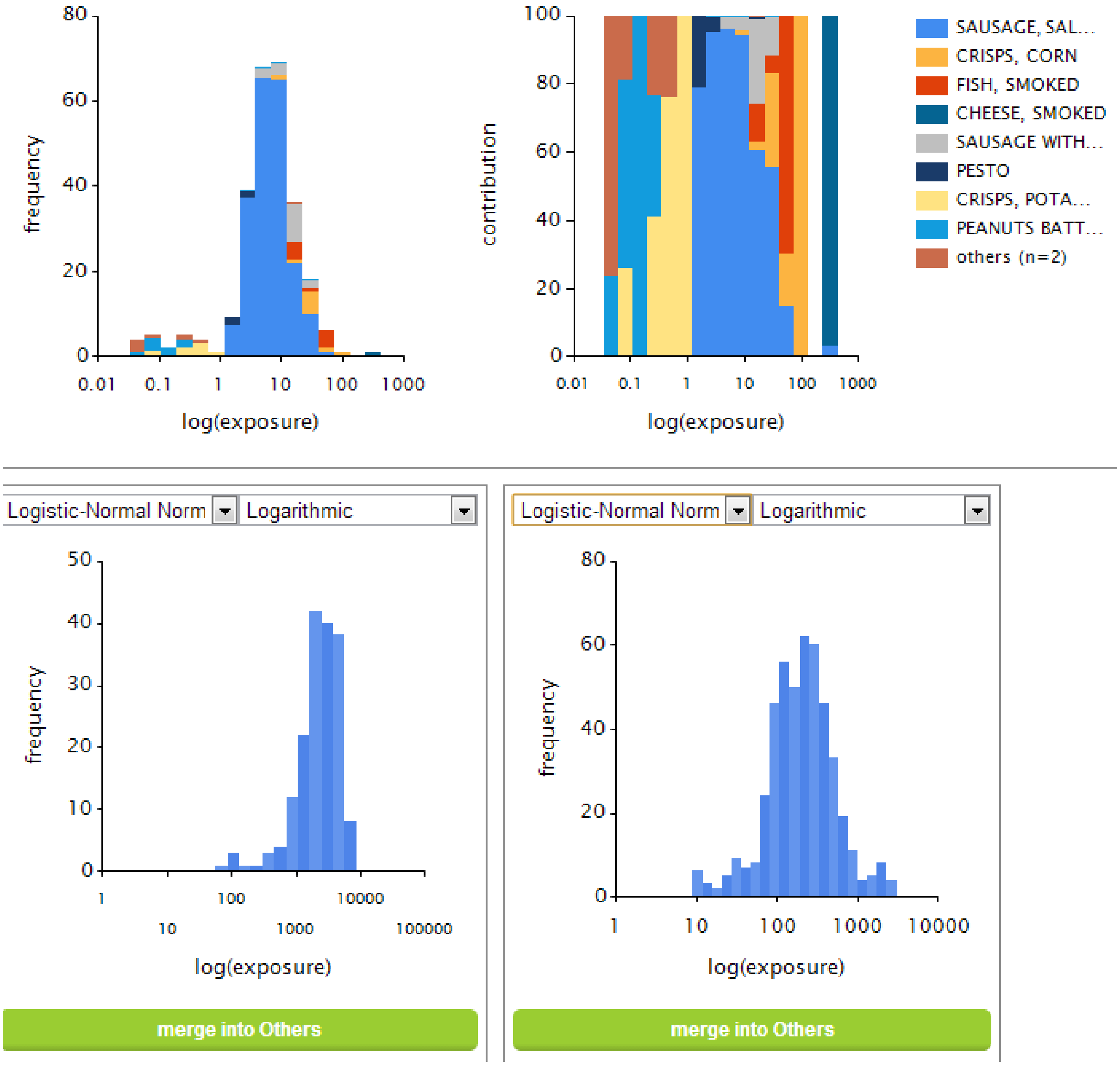


Figure 2.17: Result of a selection into two split-off groups and a rest group. The graph bottom left represents the exposure via a food group containing *'Sausage, frankfurter'* and *'Sausage, smoked cooked'*. The graph bottom right represents the exposure via a food group containing *'Sausage, luncheon meat', Herbs, mixed, main brands, not prepared', 'Soup, pea', 'Ham', and 'Bacon'*. The top graph represents the exposure via the rest group.

After a split-off selection has been made, the OIM distribution is summarised in terms of the defined grouping (Figure 2.18), and the usual exposure distribution per split-off food or food group is fitted according to the chosen modelling settings.

The Add step

Consumptions of foods may be correlated. In the traditional Add-Then-Model approach the Add step automatically reflects any correlations that are apparent in the consumptions at the individual-day or individual level. In the Model-Then-Add approach the estimated usual exposure distributions for different foods or food groups have to be combined to assess the total usual exposure. Two approaches are available for this:

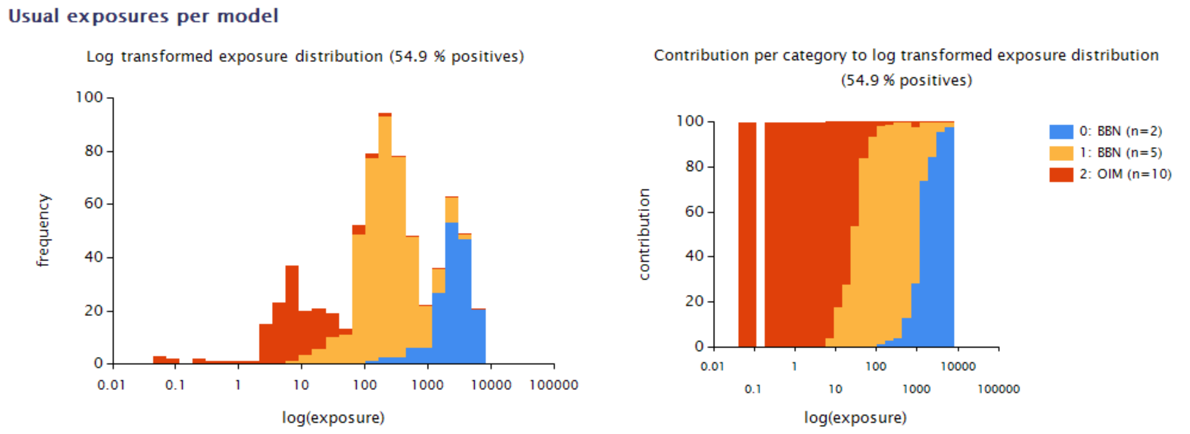


Figure 2.18: OIM usual exposure distribution showing the contributions from the three food groups as constructed in Figure 2.17.

1. Model-based approach: adds independent samples from the usual exposure distribution per food or food group, ignoring any correlations in consumption;
2. Model-assisted approach: adds the model-assisted, person-specific usual exposure estimates per food or food group, taking correlations in consumptions into account.

Before the addition is made, in the model-based approach, model-based estimates of the usual exposure amounts distribution per food or food group are back-transformed values from the normal distribution assumed for transformed amounts per food or food group, and the model-based frequency distribution is sampled to decide if a simulated individual has exposure via the food or food group or not. Model-assisted estimates of the usual exposure distribution are back-transformed values from a shrunken version of the transformed OIM distribution, also done per food or food group, where the shrinkage factor is based on the variance components estimated using the linear mixed model for amounts at the transformed scale (van Klaveren et al. 2012). For individuals with no observed exposure (OIM=0) no model-assisted estimate of usual exposure can be made and a model-based replacement is used.

The model-based approach was investigated in Slob et al. (2010) [41] and performed surprisingly well, even if correlations in consumptions of foods were present. The model-assisted approach adds exposures at the individual level, and therefore retains effects of correlations between foods in the usual exposure distribution.

MCRA calculates both the model-based and model-assisted usual intake distributions.

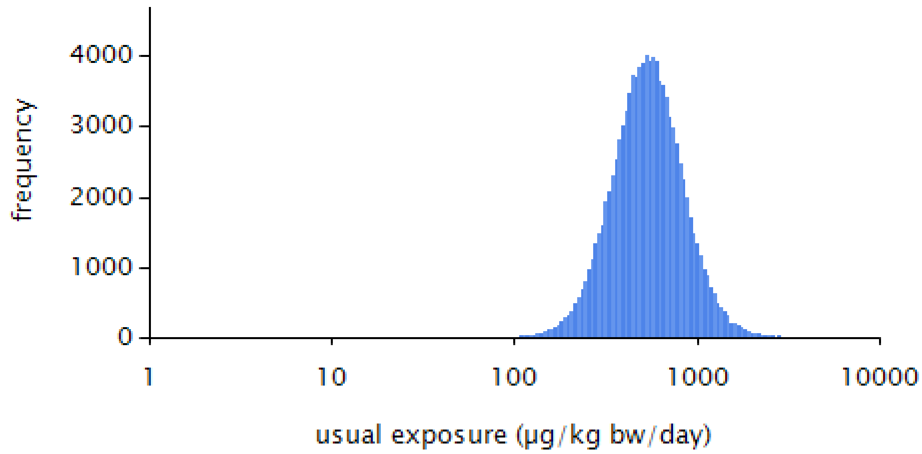


Figure 2.19: Model-assisted estimated usual exposure distributions (excluding the zero exposures).

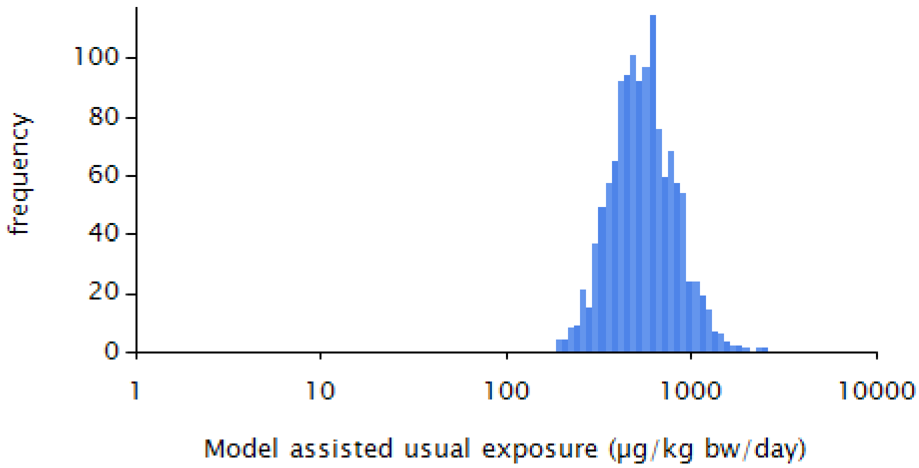


Figure 2.20: Model-based estimated usual exposure distributions (excluding the zero exposures).

Chronic exposure as a function of covariates

The intake X frequency and transformed intake amounts may be modelled as a function of covariates. MCRA allows one covariable and/or one cofactor.

Table 2.76: Intake frequencies and amounts, modelled as a function of covariates.

	Frequencies	Amounts
cofactor	$logit(\pi) = \beta_{0l}$	$transf(y_{ij}) = \beta_{0l} + c_i + u_{ij}$
covariable	$logit(\pi) = \beta_0 + \beta_1 f(x_1; df)$	$transf(y_{ij}) = \beta_0 + \beta_l f(x_1; df) + c_i + u_{ij}$
both	$logit(\pi) = \beta_{0l} + \beta_1 f(x_1; df)$	$transf(y_{ij}) = \beta_{0l} + \beta_l f(x_1; df) + c_i + u_{ij}$
interaction	$logit(\pi) = \beta_{0l} + \beta_{1l} f(x_1; df)$	$transf(y_{ij}) = \beta_{0l} + \beta_{1l} f(x_1; df) + c_i + u_{ij}$

Here $l = 1 \cdots L$ and L is the number of levels of the cofactor, y_{ij} , the intake amount, x_1 is the covariable, f is a polynomial function with the degrees of freedom df, c_i and u_{ij} are the individual effect and interaction effect, respectively. These effects are assumed to be normally distributed $N(0, \sigma^2_{between})$ resp. $N(0, \sigma^2_{within})$. The degree of the function is determined by backward or forward selection. In the output, the usual intake is displayed for a specified number of values of the covariable and/or the levels of the cofactor.

Total Diet Study

In Total Diet Studies (TDS), substance occurrence data is obtained from measuring food products as consumed. TDS offers a more direct measure of substance concentrations compared to traditional monitoring and surveillance programs that are concerned with contamination of raw agricultural commodities. In a TDS, food selection is based on national consumption data in such a way that 90 to 95% of the usual diet is represented by the samples. Selected foods are collected, prepared as consumed and related foods are pooled prior to analysis. The compositions these TDS food samples are described by the *TDS food sample compositions* data module.

In MCRA, TDS concentration data can also be used in *dietary exposure assessments*, using it as an alternative type of concentration data where the foods-as-measured are not the raw primary commodities (RACs), but these are TDS food compositions. To link the concentration data to the consumed foods, the *TDS food sample composition information* is used in the *food conversion algorithm* in a manner analogous to the use of *food recipes* describing the composition of a composite food. The main difference is that the translation proportion is always 100% (default). Take, as an example, a TDS food *FruitMix* that is composed of *apple, orange* and *pear*, then a consumed food (food-as-eaten) *apple-pie* is converted to *apple*, *wheat* and *butter* (in some specific proportions) and subsequently, *apple* is converted to food-as-measured *FruitMix* (100%). Not necessarily all foods as consumed are represented in a TDS food sample. In addition to the TDS food sample compositions, there may be additional foods that are not officially part of a TDS food, but which can be extrapolated to a TDS food sample. Through the use of *food extrapolations* (read across translations), these foods may be directly linked to a TDS food sample, e.g., by specifying that *pineapple* is translated to *FruitMix*, *pineapple* or foods containing *pineapple* will also be matched to a *FruitMix* concentration.

Because TDS samples only contain one single, average measurement, TDS occurrence data can currently only be used for only applicable for chronic exposures assessments. However, when variability information is available for the raw primary foods in the TDS food samples (e.g., from monitoring), this information may be used to approximate the variance of TDS samples.

For more information about Total Diet Studies, visit the TDS-Exposure website http://www.tds-exposure.eu.

Deriving the variance of TDS samples from monitoring

Variability of TDS food sample concentrations can be derived using *concentration distributions* for the sub-foods of the TDS food samples. For each sub-food, e.g. *apple* (sub-food of TDS food *FruitMix*), a coefficient of variation (CV) is specified that is derived using the available monitoring samples. Note that monitoring samples may be composite samples. For *apple*, composite food samples are measured and each sample contains, for instance, 12 apples with unit weight 200 g. So monitoring concentrations, c_{mi}, are based on composite samples with a total weight w_{mi} = 2400 g each.

A TDS food sample is composed of w_i g of food i with i = 1…k, wi represents the *PooledAmount* in *TDS food sample compositions table*. Then, the concentration of a TDS food sample may be represented as:

$$c_{TDS} = \sum_{i=1}^{k}(w_i \cdot c_i) / \sum_{i=1}^{k} w_i$$

with variance:

$$var(c_{TDS}) = \sum_{i=1}^{k}(w_i \cdot var(c_i)) / \sum_{i=1}^{k} w_i$$

and $var(c_i)$ is the variance of concentrations c_i of food i with portion sample size w_i.

It is expected that increasing the number of units in a composite sample will have a reverse effect on the variation between concentrations. Suppose TDS food *FruitMix* is composed of 2 x 200 = 400 g *apple*. The expected variation between portion sizes of 400 g will be larger than between portion sizes of 2400 g:

$$var(c_i) = var(c_{mi}) \cdot w_{mi}/w_i$$

The variance of the monitoring samples are corrected as follows, calculate:

1. $var(c_{mi}) = \log(CV_{mi}^2 + 1)$
2. $var(c_i) = var(c_{mi}) \cdot w_{mi}/w_i$
3. $CV_i = \sqrt{\exp(var(c+i) - 1)}$

Scenario analysis

The outcome of a MCRA risk assessment may be that some foods dominate the right upper tail of the exposure distribution. A scenario analysis answers the question to what extent the risk of foods with a high exposure would have been diminished by an intervention or by taking any precautions. To be able to do so, some information is needed about the variability of the concentration distribution of the raw agricultural commodities that make up the TDS food sample. These distributions may be characterised by a mean and a dispersion factor, the standard deviation or, preferably, a percentile point e.g. p95. Monitoring samples may be used for this purpose. In addition, for each subsample food an upper concentration limit is needed. This value is interpreted as the concentration that is considered a high risk. The decision to intervene or not is based on the comparison between this upper limit and p95.

- For p95 ≤ limit, most concentration values are below the value that is considered as a potential risk, so there is no urgency to take any precautions.
- When the opposite is true, i.c. p95 > limit, there may be an argument to intervene for this specific food.

In MCRA, limits and p95's are supplied in *the concentration distributions table*. In the MCRA interface, a scenario analysis is checked (optionally) and in the scroll down menu only foods are shown with p95 > limit. Selected foods enter the risk assessment with a reduced concentration value:

$$c_{TDS}/reductionfactor,$$

where c_{TDS} is the concentration value of the TDS food with *reductionfactor = p95 / limit*.

Substance concentrations generation

Both chronic and acute dietary exposure assessments rely on assigning substance concentrations to consumed foods-as-measured. For chronic exposure assessments, this concentration should be the mean concentration of the food and substance, as obtained from the concentration models. For acute, these concentrations are obtained through random sampling, for which there are two distinct approaches: sample-based and substance-based.

Sample-based concentrations generation

In the sample-based approach, the analytical samples from the concentration data form the basis for generating concentrations. For each identified food-as-measured of a consumption, substance concentrations are generated by drawing a random sample from the set of all samples available for that food-as-measured. Assuming that for the drawn sample, substance concentration values are known for all substances of interest (i.e., all missing values and non-detects are imputed with either a zero concentration or a positive concentration at or below LOR), the substance concentrations for all substances of the assessment group are set to the substance concentrations of the drawn samples. The rationale behind this approach is that it maintains correlations between substance concentrations on the same food.

As mentioned, the sample based approach relies on all samples being analysed for all substances of interest. Often, this is not the case and for a given sample, concentration may missing for one or more substances. Also, this approach requires non-detect values to be imputed with either positive concentration or a zero concentration.

For imputation of missing values there are two approaches:

1. **Imputation by zero:** all missing values are assumed zero.
2. **Imputation using substance-based concentration models:** all missing values are imputed by drawing a concentration value from the substance-based concentration models.

For imputation of non-detects, two approaches exist:

1. **Replace by zero:** Non-detect values are imputated by a zero concentration value. This is an optimistic approach.
2. **Replace by factor times LOR:** Each non-detect value is replaced by a factor (e.g., 1 or 1/2) times its LOR.

Substance-based concentrations generation

In the substance-based approach, substance concentrations for a given food are drawn independently per substance from the food/substance concentration models.

Processing factor correction

Processing factors can be specified as fixed factors (nominal) or as statistical distributions for the variability across samples.

Concentrations in the consumed food (food as eaten) may be different from concentrations in the food as measured in monitoring programs (typically raw food) due to processing, such as peeling, washing, cooking etc. Concentrations are therefore corrected according to

$$c'_{jhk} = pf_{jhk} \cdot c_{jhk} = \left(\frac{PF_k}{cf_k}\right) \cdot c_{jhk}$$

where c_{jhk} is the concentration of substance k in the food j with processing type h, and where $pf_{jhk} = \frac{PF_{jhk}}{cf_{jhk}}$ is a factor indicating the mass change for a specific combination k of food as measured and processing. The processing correction factor cf_{jhk} is used to correct for the fact that the processing factors PF_{jhk} as commonly available from the input data describe both the effects of chemical alteration and weight change. E.g. for a dried food with a consumption of 100 gram which is translated to 300 gram raw agricultural commodity, the correction factor is 3. Note that the weight change is already included when calculating the consumption amounts of the foods-as-measured.

- The distribution is either the logistic-normal distribution for processing types with factors restricted between 0 and 1 (e.g. washing),
- or the lognormal distribution for processing types with non-negative factors (e.g. drying).

Variability distribution parameters are specified indirectly via the 50th and 95th percentile. Uncertainty for processing factors can be specified using uncertainty distributions of the same form as for variability. Uncertainty distribution parameters are specified indirectly via the 95th uncertainty percentiles on the 50th and 95th variability distribution percentiles.

For distribution based processing factors specify $f_{k,nominal}$ and $f_{k,upper}$ (*Nominal* and *Upper* in table **Processing-Factors**). Two situations are distinguished depending on the type of transformation.

Nonnegative processing factors

Equate the logarithms of $f_{k,nominal}$ and $f_{k,upper}$ to the mean and the 95% one-sided upper confidence limit of a normal distribution. This normal distribution is specified by a mean

$$ln(f_{k,nominal})$$

and a standard deviation

$$ln(f_{k,upper})–ln(f_{k,nominal})/1.645$$

Processing factors between 0 and 1

Equate the logits of $f_{k,nominal}$ and $f_{k,upper}$ to the mean and the 95% one-sided upper confidence limit of a normal distribution. This normal distribution is specified by a mean

$$logit(f_{k,nominal})$$

and a standard deviation

$$logit(f_{k,upper})–logit(f_{k,nominal})/1.645.$$

Dietary exposures settings

Calculation settings

Table 2.77: Calculation settings for module Dietary exposures.

Name	Description
Dietary exposure calculation tier (Note: also set tier in sub-modules)	A tier is a pre-specified set of model configurations. By selecting a model tier, MCRA automatically sets all model settings in this module according to this tier. Note that currently tier setting may need to be poerformed separately in sub-modules. Use the Custom tier when you want to manually set each model setting.
Risk type	The type of exposure considered in the assessment; acute (short term) or chronic (long-term).
Cumulative	Specifies whether the assessment involves multiple substances and results should be cumulated over all substances.
Dietary intake calculation method	Dietary intake calculation method: choose between point estimates and probabilistic.
Refined method choice	Choose foods/substances for which a dietary intake method other than the default should be specified.
Sample based	Include co-occurrence of substances in samples in simulations. If checked, substance residue concentrations are sampled using the correlations between values on the same sample. If unchecked, any correlation between substances is ignored, substance residue concentrations are sampled ignoring the correlations between values on the same sample.
Consumptions on the same day come from the same sample	if checked, in procedure of EFSA Guidance 2012, section 4.1.1, all consumptions of a raw commodity of an individual on the same day are assumed to come from the same sample. If unchecked, all consumptions of a raw commodity of an individual on the same day are assumed to come from different samples.
Maximize co-occurrence of high values in simulated samples	Within each pattern of substance presence. If checked, substance residue concentrations are sorted within co-occurrence patterns of substances on the same samples. After sorting, high residue values occur more frequently on the same sample. This choice is conservative. If unchecked, substance residue concentrations are sampled at random, ignoring any co-occurrence patterns of substances on the same samples. This choice is less conservative.
Apply processing factors	Specified in table ProcessingFactor. If checked, processing factors are applied. Concentrations in the consumed food may be different from concentrations in the food as measured in monitoring programs (typically raw food) due to processing, such as peeling, washing, cooking etc. If unchecked, no processing information is used. This is in most (though not all) cases a worst-case assumption
Processing factor model	
Unit variability model	Describes variation between single units when concentration data are from composite samples.
Estimates nature	Simulated unit concentrations can be higher or lower than composite value (realistic) or only equal or higher (conservative).
Unit variability parameter	Use Coefficient of variation or Variability factor, specified in VariabilityFactor table.
Mean of LogNormal simulated values (biasing)	Unbiased: correct unit simulations for difference between median and mean.
Default variability factor for unit weight <= 25g	Default variability factor 1 (unit weight <= 25 g, small crops). Still requires specification of unit weight (FoodProperties table) and, in case of beta model, also the Number of units in a composite sample (UnitVariability table).
Default variability factor for unit weight > 25g	Default variability factor 5 (unit weight > 25 g, medium/large crops). Still requires specification of unit weight (FoodProperties table) and, in case of beta model, also the Number of units in a composite sample (UnitVariability table).
Model type	The parametric model for between-and within-individual variation, and possibly covariates.
Model-then-add	Specifies whether to create separate exposure models for specific groups of foods as measured (model-then-add).

Output settings

Table 2.78: Output settings for module Dietary exposures.

Name	Description
Include drill-down on 9 individuals around specified percentile.	Specifies whether drilldown on 9 individuals is to be included in the output.
Summarize simulated data	Specifies whether a summary of the simulated consumptions and concentrations should be included in the output.
Store simulated individual day exposures	Store the simulated individual day exposures. If unchecked, no additional output will be generated. If checked, the output will contain an additional section with the simulated individual day exposures.
Show percentiles for	Give specific percentiles of exposure distribution (%), e.g. 50 90 95 97.5 99 (space separated).
Percentage for drilldown	Gives detailed output for nine individuals near this percentile of the exposure distribution.
Percentage for upper tail	Gives detailed output for this upper percentage of the exposure distribution.
Show % of population below level(s)	Exposure levels can be generated automatically or by explicit specification (Manual).
Exposure levels	Specify exposure levels for which to give the percentage of exposure below these levels, e.g. 1 10 50 100 200 500. Specify below whether these levels are absolute or relative to ARfD/ADI.
Exposure levels are	Specify whether exposure levels are absolute or percentages of ARfD/ADI.
Number of levels of covariable to predict exposure	Specify the number of levels, e.g. 20. The range of the covariable is divided by the number of levels: range = (max - min)/levels. For these covariable levels exposures are predicted.
Predict exposure at extra covariable levels	Specify specific prediction levels in addition to the automatically generated prediction levels (space separated).
Lower percentage for variability (%)	The default value of 25% may be overruled.
Upper percentage for variability (%)	The default value of 75% may be overruled.
Report consumptions and exposures per individual instead of per kg body weight	Specifies whether body weights should be ignored and consumptions and exposures should be expressed per individual. Otherwise, the consumptions and exposures are per kg body weight.

Uncertainty settings

Table 2.79: Uncertainty settings for module Dietary exposures.

Name	Description
Resample imputation exposure distributions	Specifies whether to resample the imputated exposure distributions.

Dietary exposures tiers

Overview

Table 2.80: Tier overview for module Dietary exposures.

Name	EFSA 2012 Optimistic	EFSA 2012 Pessimistic	EC 2018 Tier 1	EC 2018 Tier 2
Dietary intake calculation method	DistributionEstimates	DistributionEstimates	DistributionEstimates	DistributionEstimates
Refined method choice	false	false	false	false
Sample based	true	true	true	true
Consumptions on the same day come from the same sample	false	true	true	true
Processing factor model	Fixed	FixedAllowHigher	Fixed	Fixed
Unit variability model	NoUnitVariability	BetaDistribution	BetaDistribution	BetaDistribution
Estimates nature		Realistic	Realistic	Realistic
Unit variability parameter		VariabilityFactor	VariabilityFactor	VariabilityFactor
Model type	OIM	OIM	OIM	OIM
Model-then-add	false	false	false	false
Covariate modelling	false	false	false	false
Iterate survey	false	false	false	false
Report consumptions and exposures per individual instead of per kg body weight	false	false	false	false

EFSA 2012 Optimistic

Use the optimistic model settings according to the EFSA Guidance 2012. Concentration values are sampled using a sample-based empirical distribution. Available processing factors are applied. No unit vatiability model should be applied.

Table 2.81: Tier definition for EFSA 2012 Optimistic.

Name	Setting
Dietary intake calculation method	DistributionEstimates
Refined method choice	false
Sample based	true
Consumptions on the same day come from the same sample	false
Processing factor model	Fixed
Unit variability model	NoUnitVariability
Model type	OIM
Model-then-add	false
Covariate modelling	false
Covariate modelling	false
Iterate survey	false
Report consumptions and exposures per individual instead of per kg body weight	false
Report consumptions and exposures per individual instead of per kg body weight	false

Input tiers

Table 2.82: Input tiers for EFSA 2012 Optimistic.

Module	Input tier
Concentration models	*EFSA 2012 Optimistic*

EFSA 2012 Pessimistic

Use the pessimistic model settings according to the EFSA Guidance 2012. Only processing factors > 1 are applied. For unit variability, the Beta or Lognormal distribution should be applied.

Table 2.83: Tier definition for EFSA 2012 Pessimistic.

Name	Setting
Dietary intake calculation method	DistributionEstimates
Refined method choice	false
Sample based	true
Consumptions on the same day come from the same sample	true
Processing factor model	FixedAllowHigher
Unit variability model	BetaDistribution
Estimates nature	Realistic
Unit variability parameter	VariabilityFactor
Model type	OIM
Model-then-add	false
Covariate modelling	false
Covariate modelling	false
Iterate survey	false
Report consumptions and exposures per individual instead of per kg body weight	false
Report consumptions and exposures per individual instead of per kg body weight	false

Input tiers

Table 2.84: Input tiers for EFSA 2012 Pessimistic.

Module	Input tier
Concentration models	*EFSA 2012 Pessimistic*

EC 2018 Tier 1

Table 2.85: Tier definition for EC 2018 Tier 1.

Name	Setting
Dietary intake calculation method	DistributionEstimates
Refined method choice	false
Sample based	true
Consumptions on the same day come from the same sample	true
Processing factor model	Fixed
Unit variability model	BetaDistribution
Estimates nature	Realistic
Unit variability parameter	VariabilityFactor
Model type	OIM
Model-then-add	false
Covariate modelling	false
Covariate modelling	false
Iterate survey	false
Report consumptions and exposures per individual instead of per kg body weight	false
Report consumptions and exposures per individual instead of per kg body weight	false

Input tiers

Table 2.86: Input tiers for EC 2018 Tier 1.

Module	Input tier
Concentration models	*EC 2018 Tier 1*

EC 2018 Tier 2

Table 2.87: Tier definition for EC 2018 Tier 2.

Name	Setting
Dietary intake calculation method	DistributionEstimates
Refined method choice	false
Sample based	true
Consumptions on the same day come from the same sample	true
Processing factor model	Fixed
Unit variability model	BetaDistribution
Estimates nature	Realistic
Unit variability parameter	VariabilityFactor
Model type	OIM
Model-then-add	false
Covariate modelling	false
Covariate modelling	false
Iterate survey	false
Report consumptions and exposures per individual instead of per kg body weight	false
Report consumptions and exposures per individual instead of per kg body weight	false

Input tiers

Table 2.88: Input tiers for EC 2018 Tier 2.

Module	Input tier
Concentration models	*EC 2018 Tier 2*

Calculation of dietary exposures

Dietary exposures are calculated from consumptions per food-as-measured and concentration models. Optionally, also processing factors and unit variability models are applied.

- *Dietary exposures calculation*

Inputs used: *Consumptions by food as measured Concentration models Processing factors Unit variability factors Dietary exposures with screening Active substances Relative potency factors*

Settings used

- *Calculation Settings*

2.4.3 Dietary exposures with screening

Dietary exposures with screening are just dietary exposures, but the calculation includes a prior screening step to identify the main risk drivers (food-substance combinations). This allows computations with more substances by suppressing some details for less important food-substance combinations.

This module has as primary entities: *Foods Substances Effects*

Output of this module is used by: *Dietary exposures*

Dietary exposures with screening calculation

A full Monte Carlo analysis can be unwieldy for large cumulative assessment groups (CAGs) and/or large number of foods or concentration data. An algorithmic approach was developed to handle large CAGs. Two unique features of MCRA are:

- contributions to the exposure results can be seen both in terms of food-as-eaten (e.g. white bread) and foods-as-measured (e.g. wheat), and
- a drill-down can be made into the exact foods and compounds contributing for simulated individuals or individual-days in the upper tail.

The number of combinations of simulation, compound, food-as-measured and food-as-eaten can be very large. To avoid memory problems with very large datasets, an additional optional modelling step, named Screening, was added to MCRA. Screening should be used if the data dimensions are too large for a direct analysis. Screening identifies risk drivers. A full analysis based on screened risk drivers will still retain all food/compound combinations in the exposure calculation, and will therefore produce exactly the same cumulative exposure distribution, and allow to see contributions of all compounds and all foods-as-measured. Details with respect to foods-as-eaten are however restricted to the risk drivers selected in the screening step.

The two-step approach consists of:

- **Step 1: Data screening and selection of risk drivers** Run a simple analysis for each potential source/compound combination (SCC). Here source means the combination of food-as-eaten and food-as-measured, for example apple in apple pie. The screening is based on this combination, and not just foods-as-measured, to avoid problems with potentially multi-modal consumption distributions as much as possible (see van der Voet et al. 2014). SCCs are also referred to as risk driver components. The screening step in MCRA implements a simple model that is applied to each SCC. The model calculates a percentile of interest in a distribution, consisting of a spike of zeroes (non-consumptions), and a mixture of two lognormal distributions for the exposure related to non-detects and positive concentrations, respectively. SCCs (risk driver components) can be combined to measured source/compound combinations (MSCCs, risk drivers). For example APPLE/apple juice/captan and APPLE/apple pie/captan combine to APPLE/captan. MCRA has an interface which identifies the Top-N SCCs (based on a chosen exposure percentile, e.g. p95) with an option to select N based on cumulative importance according to some criterion. Remark: Screening is performed before concentration modelling. Therefore there is no correction for processing factors at the screening stage.

- **Step 2: Full MC analysis** Perform the standard MC to all combinations of compounds and foods, but restrict the stored information regarding foods-as-eaten to the SCCs selected in step 1.

The screening method requires to specify:

- Which percentile to consider for each single Source-Compound Combination (SCC, potential risk driver component) (default p95)
- Which percentage of all exposures (according to the screening model) should be covered by the selected set of SCCs (default 95%)
- How to impute non-detect concentrations (c < LOR) in the screening step. The choice of a factor 0 (default) represents optimistic imputation, the choice of a factor 1 represents a pessimistic imputation. It may be noted that a factor 1 (pessimistic imputation) may select many SCCs (risk driver components) with relatively high LORs and high RPFs, but with only nondetect measurements. Choosing a lower fraction, e.g. 0.25 can be useful if a more realistic method is sought.

Based on limited experience with the EFSA test data, useful settings of these three screening parameters were found to be (95, 95, 0) in preparation for an EFSA optimistic run, and (50, 95, 0.25) in preparation for an EFSA pessimistic run. See also screening calculatuion *acute exposure* and *chronic exposure* .

Dietary exposures with screening settings

Calculation settings

Table 2.89: Calculation settings for module Dietary exposures with screening.

Name	Description
Percentage defining the exposure percentile of interest per food-as-eaten/food-as-measured/substance combination	Percentage defining the exposure percentile of interest per food-as-eaten/food-as-measured/substance combination.
Include risk drivers to include minimally a percentage	The selection criterion for the risk drivers. The cumulative contribution percentage of the selected risk drivers will be this percentage.
Non-detect replacement: factor x LOR	A constant between 0 and 1. A value 0 can be used for an optimistic screening (LOR not used). Note that a factor 0.5 (pessimistic) may result in many and often high contributions from food-substance combinations with only non-detects.

Calculation of dietary exposures with screening

Screening results are computed for each combination of source (being a specific combination of food-as-eaten/food-as-measured) and substance by combining simple approximations of the consumption and the concentration distribution.

- *Dietary exposures with screening calculation*

Inputs used: *Consumptions by food as measured Concentration models Active substances Relative potency factors*

Settings used

- *Calculation Settings*

2.4.4 Exposures

Exposures are amounts of substances, typically expressed per mass unit and per day, to which individuals in a population are exposed at a chosen target level. This target level may be external exposure (dietary exposure, expressed per unit body weight, or per person) or internal exposure (expressed per unit organ weight). Internal exposures may be aggregated from dietary and non-dietary exposures using either absorption factors or kinetic models to translate the external exposures to internal exposures. Exposures can be short-term/acute exposures and then contain exposures for individual-days, or they can be long-term/chronic exposures, in which case they represent the average exposure per day over an unspecified longer time period.

This module has as primary entities: *Populations Foods Substances*

Output of this module is used by: *Exposure mixtures Human monitoring analysis Risks*

Exposures calculation

Calculation of exposures comprises two main steps:

1. *Linking dietary and non-dietary individual/individual-day exposures*.
2. *Computing the (aggregated) internal exposures at the specified target compartment*.

Both steps are optional in this module. If none is selected, exposures are external dietary exposures, i.e the target level is external/dietary. However, when multiple routes of exposure are considered, then the target level should be an internal compartment (organ). In the latter case, absorption factors or kinetic model are needed to aggregate the exposures from multiple routes into exposure at the target compartment. It is also possible to only provide dietary exposures and compute internal exposures at some target compartment.

In cumulative exposure calculations two simple approaches are used to identify and select mixtures contributing to the exposure of a target population:

1. qualitative approach: *counting of co-exposure*. To which combinations of compounds are individuals exposed?
2. quantitative approach: *maximum cumulative ratio (MCR)*. To what degree are mixtures more important than single compounds?

A quantitative approach is available in the *exposures mixtures module*.

Combining dietary and non-dietary exposures

If dietary and non-dietary exposures are available for the same individuals or individual-days, the non-dietary exposures can be matched to specific individuals of the food survey from which the dietary exposures are generated. More commonly, dietary and non-dietary exposures are available from separate surveus, in which case they can be randomly combined. If both dietary and non-dietary information is available for a known population of individuals, the user may select the *matching option* such that specific dietary and non-dietary estimates are aggregated for each individual in the food survey population. If matching is enabled, any non-dietary exposures that do not correspond to individuals from the food survey will be ignored (see *Example 2*), unless an individual is specified with *id = General*. In that case, the dietary individual should meet the criteria of the non-dietary survey, specified by the survey properties, to be assigned. If the non-dietary data relates instead to a population in which individuals have no corresponding records in the food survey (unmatched case), the user may choose to randomly assign the non-dietary exposures to the individuals from the food survey.

When multiple non-dietary surveys are available, the options with or without correlation are important (not relevant when matching is switched on). When correlation is chosen, the exposure contributions of non-dietary individuals with identical ids in different surveys are combined and allocated to a randomly selected dietary individual. When the correlation is not chosen, the non-dietary exposures of randomly selected individuals from different surveys are combined and allocated to a dietary individual.

The user may also define demographic criteria for the assignment (for each source of non-dietary exposure) to indicate that those exposures are relevant only to a defined sub-population. Only those individuals in the food survey who meet the criteria of the non-dietary survey will be assigned non-dietary exposures from that source e.g. only males aged 18 to 65 (see *Example 1*). The simplest assessment consists of a single (deterministic) non-dietary exposure estimate which is assigned to all individuals in the food survey (*idIndividual* = *General*). This case, and more complex possibilities are illustrated below using hypothetical examples.

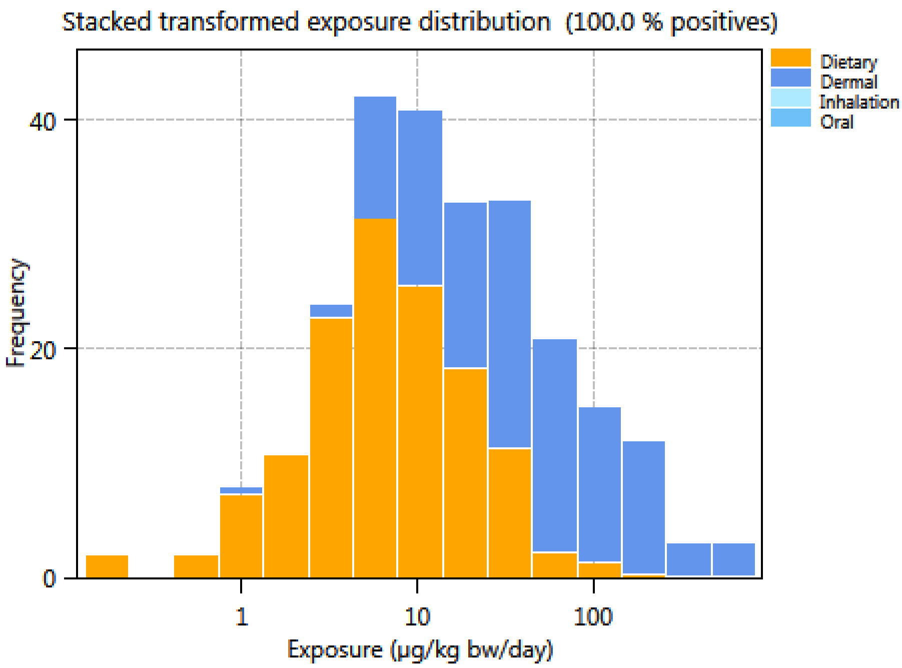


Figure 2.21: Aggregate exposure distributions.

Example 1

Deterministic cumulative (multi-substance) non-dietary exposure input, adult male sub-population. Unmatched case.

Table 2.90: NonDietaryExposures

idIndividual	idNonDietarySurvey	idSubstance	Dermal	Oral	Inhalation
General	1	011003001	10	5	17
General	1	011003002	34	20	18
General	1	011003002	56	43	19

Table 2.91: NonDietarySurveys

idNonDietary-Survey	Description	Location	Date	NonDietary-IntakeUnit
1	BROWSE, acute, cumulative, operators	York	09/10/2012	$\mu g/day$

Table 2.92: NonDietarySurveyProperties

idNonDietary-Survey	Property-IndividualName	Individual-PropertyText-Value	Individual-Property-DoubleMin-Value	Individual-Property-DoubleMax-Value
1	Age		18	65
1	Gender	Male		

In this example, there are exposure values for multiple substances in Table 2.90 and the user has provided Table 2.92 which specifies that the non-dietary exposures given in survey number 1 relate to males aged 18 to 65.

When this assessment is performed, only those individuals whose property values fit the criteria in Table 2.92 will receive the non-dietary exposures in survey 1. The use of *idIndividual* = *General* indicates that this is the default exposure. All individuals in the dietary survey meeting the criteria receive all exposures given in the 3 rows, corresponding to 3 substances. The following should be noted:

- There should only ever be one *General* entry in the dietary exposures table per substance, survey combination.
- The property names and the values of any text properties must precisely match those given in the **Individual-Properties** and **IndividualPropertyValues** tables for this to work.
- The minimum and maximum values for numeric properties are both inclusive boundaries.

So in this example, all males aged 18 to 65 will receive the given exposures of all three substances and the other members of the population will receive no non-dietary exposure. Note that example 1 describes the unmatched case.

Example 2

Variability (but no uncertainty) in cumulative non-dietary exposure input (matched to dietary survey individuals).

Table 2.93: NonDietaryExposures

idIndividual	idNonDietarySurvey	idSubstance	Dermal	Oral
5432	1	011003001	10	5
5432	1	011003002	33	22
5433	1	011003001	12	7
5433	1	011003002	34	23
5434	1	011003001	18	9
5434	1	011003002	35	25
5435	1	011003001	10	5
5435	1	011003002	33	21

Table 2.94: NonDietarySurveys

idNonDietary-Survey	Description	Loca-tion	Date	NonDietaryIntakeU-nit
1	BROWSE, acute, cumulative, operators	York	09/10/2012	$\mu g/day$

In this example, the non-dietary exposures are being specified explicitly for individuals in the dietary population. Switch 'matching' on to allow exposure variability to be specified at the individual level. For the purposes of illustration, the population is extremely small, consisting of only four individuals. The values in the *idIndividual* column of Table 2.93 match those in the **Individuals** table (dietary population).

It is not mandatory to specify exposures for every individual in the population. Those not included will simply receive a zero non-dietary exposure, unless there is also a default exposure value (*General* row(s) in Table 2.93) and the individual matches the specified demographic criteria for the survey, as specified in Table 2.92. (In this example, a default exposure would apply to all individuals not listed in Table 2.93 because Table 2.92 has not been used).

There is variability between individuals in this example, but no uncertainty in exposure. Note that these data could also be used with matching switched off. This would be the same as treating the *idIndividual* values as generic individuals, so that each pair of exposure lines would be assigned at random to individuals meeting the criteria.

Example 3

Variability (no uncertainty) in cumulative non-dietary exposure input (unmatched individuals).

Table 2.95: NonDietaryExposures

idIndividual	idNonDietarySurvey	idSubstance	Dermal	Oral	Inhalation
ND1	1	011003001	10	5	17
ND1	1	011003002	33	22	45
ND2	1	011003001	12	7	18
ND2	1	011003002	34	23	47
ND3	1	011003001	18	9	19
ND3	1	011003002	35	25	49
ND4	1	011003001	10	5	17
ND4	1	011003002	33	21	45

Table 2.96: NonDietarySurveys

idNonDietary-Survey	Description	Loca-tion	Date	NonDietaryIntakeU-nit
1	BROWSE, acute, cumulative, operators	York	09/10/2012	$\mu g/day$

Table 2.97: NonDietarySurveyProperties

idNonDi-etarySurvey	PropertyIndi-vidualName	IndividualProper-tyTextValue	IndividualProperty-DoubleMinValue	IndividualProperty-DoubleMaxValue
1	Age		50	65
1	Gender	Female		

This example is similar to example 2, except that the values in the *idIndividual* column of Table 2.95 do not match those in the **Individuals** table. In this instance, 'matching' would not be an option, and the non-dietary exposures would be randomly assigned to individuals who meet the criteria in Table 2.97. (In fact for the same result rather than changing the values in the *idIndividual* column in Table 2.93 from the previous example may be used with matching switched off). Exposures in Table 2.95 will be recycled if the number of exposure rows is less than the number of dietary records with which to aggregate exposures.

Again, there is variability between individuals in this example, but no uncertainty in exposure.

By allowing generic *idIndividual* values in this way, correlations between different sources (within individual) can be accounted for even in the unmatched case. If the same *idIndividual* value is used in different surveys, then the corresponding exposure values will be kept together and assigned to an eligible individual as a combined exposure.

So for option matching switched of, it is relevant whether individuals are correlated or not. In the following example, two non-dietary surveys are available, per survey three individuals are specified.

Table 2.98: matching switched of, with correlation or without.

idIndividual	idNonDietarySurvey	idSubstance	Dermal	Oral	Inhalation
ND0	1	011003001	10	5	17
ND1	1	011003001	23	22	45
ND2	1	011003001	12	7	18
ND0	1	011003001	34	23	47
ND3	1	011003001	18	9	19
ND4	1	011003001	33	16	35

- When a correlation is applied, the non-dietary exposure for individual ND0 from survey 1 and 2 are combined and allocated to a dietary individual. For individual ND1, ND2, ND3 and ND4 just a single non-dietary exposure is found and allocated to a dietary individual.
- When no correlation is applied, the exposure for individual ND0 from survey 1 is combined with one of the exposures of ND0, ND3 or ND4 from survey 2; exposure of ND1 from survey 1 is combined with one of the exposures of ND0, ND3 or ND4 from survey 2, etc.

When the intention is to sample just one exposure for a dietary individual, specify per survey different codes, e.g. ND1, ND2, ND3 for survey 1, ND4, ND5, ND6 for survey 2 and apply correlation, or specify 6 different individual codes and just one *idNonDietarySurvey*. Then, options with or without correlation are irrelevant and sampling results are identical no matter which option is chosen.

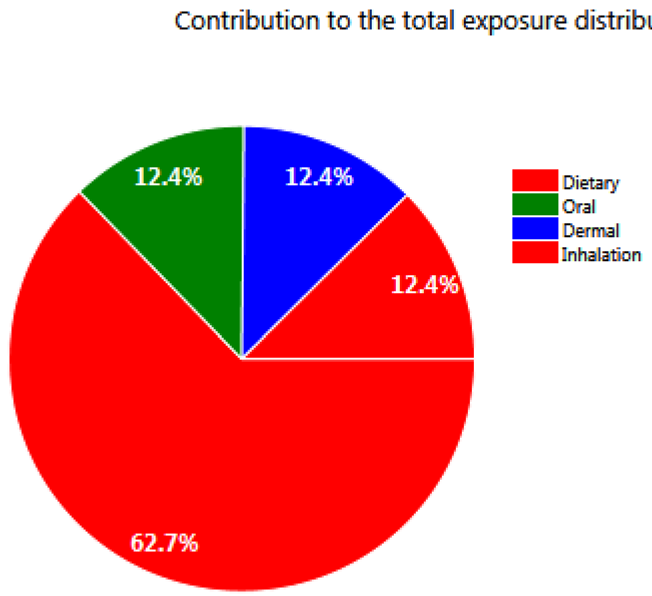


Figure 2.22: Contributions by route to aggregate exposure distributions.

Internal exposures calculation

Computation of internal exposures (internal substance amounts and concentrations) requires a *kinetic model* to translate external doses, possibly from multiple routes, to internal doses at the target compartment/organ of interest.

Calculation of internal concentrations using absorption factors

In the simplest form, internal concentrations are derived from external exposure concentrations using multiplication factors (or, absorption factors) that can be specified by substance and by route. That is, for a given substance, the internal exposure $exp_{\mathtt{int}}$ is computed as

$$exp_{\mathtt{int}} = \sum_{r \in Routes} f_{\mathtt{abs},r} \cdot exp_{\mathtt{ext},r}$$

Here, *Routes* denotes the set external exposure routes, $exp_{\mathtt{ext},r}$ denotes the external exposure for route r and $f_{\mathtt{abs},r}$ denotes the absorption factor of route r. Note that this model assumes that both external and internal exposures refer to amounts or concentrations depending on the *dietary exposures* setting (External exposure: substance amount per individual, or substance amount divided by body weight; internal exposure: substance amount per organ, or substance amount divided by organ weight.) Also, both external and internal exposures are expressed per day.

Calculation of internal concentrations using kinetic models

A more detailed alternative to using absorption factors is to use one of the *advanced kinetic models* available in MCRA. In this approach, for each substance independently, the external exposures of an individual (chronic) or individual-day (acute) are presented for a number of simulated day to a PBK model of the individual. This yields a time course of the internal substance amount at the specified target compartment/organ from which a long term average substance amount (chronic) or peak substance amount (acute) can be obtained. An example of such a time course is given in Figure 2.23 for acute exposure assessments, and in Figure 2.24 for chronic exposure assessments. By dividing this substance amount by the weight of the compartment, an internal concentration is obtained. Notice that this procedure also changes the unit of the exposures from exposure per day to long term exposure.

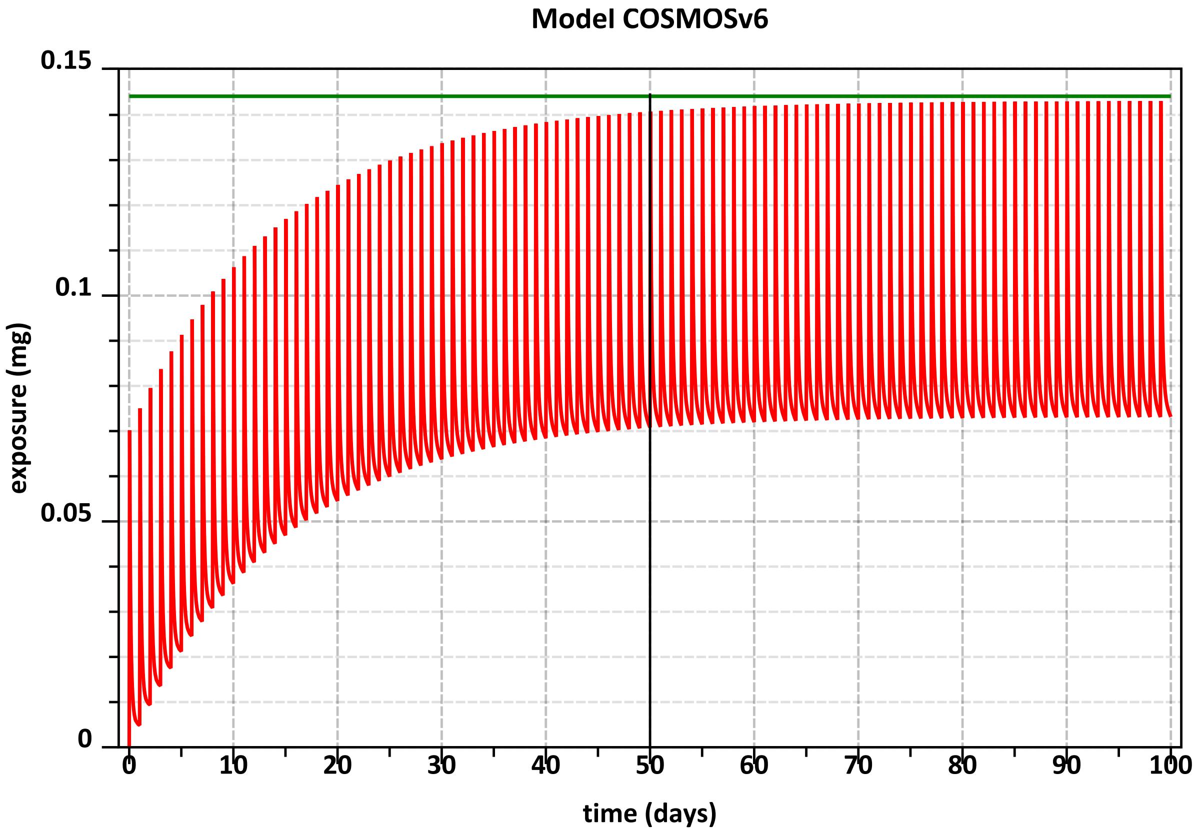


Figure 2.23: Time course of the internal substance amount when applying the same single dose on each day. The acute internal concentration is derived as the peak substance amount (the green line in the figure) divided by the compartment weight. The vertical line at 50 indicates the selected end of an assumed non-stationary period, defining a burn-in period that is to be ignored for computing the peak substance amount.

Mathematically, the calculation of the peak substance amount ($d_{\texttt{peak}}$) for deriving acute internal exposures is as follows:

$$d_{\texttt{peak}} = \max_{i=0,\ldots,n_{\texttt{stop}}} \{d(t_{\texttt{start}} + i\Delta t)\}\,.$$

Here, $d(t)$ denotes the substance amount at time t, $t_{\texttt{start}}$ denotes the starting time of the evaluation window (defined by the *non-stationary period*), Δt denotes the time resolution of the kinetic model (e.g., hours or minutes), and $n_{\texttt{stop}}$ denotes the total number of time-points, marking the end of the evaluation window (defined by the specified number of simulation days), which is computed as

$$n_{\texttt{stop}} = \left\lfloor \frac{t_{\texttt{stop}} - t_{\texttt{start}}}{\Delta t} \right\rfloor.$$

Likewise, chronic long term average substance amounts ($d_{\texttt{avg}}$) are computed as:

$$d_{\texttt{avg}} = \frac{\sum_{i=0}^{n_{\texttt{stop}}} d(t_{\texttt{start}} + i\Delta t)}{n_{\texttt{stop}}}.$$

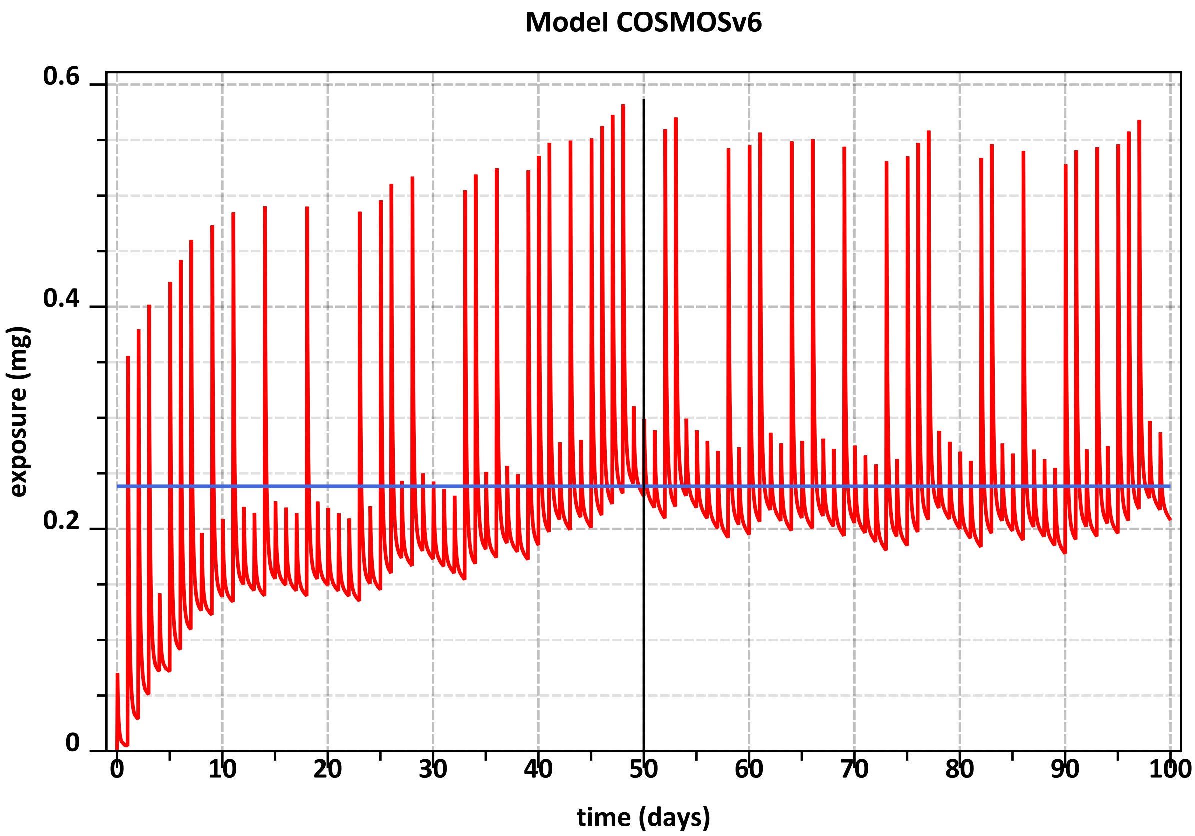


Figure 2.24: Time course of the internal substance amount when randomly applying one of the individual-day doses for a number days. The chronic internal concentration is derived as the average substance amount (the blue line in the figure), divided by the compartment weight. The vertical line at 50 indicates the selected end of an assumed non-stationary period, defining a burn-in period that is to be ignored for computing the average substance amount.

Dosing patterns

In MCRA, the dietary and non-dietary exposures are computed at the level of exposures per day. However, when applying advanced PBK models, dosing patterns may be specified at a much finer resolution (e.g., hours or minutes). For this, a method is needed to translate external exposures provided per day to dosing patterns of substance amounts during the day. The simplest, yet not very realistic model is to apply, per route, the full exposure amount in one single dose at the beginning of the day. Alternatively, MCRA offers the possibility to specify, per route, the *number of exposure events per day*. If it is specified to use multiple doses per day, then the total substance amount of each day is divided into equal portions which are applied at regular time-intervals during the day.

Non-stationary period

Especially in the case of chronic exposure assessments, where a long term average exposure is computed based on the simulated time-course, it is important to realise that at time zero, the substance is commonly considered to be completely absent in the simulated system. However, this is not a realistic assumption. It is much more likely that the substance was already present in the system, and that the level is equal to the level obtained from applying the same chronic exposures to the system. For this, a specification of the *number of days skipped* (or burn-in period) is required in order to come to these initial concentration levels. This period is not used for computing the long term average or peak exposures, but just to determine initial (background) concentration levels.

Counting of co-exposure

In this qualitative approach, the number of combinations of substances to which an individual is exposed are recorded, see Table 2.99. There is no cut-off level, the only criterion is the presence of a substance in the simulated daily diet or not. For an acute or short term exposure assessment, a simulated individual day is the smallest entity to determine co-exposure. For a chronic or long term exposure assessment, co-exposures are summarized at the individual level, e.g. co-exposure is determined combining all consumption days of an individual.

Table 2.99: Counting combinations of substances in the exposure matrix: for example, on day 1 there is coexposure to substances Tebuconazole, Bitertanol and Triadimefon

Substance	day 1	day 2	day 3	…	day n
Tebuconazole	x	x		…	
Bitertanol	x		x	…	x
Triadimefon	x		…	…	x
…	…	…	…	…	…

In Table 2.100, the frequency and percentage for the number of substances occurring together are shown.

Table 2.100: Co-exposure of substances

Number of substances	Frequency	Percentage
0	337	3.4
1	959	9.6
2	1207	12.1
3	1275	12.8
…	…	…

In Table 2.101, the mixtures containing the substance(s) including all other combinations with the specified combination of substance(s), (a maximum number of 15 records are shown)

Table 2.101: Mixtures containing substances

Number of substances	Percentage	Substances
1	5.88	Tebuconazole
2	3.88	Imazalil (aka enilconazole), Tebuconazole
0	3.37	
3	2.20	Difenoconazole, Imazalil (aka enilconazole), Tebuconazole
1	1.78	Imazalil (aka enilconazole)
3	1.76	Imazalil (aka enilconazole), Tebuconazole, Triadimenol
…	…	…

Maximum Cumulative Ratio

Price and Han [38] propose the Maximum Cumulative Ratio (MCR) which is defined as the ratio of the cumulative exposure received by an individual on an intake day to the largest exposure received from a single substance:

MCR = Cumulative exposure/ Maximum exposure

This MCR statistic is also picked up as a practical device in a recent JRC report [8] to investigate cumulative exposure. If MCR is large, it is important to consider cumulative effects, if MCR is close to 1, the individual exposure will not be much different from a single-substance assessment. The MCR can therefore be interpreted as the degree to which the risk of being exposed is underestimated by not performing a cumulative risk assessment.

The MCR statistic is implemented in MCRA for both the acute risk and the chronic risk cases. In the acute risk case the short-term (single-day) exposures are used, in the chronic case the long-term individual exposures (estimated by aggregating over the available survey days of each individual).

Table 2.102 shows an artificial example how the MCR is calculated in the acute risk case. First the cumulative exposure per day is calculated by cumulating the exposure of each substance multiplied by the relative potency factors (RPF). Then, for each day, the cumulative exposure (in equivalents of the reference substance) is divided by the maximum exposure of a single substance on that day. The last column shows the MCR values within parenthesis the substance with the highest exposure. The MCR has a value of 1 or close to 1 for mixtures where the exposure is dominated by one substance (e.g. day 1, substance B). When all substances have approximately equal exposure (e.g. day 3) the MCR value is equal or close to the number of substances, here 4. Day 2 represents an intermediate case. The MCR suggest that for exposure days (or persons) with MCR values close to 1, the need for a cumulative risk assessment is low.

Table 2.102: Maximum Cumulative Ratios

	Substance A	Substance B	Substance C	Substance D	total exposure	ratio
day 1	0.01	0.99	0	0	1	1.01 (B)
day 2	0.1	0.2	0.3	0.4	1	2.50 (D)
day 3	0.25	0.25	0.24	0.26	1	3.99 (D)

In the example, all days have equal values for total exposure. For real data, total exposure will vary. It is obviously of interest to know if the MCR is high or low at those days (or individuals) where the total exposure is highest.

In Figure 2.25, French steatosis data (39 substances, 4079 persons) are used to calculate the chronic exposure matrix. For each indiviual the MCR is calculated and plotted against the total exposure. The different colors are used to identify the single substances with maximum exposure. From the original 39 substances, 10 different substances have the largest exposures. For the total exposure and MCR, the p5, p50 and p95 percentiles are indicated with the black line segments. The red line indicates the ratio with value 5. The dashed green lines indicate the p95 percentiles for the MCR value for different ranges of the total exposure.

The plot shows that MCR values with Imazalil as risk driving substance (purple) are predominantly found in the lower part of the plot for relatively high values of the total exposure. A second finding is that MCR values decline when total exposure increases. This implies that cumulative exposure for most individuals is driven by multiple substances. At the right site of the plot, individuals are found with high exposure. Because MCR values tend to be lower here, higher exposures are received from one predominant substance and not because many substances are above the average level. For those individuals a cumulative risk assessment has less value.

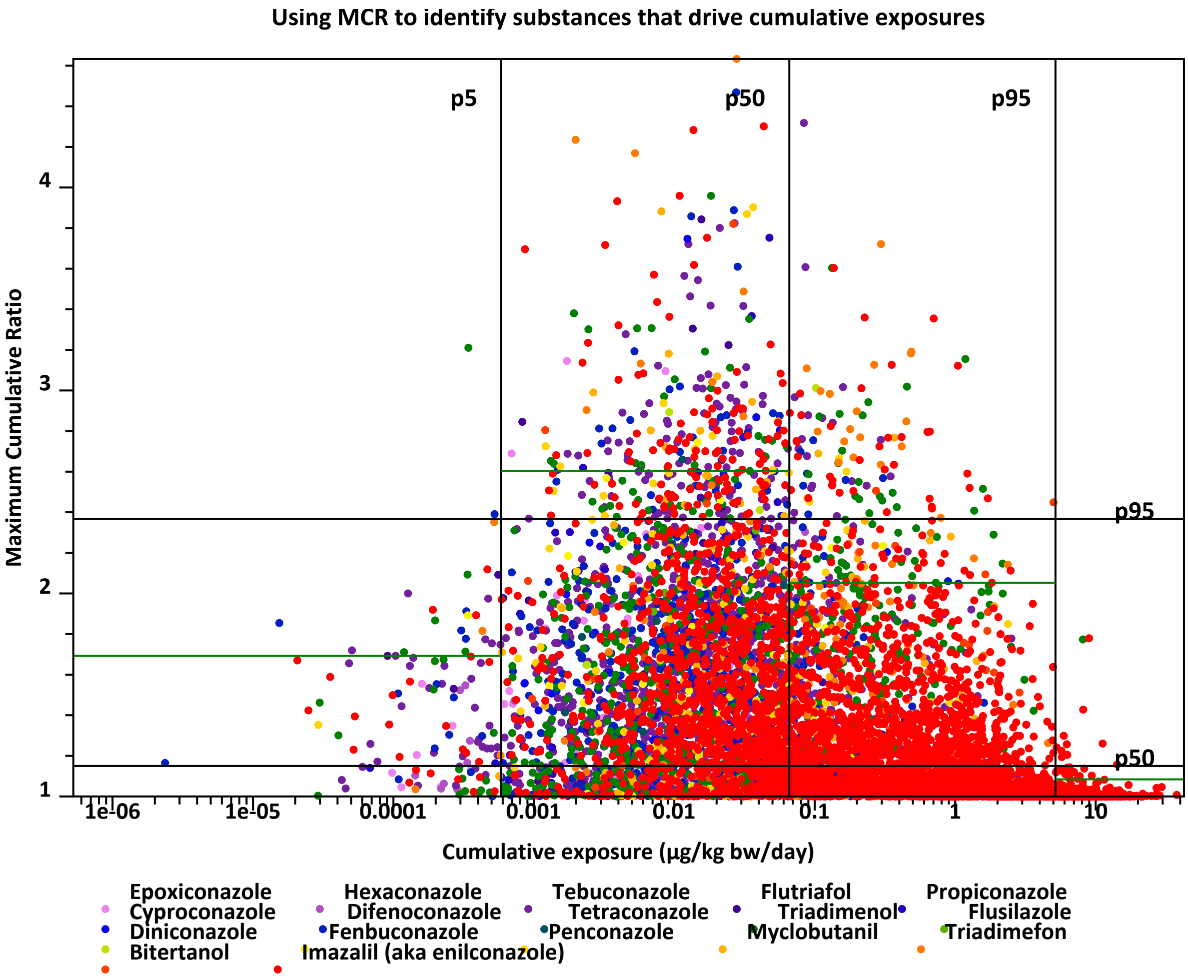


Figure 2.25: Maximum Cumulative Ratios vs total exposure

Because Figure 2.25 can be very dense, in Figure 2.26, 95% confidence regions representing bivariate lognormal distributions of MCR and total exposure are plotted. The latter figure facilitates interpretation of the first figure. Note that substances with just one or two observations cannot be plotted in this display (substances with 2 observations are represented by a line).

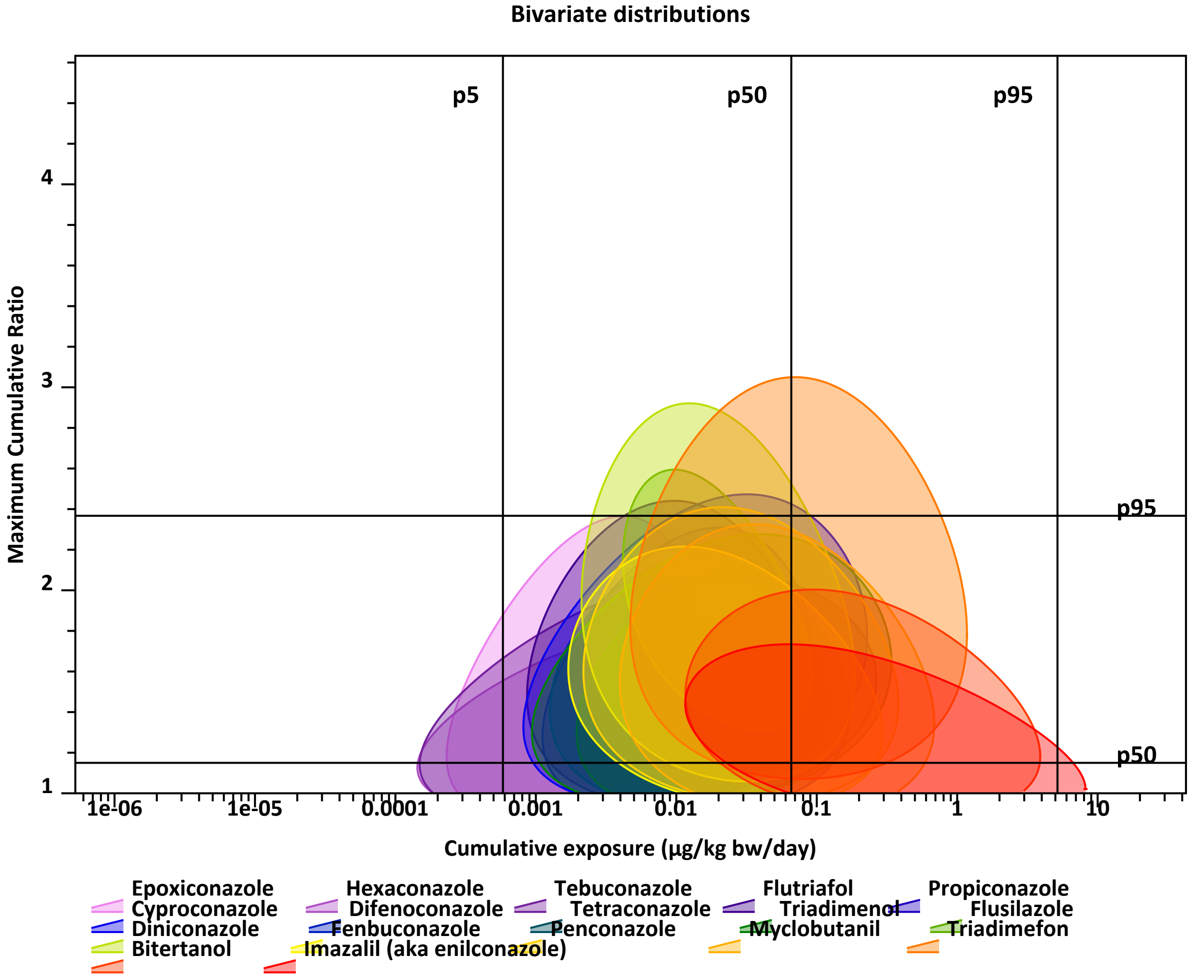


Figure 2.26: Bivariate distributions MCR vs total exposure

Exposures settings

Calculation settings

Table 2.103: Calculation settings for module Exposures.

Name	Description
Risk type	The type of exposure considered in the assessment; acute (short term) or chronic (long-term).
Cumulative	Specifies whether the assessment involves multiple substances and results should be cumulated over all substances.
Include dietary and non-dietary routes of exposure	Specifies whether the assessment involves both dietary and non-dietary (oral, inhalatory or dermal) routes of exposure.
Target level	Select to express hazard characterisations at external or internal exposure level.
Match non-dietary to dietary survey individuals	Specifies whether the individuals of one or more non-dietary surveys should be matched to individuals in the dietary survey according to the individual codes (idIndividual). If unchecked, nondietary exposures are randomly allocated to dietary survey individuals.
Match individuals of multiple non-dietary surveys	If checked, exposures from identical individuals in non-dietary surveys are aggregated to obtain the overall nondietary exposures. If unchecked, exposures from random individuals in all non-dietary surveys are aggregated.

Output settings

Table 2.104: Output settings for module Exposures.

Name	Description
Include drill-down on 9 individuals around specified percentile.	Specifies whether drilldown on 9 individuals is to be included in the output.
Summarize simulated data	Specifies whether a summary of the simulated consumptions and concentrations should be included in the output.
Store simulated individual day exposures	Store the simulated individual day exposures. If unchecked, no additional output will be generated. If checked, the output will contain an additional section with the simulated individual day exposures.
Show percentiles for	Give specific percentiles of exposure distribution (%), e.g. 50 90 95 97.5 99 (space separated).
Percentage for drilldown	Gives detailed output for nine individuals near this percentile of the exposure distribution.
Percentage for upper tail	Gives detailed output for this upper percentage of the exposure distribution.
Show % of population below level(s)	Exposure levels can be generated automatically or by explicit specification (Manual).
Exposure levels	Specify exposure levels for which to give the percentage of exposure below these levels, e.g. 1 10 50 100 200 500. Specify below whether these levels are absolute or relative to ARfD/ADI.
Exposure levels are	Specify whether exposure levels are absolute or percentages of ARfD/ADI.
Number of levels of covariable to predict exposure	Specify the number of levels, e.g. 20. The range of the covariable is divided by the number of levels: range = (max - min)/levels. For these covariable levels exposures are predicted.
Predict exposure at extra covariable levels	Specify specific prediction levels in addition to the automatically generated prediction levels (space separated).
Lower percentage for variability (%)	The default value of 25% may be overruled.
Upper percentage for variability (%)	The default value of 75% may be overruled.

Uncertainty settings

Table 2.105: Uncertainty settings for module Exposures.

Name	Description
Resample kinetic model parameter values	Specifies whether kinetic model parameter values are resampled.

Calculation of exposures

Exposures are computed by linking dietary and (if available) non-dietary individual/individual-day exposures and computing the (aggregated) internal exposures at the specified target compartment.

- *Exposures calculation*

Inputs used: *Dietary exposures Non-dietary exposures Active substances Relative potency factors Kinetic models*

Settings used

- *Calculation Settings*

2.4.5 Exposure mixtures

Exposure mixtures are mixtures of substances that contribute relatively much to the overall cumulative exposure (potential risk drivers).

This module has as primary entities: *Foods Substances Effects*

Exposure mixtures calculation

The most common model of cumulative risk assessment is to focus on substances that belong to the same common assessment groups (CAG). Substances in such a group belong to the same chemical family and may or may not have a similar mode of action. In assessing the risk, possible interactions between substances are often ignored and, moreover, little information is available about synergistic effects at low doses. More information is needed about the combined effects of such substances, but it is impractical to investigate all possible mixtures, and therefore instruments are needed to select the most relevant compounds for further investigation. This selection should not only be based on the hazard (highest relative potencies) but also on the exposure of the population to these substances. The potential risk of being exposed to chemicals in a mixture depends on the food consumption patterns of individuals in a population. A regular diet can contain hundreds of substances, so the number of combinations of compounds to which an individual in a population is exposed can be numerous. The exposures m,ixtures module can be used to identify the most relevant mixtures to which a population is exposed.

Exposure mixtures are identified using a quantitative approach: *sparse non-negative matrix underapproximation (SNMU)*. What mixtures predominantly determine the underlying patterns in the exposure matrix (compound x person (day))?

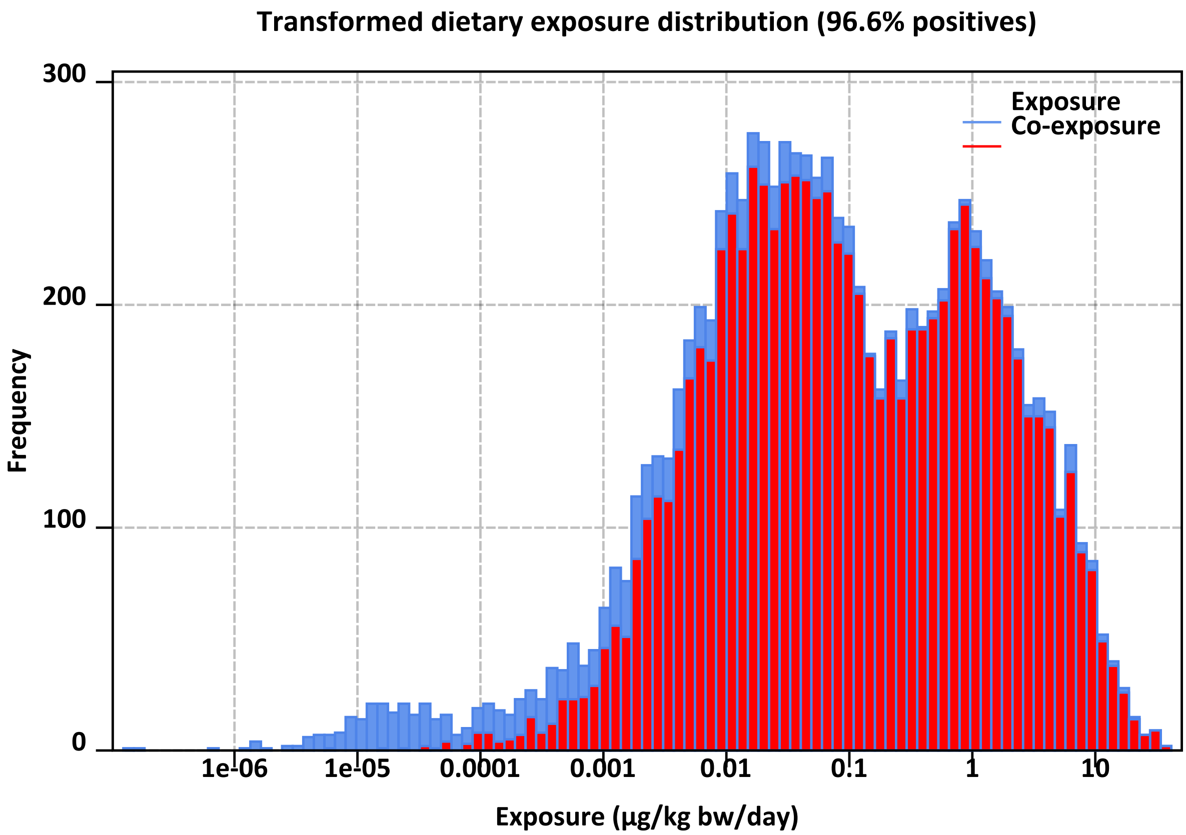


Figure 2.27: Example of co-exposure distribution (from >1 compound per individual-day, red) super-imposed on the total exposure distribution (blue).

Sparse nonnegative matrix underapproximation

Starting point to identify major mixtures of substances using exposure data was to use Non-negative Matrix Factorization (NMF). Non-negative Matrix Factorization was proposed by Lee & Seung [32] and Saul & Lee [39] and deals specifically with non-negative data that have excess zeros such as exposure data. Zetlaoui et al. [48], introduced this method in food risk assessment to define diet clusters.

The NMF method was then implemented by Béchaux et al. [11] in order to identify food consumption patterns and main mixtures of pesticides to which the French population was exposed using TDS exposure to 26 priority pesticides.

At the start of the Euromix project ideas had evolved: to obtain less components per mixture experiments were made using Sparse Nonnegative Matrix Factorization (SNMF) [23]. This method was found not to give stable solutions. Better results were obtained with Sparse Nonnegative Matrix Underapproximation (SNMU) [20]. This model also fits better to the problem situation because also the residual matrix after extracting some mixtures should be nonnegative. The SNMU method has been implemented in MCRA.

Indeed, NMF may be used to identify patterns or associations between substances in exposure data. NMF can be described as a method that finds a description of the data in a lower dimension. There are standard techniques available such as principal components analysis or factor analysis that do the same, but their lower rank representation is less suited because they contain negative values which makes interpretation hard and because of the modelling with a Gaussian random vectors which do not correctly deal with the excess of 0 and non-negative data. The NMF solution approximates a non-negative input matrix (i.c. exposure data) by two constrained non-negative matrices in a lower dimension such that the product of the two is as close as possible to the original input matrix. In Figure 2.28, the exposure matrix M with dimensions m (number of compounds) and n (number of intake days or persons) is approximated by matrix U and V with dimensions $(m \times k)$ and $(k \times n)$ respectively, where k represents the number of mixtures. Matrix U contains weights of the compounds per mixture, matrix V contains the coefficients of presence of mixtures in exposure per intake day or person. M is non-negative (zero or positive) and U and V are constraint to be non-negative. The minimization criterium is: $||M – UV||2$ such that $U \geq 0$ and $V \geq 0$.

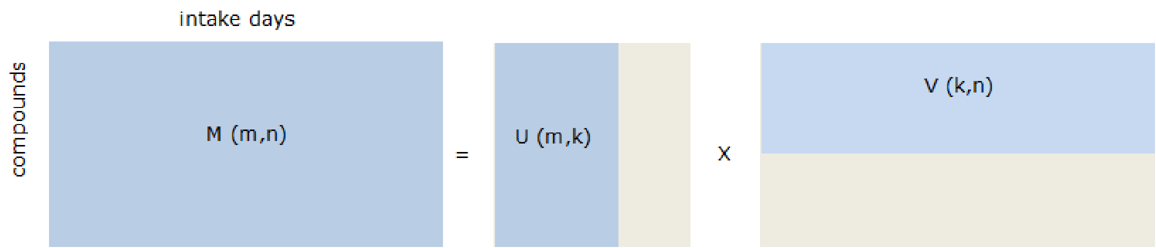


Figure 2.28: NMF approximation of exposure data

The solution found by NMF contains zeros, but mixtures still contain many components which complicates interpretability. Therefore, the Sparse Nonnegative Matrix Underapproximation (SNMU) [20] which also provide sparse results was investigated. The SNMU has also some nice features well adapted to the objective of the Euromix project: the solution is independent of the initialization and the algorithm is recursive. Consequently, the SNMU method which is based on the same decomposition process as the NMF was the one implemented in MCRA.

SNMU is initialized using an optimal nonnegative rank-one approximation using the power method (https://en.wikipedia.org/wiki/Power_iteration). This initialization is based on a singular value decomposition and results in general in a unique solution that is sparse. The SNMU algorithm is called recursive because after identifying the first optimal rank-one underapproximation $u_1 v_1$, the next rank-one factor is recovered by subtracting $u_1 v_1$ from M and applying the same factorization algorithm to the remainder $M - u_1 v_1$. The solution $u_1 v_1$ is called a rank-one underapproximation because an upper bound constraint is added to ensure that the remainder $M - u_1 v_1$ is non-negative. For Matlab code see: https://sites.google.com/site/nicolasgillis/code.

For each mixture, a percentage of explained variance is calculated. M is the exposure matrix with m rows (substances) and n columns (individuals or individual days) S_t is sum of squared elements of M:

$$S_t = ||M||^2 = \sum_{i,j}^{m,n} e_{i,j}^2$$

Apply SNMU on M, then for the first mixture:

- u is $m \times 1$ vector, contains weights of the mixture.
- v is $1 \times n$ vector, contains presence of mixture in exposure per individual.

Calculate residual matrix R:

$$R = M - uv$$

Calculate S_r, residual sum of squared elements of R:

$$S_r = ||R||^2 = \sum_{i,j}^{m,n} e_{i,j}^2$$

Percentage explained variance first mixture $(k = 1)$ is:

$$V_k = (1 - S_r)/S_t) \cdot 100$$

For the second mixture:

1. continue with residual matrix R (replace M by R),
2. estimate u and v,
3. calculate new residual matrix R
4. calculate S_r, residual sum of squared elements of R

Percentage explained variance second mixture $(k = 2)$ is:

$$V_k = (1 - S_r)/S_t) \cdot 100 - \sum_{l=1}^{k-1} V_l$$

The last term is de percentage explained variance of the first mixture. Continue with the third mixture etc….

Exposure matrix

Basically, exposure is calculated as consumption x concentration. By summing the exposures from the different foods for each compound per person day separately, the exposure matrix for mixture selection is estimated:

$$y_{ijc} = \frac{\sum_{k=1}^{p} x_{ijk} c_{ijkc}}{bw_i}$$

where y_{ijc} is the exposure to compound c by individual i on day j (in microgram substance per kg body weight), x_{ijk} is the consumption by individual i on day j of food k (in g), c_{ijkc} is the concentration of compound c in food k eaten by individual i on day j (in mg/kg), and bw_i is the body weight of individual i (in kg). Finally, p is the number of foods accounted for in the model. More precisely, for an acute or short term risk assessment, the exposure matrix summarises the y_{ijc} with columns denoting the individual-days (ij) and rows denoting the compounds (c). Each cell represents the sum of the exposures for a compound on that particular day. For a chronic or long term risk assessment, the exposure matrix is constructed as the sum of all exposures for a particular compound averaged over the total number of intake days of an individual (compounds x persons). So each row represents a compound and a column an individual. Each cell represents the observed individual mean exposure for a compound for that individual. Note that in the exposure calculation, the concentration is the average of all residue values of a compound.

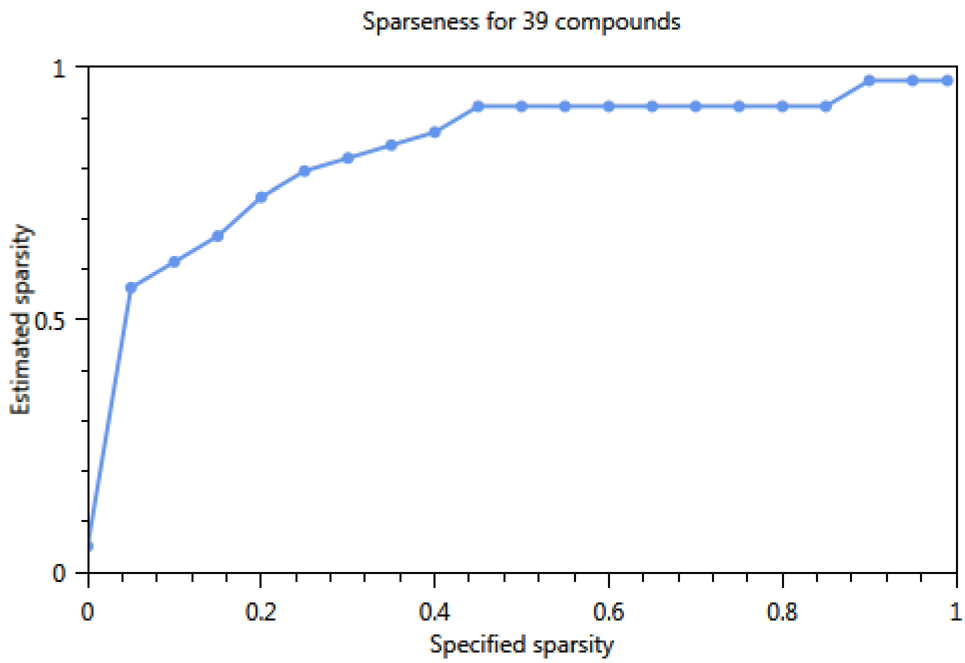
When relative potency factors (RPF) are available then exposures are multiplied by the RPF and thus exposures to the different substances are on the same and comparable scale (toxicological scale). In this case, the selection of the mixture is risk-based. In some cases, RPFs may not be available. In this case exposure to different substances, even in the same unit, may lead to very different effect. To give all compounds an equal weight a priori and avoid scaling effect, a normalization of the data can be applied as done in Béchaux et al. [11]. In this case, all compounds are assigned equal mean and variance, and the selection of the mixtures will work on patterns of correlation only. Then mixture selection is not risk-based anymore but, what could be called, co-exposure-based.

Finally, in the mixture selection module of MCRA, the SNMU expects RPF data for a risk-based selection. If not available, the user should provide alternative RPF data, e.g. values 1 for a purely exposure-based selection, or lower-tier estimates (e.g. a maximum value on RPF thought possible).

Mechanisms to influence sparsity

Two mechanisms to influence sparsity are available. The SNMU algorithm incorporates a sparsity parameter and by increasing the value, the final mixtures will be more sparse (sparsity close to 0: not sparse; sparsity close to 1: sparse). The other approach is by using a subset of the exposure matrix based on a cut-off value for the MCR. High ratios correspond to high co-exposure, so it is reasonable to focus on these (person) days and not include days where exposure is received from a single compound (ratio close to 1). To illustrate the combined use of MCR and the sparsity parameter, the French steatosis data (39 compounds, 4079 persons) are used and person days with a ratio > 5 (see Figure 2.25) are selected yielding a subset of 139 records.

In Figure 2.29, the effect of using a cut-off level is immediately clear. The number of compounds of the first mixture is 17 whereas in the unselected case only 4 compounds were found The three plots show the influence of increasing the sparsity parameter from 0 to 1 on the number of compounds in the mixture. For values close to 0, the mixture contains 17 compounds. For values > 0.4 the number of compounds in the mixture drops to 3.



In Figure 2.30 and Figure 2.31 the sparsity parameter is set to 0.0001 (not sparse) and 0.4 (sparse), respectively. This leads to mixtures containing different number of substances.

As mentioned before, one of the nice features of the SNMU algorithm is its recursive character which results in identical mixtures. In Figure 2.32 and Figure 2.33, two U matrices are visualized. In the upper plot 4 mixtures are estimated, in the lower plot the solution for 5 mixtures is shown. Because of the ordering the plots look different, but a closer inspection of the first 4 mixtures of each solution shows that they are the same. In both figures, mixture 1 contains Imazalil, Thiacloprid, Deltamethrin (cis-deltamethrin) and Deltamethrin including other mixture.

Exposure mixtures settings

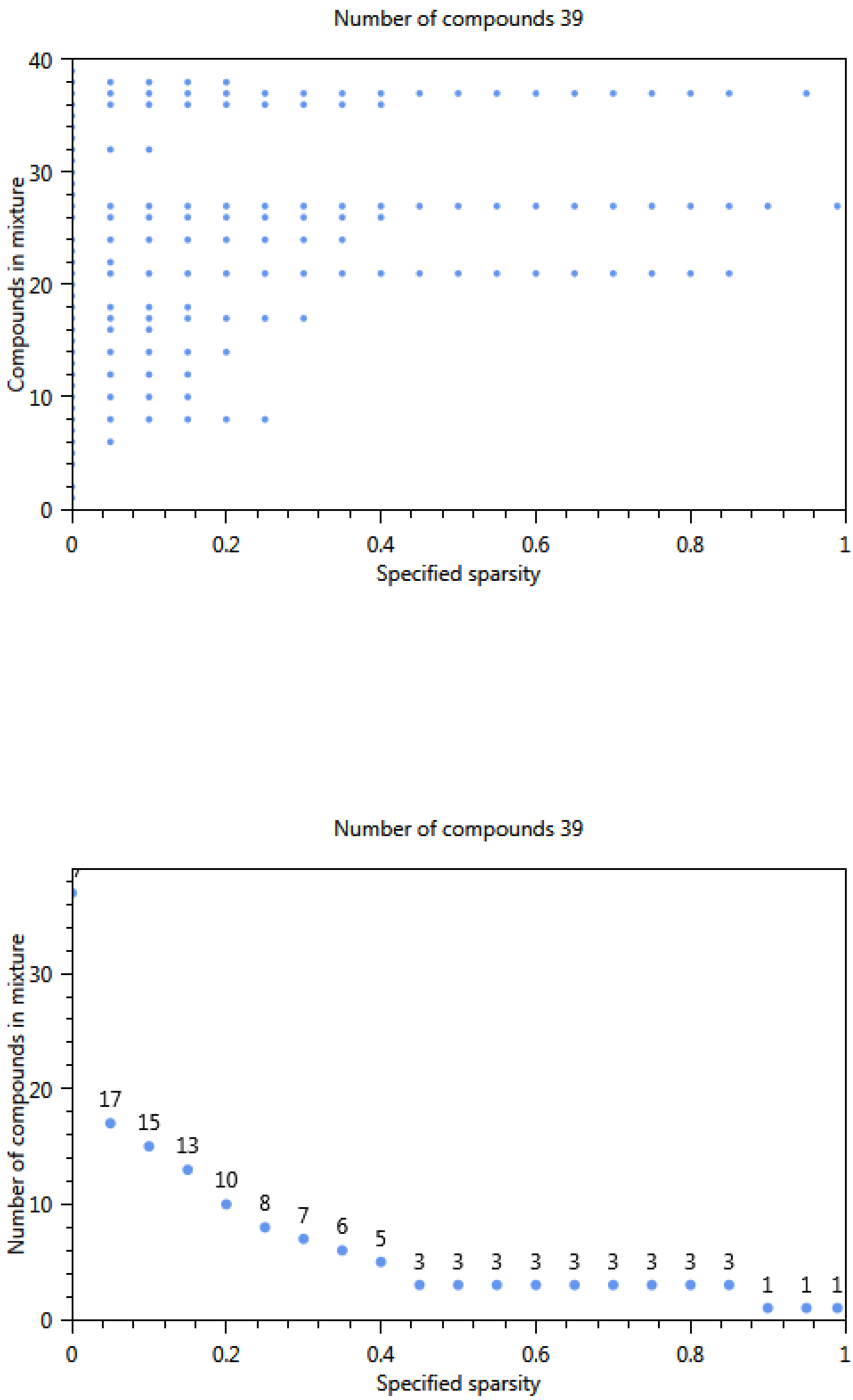


Figure 2.29: Influence of the specified sparsity parameter on the realized sparsity, n = 139

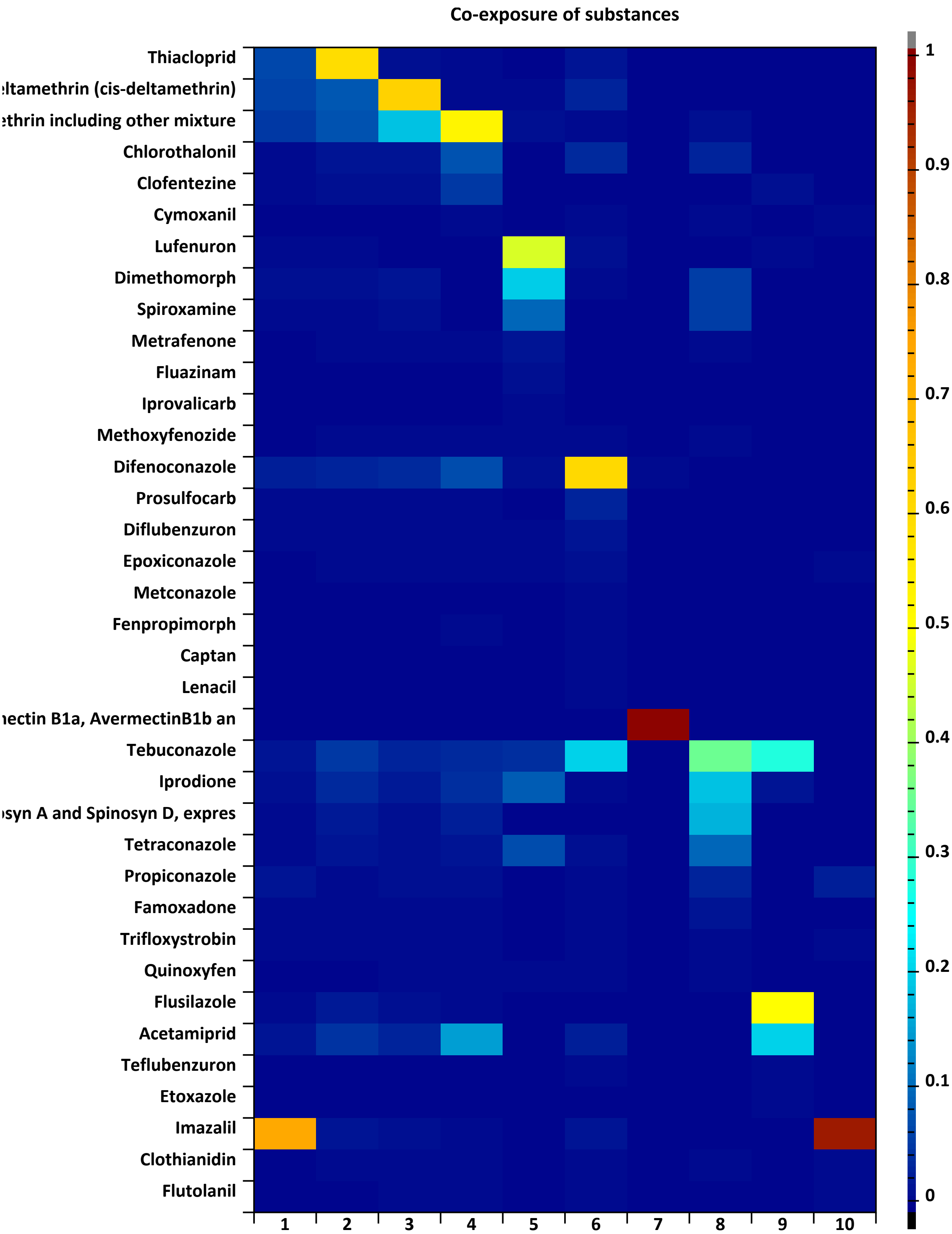


Figure 2.30: Heatmap for a solutions with 10 mixtures. The sparsity is set to 0 (not sparse). Each mixture contains many substances (see also Figure 2.31).

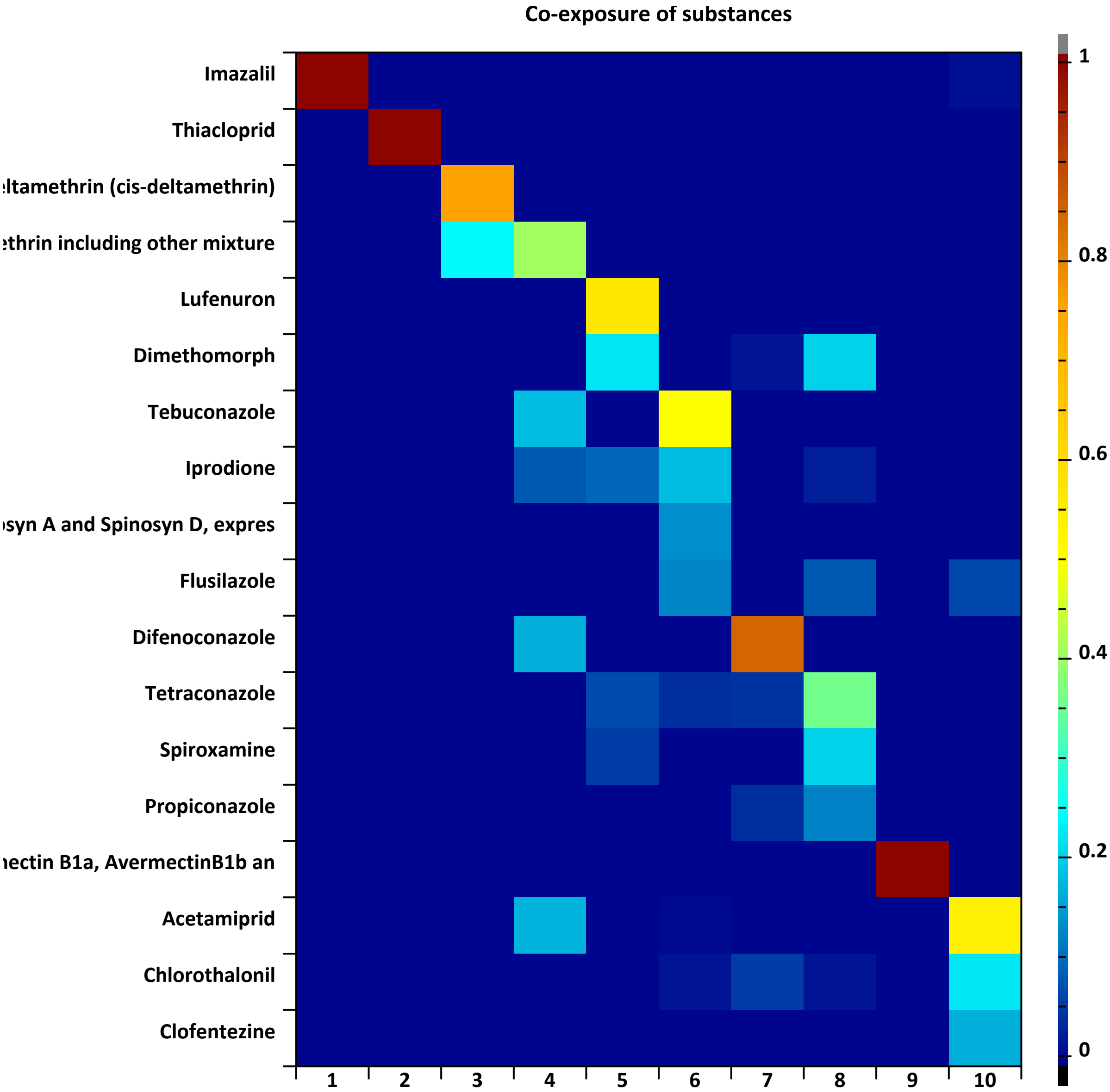


Figure 2.31: Heatmap for a solutions with 10 mixtures. The sparsity is set to 0.4 (sparse). Mixtures contain less substances compared to Figure 2.30.

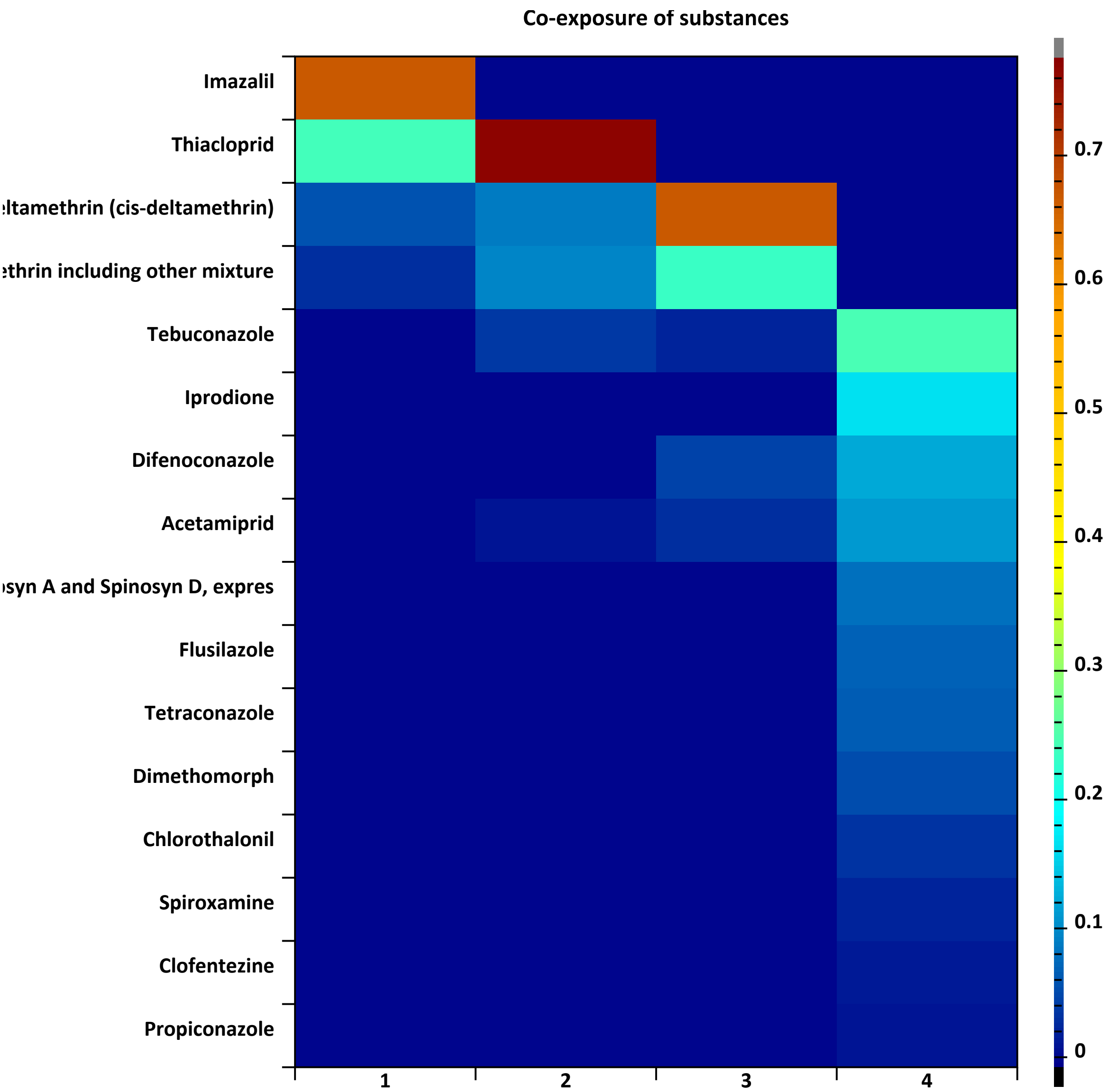


Figure 2.32: Heatmap for solution with 4 mixtures. The first 4 mixtures in Figure 2.32 and Figure 2.33 are identical.

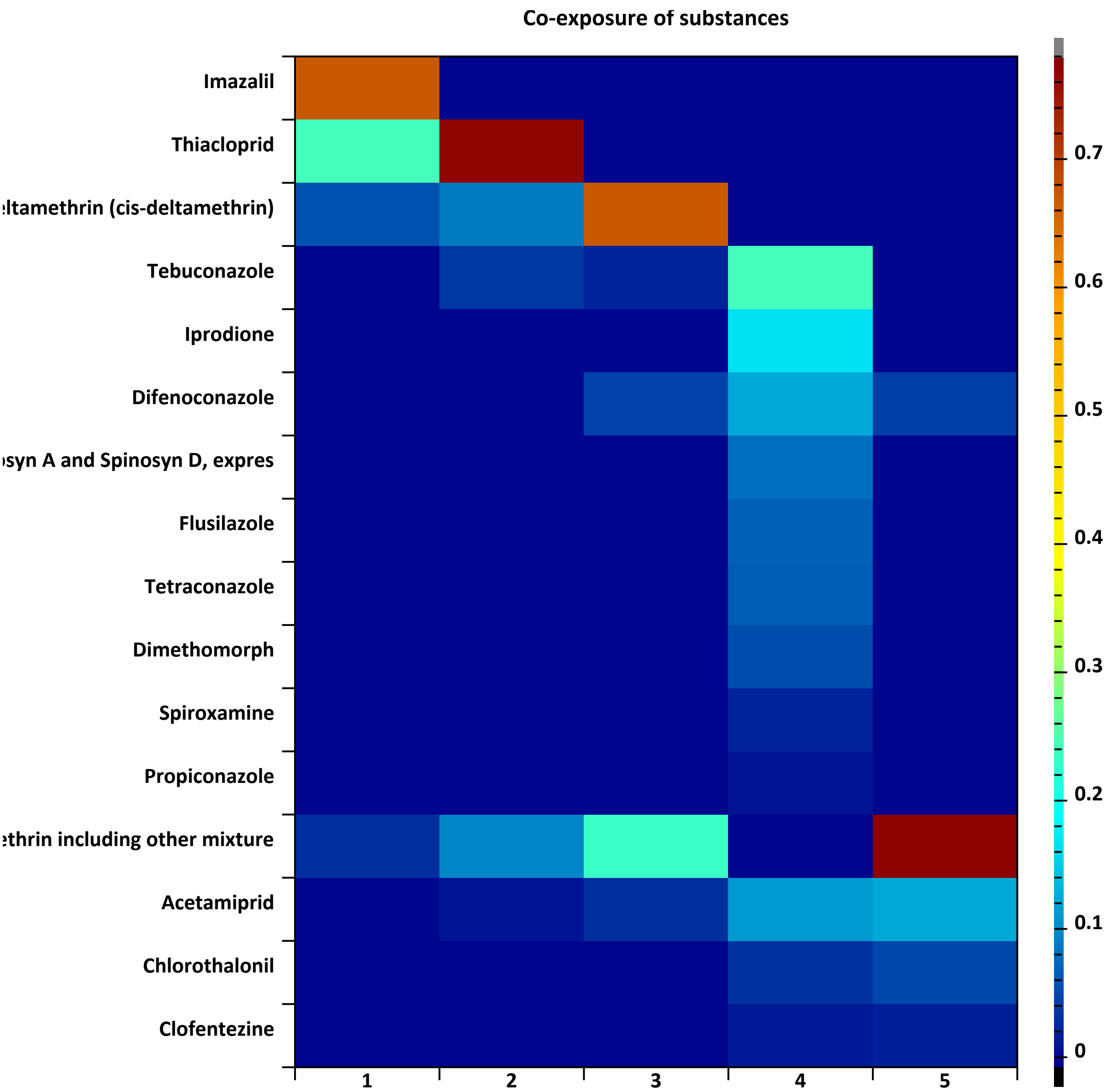


Figure 2.33: Heatmap for solution with 5 mixtures. The first 4 mixtures in Figure 2.32 and Figure 2.33 are identical.

Calculation settings

Table 2.106: Calculation settings for module Exposure mixtures.

Name	Description
Sparseness constraints	Sparseness parameter. Should be a value between 0 (not sparse, many substances) and 1 (sparse, few substances).
Number of mixtures	The number of mixtures.
Number of iterations	Number of iterations, e.g. 1000.
Seed for pseudo-random number generator.	Random seed for initializing matrix W and H.
Exposures are	Exposures are risk based (expressed in equivalents of the reference substance) or standardized.
Convergence criterium	Convergence criterium for factorization algorithm.
Cutoff for ratio total exposure/ maximum	For selection of individual(day) exposures specify cutoff for ratio total exposure/ maximum.
Cutoff percentage (%) for total exposure	For selection of individual(day) exposures specify cutoff percentage (%) for total exposure.

Calculation of exposure mixtures

A multivariate decomposition method, sparse non-negative matrix underestimation (SNMU), is applied to the matrix of exposures per substance and per individual (chronic) or individual-day (acute) to find substance combinations that contribute most to the cumulative exposure.

- *Exposure mixtures calculation*

Inputs used: *Exposures*

Settings used

- *Calculation Settings*

2.4.6 Food conversions

Food conversions relate foods-as-eaten, as found in the consumption data, to foods-as-measured, which are the foods for which concentration data are available. A food-as-eaten can be linked to one, or multiple food-as-measured using various conversion steps (e.g., using food recipes to translate a composite food to its ingredients, or using processing information to relate a processed food to its unprocessed form). There are several types of conversion steps, and a conversion path may comprise multiple conversion steps between a food-as-eaten and a food-as-measured.

This module has as primary entities: *Foods Substances*

Output of this module is used by: *Consumptions by food as measured*

Food conversions calculation

Food conversions are computed using a recursive search algorithm to link foods-as-eaten to foods-as-measured, possibly through intermediate conversion steps. For instance, if (unpeeled) apple and grapes are the foods-as-measured, the food-as-eaten apple pie contains peeled apple and raisins, peeled apple is linked to unpeeled apple, and raisins are dried grapes. Hence, for this apple pie, there are two conversions, one to apple and one to grapes, each with its own conversion path of intermediate conversion steps.

For each food-as-eaten, the food conversion algorithm recurstively builds up the conversion paths using the the following 7-step procedure:

1. **Check food-as-measured (step 1):** Check whether the current food is considered a food-as-measured, i.e., it is a food for which substance concentration measurements are considered to be available. If successful, a food-as-measured has been found, and the current search stops.

2. **Find processing link (step 2):** Check whether the current food can be considered to be a processed variant (e.g., cooked or peeled) of another food.

   a. **Match processing factor (step 2a):** try to find the code in the processing factors table.

   b. **Processing link wildcard match (step 2b):** try to find a wildcard match in the processing table. Wildcard match codes consist of an initial string (startcode, may be empty), an asterisk (*), and possibly a processing part (-processingtype). * may be any string endcode (not containing a -) such that code equals startcodeendcode or startcodeendcode-processingtype.

   i. If code contains a processing part (-processingtype), then the wildcard match code should also end with -processingtype. Convert to the code specified in the field foodunprocessed, where endcode is substituted for any * in the new code.

   ii. If code contains no processing part, then the wildcard match code should also contain no processing part. Convert to the code specified in the field foodunprocessed, where endcode is substituted for any * in the new code.

   If successful, try to find find food translation information in the food recipes data to correct for weight reduction or increase. Then, restart at step 1 with the new code of the unprocessed food.

3. **Food translation link (step 3):** Check whether the current food translates to one or more other foods through composition or read-across.

   a. **Food recipe link (step 3a):** Try to find food translations for the current food (i.e., the ingredients of a composite food). This may result in one or more food codes for ingredients, and the iterative algorithm will proceed with each of the ingredient food codes in turn.

   b. **TDS food sample composition link (step 3b):** Try to find the code in the TDSFoodSampleCompositions table (column idFood), a default translation proportion of 100% is assumed. The iterative algorithm will proceed with a TDS food (column idTDSFood) sample.

   c. **Read-across link (step 3c):** Try to find a food extrapolation rule for the current food, a default translation proportion of 100% for 'idToFood' is assumed.

If successful, restart at step 1 with each of the new codes of the ingredient foods, TDS foods or Read Across foods.

4. **Subtype link (step 4):** try to find subtype codes, e.g. 'xxx$*' in the MarketShares table. In general, marketshares should sum to 100%. Foods with marketshares not summing to 100% are ignored in the analysis unless the checkbox 'Allow marketshares not summing to 100%' is checked. This step is optional, see advanced settings if you want to use this. If successful, restart at step 1 with each of the new codes of the subtype foods.

5. **Supertype link (step 5):** try to find supertypes, e.g. 'xxx$yyy' is converted to 'xxx'. This step is optional, see advanced settings if you want to use this. If successful, restart at step 1 with the new code of the supertype food.

6. **Default processing factor (step 6):** remove processing part (-xxx) of the code. If successful, restart at step 1 with the new code without processing part.

7. **Maximum residue limit (step 7):** try to find the code in the MaximumResidueLimits table. If successful, the current search stops. If not successful, then stop anyway and the search is marked as failed food conversion.

Food conversion settings

Calculation settings

Table 2.107: Calculation settings for module Food conversions.

Name	Description
Include foods with only non-detect measurements	Specifies whether foods with only non-detect measurements are part of the exposure assessment (default yes).
Include substances with only non-detect measurements	Specifies whether substances with only non-detect measurements are part of the exposure assessment (default yes).
Include substances without measurements	Specifies whether substances without any measurements should be included.
Step 2: allow conversion using processing info	Try to find the code in the processing table (step 2a). Processing link wildcard match: try to find a wildcard match in the processing table (step 2b) e.g. code FP026 matches FP* in column FoodProcessed where '*' is used as a wildcard match for '026'. Try to find the code in the FoodTranslation table (step 3a) to account for weight reduction/increase. If unchecked, processing table is ignored, default is 'Use processing info'. If successful, restart at step 1.
Step 3a: allow conversion using food translations	Step 3a: Try to find food translations for the current food (i.e., the ingredients of a composite food). This may result in one or more food codes for ingredients, and the iterative algorithm will proceed with each of the ingredient food codes in turn.
Step 3b: allow conversion using TDS food sample compositions	Step 3b: Try to find the code in the TDS food sample compositions table (idFood), a default translation proportion of 100% is assumed. The iterative algorithm will proceed with a TDS food (column idTDSFood) sample.
Step 3c: allow conversion using food extrapolations	Try to find read accross codes. If unchecked, read across table is ignored, default is 'Use read across info'. E.g. for pineapple no measurements are found but by specifying that pineapple is converted to FruitMix (with a default proportion of 100%), the TDS sample concentration value of FruitMix will be used for pineapple (as-eaten or as ingredient). If successful, restart at step 1.
Step 4: allow conversion using market shares	Try to find subtype codes, e.g. 'xxx$*' in the market shares table.
Allow marketshares not summing to 100%	Specify whether to rescale market share percentages that do not sum to 100%. If true, then foods with marketshares not summing to 100% are allowed. If not, then these foods are ignored in the analysis.
Step 5: allow conversion to supertypes	Try to find supertypes, e.g. 'xxx$yyy' is converted to 'xxx' (optional, check box if you want to use this). If checked, allows for linkage of consumed foods coded at a lower hierarchical level to foods with measured concentrations at a higher hierarchical level e.g. consumed is Apple (code PF$Apple) –> measured is Pome Fruit (code PF). Note: food codes are split on '$'. Measurements of substances on food are available at a less detailed food coding level than consumption data. MCRA allows to use the concentration data of a supertype for all underlying food codes. If successful, restart at step 1.
Step 6: allow conversion using default processing factors	Remove processing part. If unchecked, no default processing factors are assumed, default is 'Use default processing factors'. If successful, restart at step 1.
Include foods without occurence data but with specified maximum residue limits	Include foods without concentration data but for which for which concentration limits such as MRLs are defined (default: no).

Calculation of food conversions

Food conversions are computed recursively, starting with a food-as-eaten and following a path to ingredients (food recipes), unprocessed foods (processing), super/sup-type foods, etc. until either arriving at a food-as-measured (commonly the raw primary commodities) or concluding that the path does not lead to a food-as-measured.

- *Food conversions calculation*

Inputs used: *Consumptions Foods as measured Processing factors Food recipes Market shares Food extrapolations Total diet study sample compositions Active substances*

Settings used

- *Calculation Settings*

2.4.7 Human monitoring analysis

Human monitoring analysis compares observed human monitoring data with predictions made for the same population of individuals from dietary survey data, concentration data and (optionally) non-dietary exposure data.

This module has as primary entities: *Populations Substances*

Human monitoring analysis calculation

Human monitoring analysis computes internal substance concentration estimates based on provided human monitoring data. These estimates are specified at the level of long term average concentrations for individuals in case of chronic assessments, or the average concentrations for individual-days in vase of acute assessments. The internal concentrations are computed independently for each substance, compartment, and sampling type.

The main steps for computing the human monitoring concentration estimates are:

1. Imputation of non-detects.
2. Imputation of missing values.
3. Calculation of individual concentrations (chronic) or individual day concentrations (acute).
4. Comparison of monitoring versus modelled exposures by substance and compartment (optional).

Imputation of non-detects

Similar to *concentrations measurements in food*, human monitoring measurements can also contain measurements below the limit of reporting and similar to *concentrations modelling in foods*, human monitoring analysis needs to address these non-detects and replace them with imputed concentration values. For this, two approaches are available:

1. Replace non-detects by zero.
2. Replace non-detects by a factor times LOR, in which the factor is set between zero and one.

Imputation of missing values

Concentration measurements may be missing. The following imputation methods are available for imputation of missing values:

1. Replace missing values by zero.
2. For each substance, sampling type, and compartment, replace missing values by a random other sample of this substance, sampling type, and compartment.

Note: For the second imputation method, more refined methods could be useful as well. E.g., when for a given day multiple samples are available, of which one is missing, then it may alternatively be sensible to leave this sample out when computing an average exposure. Also, when samples have been taken at different times during the day, it may be better to impute missing records using samples approximately from the same time-slot.

Calculation of acute human monitoring concentrations

For acute assessments, the monitoring concentrations are computed for each substance, compartment, and sampling type as average individual-day concentrations. That is, for a given substance, compartment, and sampling type, the acute individual-day concentration c_{ij} for individual i on day j is:

$$c_{ij} = \frac{\sum_{k=1}^{n_{\texttt{samples}}} c_{ijk} \cdot sg_{ijk}}{n_{\texttt{samples}}},$$

where $n_{\texttt{samples}}$ is the number of samples available for individual i on day j, and c_{ijk} and sg_{ijk} denote the concentration and specific gravity, respectively, of the k-th sample of the individual day.

Note: Note that currently, the acute concentrations are computed as mean concentrations when multiple samples are available for one day. In acute scenarios, one may be more interested in peak concentrations. I.e., the highest concentration of a day.

Calculation of chronic human monitoring concentrations

Note: The implementation for chronic is not yet available. Below is a description of the forseen implementation.

For chronic assessments, the monitoring concentrations are computed as the avarage monitoring concentrations of multiple individual-days for each substance, compartment, and sampling type. That is, for a given substance, compartment, and sampling type, the chronic concentration c_i for individual i is:

$$c_i = \frac{\sum_{j=1}^{n_{\texttt{days}}} c_{ij}}{n_{\texttt{days}}},$$

where $n_{\texttt{days}}$ is the number of days that individual i was monitored, and c_{ij} denotes the average monitoring concentration of individual i on day j.

Compare measured and modelled exposures

An optional step of the human monitoring analysis is to compare the monitoring concentrations with *modelled exposures* that were obtained from dietary (and optionally non-dietary) exposure assessments. This comparison may provide insight in the coherence between modelled exposures and the measured reality. A requirement is that both monitoring data and dietary/non-dietary use data is available for the same individuals or individual-days. An example of a graphical output of these comparison is given in Figure 2.34.

Human monitoring analysis settings

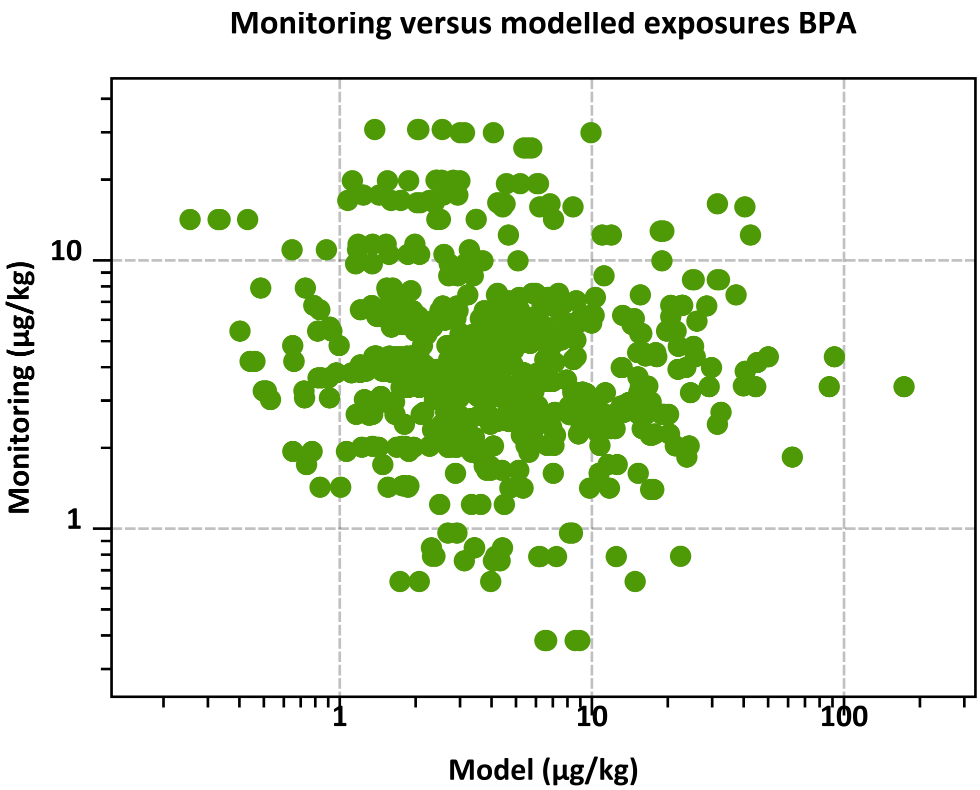


Figure 2.34: Measured exposures from monitoring versus modelled exposures

Calculation settings

Table 2.108: Calculation settings for module Human monitoring analysis.

Name	Description
Non-detects handling method	Method for dealing with non-detects samples in human monitoring data.
Fraction of LOR	Factor for replacing non-detects with factor times LOR.
Missing value imputation method	Imputation method for missing values.
Correlate monitoring with modelled exposures	Correlate monitoring with modelled exposures.

Calculation of human monitoring analysis

Human monitoring analysis calculations comprise two parts. The first part is to compute estimates of the human monitoring concentrations based on the human monitoring data. The second part, which is optional, is to relate the human monitoring concentrations to modelled concentrations from exposure assessments.

- *Human monitoring analysis calculation*

Inputs used: *Human monitoring data Exposures*

Settings used

- *Calculation Settings*

2.4.8 Human monitoring data

Human monitoring data quantify substance concentrations found in humans collected in human monitoring surveys.

This module has as primary entities: *Substances*

Output of this module is used by: *Human monitoring analysis*

Human monitoring data data formats

Data are provided on the survey, the individuals in the survey, the samples taken, the analyses performed, the analytical methods used, the properties for substances analysed, and the concentrations found.

Data are provided in the following relational tables:

- Human monitoring surveys
- Human monitoring individuals
- Human monitoring samples
- Human monitoring sample analyses
- Sample concentrations
- Analytical methods
- Analytical method properties for substances

Human monitoring samples

Suggested table definitions for human monitoring data.

Human monitoring surveys

Contains the survey definitions.

Table 2.109: Table definition for HumanMonitoringSurveys.

Name	Type	Description	Aliases	Required
idSurvey	AlphaNumeric(50)	Unique identification code of the survey.	idSurvey	Yes
Name	AlphaNumeric(100)	Name of the survey.	Name	No
Description	AlphaNumeric(200)	Description of the survey.	Description	No
Location	AlphaNumeric(50)	The location or country where survey is held. It is recommended to use ISO Alpha-2 country codes.	Location, Country	No
BodyWeight-Unit	AlphaNumeric(50)	The unit of bodyweight of the individuals of the survey: kg (default) or g.	BodyWeight-Unit, UnitBody-Weight, WeightIn	No
AgeUnit	AlphaNumeric(50)	The unit of age, i.e., year or month.	UnitAge, agein, AgeUnit	No
StartDate	DateTime	The starting date of the survey.	StartDate	No
EndDate	DateTime	The end date of the survey.	EndDate	No
NumberOf-SurveyDays	Integer	The number of days each individual participated in the survey.	NumberOf-SurveyDays, NDaysInSurvey	Yes
idPopulation	AlphaNumeric(50)	Unique identification code of the population.	IdPopulation, PopulationId	No

Table aliases: HumanMonitoringSurveys, HumanMonitoringSurvey, RawHumanMonitoringSurveys.

Human monitoring individuals

The individuals of a survey are recorded in the individuals table.

Table 2.110: Table definition for HumanMonitoringIndividuals.

Name	Type	Description	Aliases	Required
idIndividual	AlphaNumeric(50)	Unique identification code of the individual.	idIndividual, IndividualId, Individual, Id	Yes
idSurvey	AlphaNumeric(50)	The identification code / short name of survey.	idSurvey	Yes
BodyWeight	Numeric	The body weight of the individual.	BodyWeight, Weight	Yes
Sampling-Weight	Numeric	The sampling weight for an individual (default = 1).	SamplingWeight	No
NumberOf-DaysInSurvey	Integer	The number of days the individual participated in the survey.	NumberOf-SurveyDays, NumberOfDays-InSurvey, DaysInSurvey, NDaysInSurvey	No
Age	Numeric	The age of the individual.	Age	No
Gender	AlphaNumeric(50)	The gender of the individual. Recommendation: use the codes Male/Female for coding the gender.	Gender	No
Other inidividual properties		Other individual properties can be added just like the fields age and gender. These properties are automatically parsed as co-factors or co-variables.		No

Table aliases: HumanMonitoringIndividuals, HumanMonitoringIndividual, RawHumanMonitoringIndividuals.

Human monitoring samples

Contains the samples taken during the study.

Table 2.111: Table definition for HumanMonitoringSamples.

Name	Type	Description	Aliases	Required
idSample	AlphaNumeric(50)	Unique identification code of the monitoring sample.	idSample, Sample	Yes
idIndividual	AlphaNumeric(50)	Unique identification code of the individual.	idIndividual, IndividualId, Individual, Id	Yes
DateSampling	DateTime(50)	Date of sampling.	DateSampling, DateOf-Sampling, SamplingDate	No
DayOfSurvey	AlphaNumeric(50)	Identification code of the day of measurement.	Day, idDay, DayId, DayOfSurvey	Yes
TimeOf-Sampling	AlphaNumeric(50)	Identification code of the time of sampling.	TimeOf-Sampling, SamplingTime, TimeSampling	No
SampleType	AlphaNumeric(50)	Type of sample (e.g., pooled, 24h urine, spot urine, serum from blood, etc.).	SampleType, SamplingType	No
Compartment	AlphaNumeric(50)	If applicable, the measured compartment of the human body (e.g., blood, urine). When specified, the measurements are considered at the level of internal doses.	Compartment	No
ExposureRoute	AlphaNumeric(50)	If applicable, the measured exposure route, e.g., dermal (in case of skin wipes). When specified, the measurements are considered at the level of external doses.	ExposureRoute	No
SpecificGravity	Numeric	Specific gravity of the measured person for this particular sample.	SpecificGrafity, SpecificGravity	No
SpecificGravity-Correction-Factor	Numeric	Specific gravity of the measured person for this particular sample.	SpecificGravity-Correction-Factor	No

Table aliases: HumanMonitoringSamples, HumanMonitoringSample, RawHumanMonitoringSamples.

Human monitoring sample analyses

Contains the measurements of the samples of human monitoring studies.

Table 2.112: Table definition for HumanMonitoringSampleAnalyses.

Name	Type	Description	Aliases	Required
idSample-Analysis	AlphaNumeric(50)	Unique identification code of the sample analysis.	idSample-Analysis, SampleAnalysis	Yes
idSample	AlphaNumeric(50)	Code of the measured monitoring sample.	idSample, Sample	Yes
idAnalytical-Method	AlphaNumeric(50)	The code of method of analysis.	idAnalytical-Method, Analytical-MethodName, Analytical-MethodId	Yes
AnalysisDate	AlphaNumeric(50)	Date of analysis.	AnalysisDate, DateAnalysis	No
Substance concentration(s)	AlphaNumeric(100)	One or more columns with the measured concentrations of the substances in the unit as specified by the analytical method. The column name(s) should match the substance codes of the substances measured by the analytical methods. Empty fields for substances that should have been measured by the analytical method are considered to be non-detects with measurement values below LOR.		Yes

Table aliases: HumanMonitoringSampleAnalyses, HumanMonitoringSampleAnalysis, RawHumanMonitoringSampleAnalyses.

Sample concentrations

The positive concentration values for substances from analysis in the unit specified in table human monitoring sample analyses. Non-detects (i.e. results 'less than LOR') are not included, their existence can be inferred from the tables AnalysisSamples and AnalyticalMethodSubstances, and the LOR itself from the analytical method.

Table 2.113: Table definition for HumanMonitoringSampleConcentrations.

Name	Type	Description	Aliases	Required
idAnalysis-Sample	AlphaNumeric(50)	The identification number of the analysed sample.	idAnalysis-Sample, AnalysisSample-Id	Yes
idSubstance	AlphaNumeric(50)	The substance code.	idSubstance, SubstanceId, Substance, idCompound, CompoundId, Compound	Yes
Concentration	Numeric	The measured concentration.	Concentration	Yes

Table aliases: HumanMonitoringSampleConcentrations, HumanMonitoringSampleConcentration, RawHumanMonitoringSampleConcentrations.

Analytical methods

The analytical methods used for analyzing the samples are recorded in the analytical methods table. Each analytical method should have a unique identification code (idAnalyticalMethod). The description field may be used for a more detailed description of the analytical method. The records of this table should be linked to one or more analytical method substance properties table, which record the substances that are measured by this method (and their limits of reporting).

Table 2.114: Table definition for AnalyticalMethods.

Name	Type	Description	Aliases	Required
idAnalyticalMethod	AlphaNumeric(50)	The code for the method of analysis.	idAnalyticalMethod, AnalyticalMethodId, AnalyticalMethodName, Id	Yes
Description	AlphaNumeric(200)	Additional description of method of analysis.	Description	No

Table aliases: AnalyticalMethod, AnalyticalMethods, RawAnalyticalMethods.

Analytical method properties for substances

Table 2.115: Table definition for AnalyticalMethodCompounds.

Name	Type	Description	Aliases	Required
idAnalytical-Method	AlphaNumeric(50)	The code of method of analysis.	idAnalytical-Method, Analytical-MethodName, Analytical-MethodId	Yes
idSubstance	AlphaNumeric(50)	The substance code.	idSubstance, SubstanceId, Substance, idCompound, CompoundId, Compound	Yes
LOR	Numeric	The limit of reporting (LOR). In MCRA, LOR just means the limit below which no quantitative result has been reported. Depending on a laboratory's format of reporting, LOR may be a limit of detection (LOD), a limit of quantification (LOQ) or another limit.	LOR	Yes
Concentration-Unit	AlphaNumeric(50)	The code of the unit as used for substance concentration data. Allowed code: kg/kg or kilogram/kilogram; g/kg or gram/kilogram; mg/kg or milligram/kilogram (default); µg/kg or microgram/kilogram; ng/kg or nanogram/kilogram; pg/kg or picogram/kilogram.	Concentration-Unit, Units, Unit	No

Table aliases: AnalyticalMethodSubstances, AnalyticalMethodSubstance, AnalyticalMethodCompounds, AnalyticalMethodCompound, RawAnalyticalMethodCompounds.

Human monitoring data settings

Selection settings

Table 2.116: Selection settings for module Human monitoring data.

Name	Description
Surveys	The surveys that should be included in the action.
Sampling methods	The sampling methods that should be included in the action.

Human monitoring data as data

Data are provided in the form of surveys consisting of individuals from which the human monitoring samples taken. Substance concentration measurements are linked to analyses performed on the human monitoring samples. The data should also include information about the analytical methods that were used used.

- *Human monitoring data data formats*

2.4.9 Non-dietary exposures

Non-dietary exposures are the amounts of substances to which individuals in a population are exposed via any of three non-dietary routes: dermal, inhalation or oral, per day. Non-dietary exposures can be used for *computing aggregate exposure distributions* from both dietary and non-dietary routes of exposure. Depending on the exposure type, non-dietary exposures can be short-term/acute exposures and then contain exposures for individual-days, or they can be long-term/chronic exposures, in which case they represent the average exposure per day over an unspecified longer time period. Examples are presented as case studies in Kennedy et al. ([28], [26], [27], [29]) and R code to generate these examples is available for general use.

Datasets are typically generated by external programs, e.g. Browse, Bream2 or PACEM. The Browse and Bream2 models both simulate distributions of potential exposure of residents and bystanders to pesticides sprayed on crops. Probability distributions are included to quantify variations in input parameters representing conditions during a spray event. PACEM is a probabilistic exposure model for substances present in consumer products. Browse was an EU FP7 project (https://secure.fera.defra.gov.uk/browse/software), that in addition to bystanders and residents from boom-sprayers includes various arable and orchard scenarios. It includes dermal, oral and inhalation routes of exposure and can generate exposure files in the correct format for MCRA non-dietary exposure. The underlying simulation of dermal spray deposits on bystanders and residents was taken from Bream, although Browse includes post-processing to model indirect exposures, multiple routes and long-term exposure [29]. Volatilisation is also included through the PEARL-OPS model [44] to account for inhalation of vapours. Bream2 is an updated version of the original Bream model [28] and software is available online (http://www.ssau.co.uk/bream2-calculator). Results from Bream had been used as part of EFSA guidance on bystander and resident exposure. Bream2 was recently shown to produce more realistic exposure distributions, when compared to measured dermal exposure [10].

This module has as primary entities: *Populations Substances*

Output of this module is used by: *Exposures*

Non-dietary exposures data formats

Non-dietary exposures may be specified for multiple routes of exposure (dermal, oral and inhalation), for multiple substances, and for multiple exposure sources. Also, they can be provided as single deterministic exposure levels or as probabilistic exposure estimates and it is possible, but not mandatory, to specify uncertainty. The non-dietary exposures may be short term (acute) or longer term averages (chronic), and the user must ensure to supply appropriate non-dietary data for the type of exposure assessment of interest. For chronic assessments this means the non-dietary exposure is averaged over an appropriate time interval.

Non-dietary exposures are defined by non-dietary surveys to which dietary exposures are linked. For these surveys, individual properties can be specified to define non-dietary exposures for particular sub-groups of the population (e.g., specific age groups, or a specific gender). For each non-dietary survey a percentage of the target population that is not exposed from this source can be specified by means of a percentage. Uncertainy about non-dietary exposures can be specified by specifying multiple records for each individual in an additional table.

The use of multiple surveys can be used when multiple sources are relevant. For example, when modelling individuals taking part in various activities involving pesticide use or incidental exposures as a resident. Each non-dietary source is characterised in a particular user-selected or user-supplied non-dietary survey. By default, exposures from separate non-dietary surveys (sources) are considered to be independent events, but as explained below correlations between substances and/or activity types per individual can be represented if generated prior to uploading to MCRA. When including multiple non-dietary surveys it is possible to supply some with uncertainty/variability and others without variability/uncertainty according to the requirements and data availability.

When the user supplies probabilistic non-dietary exposure estimates (i.e., there is a distribution for the non-dietary exposure rather than a single nominal value), then this information will be propagated as part of the *exposure assessment*. Distributions may be included to represent variability, uncertainty or both, and in these cases the aggregate exposure estimates are reported with variability and/or uncertainty as appropriate. Multiple (uncertain) values from the non-dietary exposure distribution may be supplied per individual and per substance.

Exposures within a non-dietary survey may be expressed as correlated or independent for the different compounds. For example, if the exposures are a mixture of substances in a known ratio (e.g. from a specific tank mix of pesticides), or if exposure to one substance strongly implies that exposure to another is likely, these relationships may be included in the non-dietary data supplied by the user. Inference for the matched-case scenario with uncertainty analysis can use exposure sets. These are specific sets of exposures defined for each individual and in any uncertainty iteration an individual will receive exactly one of the exposure sets for that individual. Alternatively, independence may be represented by generating sets drawn from independent distributions when generating these tables.

Non-dietary exposures

Non-dietary exposure data is provided per non-dietary surveys. Each non-survey has some general information about the exposed population and the origin of the non-dietary exposure data. Also, a number of properties, such as specific age groups, can be specified for a survey. To each non-dietary survey, non-dietary exposures can be linked. These exposures may originate from dermal, oral and/or inhalatory exposure routes.

Non-dietary surveys

This table provides detail about non-dietary surveys (source of non-dietary exposure): description, location, date and unit of exposure).

Table 2.117: Table definition for NonDietarySurveys.

Name	Type	Description	Aliases	Required
idNonDietary-Survey	AlphaNumeric(50)	The survey identification number.	idNonDietary-Survey	Yes
Description	AlphaNumeric(200)	Description of non-dietary survey.	Description	No
Location	AlphaNumeric(50)	The location of survey.	Location	No
Date	DateTime	The date of survey.	Date	No
NonDietary-IntakeUnit	AlphaNumeric(50)	The unit of the non-dietary exposure.	Unit, NonDietary-IntakeUnit, NonDietary-ExposureUnit	Yes
Percentage-Zeros	Numeric	The proportion zeros, specified as a percentage (%).	PercentageZeros	No
idPopulation	AlphaNumeric(50)	Unique identification code of the population.	IdPopulation, PopulationId	No

Table aliases: NonDietarySurveys, NonDietarySurvey, RawNonDietarySurveys.

Non-dietary survey properties

This table specifies demographic properties that apply to the individuals in the surveys. These properties could be used to link the individuals of a non-dietary survey with individuals from dietary surveys. That is, if demographic criteria are defined, only those individuals in the dietary survey that meet these criteria will be assigned non-dietary exposures. This table is not relevant when matching is switched on (i.e., when individuals are matched based on individual id).

Table 2.118: Table definition for NonDietarySurveyProperties.

Name	Type	Description	Aliases	Required
Individual-PropertyName	AlphaNumeric(50)	Name of demographic criteria for non-dietary exposures in a particular survey e.g. age, gender, height (must correspond to a column name in Individuals table).	Individual-PropertyName	Yes
idNonDietary-Survey	AlphaNumeric(50)	The code of survey (must correspond to values in id column of the non-dietary surveys table).	idNonDietary-Survey	Yes
Individual-PropertyText-Value	AlphaNumeric(50)	Text value of the property e.g. male or female, smoker or non-smoker.	Individual-PropertyText-Value	No
Individual-Property-DoubleValue-Min	Numeric	Inclusive lower bound value of the property. E.g., a value of “18” for an individual property name called Age would mean that only individuals aged 18 and above receive the non-dietary exposures.	Individual-PropertyDouble-ValueMin	No
Individual-Property-DoubleValue-Max	Numeric	Inclusive upper bound value of property e.g. a value of “65” for an IndividualPropertyName called Age would mean that only individuals aged 65 and below receive the non-dietary exposures.	Individual-PropertyDouble-ValueMax	No

Table aliases: NonDietarySurveyProperties, NonDietarySurveyProperty, RawNonDietarySurveyProperties.

Non-dietary exposures

This table defines nominal non-dietary exposure values (such as means) for individuals within the non-dietary surveys. It can also be used to specify non-dietary exposures for individuals within the food surveys. Each exposure comprises a non-dietary survey (source of exposure); a string identifying an individual, which may or may not correspond to the ID of an individual in a food survey; a substance; and dermal, oral and inhalation exposure values. Exposures are assumed to be external doses.

Table 2.119: Table definition for NonDietaryExposures.

Name	Type	Description	Aliases	Required
idIndividual	AlphaNumeric(50)	Non-dietary individual identification number. This id may 1) match with the individual ids of the dietary survey (dietary exposures matched to food survey individuals), 2) not match with the individual ids of the dietary survey (unmatched individuals), or contain a default exposure (indicated by idIndividual = 'General') linking the dietary exposures to individuals based on the demographic criteria defined in the non-dietary survey properties table.	idIndividual	Yes
idNonDietary-Survey	AlphaNumeric(50)	The code of the survey (must correspond to values in id column of non-dietary surveys table).	idNonDietary-Survey	Yes
idSubstance	AlphaNumeric(50)	The substance code.	idSubstance, SubstanceId, SubstanceCode, Substance, idCompound, CompoundId, Compound-Code, Compound	Yes
Dermal	Numeric	The dermal (non-dietary) exposure value.	Dermal	No
Oral	Numeric	The oral (non-dietary) exposure value.	Oral	No
Inhalation	Numeric	The inhalation (non-dietary) exposure value.	Inhalation	No

Table aliases: NonDietaryExposures, NonDietaryExposure, RawNonDietaryExposures.

Non-dietary exposure uncertainty records

This table may be used to supply uncertainty sets of multiple (uncertain) non-dietary exposure values for individuals within the non-dietary surveys. Multiple non-dietary values are generated by probabilistic exposure calculations i.e. when there is a distribution for the non-dietary exposure rather than a single nominal value. If this table is supplied, aggregate exposure estimates will be reported with uncertainty using the uncertainty set approach. Each exposure set comprises a non-dietary survey (source of exposure); an individual ID; a substance; and dermal, oral and inhalation exposure values. In addition, the id column is used to define the uncertainty set. Summarizing, an uncertainty set is identified by column id and contains all exposure sets defined for each individual. In each uncertainty run (outer loop) an uncertainty set is sampled and in each iteration (inner loop) nondietary individuals are sampled from this set.

Table 2.120: Table definition for NonDietaryExposuresUncertain.

Name	Type	Description	Aliases	Required
idIndividual	AlphaNumeric(50)	Non-dietary individual identification number. The idIndividual value may correspond to an id in the Individuals table (dietary exposures matched to food survey individuals), may not correspond to an id in the Individuals table (unmatched individuals), or may contain a default exposure (indicated by idIndividual = 'General' - demographic criteria for the assignment of exposures are defined in the NonDietarySurveyProperties table). For matching to occur, the user will need to tick the option to 'match specific dietary survey individuals' in the user-interface. The software will then assign non-dietary exposures to the dietary individuals according to the values in this column. Any idIndividual values that do not correspond to individuals in the food survey will be ignored, unless a value 'General' is specified. Then the individual should meet the demographic criteria as defined in the NonDietarySurveyProperties table. If this box is left unticked, the non-dietary exposures will be randomly allocated to the dietary population provided they meet the demographic criteria.	idIndividual	Yes
idNonDietary-Survey	AlphaNumeric(50)	code of survey (must correspond to values in id column of NonDietarySurveys table)	idNonDietary-Survey	Yes
idCompound	AlphaNumeric(50)	Substance code (must correspond to values in id column of Substances table).	idCompound	Yes
id	AlphaNumeric(50)	Uncertainty set identification number.	id	Yes
Dermal	Numeric	Dermal non-dietary exposure value.	Dermal	No
Oral	Numeric	Oral non-dietary exposure value.	Oral	No
Inhalation	Numeric	Inhalation (non-dietary) exposure value.	Inhalation	No

Table aliases: NonDietaryExposuresUncertain, NonDietaryExposureUncertain, RawNonDietaryExposuresUncertain.

Non-dietary exposures settings

Uncertainty settings

Table 2.121: Uncertainty settings for module Non-dietary exposures.

Name	Description
Resample non-dietary exposures	Specifies whether non-dietary exposures are resampled. Note that non-dietary uncertainty is only ignored when individual uncertainty is set to false (uncheck box: do NOT resample individuals).

Non-dietary exposures uncertainty

In an aggregate exposure assessment, dietary and nondietary data are combined into an aggregate exposure distribution. The nondietary data are supplied in table NonDietaryExposures. In an uncertainty analysis, MCRA provides two ways to assess the uncertainty:

1. the uncertainty set approach
2. the bootstrap algorithm.

When table **NonDietaryExposuresUncertain** is not supplied, the nondietary data in table **NonDietaryExposures** is resampled and the bootstrapped sets are used in the uncertainty run. More precisely, in each outer loop of the 2D Monte Carlo, within each nondietary survey (multiple surveys may be supplied), the nondietary individuals are resampled. Each individual represents a nondietary exposure set containing dermal and/or oral and/or inhalation exposure values for multiple substances. Bootstrapping is the default behaviour when the **NonDietaryExposuresUncertain** table is missing. When uncertainty distributions supplied in this table represent sampling uncertainty (individual exposure sets are repeatedly sampled using the same nondietary exposure generator without changing the input parameters), then bootstrapping the data performs equally well and is more efficient.

Non-dietary exposures as data

Non-dietary exposures are collected in non-dietary surveys. Data may be specified on population level or individual level, and may or may not include variability and uncertainty.

- *Non-dietary exposures data formats*

Inputs used: *Active substances*

2.5 Hazard modules

Hazard data exist at two levels: at a lower level *dose response data* give *responses* measured in *test systems* from doses of *active substances*. Such data can be modellled with *dose response models*.

At a higher level *responses* can be linked to *effects*, optionally via *AOP networks*, using *effect representations*. If benchmark responses (BMRs) have been specified, *dose response models* can calculate Benchmark Doses (BMDs), which are the preferred Points of departure in hazard assessments. In addition, or alternatively, external *points of departure* can be specified for *active substances* and *effects*.

BMDs from *dose response models* and/or other *points of departure* can be converted to *hazard characterisations* at the intended level (external or internal dose, without or with safety factors), using *kinetic models*, *inter-species conversions* and/or *intra-species factors*. Finally, *hazard characterisations* can be translated to *relative potency factors*.

2.5.1 Active substances

Active substances are the substances that may lead with non-zero probability (P (AG)>0) to a specific health effect (adverse outcome). In the simplest case, all substances in the scope of the action will form one assessment group (AG) of active substances. In more advanced cases, the list of active substances is derived from possibly multiple assessment group memberships, which are scores for substances that determine whether a substance is included (score > 0) or excluded (score = 0) in the set of active substances. Substances with membership 0 are excluded from the list of active substances. Memberships scores between 0 and 1 are treated as probabilities of being in the set of active substances. Assessment group memberships can be either specified directly as data or derived from *QSAR membership models*, *molecular docking models*, or from availability of *points of departure*.

This module has as primary entities: *Effects Substances*

Output of this module is used by: *Concentrations Occurrence patterns Substance conversions Non-dietary exposures Kinetic models Food conversions Dietary exposures with screening Dietary exposures Exposures Hazard characterisations*

Active substances data formats

Active substances as data have to be specified via assessment group (AG) memberships in an AG membership model. For each effect one or more AG membership models can be available, one of which should be chosen in assessments. The AG memberships can be crisp, i.e. a positive list of active substances (with default memberships 1, although it is also allowed to include the negative memberships with membership 0 explicitly) or probabilistic ($0 \leq P \leq 1$).

Assessment group membership models

Assessment group membership models contain substance membership definitions for a given (health) effect. This data is described using two tables: the assessment group membership models table and the assessment group memberships table. The groups for a specified health effect are defined in the assessment group membership models table. The assessment group memberships table describes the substance memberships (or membership probabilities) in each group.

Assessment group membership models

This table contains the definitions of the assessment group membership models. Each model contains a id, name, an optional description, and refers to its related health effect.

Table 2.122: Table definition for AssessmentGroupMembershipModels.

Name	Type	Description	Aliases	Required
id	AlphaNumeric(50)	The unique identification code of the assessment group membership model.	id, idModel, Model, idAssessment-GroupModel, Assessment-GroupModel, idGroup-Membership-Model, Group-Membership-Model	Yes
Name	AlphaNumeric(100)	The name of the assessment group membership model.	Name	No
Description	AlphaNumeric(200)	Description of the assessment group membership model.	Description	No
idEffect	AlphaNumeric(50)	The effect code.	idEffect, EffectId, Effect	Yes
Accuracy	Numeric	If applicable, the accuracy of the assessment group membership model memberships.	Accuracy	No
Sensitivity	Numeric	If applicable, the sensitivity of the assessment group membership model.	Sensitivity	No
Specificity	Numeric	If applicable, the specificity of the assessment group membership model.	Specificity	No
Reference	AlphaNumeric(200)	External reference(s) to sources containing more information about the assessment group model.	References	No

Table aliases: AssessmentGroupMembershipModels, AssessmentGroupMembershipModel, RawAssessmentGroupMembershipModels.

Assessment group memberships

Substances can belong to an assessment group with certainty (probability 1), or the membership can be uncertain. This table allows to specify membership probabilities for assessment group membership models. The probability should be a value between zero and one. For example, set to 1 or 0, or prior probabilities, or probabilities or 0/1 values estimated from QSAR, from Molecular Docking or from expert elicitation. The table can contain prior or posterior memberships. Default membership are specified with an empty idSubstance field.

Table 2.123: Table definition for AssessmentGroupMemberships.

Name	Type	Description	Aliases	Required
idGroup-Membership-Model	AlphaNumeric(50)	The id of the assessment group memberships model or source.	Model, idModel, idAssessment-Group-Membership-Model, Assessment-Group-Membership-Model, idGroup-Membership-Model, Group-Membership-Model, idGroup	Yes
idSubstance	AlphaNumeric(50)	The code of the substance.	idSubstance, SubstanceId, SubstanceCode, Substance, idCompound, CompoundId, Compound-Code, Compound	Yes
Group-Membership	Numeric	Probability of the substance for belonging to the assessment group for the effect. If omitted, the default is 1, i.e. certain membership.	Group-Membership, Membership, Membership-Probability, Probability, Assessment-Group-Membership	Yes

Table aliases: AssessmentGroupMemberships, AssessmentGroupMembership, RawAssessmentGroupMemberships.

Active substances calculation

Depending on the *model settings*, the set of active substances for a specified effect can be computed in several ways:

1. From the list of substances with available *points of departure (POD) data* for the specified effect. If there is a POD, then the substance is considered an active substance, with membership 1. If not, the membership is 0, and the substance is excluded from the list of active substances.
2. From one or more in-silico (QSAR and/or molecular docking) models. The results of the in-silico models should be provided as *QSAR membership models data* and/or *molecular docking models data*. Binding energies from molecular docing models are first translated to crips memberships using a threshold value. The results from multiple in-silico models can be combined in any of four membership calculation methods:
    1. (crisp, any) the substance is considered an active substance if any in-silico model indicates activity;
    2. (crisp, majority) the substance is considered an active substance if the majority of in-silico models indicates activity;
    3. (probabilistic, ratio) the membership probability is the fraction of in-silico models that indicate activity;

4. (probabilistic, Bayesian) the membership probability is calculated using a Bayesian model according to Kennedy et al. [30] and a specified prior probability (which is by default 0.5).

For substances within the scope of the assessment but without in-silico data, the default is to omit them in the AG. There is an option however to include such substances with a default membership probability.

3. From a combination of 1 and 2, using either the union (OR) method or the intersection (AND method) of results.

Active substances settings

Calculation settings

Table 2.124: Calculation settings for module Active substances.

Name	Description
Derive memberships from POD presence	Determine assessment group membership based on presence/absence of points of departure.
Derive memberships from QSAR membership data	Specifies whether QSAR membership data is used for computing the assessment group memberships.
Derive memberships from molecular docking data	Specifies whether molecular docking data is used for computing the assessment group memberhips.
Include substances without membership information	For non-probabilistic methods: specifies whether substances for which no membership information is available in the specified inputs should be included in the assessment group.
Combination method memberships from available PODs and in-silico data	Specifies whether to take the intersection or the union of the set of substances with available PoDs and the set of substances with positive/probable (in-silico) membership score.
Membership calculation method	Calculation method for computing assessment group memberships: majority/any (crisp methods), ratio/Bayesian (probabilistic methods)
Default/prior membership probability	Default substance membership probability for which no membership information is available in the specified inputs. Prior probability for Bayesian method.

Uncertainty settings

Table 2.125: Uncertainty settings for module Active substances.

Name	Description
Resample assessment group memberships	Specifies whether assessment group memberships of substances should be resampled using the assessment group membership probabilities.

Active substances as data

When provided as data, in the form of assessment group memberships, the active substances are derived from the specified memberships.

- *Active substances data formats*

Inputs used: *AOP networks Points of departure*

Calculation of active substances

Active substances and assessment group memberships may be computed from PoD presence of in-silico data.

- *Active substances calculation*

Inputs used: *Molecular docking models QSAR membership models*

Settings used

- *Calculation Settings*

2.5.2 AOP networks

Effects can be related to each other using the toxicological concept of adverse outcome pathways (AOPs) and adverse outcome pathway networks (see https://aopwiki.org). Adverse Outcome Pathway (AOP) Networks specify how biological events (effects) can lead to an adverse outcome (AO) in a qualitative way through relations of upstream and downstream key events (KEs), starting from molecular initiating events (MIEs). Using AOPs, the adverse outcome (AO), e.g., liver steatosis, is linked to key events (KEs), e.g., triglyceride accumulation in the liver, and to molecular initialing events (MIEs), e.g., PPAR-alpha receptor antagonism. In general, multiple AOPs may lead to the same AO, and therefore AOP networks can be identified.

This module has as primary entities: *Effects*

Output of this module is used by: *QSAR membership models Molecular docking models Active substances Relative potency factors Hazard characterisations Points of departure Effect representations*

AOP networks data formats

AOP networks

AOP networks are described using two tables: the AOP networks table, and the effect relations table. The AOP networks table records the ids, names, descriptions, and other metadata of the AOP networks. The effect relations table describes the effects and effect relations (i.e., upstream and downstream key event relations) that are part of the AOP network.

AOP networks

Data format for specification of adverse outcome pathway (AOP) networks.

Table 2.126: Table definition for AdverseOutcomePathwayNetworks.

Name	Type	Description	Aliases	Required
idAdverse-Outcome-Pathway-Network	AlphaNumeric(50)	Unique identification code of the AOP network.	idAOPN, idAOPNetwork, AOPN, AOPNetwork, Id	Yes
Name	AlphaNumeric(100)	Name of the AOP network.	Name	No
Description	AlphaNumeric(200)	Additional description or label of the AOP network.	Description	No
Reference	AlphaNumeric(200)	External reference(s) to sources containing more information about the AOP network. E.g., the AOP wiki, and the associated AOP wiki Ids.	Reference, References	No
idAdverse-Outcome	AlphaNumeric(50)	The identification code of the effect representing the adverse outcome of this AOP network.	idAdverse-Outcome, idAO, idEffect, Adverse-Outcome	Yes
RiskType	AlphaNumeric(100)	The risk type of the adverse outcome.	RiskType	No

Table aliases: AOPNetworks, AOPNetwork, RawAdverseOutcomePathwayNetworks.

Effect relations

Dataformat for specification of the effect (key event) relationships of adverse outcome pathway (AOP) networks.

Table 2.127: Table definition for EffectRelations.

Name	Type	Description	Aliases	Required
idAdverse-Outcome-Pathway-Network	AlphaNumeric(50)	Identification code of the AOP network for which this link is defined.	idAdverse-Outcome-Pathway-Network, idAOPN, idAOPNetwork, AOPN, AOPNetwork	Yes
idDownstream-KeyEvent	AlphaNumeric(50)	Identification code of the (triggered) effect of this relationship.	idDownstream-KeyEvent, idEffect, idKeyEvent, Effect, KeyEvent	Yes
idUpstream-KeyEvent	AlphaNumeric(50)	Identification code of the triggering effect of this relationship.	idTrigger, idUpstreamKey-Event, Trigger	Yes
Reference	AlphaNumeric(200)	External reference(s) to sources containing more information about the effect (key event) relationships.	Reference, References	No

Table aliases: EffectRelations, EffectRelation, EffectRelationships, EffectRelationship, KeyEventRelationships, KeyEventRelationship, RawEffectRelations.

AOP networks settings

Selection settings

Table 2.128: Selection settings for module AOP networks.

Name	Description
AOP Network	The AOP networks of interest.
Restrict AOP network by focal upstream event	Restrict the AOP network to a specific sub-network, containing only the AOPs that include both the focal key event (KE) defined here (which must be upstream of the AO) and the focal effect (adverse outcome, AO)
Focal upstream event	The focal key event used for restricting the AOP network to a specific sub-network of interest.

AOP networks as data

AOP networks can only be provided as data in the form of network definitions containing effect relations (key-event relationships) collections.

- *AOP networks data formats*

2.5.3 Dose response data

Dose response data are data on response values of test systems at specified doses of substances (or mixtures of substances) from dose response experiments.

This module has as primary entities: *Substances Test systems Responses*

Output of this module is used by: *Dose response models*

Dose response data data formats

The meta-data of dose response experiments (such as name, description, etc.) are specified in the DoseResponseExperiments table.

For presenting the data of these experiments to the system, there are two formats: a single table format (DoseResponseData) and a relational data format (three tables DoseResponseExperimentDoses, ExperimentalUnitProperties, DoseResponseExperimentMeasurements). Usually, the single table format will be the easier one. For internal use in MCRA, this single table data is converted to the relational data format.

Dose response data

Dose response data can be used to extract assessment group membership or hazard doses . The meta-data of dose response experiments (such as name, description, etc.) are specified in the DoseResponseExperiments table. For presenting the data of these experiments to the system, there are two formats: a single table format (DoseResponseData) and a relational data format (three tables). Usually, the single table format will be the easier one. For internal use in MCRA, this single table data is converted to the relational data format.

Dose response experiments

General information about the dose response experiments, such as the (unique) identifier, name, description, the used test-system, and the dose unit is stored in the table DoseResponseExperiments. If the data of an experiment is provided in a single table format, then the fields Time, Covariates, Substances, and Responses can be used to map the column header names of the columns of the single data table to these their respective types.

Table 2.129: Table definition for DoseResponseExperiments.

Name	Type	Description	Aliases	Required
idExperiment	AlphaNumeric(50)	Unique identification code of the dose effect experiment.	idExperiment, Id, Code	Yes
Name	AlphaNumeric(100)	Name of the dose effect experiment.	Name	No
Description	AlphaNumeric(200)	Description of the dose effect experiment.	Description	No
Date	DateTime	The starting date of the experiment.	Date	No
Reference	AlphaNumeric(200)	External reference, for instance, to the experiment protocol and/or supporting material.	Reference	No
Experimental-Unit	AlphaNumeric(100)	The name of the experimental unit of the experiment, e.g., rat, cage, litter, vial, cup, petridish.	Experimental-Unit	No
DoseRoute	AlphaNumeric(100)	For in-vivo test systems, the route in which the dose was administered	DoseRoute	No
Substances	AlphaNumeric	Code or comma separated list of the codes of the substances measured in the experiment. E.g., 'Cyproconazole, Thiram'. Required when presenting the dose-response data in a single table. Make sure that in table DoseResponseData the column headers exactly match these names.	idSubstance, SubstanceId, SubstanceCode, Substance, idSubstances, SubstanceIds, SubstanceCodes, Substances, idCompound, CompoundId, Compound-Code, Compound	Yes
DoseUnit	*DoseUnits*	Unit of the doses administered in this experiment.	DoseUnit	Yes
Responses	AlphaNumeric	Code or comma separated list of codes of the responses measured in the experiment. E.g., 'AngleM_PQ, Mortality'. Required when presenting the dose-response data in a single table. Make sure that in table DoseResponseData the column headers exactly match these names.	Responses, Response, idResponses, idResponse	Yes
Time	AlphaNumeric(100)	Identifier of the time field of the experiment. Required when presenting the dose-response data in a single table and responses are measured at multiple times. Make sure that in the table DoseResponseData the column header of the time-column exactly matches this name.	Time, Times	No
TimeUnit	TimeUnit	Unit of the time scale used in the experiments.	TimeUnit	No
Covariates	AlphaNumeric(200)	Comma separated list of the names/codes of the covariates	Covariates, Covariate	No

Table aliases: DoseResponseExperiments, DoseResponseExperiment, RawDoseResponseExperiments.

Dose response data

Single (two-way) table data format for specifying data of dose response experiments (as alternative for the relational format). The column headers are dynamic and should be defined in the table DoseResponseExperiments through fields Substances and Responses (and, optionally, Covariates and Time). For responses given as aggregated statistics, also SD, CV, N and Uncertainty can be specified as [Datatype:Response]. E.g., 'SD:Y', 'CV:Y', 'N:Y'. Uncertainty upper 95%limits can be specified as 'UncertaintyUpper:Y'. For each quantal response an additional column 'N:[responsename]'is required with binomial totals (e.g. Mortality = 3, N:Mortality = 10).

Table 2.130: Table definition for DoseResponseData.

Name	Type	Description	Aliases	Required
idExperiment	AlphaNumeric(50)	Unique identification code of the dose effect experiment.	idExperiment, Experiment, Code	No
Experimental unit	AlphaNumeric(50)	Experimental unit numbers or identifiers. The column name of the experimental unit should be as specified in the dose response experiment record.	ExperimentalUnit, ExperimentalUnits, Experimental unit	No
Substance(s)	AlphaNumeric(100)	One or more columns with doses for each substance, in the unit as specified in the dose response experiment table. The column name(s) should match the substance codes listed in the comma-separated list of the substances field of the dose response experiment record.		Yes
Response(s)	AlphaNumeric(100)	One or more columns with results for each response, in the unit(s) as specified in the dose response experiment table. The column name(s) should match the response codes listed in the comma-separated list of the responses field of the dose response experiment record.		Yes
Time	Numeric	The column containing the observed response times. The column name (header) should match that of the Time column in the dose response experiment record.		No
Covariate(s)	AlphaNumeric(100)	The column(s) containing additional properties of the experimental unit. The column name (header) should match the codes of the comma-separated covariates list in the dose response experiment record.		No

Table aliases: TwoWayDoseResponseData, DoseResponseDataTwoWay, DoseResponseData, RawDoseResponseData.

Relational dose response data

In the relational data format, dose response experiment data can be specified using the triplet of tables: DoseResponseExperimentDoses, DoseResponseExperimentMeasurements, and ExperimentalUnitProperties. These tables describe, respectively, the experiment designs (including the administered substance doses), the response measurements, and additional properties of the experimental units of the experiment.

Dose response experiment doses

The table DoseResponseExperimentDoses describes the experiment design, being a complete specification of which doses of which substances were applied to which experimental unit and if relevant at what time.

Table 2.131: Table definition for DoseResponseExperimentDoses.

Name	Type	Description	Aliases	Required
idExperiment	AlphaNumeric(50)	Identification code of the experiment to which this design record belongs.	idExperiment, Experiment	Yes
idExperimental-Unit	AlphaNumeric(50)	Identification code of the experimental unit to which the dose is applied.	idExperimental-Unit, Experimental-Unit	Yes
Time	Numeric	The time of administration of the dose.	Time	No
idSubstance	AlphaNumeric(50)	Code of the substance that was administered.	idSubstance, SubstanceId, SubstanceCode, Substance, idCompound, CompoundId, Compound-Code, Compound	Yes
Dose	Numeric	The dose that was administered.	Dose	Yes

Table aliases: DoseResponseExperimentDoses, DoseResponseExperimentDose, RawDoseResponseExperimentDoses.

Experimental unit properties

The table ExperimentalUnitProperties can be used to specify additional properties of the experimental units of the experiment. For instance, the gender of the rat, in case rats are the experimental units.

Table 2.132: Table definition for ExperimentalUnitProperties.

Name	Type	Description	Aliases	Required
idExperiment	AlphaNumeric(50)	Identification code of the experiment.	idExperiment, Experiment	Yes
idExperimental-Unit	AlphaNumeric(50)	Identification code of the experimental unit.	idExperimental-Unit, Experimental-Unit	Yes
PropertyName	AlphaNumeric(50)	Name of the experimental unit property.	Property, Name	Yes
Value	AlphaNumeric(100)	Value of the experimental unit property.	PropertyValue	No
OtherProperty		Other properties of experimental units are automatically parsed, using the column name (header) as property name.		No

Table aliases: ExperimentalUnitProperties, ExperimentalUnitProperty, RawExperimentalUnitProperties.

Dose response experiment measurements

The table DoseResponseMeasurements describes the measurements that were done in the experiments. That is, for each response and experimental unit, at each observation time, one measurement should be recorded. If the response is an aggregated statistic, then this record may also include a standard deviation and number of units over which was aggregated.

Table 2.133: Table definition for DoseResponseExperimentMeasurements.

Name	Type	Description	Aliases	Required
idExperiment	AlphaNumeric(50)	Identification code of the experiment to which this measurement belongs.	idExperiment, Experiment	Yes
idExperimental-Unit	AlphaNumeric(50)	Identification code of the experimental unit from which the measurement is taken.	idExperimental-Unit, Experimental-Unit	Yes
idResponse	AlphaNumeric(50)	Identifier of the response that is measured.	idResponse, Response	Yes
Time	Numeric	Time of observation.	Time	No
ResponseValue	Numeric	The measured response.	ResponseValue, Value	Yes
SD:Response	Numeric	For aggregated responses, the standard deviation of the measurement.	SD:Response, ResponseSD	No
CV:Response	Numeric	For aggregated responses, the coefficient of variation (cv) of the measurement.	CV:Response, ResponseCV	No
N:Response	Numeric	For aggregated responses, the number of units over which was aggregated.	N:Response, ResponseN	No
Uncertainty-Upper:Response	Numeric	Optionally, measurement uncertainty quantification in terms of the upper value (i.e., an estimate of 95th percentile).	Uncertainty-Upper:Response, Response-Uncertainty-Upper, Uncertainty-Upper, Upper	No

Table aliases: DoseResponseExperimentMeasurements, DoseResponseExperimentMeasurement, DoseResponseMeasurements, DoseResponseMeasurement, RawDoseResponseExperimentMeasurements.

Dose response data settings

Selection settings

Table 2.134: Selection settings for module Dose response data.

Name	Description
Experiments	The dose response experiments of interest.
Merge dose response data of multiple experiments	Specifies whether the dose response data of multiple experiments should be merged into one large dose response data set.

Dose response data as data

Dose response data can be provided per experiment or study in which several responses (on in-vitro or in-vivo test systems) are measured from several administered substance doses.

- *Dose response data data formats*

2.5.4 Dose response models

Dose response models are models fitted to dose response data and can be provided as data or calculated using a local or remote version of PROAST. The main results for hazard and risk assessment are benchmark doses (BMDs), related to a specified substance, response, optionally covariate value, and the benchmark response (BMR). Dose response models can be uploaded as data, retrieved from PROASTweb through *linked remote repositories*, or *calculated using an internal version of PROAST*.

This module has as primary entities: *Test systems Responses Substances*

Output of this module is used by: *Hazard characterisations*

Dose response models data formats

Dose response models

Dose response models are specified using three tables: the dose response models table holds the dose response model definitions (id, name, description) and other information about the dose response models. The dose response model benchmark doses table records the benchmark doses and (optionally) the model parameters for specific substances and covariates. The dose response model benchmark doses uncertainty table records results from bootstrap runs for the benchmark doses per substance/covariate combination.

Dose response models

Each dose response model has a unique id, a name (optional), and description (optional). Also, each dose response model is associated with a specific dose response experiment (idExperiment) from which the data used to create the model is obtained, a response (idResponse), one or more substances, and, optionally, specific covariates considered by the dose response model. The combination of the benchmark response type and the associated value define the benchmark response of the model. The dose unit specifies the unit used for the doses, and if applicable, the model equation can be specified.

Table 2.135: Table definition for DoseResponseModels.

Name	Type	Description	Aliases	Required
idDose-ResponseModel	AlphaNumeric(50)	The unique identification code of the fitted dose response model.	idDose-ResponseModel, idModel	Yes
idExperiment	AlphaNumeric(50)	The identification code of the experiment from the dose response model.	experiment-Code, experimentId	Yes
Name	AlphaNumeric(100)	The name of the dose response model.	Name	No
Description	AlphaNumeric(200)	Description of the dose response model.	Description	No
Substances	AlphaNumeric	Code or comma separated list of the codes of the substances in the Dose Response Model. E.g., 'Cyproconazole, Thiram'.	Substances	Yes
idResponse	AlphaNumeric	The response of the dose response model.	idResponse, Response	Yes
Covariates	AlphaNumeric	The covariates considered by the dose response model.	Covariates, Covariate	No
Benchmark-Response	Numeric	The value of the benchmark response or critical effect size.	Benchmark-Response, CriticalEffect-Size, CES	Yes
Benchmark-ResponseType	*Benchmark-ResponseTypes*	Specifies how the benchmark response is expressed. E.g., using a percent change in mean response or, for quantal response types, in terms of extra risk, additional risk, or ED50.	Benchmark-ResponseType, HazardEffect-SizeType, CriticalEffect-SizeType	No
LogLikelihood	Numeric	Loglikelihood of the model fit.	LogLikelihood	No
DoseUnit	AlphaNumeric(50)	The dose unit (if not specified, then mg/kg is assumed).	DoseUnit, UnitDose	No
ModelEquation	AlphaNumeric(500)	If available, the model equation of the dose response model (R model equation) or the identifier of the dose response model type.	ModelEquation, DoseResponse-ModelEquation, Equation	No

Table aliases: DoseResponseModels, DoseResponseModel, RawDoseResponseModels.

Dose response model benchmark doses

The benchmark responses and benchmark doses belonging to the dose response models are recorded per substance/covariate in the dose response model benchmark doses table. Optionally, if the model equation of the dose response model has been specified in the dose response models table, the model parameter values for this specific substance/covariate can be specified here.

Table 2.136: Table definition for DoseResponseModelBenchmarkDoses.

Name	Type	Description	Aliases	Required
idDose-ResponseModel	AlphaNumeric(50)	The identification code of the dose response model to which this record belongs.	idDose-ResponseModel	Yes
idSubstance	AlphaNumeric(50)	The code of the substance.	idSubstance, SubstanceId, SubstanceCode, Substance, idCompound, CompoundId, Compound-Code, Compound	Yes
Covariates	AlphaNumeric(500)	Comma separated list of the covariate values for which this benchmark dose applies.	Covariates, Covariate	No
Benchmark-Dose	Numeric	The (nominal) benchmark dose (BMD).	Benchmark-Dose, BMD, CED	Yes
Benchmark-DoseLower	Numeric	Benchmark dose lower uncertainty bound (BMDL).	Benchmark-DoseLower, BMDL, CEDL	No
Benchmark-DoseUpper	Numeric	Benchmark dose upper uncertainty bound (BMDU).	Benchmark-DoseUpper, BMDU, CEDU	No
Model-Parameter-Values	AlphaNumeric(500)	Parameter values for dose response models.	ParameterValues	No

Table aliases: DoseResponseModelBenchmarkDoses, RawDoseResponseModelBenchmarkDoses.

Dose response model benchmark dose bootstraps

Empirical uncertainty values of the benchmark benchmark doses of dose response models can be recorded in the dose response model benchmark doses bootstraps table. The uncertainty set identifier (idUncertaintySet) can be specified to retain correlations between uncertainty records that originate from the same bootstrap run.

Table 2.137: Table definition for DoseResponseModelBenchmarkDosesUncertain.

Name	Type	Description	Aliases	Required
idDoseResponseModel	AlphaNumeric(50)	The identification code of the dose response model to which this record belongs.	idDoseResponseModel	Yes
idUncertaintySet	AlphaNumeric(50)	The uncertainty set identifier.	idUncertaintySet, UncertaintyId	Yes
idSubstance	AlphaNumeric(50)	The code of the substance.	idSubstance, SubstanceId, SubstanceCode, Substance, idCompound, CompoundId, CompoundCode, Compound	Yes
Covariates	AlphaNumeric(500)	Comma separated list of the covariate values for which this benchmark dose applies.	Covariates	No
BenchmarkDose	Numeric	Benchmark dose (BMD).	BenchmarkDose, BMD, CED	Yes

Table aliases: DoseResponseModelBenchmarkDosesBootstraps, DoseResponseModelBenchmarkDosesUncertain, RawDoseResponseModelBenchmarkDosesUncertain.

Dose response models calculation

Besides uploading dose response models as data or retrieving them from PROASTweb through *linked remote repositories*, there is also a possibility to compute dose response models using an integrated version of PROAST. When computing dose response models using the integrated version, MCRA will attempt to fit a dose response model for each response of each dose response experiment. Depending on the type of data (e.g., response type, covariates y/n, single or multiple substances) a PROAST run is configured and executed. If *effect representations* are provided, then benchmark responses specified by the effect representations data are used, otherwise only the model fits will be computed without benchmark doses.

Dose response models as data

Dose response models as data contain the details of fitted dose response models. The main elements for hazard and risk assessment are the benchmark doses (BMDs) related to specified substances, responses, and optionally covariate values for specified benchmark responses (BMR). These specifications can be provided in data files or can be retrieved/imported from PROAST output files on the PROAST website https://proastweb.rivm.nl/user/login using a PROASTweb user account and an application access key.

- *Dose response models data formats*

Inputs used: *Dose response data*

Calculation of dose response models

Used as a calculator, dose response models are fitted to dose response data using an MCRA-internal version of PROAST. Currently, all available models appropriate for the response type will be fitted, and for the Hill and Exponential model families, the best fitting model based on maximum likelihood will be selected. The set of results for the calculation will include BMDs etc. for all fitted models.

- *Dose response models calculation*

Inputs used: *Effect representations*

2.5.5 Effect representations

Effect representations specify the responses that can be used to measure specified effects and which response levels, the benchmark response (BMR), define the hazard limits for the effects.

This module has as primary entities: *Effects Responses*

Output of this module is used by: *Hazard characterisations Dose response models*

Effect representations data formats

Effect representations

Effect representations specify responses that may represent the effect.

Effect representations

One response can be set as the canonical response (golden standard). For a quantitive or stochastically qualitative canonical response a benchmark response should be defined.

Table 2.138: Table definition for EffectRepresentations.

Name	Type	Description	Aliases	Required
idEffect	AlphaNumeric(50)	Identifier of the effect	idEffect	Yes
idResponse	AlphaNumeric(50)	Identifier of the response	idResponse	Yes
Benchmark-Response	Numeric	The threshold response value that defines a hazard. For numeric responses (Continuous, Quantal, Count) the value that defines a hazard. For Binary responses 1 defines a hazard by default, unless redefined here.	BenchMark-Response, HazardEffect-Size, BMR, CriticalEffect-Size, CES	No
Benchmark-ResponseType	*Benchmark-ResponseTypes*	Specifies how the BenchMarkResponse is expressed, relative to the response at zero dose, or absolute. Required for numeric response types (Continuous, Quantal, Count). For qualitative responses (Ordinal, Categorical) Absolute is used.	Benchmark-ResponseType, HazardEffect-SizeType, CriticalEffect-SizeType	No

Table aliases: EffectRepresentations, EffectRepresentation, RawEffectRepresentations.

Effect representations as data

Effect representations are provided as data in the form of specified combinations of effect and response, optionally with a benchmark response that defines a hazard limit for the effect.

- *Effect representations data formats*

Inputs used: *AOP networks*

2.5.6 Hazard characterisations

Hazard characterisations are benchmark doses for active substances and for the chosen effect at the chosen target level (external or internal) of the hazard assessment. Hazard characterisations are based on points of departure, such as BMDs from dose-reponse models or externally specified points of departure (MDSs, NOAELs or LOAELs). The computation may involve inter-species conversion, intra-species factors and the use of kinetic models or absorption factors to convert external doses to internal doses.

This module has as primary entities: *Substances Effects*

Output of this module is used by: *Relative potency factors Risks*

Hazard characterisations calculation

Hazard characterisations can be defined as deterministic threshold values (e.g. ADI, ARfD) or as distributions (using probabilistic models). They are linked to an effect of interest. Hazard characterisations depend on the *risk type* (acute or chronic) and the *target level* of the human body (external via some route of exposure or internal for a specific defined organ or compartment). Hazard characterisations are derived from *points of departure* provided as data and/or from *dose-response models*. The procedure for computing hazard characterisations has two main phases: 1) collection of all available hazard characterisation candidates and alignment with the target system, and 2) aggregation over multiple available hazard characterisations and imputation of missing hazard characterisations.

Collection of available hazard characterisation candidates involves collecting the appropriate points of departure data and/or dose-response models that are used for deriving the hazard characterisations. In MCRA, a distinction is made between three *methods for computing hazard characterisations*:

1. Calculation of hazard characterisations from externally specified in-vivo PoDs (BMDs, NOAELs, other).
2. Calculation of hazard characterisations from PoDs (in this case BMDs) calculated from dose response data.
3. (in cumulative assessments) Calculation of hazard characterisations based on an *in-vivo PoD for the index substance and in-vitro RPFs from dose-response models for the other substances (IVIVE model)*.

For all three methods, the collected points of departure and benchmark doses should be aligned with the target system. This alignment may involve various conversion steps for each point of departure and specific substance, and can be formally specified as:

$$HC = f_{\texttt{expression-type}} \cdot f_{\texttt{kinetic}} \cdot \frac{1}{f_{\texttt{inter-species}}} \cdot \frac{1}{f_{\texttt{intra-species}}} \cdot PoD$$

where:

- HC denotes the hazard characterisation.
- $f_{\texttt{expression-type}}$ denotes the *expression type correction factor*, e.g., for extrapolation from LOAEL or NOAEL, or from NOAEL to BMD.
- $f_{\texttt{inter-species}}$ denotes the inter-species factor for *extrapolation from animal to human (inter-species)*.
- $f_{\texttt{intra-species}}$ denotes the intra-species factor *extrapolation from the average to the sensitive human or probabilistic calculation of the distribution of human individuals (intra-species)*.
- $f_{\texttt{kinetic}}$ denotes the kinetic conversion factor for *conversion from internal to external or external to internal hazard characterisations*.
- PoD denotes the point of departure.

It may be that for some substances multiple hazard characterisations are available (e.g., obtained from multiple experiments) and/or that for other substances hazard characterisations are still missing. Hence, two final steps remain to come to the final set of hazard characterisation:

- *Aggregation over multiple available hazard characterisations*.
- *Imputation of missing hazard characterisations*.

Hazard characterisation type extrapolation

Hazard doses, or points of departure can be of *various types*. E.g., BMDs, NOAELs, or LOAELs. When computing hazard characterisations, the type in which the hazard characterisations are expressed (i.e., the *hazard characterisation expression type*) should be specified explicitly. When points of departure from types different from the expression type are provided, these should be translated to the specified expression level. In the current implementation, the simple conversion factors shown in Table 2.139 are used, roughly based on the WHO guidance document on evaluating and expressing uncertainty in hazard characterization [6].

Table 2.139: Conversion factors for hazard characterisation types.

From	To	Conversion factor
BMD	NOAEL	1/3
BMD	LOAEL	1
NOAEL	BMD	3
NOAEL	LOAEL	1/3
LOAEL	BMD	1
LOAEL	NOAEL	1/3

Inter-species extrapolation

Hazard doses, or points of departure, are commonly only determined for animals, not for humans. In order to derive hazard characterisations for humans, the animal hazard doses need to be converted to toxicologically equivalent doses for humans. This extrapolation is usually expressed as a multiplication factor, and traditionally a factor of 10 is used (which is roughly obtained from the product of a factor of 3.2 for toxicokinetic variability and a factor 3.2 for toxicodynamic variability).

The following methods are available within the toolbox:

1. **No inter-species extrapolation:** Assume that for all available points of departure, the animal hazard dose is equal to the human hazard dose. Effectively, this is equivalent to using a convertion factor of 1.
2. **Default distribution:** Use a conversion factor drawn from a default, substance and species independent log-normal uncertainty distribution specified (as *model settings*) by a geometric mean (GM) and geometic standard deviation (GSD). In the *nominal run*, the nominal value of this distribution (i.e., the geometric mean) is used as a conversion factor. In the *uncertainty analysis loop*, provided that inter-species extrapolation uncertainty is *included in the uncertainty analysis*, a single factor is drawn from the lognormal distribution.
3. **Substance/species specific distributions:** Use conversion factors drawn from substance/species specific log-normal uncertainty distributions specified (as *data*) by a geometric mean (GM) and geometic standard deviation (GSD). In the *nominal run*, a factor equal to the geometric mean is used for all combinations of substance and species. In the *uncertainty analysis loop*, provided that inter-species extrapolation uncertainty is *included in the uncertainty analysis*, a uncertainty factor is drawn from the lognormal distribution with $\mu = 0$ and $\sigma^2 = 1$, which is used to obtain correlated draws for all available inter-species conversion factor distributions. If the distribution parameters are missing for a specific substance/species, then the default distribution is used as a fallback.

Intra-species extrapolation of hazard characterisations

There is variation between individuals concerning their individual sensitivities to experience health effects. In some scenarios the aim is to perform assessments for the sensitive individuals instead of the average individuals for which the points of departure are derived. If this is the case, then extrapolation is required to translate hazard characterisations derived for the average individual to hazard characterisations for a sensitive individual. In traditional exposure assessments, a safety of 100 is commonly used as a margin of safety, that is assumed to be composed of a interspecies extrapolation factor (factor 10), and inter-individual extrapolation factor (factor 10). I.e., the hazard characterisation defined for the sensitive individual is defined as

$$HC_{\mathtt{sens}} = \frac{1}{f_{\mathtt{intra-species}}} \cdot HC_{\mathtt{avg}}$$

Here $f_{\mathtt{inter-species}}$ denotes the intra-species factor. An alternative to using a fixed safety factor, is to define intra-species variability may be explicitly *a lognormal distribution*, characterised by a geometric mean (GM) equal to 1 and a geometric standard deviation (GSD). For *risks calculations*, this distribution could be used to sample individual hazard characterisations. This effectively converts the description of hazard characterisations to include variability, with an unbiased central value.

In-vitro in-vivo extrapolation (IVIVE)

The in-vitro in-vivo extrapolation method implemented in MCRA is based on the following prerequisites:

1. For one substance, the index substance, a reliable point of departure is available for the adverse outcome of interest obtained from an in-vivo assay (i.e., external dose).
2. There are other substances for which there is a dose-response model available from an in-vitro assay on a response representing an early key event of the adverse outcome for these substances and the index substance.

In IVIVE, these RPFs, in combination with the known hazard characterisation of the index substance, can be used to derive hazard characterisations for the other substances as well. Figure 2.35 shows the conceptual model that forms the basis of the IVIVE methodology of MCRA.

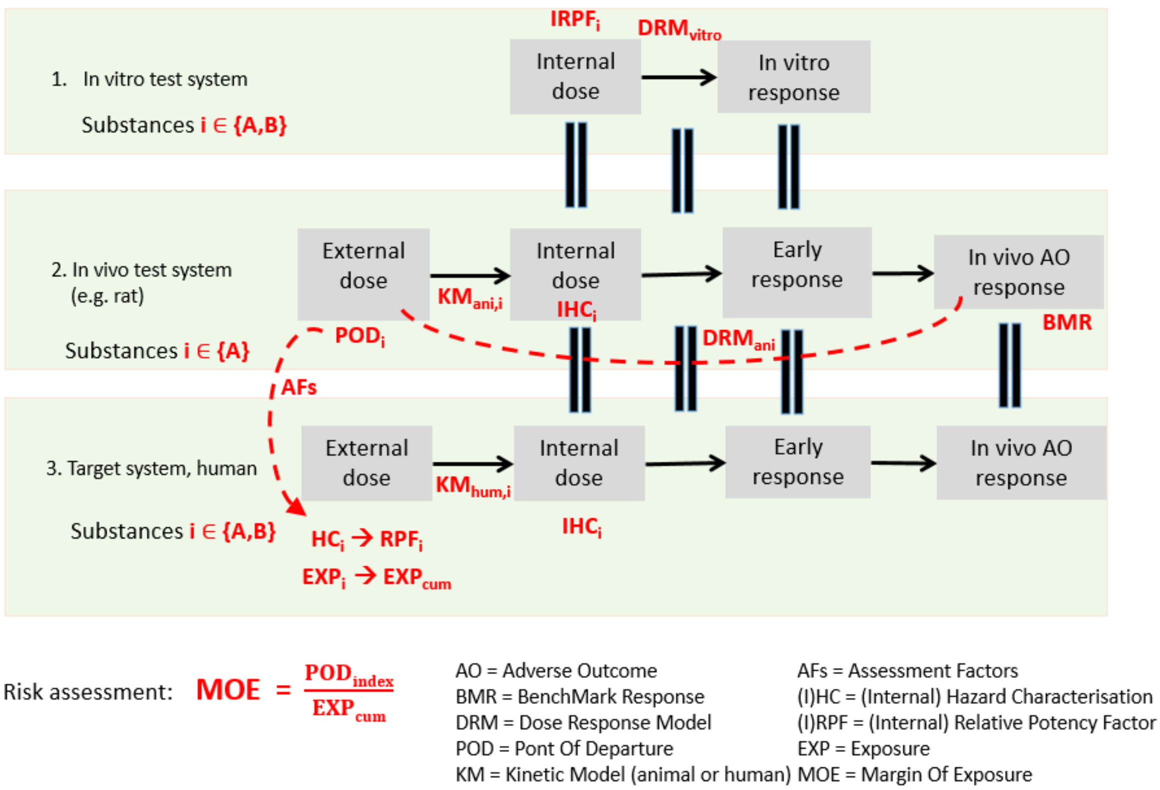


Figure 2.35: Conceptual model IVIVE.

IVIVE for calculating internal hazard characterisations

1. Translate the (external) PoD of the index substance substance to an internal hazard characterisation for the human target system/compartment.
2. If the RPFs are obtained are obtained using mol-based specification of the doses, then convert the mol-based RPFs to mass-based RPFs. I.e.,

$$RPF_{\mathtt{mass-based},i} = RPF_{\mathtt{mol-based},i} \cdot \frac{MW_{\mathtt{ref}}}{MW_i}$$

3. Derive internal hazard characterisations for the human target system for the other substances by multiplying the RPF obtained from dose-response modelling with the hazard characterisation of the index substance. I.e.,

$$HC_i = HC_{\mathtt{ref}} \cdot RPF_{\mathtt{mass-based},i}$$

IVIVE for calculating external hazard characterisations

1. Translate the PoD of the index substance to an external human hazard characterisation (dietary/oral exposure route).
2. Derive an internal hazard characterisation for the index substance, with an target organ/compartment representative for the response of the dose-response model.
3. If the RPFs are obtained are obtained using mol-based specification of the doses, then convert the mol-based RPFs to mass-based RPFs.
4. Derive internal hazard characterisations for the human target system for the other substances by multiplying the RPF obtained from dose-response modelling with the hazard characterisation of the index substance.
5. Convert the internal hazard characterisations of the other substance to external hazard characterisations for the dietary/oral exposure route using.

Kinetic conversion of hazard characterisations

When the *hazard characerisation level* is internal and points of departure are available for external exposures (e.g., NOAELs from in-vivo animal studies) or when the hazard characterisation level is external and benchmark doses are available at the internal level, then *kinetic conversion models* are needed to *translate the external doses to equivalent internal doses at the target compartment/organ* of interest or *vice-versa*.

In the toolbox, this alignment from internal to external or from external to internal is generally termed *kinetic conversion*, associated with a *kinetic conversion factor*. The kinetic conversion factor is a multiplication factor needed to obtain a hazard characterisation on the target level from a hazard characterisation of the point of departure or benchmark dose. Depending on the chosen kinetic modelling tier, this kinetic conversion factor may be 1) assumed to be one, 2) derived from absorption factors, or 3) derived using PBPK models.

An important detail in the use of kinetic conversion factors for computing hazard characterisations is the order between kinetic conversion and inter-species extrapolation. Notice that when points of departure are determined for animals, a choice should be made regarding the order of inter-species extrapolation and kinetic modelling. That is, one may first choose to convert animal external point of departure to an internal hazard characterisation for that animal, using the available animal kinetic model. Alternatively, one may first extrapolate the animal external point of departure to a human external hazard characterisation, and thereafter apply the human kinetic model to obtain internal hazard characterisations. In the toolbox, only the latter approach is currently implemented.

Extrapolation from external to internal hazard characterisations

The calculation of internal hazard characterisations based on external hazard characterisations is similar to the procedure for *computing internal exposures*. In the simplest tier, equivalence can be assumed between internal and external hazard characterisations, and in higher tiers absorption factors, respectively PBPK models can be used.

Calculation of internal doses using absorption factors

In the simplest form, internal doses are obtained from external exposure concentrations using multiplication factors (or, absorption factors) that can be specified by substance and by route. That is, for a given substance, the internal hazard characterisation $HC_{\mathtt{int}}$ can be derived from an external hazard characterisation $HC_{\mathtt{ext}}$ as

$$HC_{\mathtt{int}} = f_{\mathtt{abs},r} \cdot HC_{\mathtt{ext},r}$$

Here, r denotes the route of the external exposure $HC_{\texttt{ext}}$, and $f_{\texttt{abs},r}$ denotes the absorption factor of route r for the specified compartment. Note that this model assumes that the external hazard characterisations are specified as concentrations (i.e., substance amount divided by the body weight).

Calculation of internal doses using human PBPK models

A more detailed alternative to using absorption factors is to use one of the *advanced PBPK models* available in MCRA. In this approach, for each substance independently, an external exposure equivalent to the dose of the external hazard characterisation is presented to a representative simulated individual for a number of simulated days to the PBPK model of the individual. This representative individual should represent the "average" individual of the population, with nominal physiological properties (e.g., an average bodyweight of 70kg). This yields a time course of the internal substance amount at the specified target compartment/organ from which a long term average substance amount (chronic) or peak substance amount (acute) can be obtained. By dividing this substance amount by the weight of the compartment, an internal concentration is obtained, which then represents the internal hazard characterisation.

More details on computing internal doses from external doses can be found in the description of the *calculation of internal exposures from external exposures*. For both tasks, the procedure for computing internal exposures/doses is exactly and the same *kinetic model settings*, such as *dosing patterns* and *non-stationary period* period apply for calculation of internal hazard characterisations as well.

Calculation of internal doses using animal PBPK models

In the above methods, the assumption is that the external points of departure (often obtained from experiments on animals) are first converted to external hazard characterisations for humans, and a human kinetic model is used for obtaining the internal hazard characterisations. As mentioned, an alternative approach is to use first the animal PBPK models to derive an internal hazard characterisation specific for the tested animal species and thereafter extrapolate to humans. When there are more precise kinetic models available for the animal used in the experiments for obtaining the point of departure, this could be a prefered path.

Note: Notice that this procedure is not yet implemented.

Extrapolation from internal to external hazard characterisations

In some cases, hazard characterisations are available at the internal level whereas the specified *hazard characerisation level* is external. This situation may occur, for instance, in *in-vitro in-vivo extrapolation (IVIVE)*. In this case, conversion is needed from the internal level to the external level, where the external level is implicitly defined as comming from the dietary/oral route of exposure.

When using absorption factors, the external (dietary) hazard characterisation of a substance is simply computed by dividing the internal hazard characterisation by the dietary absorption factor. I.e.,

$$HC_{\texttt{ext,diet}} = \frac{HC_{\texttt{int}}}{f_{\texttt{abs,diet}}}$$

When using PBPK models, reverse dosimetry is needed to find for the available internal hazard characterisation, the corresponsing external (dietary) doses that yield the internal concentrations specified by the internal hazard characterisation. In MCRA, this is done using a bisection method, in which external doses are systematically fed to the PBPK model in order to converge to an external dose that yields the specified internal hazard characterisation with some level of precision.

Hazard characterisation imputation

In some cases it may be that there are substances that are known to cause (or may possibly cause) the effect of interest, but for which there are no data available for obtaining hazard characterisations. I.e., for these substances, there are no points of departure or dose response models. Instead of excluding these substances in quantitative analyses, it is also possible to impute hazard characterisations for these substances based on hazard characterisations of other (similar) substances, and use these for calculating, e.g., relative potency factors or for risk assessment.

Munro P5 (TTC approach)

The Threshold of Toxicological Concern (TTC) is an example of a tier for extrapolation of hazard characterisations from other compounds that is already in common use (see [34]). The *Munro collection of NOELs/LOAELs* is a collection of NOELs/LOAELs for chemicals for the critical (i.e., first occurring) effect. In the TTC approach, the toxicity of an unknown substance is, depending on its Cramer class (see [13]), imputed by the 5th percentile NOAEL of the sub-collection of the corresponding Cramer class.

Two variations of this approach are to use the empirical NOAEL distribution itself (just sample from the NOAEL data), or to fit a distribution (e.g. lognormal) to the empirical data and sample from the parametric distribution. MCRA provides an implementation of the TTC approach that uses the empirical distribution. In the nominal run, this implementation imputes the hazard characterisations with a value equivalent to the TTC. In the uncertainty runs, NOAELs are sampled from the empirical distribution.

The TTC is a conservative estimate of the NOAEL for at least two reasons:

1. TTCs are calculated from a collection of NOELs for the critical (i.e., first occurring) effect within each study and often the effect of interest will not be the critical effect, and therefore higher NOAELs are expected.
2. The TTC is a low percentile and therefore a conservative estimate for a random class member with unknown NOAEL.

Munro central value

To avoid the conservatism of taking the 5th percentile in the Munro P5 approach, alternatively, a nominal (or central) value could be taken from the Munro collection for each Cramer class. For a nominal run without uncertainty, the expected contribution of a substance with missing hazard characterisation to the risk as quantified in the hazard index is obtained from

$$HI = SF \cdot \sum_{i}^{n} \frac{\exp_i}{HC_i}$$

Here SF are all combined safety factors. It follows from this equation that an unbiased estimate for the contribution from a substance with missing hazard characterisations is obtained by taking the harmonic mean from the available NOAELs:

$$NOAEL = \left(\sum_{i=1}^{n} \frac{1}{NOAEL_i} \right)^{-1}$$

This is the value to use in a nominal run without uncertainty for the Munro central value approach. For the uncertainty runs, this approach also uses random sampling from the empirical distribution of the corresponding Cramer class.

Available hazard characterisations distribution P5

Another conservative aspect of the TTC approach is the fact that the Munro set lists NOELs/LOAELs for critical effects, not for the specific effect under study. Therefore an alternative is to use the effect-specific hazard characterisations of the substances for which these are available. This collection will have on average higher NOAELs than those of the Munro NOEL collection, because for many substances, the effect of interest will not be the critical effect.

Available hazard characterisations distribution central value

Similar to the Munro central value approach, a central value could also be obtained from the set of effect-specific hazard characterisations distribution for imputation of hazard characterisations. This approach may yield the most realistic, or unbiased imputation value for missing hazard characterisations.

Aggregation over multiple available hazard characterisations

In some scenarios, it may be that for a given substance and effect there are multiple available hazard characterisations. This can happen, for instance, if there are two different NOAELs originating from different studies. In such cases, a single hazard characterisation should be derived from the available candidates.

A conservative approach is to choose the lowest hazard characterisation of the available hazard characterisations. I.e.,

$$HC = \min_{i=1,\ldots,n} HC_i$$

Alternatively, it is possible to aggregate the candidates into a new "average" hazard characterisation. For this, the harmonic mean, also used for obtaining central value estimates in the *imputation of missing hazard characterisations*, is a suitable approach.

$$HC = \left(\sum_{i}^{n} \frac{1}{HC_i} \right)^{-1}$$

Hazard characterisations settings

Calculation settings

Table 2.140: Calculation settings for module Hazard characterisations.

Name	Description
Risk type	The type of exposure considered in the assessment; acute (short term) or chronic (long-term).
Target level	Select to express hazard characterisations at external or internal exposure level.
Method	Choose method for computing the hazard characterisations: from in-vivo or in-vitro points of departure or both.
Expression type	Specifies how hazard characterisations are expressed: as BMD, as NOAEL, or the expression type is ignored.
Selection method in case of multiple candidate hazard characterisations	Choose either the most toxic (default) or an aggregated hazard characterisation when in nominal runs there are multiple available candidates. In uncertainty runs, multiple candidates are always resampled.
Impute missing hazard characterisations	If selected, missing hazard characterisations are imputed based on Munro NOELs or on other available points of departure.
Imputation method	Imputation of Hazard characterisations: use low percentile (P5) or unbiased central estimate from either the Munro set or the available POD collection.
Use BMDs from dose response models	If selected, preferably BMDs from dose response models will be used. Only if these are not available, other POD data are used.
Use inter-species conversions	If selected, inter-species conversion factors ill be used (default value, e.g. 10, or data).
Use intra-species factors	If selected, intra-species conversion factors will be used (default value, e.g. 10, or data).

Uncertainty settings

Table 2.141: Uncertainty settings for module Hazard characterisations.

Name	Description
Resample intra-species factor	Specifies whether intra-species factors are resampled from a parametric uncertainty distribution.
Resample hazard characterisations or RPFs	Specifies whether to resample the hazard characterisations or relative potency factors. Requires hazard characterisation or RPF uncertainty to be quantified in DoseResponseModelsUncertain or RelativePotencyFactorsUncertain tables.

Calculation of hazard characterisations

- *Hazard characterisations calculation*

Inputs used: *Active substances Dose response models Effect representations Points of departure Inter-species conversions Intra species factors Kinetic models*

Settings used

- *Calculation Settings*

2.5.7 Inter-species conversions

Inter-species conversions specify how to convert a hazard characterisation for a given species to a hazard characterisation for humans. In the simplest approach, this specifies a fixed inter-species factor. In a higher tier, this specifies a geometric mean (GM) and geometric standard deviation (GSD) for a lognormal uncertainty distribution of the interspecies factor. Inter-species conversion are specified per effect and can be general or substance-specific.

This module has as primary entities: *Substances Effects*

Output of this module is used by: *Hazard characterisations*

Inter-species conversions data formats

Inter-species conversions

Inter-species conversion models specify how to convert a hazard dose for a given species to a hazard dose for humans.

Inter-species model parameters

Inter-species extrapolation factors are described using a lognormal distribution specified by a geometric mean (GM) and geometic standard deviation (GSD). Inter-species factors are defined for an effect and a species and may optionally be specified specifically for a substance.

Table 2.142: Table definition for InterSpeciesModelParameters.

Name	Type	Description	Aliases	Required
idEffect	AlphaNumeric(50)	The code of the effect for which this inter-species model is defined.	idEffect, EffectId, Effect	Yes
idSubstance	AlphaNumeric(50)	The code of the substance for which this inter-species model is defined.	idSubstance, SubstanceId, SubstanceCode, Substance, idCompound, CompoundId, Compound-Code, Compound	No
Species	AlphaNumeric(50)	Species	Species	Yes
InterSpecies-GeometricMean	Numeric	Interspecies geometric mean.	InterSpecies-GeometricMean, InterSpeciesGM	Yes
InterSpecies-Geometric-Standard-Deviation	Numeric	Interspecies geometric standard deviation.	InterSpecies-Geometric-Standard-Deviation, InterSpeciesGS-D	Yes
Standard-HumanBody-Weight	Numeric	The standard human body weight.	Standard-HumanBody-Weight	Yes
HumanBody-WeightUnit	AlphaNumeric(50)	The unit of the human body weight speciication (kg is assumed if not defined).	HumanBody-WeightUnit	No
Standard-AnimalBody-Weight	Numeric	The standard animal body weight.	Standard-AnimalBody-Weight	Yes
AnimalBody-WeightUnit	AlphaNumeric(50)	The unit of the animal body weight specification (kg is assumed if not defined).	AnimalBody-WeightUnit	No

Table aliases: InterSpeciesModelParameters, InterSpeciesModelParameter, InterSpeciesFactors, InterSpeciesFactor, RawInterSpeciesModelParameters.

Inter-species conversions settings

Selection settings

Table 2.143: Selection settings for module Inter-species conversions.

Name	Description
Default interspecies factor geometric mean	Default interspecies factor geometric mean.
Default interspecies factor geometric standard deviation	Default interspecies factor geometric standard deviation.

Uncertainty settings

Table 2.144: Uncertainty settings for module Inter-species conversions.

Name	Description
Resample inter-species factor	Specifies whether inter-species factors are resampled from a parametric uncertainty distribution.

Inter-species conversions as data

- *Inter-species conversions data formats*

2.5.8 Intra species factors

Intra-species factors specify how to convert a hazard characterisation from the average to a sensitive human individual.

This module has as primary entities: *Substances Effects*

Output of this module is used by: *Hazard characterisations*

Intra-species factors data formats

Intra-species factors

Intra-species factors.

Intra-species model parameters

Intra species factors.

Table 2.145: Table definition for IntraSpeciesModelParameters.

Name	Type	Description	Aliases	Required
idEffect	AlphaNumeric(50)	The effect code.	idEffect, EffectId, Effect	Yes
idSubstance	AlphaNumeric(50)	The code of the substance.	idCompound, CompoundId, Compound-Code, Compound, idSubstance, SubstanceId, SubstanceCode, Substance	No
IntraSpecies-Lower-VariationFactor	Numeric	The lower variability factor. The lower and upper factor are used to derive a geometric standard deviation (gsd) and degrees of freedom (df).	IntraSpecies-LowerVariation-Factor	No
IntraSpecies-UpperVariation-Factor	Numeric	The upper variability factor. The lower and upper factor are used to derive a geometric standard deviation (gsd) and degrees of freedom (df).	IntraSpecies-UpperVariation-Factor	Yes
idPopulation	AlphaNumeric(50)	Unique identification code of the population.	IdPopulation, PopulationId	No

Table aliases: IntraSpeciesModelParameters, IntraSpeciesModelParameter, IntraSpeciesFactors, IntraSpeciesFactor, RawIntraSpeciesModelParameters.

Intra species factors settings

Selection settings

Table 2.146: Selection settings for module Intra species factors.

Name	Description
Default intra-species factor	Default intra-species (safety) factor.

Intra species factors as data

In the simplest approach, intra-species factors are fixed factors. In a higher tier, lower and upper values for the intra-species factor are used to derive a variability distribution (log-normal around 1) and an uncertainty distribution for the geometric standard deviation related to human variability in sensitivity.

- *Intra species factors data formats*

2.5.9 Points of departure

Externally specified points of departure can be used as an alternative to calculated BMDs from dose response models. Points of departure can be of various types, such as NOAEL, LOAEL or BMD. They can be used to construct the list of active substances, to derive relative potency factors, and to perform health impact assessments.

This module has as primary entities: *Substances Effects*

Output of this module is used by: *Active substances Hazard characterisations*

Points of departure data formats

Points of departure

Points of departure, such as NOAELS and BMDs, describe the critical/reference levels of substance dose in relation to the presence or absence of an effect. If available, the uncertainty of externally specified points of departure can be specified with uncertainty sets (empirical distributions representing possible values) in the points of departure uncertainty table.

Points of departure

Nominal points of departure should be presented in this table. Each point of departure should be linked to an effect using the effect code (idEffect) and to substances using the substance code (idSubstance).

Table 2.147: Table definition for HazardDoses.

Name	Type	Description	Aliases	Required
idModel	AlphaNumeric(50)	The dose response model code.	idDose-ResponseModel, idModel	No
idEffect	AlphaNumeric(50)	The effect code.	idEffect, EffectId, Effect	Yes
idSubstance	AlphaNumeric(50)	The code of the substance.	idSubstance, SubstanceId, SubstanceCode, Substance, idCompound, CompoundId, Compound-Code, Compound	Yes
Species	AlphaNumeric(50)	The species used to obtain this point of departure.	Species, System	No
DoseResponse-ModelEquation	AlphaNumeric(500)	The model description of the dose response model (R model equation).	DoseResponse-ModelEquation	No
DoseResponse-Model-Parameter-Values	AlphaNumeric(200)	A comma separated list of the values of the parameters of the model, format: a=1.2,b=3.4,c=5.6	DoseResponse-Model-ParameterValues	No
Point of departure	Numeric	Point of departure, can be of various types, e.g. NOAEL, LOAEL, BMD, CED	PointOf-Departure, LimitDose, HazardDose, CED	Yes
Point of departure type	*HazardDoseTypes*	The type of the point of departure: e.g. NOAEL, LOAEL, BMD (default).	PODType, HazardDose-Type, LimitDoseType	No
DoseUnit	AlphaNumeric(50)	The dose unit (if not specified, then mg/kg is assumed).	DoseUnit, UnitDose	No
Benchmark response (BMR)	AlphaNumeric(100)	The effect size.	Benchmark-Response, CriticalEffect-Size, HazardEffect-Size	No
ExposureRoute	AlphaNumeric(100)	The route of dose administration used in the study to obtain this point of departure. If not specified exposure route = Dietary is assumed.	ExposureRoute, RouteExposure	No

Table aliases: PointsOfDeparture, PointOfDeparture, HazardDoses, HazardDose, RawHazardDoses.

Points of departure uncertainty

Often, the PODs found for a substance/effect combination are uncertain. This table facilitates in specifying the POD uncertainty in the form of a set of uncertainty values that may additionally be specified for a substance/effect combination.

Table 2.148: Table definition for HazardDosesUncertain.

Name	Type	Description	Aliases	Required
idDose-ResponseModel	AlphaNumeric(50)	The dose response model code (must correspond to values in id column of DoseResponseModels table).	idDose-ResponseModel	Yes
idUncertainty-Set	AlphaNumeric(50)	The identification code of the uncertainty set. During an uncertainty iteration one set will be picked to be the POD value.	idUncertainty-Set, UncertaintyId	Yes
idEffect	AlphaNumeric(50)	The effect code.	idEffect, EffectId, Effect	Yes
idSubstance	AlphaNumeric(50)	The code of the substance.	idSubstance, SubstanceId, SubstanceCode, Substance, idCompound, CompoundId, Compound-Code, Compound	Yes
Point of departure	Numeric	Point of departure, can be of various types, e.g. NOAEL, LOAEL, BMD, CED	PointOf-Departure, HazardDose, LimitDose, CED	Yes
DoseResponse-Model-Parameter-Values	AlphaNumeric(200)	A comma separated list of the values of the parameters of the model, format: a=1.2,b=3.4,c=5.6	DoseResponse-Model-Parameter-Values, ParameterValues	No

Table aliases: PointsOfDepartureUncertain, PointOfDepartureUncertain, HazardDosesUncertain, HazardDoseUncertain, RawHazardDosesUncertain.

Points of departure settings

Uncertainty settings

Table 2.149: Uncertainty settings for module Points of departure.

Name	Description
Resample hazard characterisations or RPFs	Specifies whether to resample the hazard characterisations or relative potency factors. Requires hazard characterisation or RPF uncertainty to be quantified in DoseResponseModelsUncertain or RelativePotencyFactorsUncertain tables.

Points of departure as data

Points of departure are provided as data for combinations of susbstance and effect and each is minimally described by a reference value and a type (e.g., NOAEL or LOAEL). In addition, the exposure route, specifes, and references may be specified.

- *Points of departure data formats*

Inputs used: *AOP networks*

2.5.10 Relative potency factors

Relative potency factors (RPFs) quantify potencies of substances with respect to a defined effect, relative to the potency of a chosen index substance. RPFs can be used to express combined exposures of multiple substances in terms of a the exposure value of the chosen index substance (i.e., in index substance equivalents). In MCRA, hazard characterisations, and therefore also RPFs are based on mass units (e.g., µg), and not on mol units. RPFs can be different for different levels of the human organism (external, internal, specific compartment). RPFs can be given as data or computed from hazard characterisations. RPFs can be specified with uncertainty. Computation from uncertain hazard characterisations allows to include correlations between uncertain RPFs which originate from using the same index substance.

This module has as primary entities: *Substances Effects*

Output of this module is used by: *Concentrations Concentration models Dietary exposures with screening Dietary exposures Exposures*

Relative potency factors data formats

Relative potency factors

Relative potency factors quantify relative potencies of substances with respect to an effect and can be used to express combined exposures of multiple substances in terms of the exposure value of the chosen index substance (i.e., as index substance equivalents). Relative potency factors can be provided in case hazard characterisations are missing. If available, the uncertainty of externally specified RPFs can be specified with uncertainty sets (empirical distributions representing possible values) in an additional table.

Relative potency factors

Relative potency factors are linked to an effect using the effect code (idEffect) and to substances using the substance code (idSubstance).

Table 2.150: Table definition for RelativePotencyFactors.

Name	Type	Description	Aliases	Required
idSubstance	AlphaNumeric(50)	The code of the substance.	idSubstance, SubstanceId, SubstanceCode, Substance, idCompound, CompoundId, Compound-Code, Compound	Yes
idEffect	AlphaNumeric(50)	The effect code.	idEffect, EffectId, Effect	Yes
RPF	Numeric	The relative potency factor.	RPF, Relative-PotencyFactor	Yes

Table aliases: RelativePotencyFactors, RelativePotencyFactor, RawRelativePotencyFactors.

Relative potency factor uncertainty

This table contains sets of values representing the uncertainty for relative potency factors.

Table 2.151: Table definition for RelativePotencyFactorsUncertain.

Name	Type	Description	Aliases	Required
idUncertainty-Set	AlphaNumeric(50)	The uncertainty set identification number. During each uncertainty iteration one set is used.	idUncertainty-Set, UncertaintyId	Yes
idEffect	AlphaNumeric(50)	The effect code (must correspond to values in id column of Effects table).	idEffect, EffectId, Effect	Yes
idSubstance	AlphaNumeric(50)	The substance code (must correspond to values in id column of Substances table).	idSubstance, SubstanceId, SubstanceCode, Substance, idCompound, CompoundId, Compound-Code, Compound	Yes
RPF	Numeric	The relative potency factor.	RPF, Relative-PotencyFactor	Yes

Table aliases: RelativePotencyFactorsUncertain, RelativePotencyFactorUncertain, RawRelativePotencyFactorsUncertain.

Relative potency factors calculation

Relative potency factors (RPFs) describe the potency of substances with respect to a defined effect, relative to the potency of a chosen index substance. RPFs can be given as data or computed from *hazard characterisations*. The RPF for substance i is defined by the ratio of hazard characterisation value for the index substance (`ref`) and the hazard characterisation value for substance i. That is,

$$\mathtt{RPF}_i = \mathtt{POD}_{\mathtt{ref}}/\mathtt{POD}_i.$$

When the hazard characterisations are resampled in the uncertainty runs, RPFs are also recomputed based on the bootstrapped hazard characterisations. In this way, RPF uncertainty can also included in the uncertainty analysis.

Relative potency factors settings

Uncertainty settings

Table 2.152: Uncertainty settings for module Relative potency factors.

Name	Description
Resample hazard characterisations or RPFs	Specifies whether to resample the hazard characterisations or relative potency factors. Requires hazard characterisation or RPF uncertainty to be quantified in DoseResponseModelsUncertain or RelativePotencyFactorsUncertain tables.

Relative potency factors as data

- *Relative potency factors data formats*

Inputs used: *AOP networks*

Calculation of relative potency factors

- *Relative potency factors calculation*

Inputs used: *Hazard characterisations*

2.6 In-silico modules

Two types of in-silico models are available: QSAR models specify assessment group memberships for active substances, as numbers in the interval [0,1]. This allows both crisp (0 or 1) and probabilistic memberships. Molecular docking models specify binding energies and thresholds which can be used to convert binding energies to assessment group memberships for active substances.

2.6.1 Molecular docking models

Molecular docking models specify binding energies for substances in specific molecular docking models related to a specific health effect (adverse outcome).

This module has as primary entities: *Substances Effects*

Output of this module is used by: *Active substances*

Molecular docking models data formats

Required data tables:

- Molecular docking models, to identify models for a specified effect (receptor)
- Molecular docking binding energies, to specify the binding energies per substance for the receptor

Molecular docking models

Contains definitions of molecular docking models for a given effect (molecular initiating event), for example parameters needed in the conversion of binding energies to group memberships or to relative potency factors. Substance specific binding energies are specified in the binding energies table.

Molecular docking models

Each docking model has a unique identifier, and optionally a name and a description. Each model is linked to an effect using the idEffect field and optionally a binding threshold and the number of receptors can be added. A reference to the source of the data can be stored in the reference field.

Table 2.153: Table definition for MolecularDockingModels.

Name	Type	Description	Aliases	Required
id	AlphaNumeric(50)	The unique identification code of the molecular docking model.	idMolecular-DockingModel, idBinding-EnergyModel	Yes
Name	AlphaNumeric(100)	The name of the molecular docking model.	Name	No
Description	AlphaNumeric(200)	Description of the molecular docking model.	Description	No
idEffect	AlphaNumeric(50)	The effect code, typically for the Molecular Initiating Event that is modelled	idEffect, EffectId, Effect	Yes
Threshold	Numeric	Threshold Molecular Docking binding energy (group membership = 1 when higher).		No
NumberOf-Receptors	Integer	Example parameter needed for translating Molecular Docking binding energies to RPFs.		No
Reference	AlphaNumeric(200)	External reference(s) to sources containing more information about the molecular docking model.	References	No

Table aliases: MolecularDockingModels, MolecularDockingModel, BindingEnergyModels, BindingEnergyModel, RawMolecularDockingModels.

Molecular docking binding energies

Molecular docking model binding energies per substance

Table 2.154: Table definition for MolecularBindingEnergies.

Name	Type	Description	Aliases	Required
idMolecular-DockingModel	AlphaNumeric(50)	The id of the molecular docking model or source.	idMolecular-Docking, Molecular-DockingModel	No
idSubstance	AlphaNumeric(50)	The code of the substance.	idSubstance, SubstanceId, SubstanceCode, Substance, idCompound, CompoundId, Compound-Code, Compound	Yes
BindingEnergy	Numeric	Molecular Docking binding energy.	Molecular-Docking-BindingEnergy	Yes

Table aliases: MolecularBindingEnergies, MolecularBindingEnergy, BindingEnergies, BindingEnergy, MolecularDockingBindingEnergies, MolecularDockingBindingEnergy, RawMolecularBindingEnergies.

Molecular docking models as data

- *Molecular docking models data formats*

Inputs used: *AOP networks*

2.6.2 QSAR membership models

QSAR membership models specify assessment group memberships for active substances related to a specific health effect (adverse outcome). Memberships should be derived externally from Quantitative Structure-Activity Relationship (QSAR) models.

This module has as primary entities: *Substances Effects*

Output of this module is used by: *Active substances*

QSAR membership models data formats

Required data tables:

- QSAR membership models, to identify QSAR models for a specified health effect
- QSAR membership scores, to specify the memberships per substance per QSAR model

Note that only memberships 1 (include) and 0 (exclude) are allowed.

QSAR membership models

Substance membership models obtained from QSAR for a given (health) effect. The models are defined in the membership models table, and substance specific memberships are specified in the QSAR memberships table.

QSAR membership models

This table contains the definitions of the QSAR membership models. Each model contains a id, name, an optional description, and refers to its related health effect.

Table 2.155: Table definition for QSARMembershipModels.

Name	Type	Description	Aliases	Required
id	AlphaNumeric(50)	The unique identification code of the QSAR membership model.	id, Model, ModelCode, idModel, QSARModel, idQSARModel, QSAR-Membership-Model, idQSAR-Membership-Model, Membership-Model, idMembership-Model	Yes
Name	AlphaNumeric(100)	The name of the QSAR membership model.	Name	No
Description	AlphaNumeric(200)	Description of the QSAR membership model.	Description	No
idEffect	AlphaNumeric(50)	The effect code.	idEffect, EffectId, Effect	Yes
Accuracy	Numeric	Accuracy of the QSAR membership model.	Accuracy	No
Sensitivity	Numeric	Sensitivity of the QSAR membership model.	Sensitivity	No
Specificity	Numeric	Specificity of the QSAR membership model.	Specificity	No
Reference	AlphaNumeric(200)	External reference(s) to sources containing more information about the QSAR model.	References	No

Table aliases: QSAR, QSARMembershipModels, QSARMembershipModel, QSARModels, QSARModel, RawQSARMembershipModels.

QSAR membership scores

Substance membership score according to the QSAR model.

Table 2.156: Table definition for QSARMembershipScores.

Name	Type	Description	Aliases	Required
idQSAR-Membership-Model	AlphaNumeric(50)	The id of the QSAR model.	Model, ModelCode, idModel, QSARModel, idQSARModel, QSAR-Membership-Model, idQSAR-Membership-Model, Membership-Model, idMembership-Model	Yes
idSubstance	AlphaNumeric(50)	The code of the substance.	idSubstance, SubstanceId, SubstanceCode, Substance, idCompound, CompoundId, Compound-Code, Compound	Yes
Membership-Score	Numeric	QSAR membership score. Value should be 1 for positive membership, or 0 for negative membership.	Membership-Score, Membership, QSARScore, Score	Yes

Table aliases: QSARMembershipScores, QSARMembershipScore, QSARMemberships, QSARMembership, RawQSARMembershipScores.

QSAR membership models as data

- *QSAR membership models data formats*

Inputs used: *AOP networks*

2.7 Kinetic modules

Kinetic models convert exposures or hazard characterisations from one or more external routes or compartments to an internal (target) compartment. The reverse conversion from internal to external can also be made (reverse dosimetry).

In a simple tier, kinetic models are specified as absorption factors. In a higher tier, physiologically based toxicokinetic (PBTK) models of a specified type (currently available is the EuroMix generic PBTK model) are linked to MCRA. Kinetic model instances for specific substances and test systems (e.g. cypermethrin in the rat) are specified with parameter sets for the chosen kinetic model.

2.7.1 Kinetic models

External exposure can be from on more more exposure routes: oral (dietary or non-dietary), dermal or inhalation. Internal exposure can be systemic or related to a specific compartment in a kinetic model. There are four tiers for relating external to internal exposures (doses):

1. Assume 100% absorption: internal exposures are equal to external exposures.
2. Assume conservative absorption factors as suggested by EFSA ([4], [5]): oral and inhalation 100%, dermal 50%.
3. Use externally provided absorption factors (*absorption factors data tables*).
4. Use one of the *implemented kinetic models*, with instances for specific substances defined in data table *kinetic model instances* and model parameters specified in data table *kinetic model instance parameters*.

Given a chosen tier, the caculation will fall back to the next lower tier in case of missing data.

This module has as primary entities: *Substances*

Output of this module is used by: *Exposures Hazard characterisations*

Kinetic models data formats

Data tables:

- Absorption factors
- Kinetic model instances
- Kinetic model instance parameters

Kinetic models

Kinetic models may be specified as kinetic model instances that contain parameter specifications of built in kinetic models or as simple absorption factors.

Kinetic model instances

Kinetic model instances.

Table 2.157: Table definition for KineticModelInstances.

Name	Type	Description	Aliases	Required
idModel-Instance	AlphaNumeric(50)	Unique identification code of the kinetic model instance.	idModel-Instance, Id, Code	Yes
idModel-Definition	KineticModelType	Identifier of the kinetic model definition for which this is an instance.	idModel-Definition, ModelDefinition	Yes
idTestSystem	AlphaNumeric(200)	The species on which the experiment was performed.	System, TestSystem	Yes
idSubstance	AlphaNumeric(50)	Unique identification code of substance, Default: valid for all substances. Should be omitted for parameters in the class Physiological	idCompound	No
Reference	AlphaNumeric(100)	Reference or author.	References	No

Table aliases: KineticModelInstances, KineticModelInstance, RawKineticModelInstances.

Kinetic model instance parameters

Kinetic model parameters

Table 2.158: Table definition for KineticModelInstanceParameters.

Name	Type	Description	Aliases	Required
idModel-Instance	AlphaNumeric(50)	Unique identification code of the kinetic model instance to which this parameter belongs	Id, Code	Yes
Parameter	AlphaNumeric(100)	Name of the parameter in the kinetic model.		Yes
Description	AlphaNumeric	Description of or reference for the parameter values in the kinetic model.		No
Value	Numeric	Mean.	MEAN, mean	Yes
Distribution-Type	AlphaNumeric(20)	Distribution.	Distribution-Type, Distribution	No
CvVariability	Numeric	Variability.		No
CvUncertainty	Numeric	Uncertainty.		No

Table aliases: KineticModelInstanceParameters, KineticModelInstanceParameter, RawKineticModelInstanceParameters.

Kinetic model absorption factors

Kinetic absorption factors

Table 2.159: Table definition for KineticAbsorptionFactors.

Name	Type	Description	Aliases	Required
idCompound	AlphaNumeric(50)	code of substance (must correspond to values in id column of Substances table)	idCompound	No
Route	AlphaNumeric(50)	Non-dietary route or pathway, use 'Oral', 'Dermal', or 'Inhalation' to specify the route.	Route, Pathway	No
Absorption-Factor	Numeric	absorption factor value	Absorption-Factor, Factor	No

Table aliases: KineticAbsorptionFactors, KineticAbsorptionFactor, AbsorptionFactors, AbsorptionFactor, RawKineticAbsorptionFactors.

Kinetic models settings

Calculation settings

Table 2.160: Calculation settings for module Kinetic models.

Name	Description
Default oral absorption factor for non-dietary exposure	When there is no kinetic model and absorption factors are not specified in file, non-dietary oral exposures (external doses) are multiplied by this factor to determine the absorbed (internal) dose.
Default oral absorption factor for dietary exposure	When there is no kinetic model and absorption factors are not specified in file, dietary exposures (external doses) are multiplied by this factor to determine the absorbed (internal) dose .
Default dermal absorption factor for non-dietary exposure	When there is no kinetic model and absorption factors are not specified in file, dermal exposures (external doses) are multiplied by this factor to determine the absorbed (internal) dose.
Default inhalation absorption factor for non-dietary exposure	When there is no kinetic model and absorption factors are not specified in file, inhalation exposures (external doses) are multiplied by this factor to determine the absorbed (internal) dose.
Number of days	The number of days.
Number of events per day for the ORAL dietary dose	The daily dose is administered in equal portions (dose / number of events) per event.
Number of initial days skipped	This period is skipped in the calculation of the mean internal exposure.
Kinetic model	Code Kinetic Model.
Use parameter variability	When specified, use parameter variability.

Uncertainty settings

Table 2.161: Uncertainty settings for module Kinetic models.

Name	Description
Resample kinetic model parameter values	Specifies whether kinetic model parameter values are resampled.

Kinetic models as data

- *Kinetic models data formats*

Inputs used: *Active substances*

Available kinetic models

Physiologically based toxicokinetic (PBTK) models, or kinetic models for short, are mathematical representations of the animal or human body aimed at describing and predicting the time course distribution of chemicals in tissues and organs. Those internal dose metrics can usefully replace external exposure dose in the derivation of the quantitative dose-response relationships and following risk assessments. PBTK models can simulate both internal doses from exposure scenarios (forward dosimetry) and external dose from biomonitoring data (reverse dosimetry).

The following generic PBTK models are currently implemented in MCRA:

- *EuroMix generic PBTK model* [12].
- *bisphenol model ETHZ* [25].

EuroMix Generic PBTK model v6

Cosmos version 6 (received 3/27/2019)

Table 2.162: Exposure routes (forcings)

Id	Description	Unit	Order
Dietary	Dietary exposure	mmoles	0
Dermal	Dermal exposure	mmoles	1
Inhalation	Inhalatory exposure	mmoles	2

Table 2.163: Output

Id	Description	ScalingFactor	Multiplication-Factor	Unit	Order
CTotal	Total concentration			mM	0
CVen	Venous blood concentration	scVBlood	0.66667	mM	1
CArt	Arterial blood concentration	scVBlood	0.33333	mM	2
CFat	Fat (adipose) tissue concentration	scVFat		mM	3
CPoor	Poorly perfused tissue (muscle) concentration			mM	4
CRich	Richly perfused tissue (viscera) concentration	scVRich		mM	5
CLiver	Liver concentration	scVLiver		mM	6
CSkin_u	Viable unexposed skin concentration			mM	7
CSkin_e	Viable exposed skin concentration	BSA, Height_vs, fsA_exposed		mM	8
CSkin_sc_u	Skin unexposed stratum corneum concentration			mM	9
CSkin_sc_e	Skin exposed stratum corneum concentration	BSA, Height_vs, fsA_exposed		mM	10

Table 2.164: Input

Id	Description	Unit	Type	Order
BM	Body mass	kg	Physiological	0
BSA	Body surface area (internally scaled by an allometric scaling factor s = 70/BM^0.3)	dm2	Physiological	1
scVFat	Fat as fraction of total body volume		Physiological	2
scVRich	Richly perfused tissues (viscera) as fraction of total body volume		Physiological	3
scVLiver	Liver as fraction of total body volume		Physiological	4
scVBlood	Blood as fraction of total body volume		Physiological	5
Height_sc	Skin thickness	decimeter	Physiological	6
Height_vs	Viable skin		Physiological	7
scFBlood	Total blood flow per unit mass	L/h/kg	Physiological	8
scFFat	Fat fraction of total blood flow going to compartments		Physiological	9
scFPoor	Poorly perfused tissues (muscles) fraction of total blood flow going to compartments		Physiological	10
scFLiver	Liver fraction of total blood flow going to compartments		Physiological	11
scFSkin	Skin fraction of total blood flow going to compartments		Physiological	12
Falv	Alveolar ventilation rate	L/h	Physiological	13
mic	Microsomal proteins content	mg/gr liver	Physiological	14
Kp_sc_vs	Diffusion rate from stratum corneum to viable skin	decimeter/h	Metabolic	22
Ke	Renal excretion rate	L/h	Metabolic	23
Michaelis	Flag for Michaelis-Menten vs linear metabolism (0 = linear)		Metabolic	24
Vmax	Maximum rate of metabolism	mmoles/h/L liver	Metabolic	25
Km	Michaelis-Menten constant for metabolism	mM	Metabolic	26
CLH	Hepatic metabolic clearance		Metabolic	27
fub	Unbound fraction in blood		Metabolic	28
Frac	Fraction absorbed by the gut		Metabolic	29
kGut	Oral 1st order absorption rate constant	1/h	Metabolic	30

Model aliases: Cosmos6, CosmosV6.

EuroMix Generic PBTK model v5

Cosmos version 5 (adapted 9/11/2018)

Table 2.165: Exposure routes (forcings)

Id	Description	Unit	Order
Dietary	Dietary exposure	mmoles	0
Dermal	Dermal exposure	mmoles	1
Inhalation	Inhalatory exposure	mmoles	2

Table 2.166: Output

Id	Description	ScalingFactor	Multiplication-Factor	Unit	Order
CVen	Venous blood	scVBlood	0.66667	mM	0
CArt	Arterial blood	scVBlood	0.33333	mM	1
CFat	Fat tissues	scVFat		mM	2
CPoor	Muscle tissues			mM	3
CRich	Viscera	scVRich		mM	4
CLiver	Liver	scVLiver		mM	5
CSkin_u	Viable skin, unexposed			mM	6
CSkin_e	Viable skin, exposed	BSA, Height_vs, fsA_exposed		mM	7
CSkin_sc_u	Skin stratum corneum, unexposed			mM	8
CSkin_sc_e	Skin stratum corneum, exposed	BSA, Height_vs, fsA_exposed		mM	9

Table 2.167: Input

Id	Description	Unit	Type	Order
BM	Body mass	kg	Physiological	0
BSA	Body skin surface area	dm2	Physiological	1
scVFat	Fat as fraction of total body volume		Physiological	2
scVRich	Richly perfused tissues (viscera) as fraction of total body volume		Physiological	3
scVLiver	Liver as fraction of total body volume		Physiological	4
scVBlood	Blood as fraction of total body volume		Physiological	5
Height_sc	Skin thickness	decimeter	Physiological	6
Height_vs	Viable skin		Physiological	7
scFBlood	Total blood flow per unit mass	L/h/kg	Physiological	8
scFFat	Fat fraction of total blood flow going to compartments		Physiological	9
scFPoor	Poorly perfused tissues (muscles) fraction of total blood flow going to compartments		Physiological	10
scFLiver	Liver fraction of total blood flow going to compartments		Physiological	11
scFSkin	Skin fraction of total blood flow going to compartments		Physiological	12
Falv	Alveolar ventilation rate	L/h	Physiological	13
mic	Microsomal proteins content	mg/gr liver	Physiological	14
PCAir	Partition coefficient: blood over air		Partition coefficient	15
Kp_sc_vs	Diffusion rate from stratum corneum to viable skin	decimeter/h	Metabolic	22
Ke	Renal excretion rate	L/h	Metabolic	23
Michaelis	Flag for Michaelis-Menten vs linear metabolism (0 = linear)		Metabolic	24
Vmax	Maximum rate of metabolism	mmoles/h/L liver	Metabolic	25
Km	Michaelis-Menten constant	mM	Metabolic	26
CLH	Hepatic clearance		Metabolic	27
fup	Unbound fraction in blood		Metabolic	28
Frac	Fraction absorbed by the gut		Metabolic	29
kGut	Oral 1st order absorption rate constant	1/h	Metabolic	30
fSA_exposed	Fraction of skin surface area actually exposed		Metabolic	35

Model aliases: Cosmos4, CosmosV4, Cosmos5, CosmosV5.

Generic Model BPA

Generic model Cecile Karrer 23 juli 2018

Table 2.168: Exposure routes (forcings)

Id	Description	Unit	Order
Dietary	Dietary exposure	nmoles	0
Oral	Oral exposure	nmoles	1
Dermal	Dermal exposure	nmoles	2
Inhalation	Inhalation exposure	nmoles	3

Table 2.169: Output

Id	Description	ScalingFactor	Multiplication-Factor	Unit	Order
CPlasmaOut	Concentration in plasma			nmol/L	0
CGonadOut	Concentration in gonads			nmol/L	1
AurinebpaOut	Cumulative excretion of BPA in urine			nmol/L	2
AurinegOut	Cumulative excretion of BPA-g in urine			nmol/L	3
AurineTotalOut	Cumulative excretion of BPA and metabolites in urine			nmol/L	4

Table 2.170: Input

Id	Description	Unit	Type	Order
BW	Bodyweight	kg	Physiological	0
QCC	Cardiac output	L/min	Physiological	1
QgonadC	Fractional blood flow to gonads		Physiological	2
QliverC	Fractional blood flow to liver		Physiological	3
QfatC	Fractional blood flow to fat tissue		Physiological	4
QbrainC	Fractional blood flow to brain		Physiological	5
QskinC	Fractional blood flow to skin		Physiological	6
QmuscleC	Fractional blood flow to gonads		Physiological	7
VplasmaC	Fractional volume of plasma		Physiological	8
VfatC	Fractional volume of fat tissue		Physiological	9
VliverC	Fractional volume of liver tissue		Physiological	10

Continued on next page

Table 2.170 – continued from previous page

Id	Description	Unit	Type	Order
VbrainC	Fractional volume of brain tissue		Physiological	11
VskinC	Fractional volume of skin tissue		Physiological	12
VgonadC	Fractional volume of gonads		Physiological	13
VmuscleC	Fractional volume of muscle tissue		Physiological	14
VrichC	Fractional volume of richly perfused tissue		Physiological	15
VbodygC	Distribution volume of BPA-g		Physiological	16
MW	Molecular weight	g/mol	Chemical property	18
pliver	Partition coefficient liver to blood		Partition coefficient	19
pfat	Partition coefficient fat to blood		Partition coefficient	20
pslow	Partition coefficient slowly perfused tissue to blood		Partition coefficient	21
prich	Partition coefficient richly perfused tissue to blood		Partition coefficient	22
pgonad	Partition coefficient gonads to blood		Partition coefficient	23
pbrain	Partition coefficient brain to blood		Partition coefficient	24
pskin	Partition coefficient skin to blood		Partition coefficient	25
geC	Gastric emptying	1/h/kg bw^-0.25	Metabolic	26
k0C	Oral uptake from the stomach into the liver	1/h/kg bw^-0.25	Metabolic	27
k1C	Oral uptake from the small intestine into the liver	1/h/kg bw^-0.25	Metabolic	28
k4C	Fecal elimination from small intestine after oral administration	1/h/kg bw^-0.25	Metabolic	29
kGIingC	Transport of glucuronide from enterocytes into serum	1/h/kg bw^-0.25	Metabolic	30
kGIinsC	Transport of sulfate from enterocytes into serum	1/h/kg bw^-0.25	Metabolic	31
kmgutg	Km of Glucuronidation in the gut	nM	Metabolic	32
vmaxgutgC	Vmax of Glucuronidation in the gut	nmol/h/kg bw	Metabolic	33

Continued on next page

Table 2.170 – continued from previous page

Id	Description	Unit	Type	Order
fgutg	Correction factor of glucuronidation in the gut		Metabolic	34
kmguts	Km of Sulfation in the gut	nM	Metabolic	35
vmaxgutsC	Vmax of Sulfation in the gut	nmol/h/kg bw	Metabolic	36
fguts	Correction factor of sulfation in the gut		Metabolic	37
met1g	Fraction of glucuronide in the liver taken up directly into serum (the rest undergoes EHR)		Metabolic	38
met1s	Fraction of sulfate in the liver taken up directly into serum		Metabolic	39
enterocytes	Sum of enterocytes weights in duodenum, jejunum and ileum	L	Metabolic	40
kmliver	Km of Glucuronidation in the liver	nM	Metabolic	41
vmaxliverC	Vmax of Glucuronidation in the liver	nmol/h/g liver	Metabolic	42
fliverg	Correction factor of glucuronidation in the liver		Metabolic	43
kmlivers	Km of Sulfation in the liver	nM	Metabolic	44
vmaxliversC	Vmax of Sulfation in the liver	nmol/h/g liver	Metabolic	45
flivers	Correction factor of sulfation in the liver		Metabolic	46
EHRtime	Time until EHR occurs	h	Metabolic	47
EHRrateC	EHR of glucuronide	1/h/kg bw^-0.25	Metabolic	48
k4C_IV	Fecal elimination of glucuronide from the EHR compartment	1/h/kg bw^-0.25	Metabolic	49
kurinebpaC	Clearance, urine excretion of parent compound	L/h/kg bw^0.75	Metabolic	50
kurinebpagC	Clearance, urine excretion of glucuronide	L/h/kg bw^0.75	Metabolic	51
kurinebpasC	Clearance, urine excretion of sulfate	L/h/kg bw^0.75	Metabolic	52
vreabsorptiong-C	Vmax for renal reabsorption of glucuronide	nmol/h/kg bw^0.75	Metabolic	53

Continued on next page

Table 2.170 – continued from previous page

Id	Description	Unit	Type	Order
vreabsorptionsC	Vmax for renal reabsorption of sulfate	nmol/h/kg bw^0.75	Metabolic	54
kreabsorptiong	Km for renal reabsorption of glucuronide	nM	Metabolic	55
kreabsorptions	Km for renal reabsorption of sulfate	nM	Metabolic	56
kenterobpagC	EHR of parent compound due to biliary excretion of glucuronide	1/h/kg bw^-0.25	Metabolic	57
kenterobpasC	EHR of parent compound due to biliary excretion of sulfate	1/h/kg bw^-0.25	Metabolic	58
EoA_O	Extent of oral absorption		Physiological	61
period_O	uptake period	h	External	63
t0_O	time point at which dosing starts	h	External	65
EoA_D	Extent of dermal absorption from TP		Physiological	68
aHL_D	Half-life for dermal penetration	h	External	70
period_D	Uptake period dermal exposure from TP	h	External	72
t0_D	Time points at which dermal dosing from TP starts	h	External	74
EoA_D2	Extent of dermal absorption from PCPs		Physiological	77
aHL_D2	Half-life for dermal penetration from PCPs	h	External	79
period_D2	Uptake period dermal exposure from PCPs	h	External	81
t0_D2	Time points at which dermal dosing from PCPs starts	h	External	83
BW075	BW^0.75	kg^0.75	External	103
BW025	BW^0.25	kg^0.25	External	104

Model aliases: PBPKModel_BPA, PBPKModelBPA, ModelBPA, BPA.

Note: Additional kinetic models can be implemented, please contact the MCRA administrator.

EuroMix generic PBTK model

Reference: Tebby et al, 2019: [12]

In MCRA updated versions (version 4b, 6) of the PBTK model developed at INERIS in the framework of the COSMOS project is used. The model describes the distribution of chemicals in venous blood, arterial blood, adipose tissues, poorly perfused tissues (muscles), gut lumen, liver, richly perfused tissues (other viscera), and skin. Each of those is described as a compartment (homogeneous virtual volume) in which distribution is instantaneous and limited only by the incoming blood flow or rate of entry in the compartment. Exposure can occur through the dermal route, ingestion or inhalation. The absorbed molecules can be excreted to urine, exhaled through the lung, or metabolized in liver.

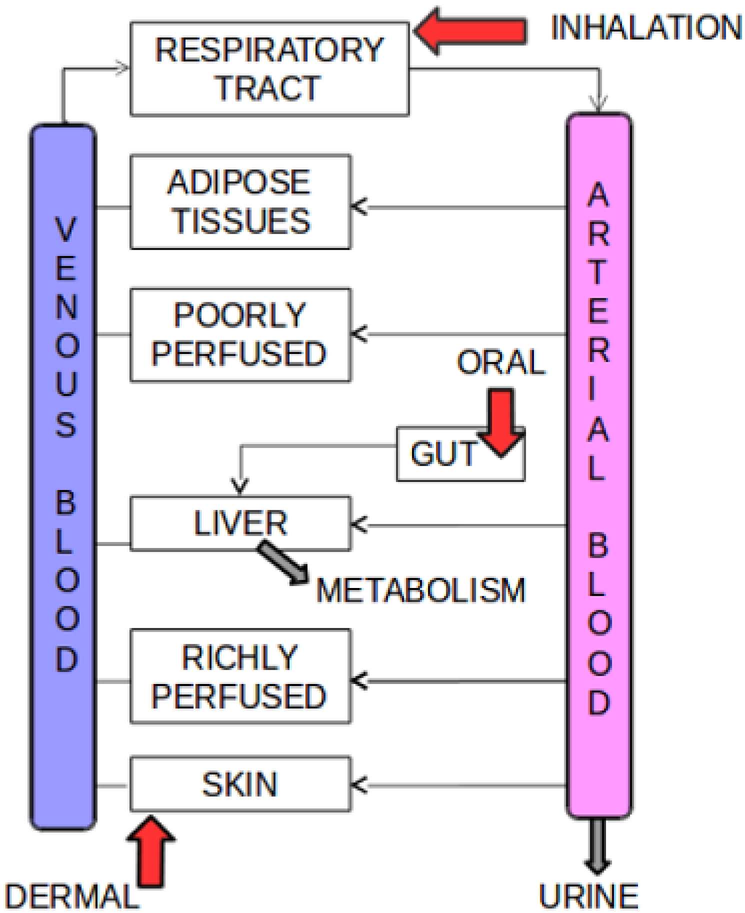


Figure 2.36: Schematic representation of the EuroMix Generic PBTK model.

The EuroMix generic PBTK model is coded as a set of ordinary differential equations. There is one such equation per time-dependent chemical quantity of the model (so-called state variables). There are 13 state variables in the model: the quantity of chemical in venous blood (Q_{ven}), in arterial blood (Q_{art}), in adipose tissues (Q_{fat}), in poorly perfused tissues (Q_p), in well perfused tissues (Q_r), in liver (Q_{liv}), in unexposed skin ($Q_{s,u}$), in exposed skin ($Q_{s,e}$), in the stratum corneum of unexposed skin ($Q_{sc,u}$), in exposed stratum corneum ($Q_{sc,e}$), in gut lumen (Q_{gut}), the quantity excreted to urine ($Q_{ex)}$, and the quantity metabolized (Q_{met}). The model can predict, as a function of time, for given oral, dermal and/or inhalation exposures, all the above quantities and the corresponding concentrations as a function of time. Concentrations are obtained by dividing quantities by compartment volumes (cited: Bois, Tebby & Brochot).

In Figure 2.37 a time course of the internal substance amount (μg) for Clothianidin in the liver is shown. For 50 consecutive days a bolus per day is submitted. The red line shows the substance amount varying over time. The green line displays the average of the peaks representing acute exposure, the blue line displays the steady state representing chronic exposure, all after skipping a nonstationary period of 10 days (the vertical black line).

From the substance amount, a concentration is computed by dividing it by the total compartment weight (i.e., the mass/volume of the compartment/organ).

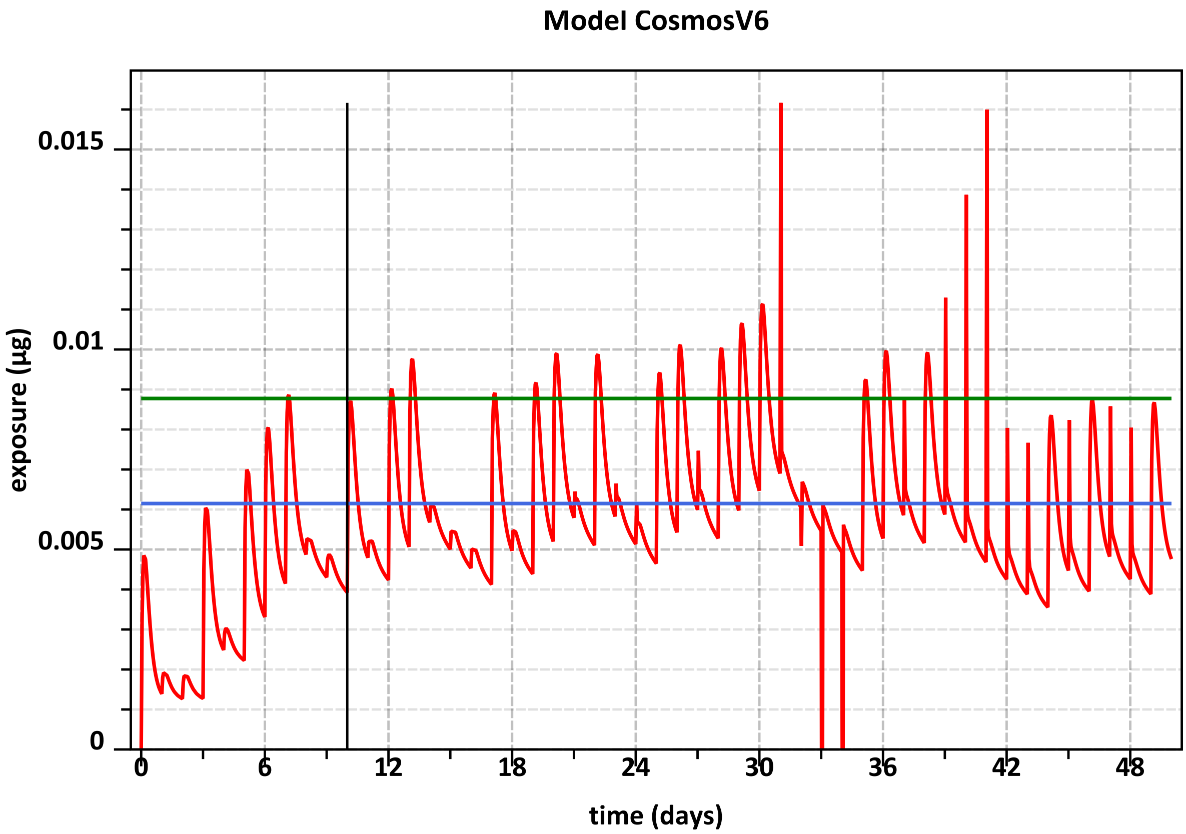


Figure 2.37: Time course of exposure (μg) for Clothianidin in the liver (EuroMix generic PBTK model version 6).

In Figure 2.38, for a large number of individuals the internal exposure (acute, green dots) in the liver is plotted versus the external exposure ($\mu g/kgbw$). The diagonal represents the 1:1 ratio of internal vs external exposure.

Bisphenol model

Reference: Karrer et al. 2019: [25]

'Structural analogs such as the bisphenols S, F, and AF (BPS, BPF, BPAF) are used to replace the endocrine disrupting chemical bisphenol A (BPA), but they exert estrogenic effects in the same order of magnitude. In order to investigate the consequences of BPA restrictions, we assessed the cumulative risk from BPA, BPS, BPF, and BPAF in Europe before and after the first BPA restrictions in 2011. We modeled external exposures from food, personal care products (PCPs), thermal paper, and dust, using the models MCRA and PACEM for food and PCPs, respectively. We calculated internal concentrations of unconjugated BPs with substance-specific PBPK models and cumulated concentrations by taking into account relative estrogenic potencies. Average cumulative exposure to unconjugated BPs was 3.8 and 2.1 ng/kg bw/day before and after restrictions, respectively. The decline was mostly caused by the replacement of BPA with BPS in thermal paper. Therefore, the margins of exposure (MOEs) for estrogenic effects were mostly higher after the restrictions. However, in high uncertainty percentiles the MOEs were partly lower than before (e.g. the MOEs for the uncertainty P97.5 of the variability P99 were 2.6 and 1.9 before and after restrictions, respectively), which shows the higher uncertainty around exposures for substitutes compared to BPA.'

Abstract: Linking probabilistic exposure and pharmacokinetic modeling to assess the cumulative risk from the bisphenols BPA, BPS, BPF, and BPAF for Europeans. Authors: Cecile Karrer, Waldo de Boer, Christiaan Delmaar, Yaping Cai, Amélie Crépet, Konrad Hungerbühler, Natalie von Goetz

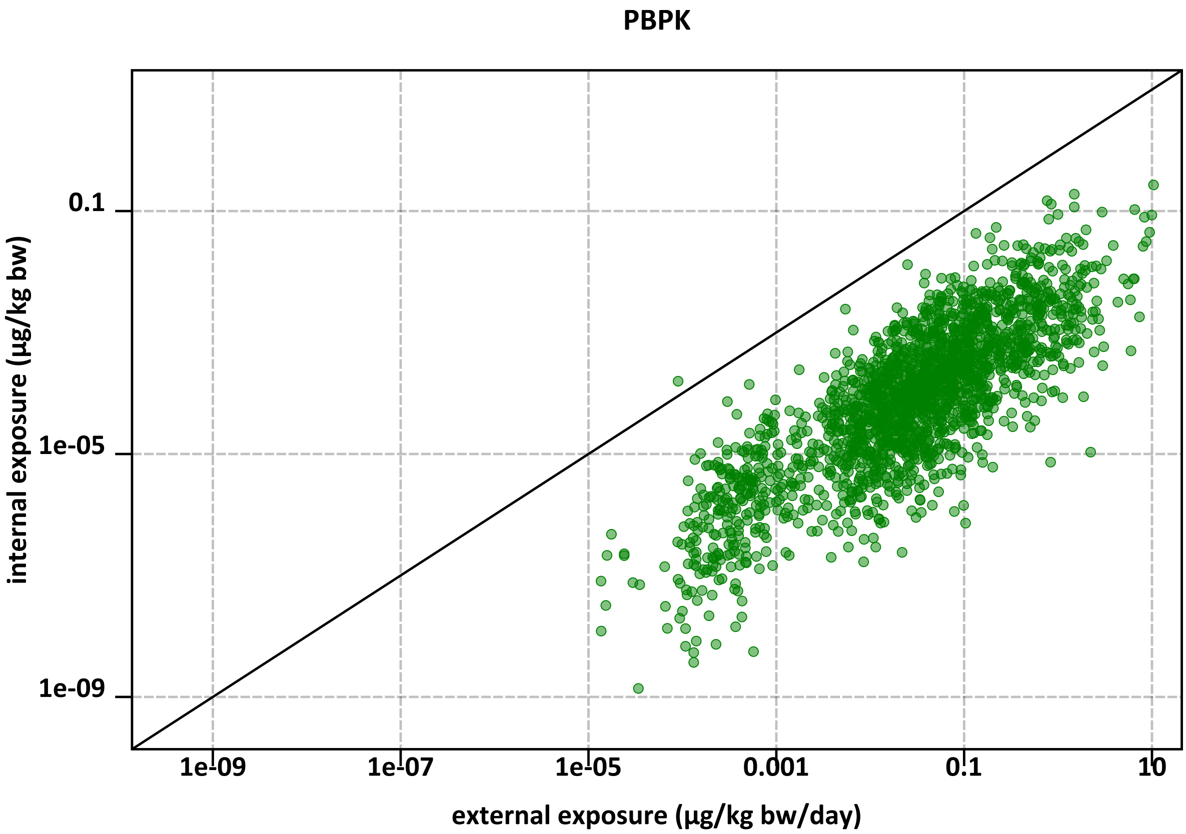


Figure 2.38: Internal versus external exposure for Clothianidin in the liver (EuroMix Generic PBTK model version 6).

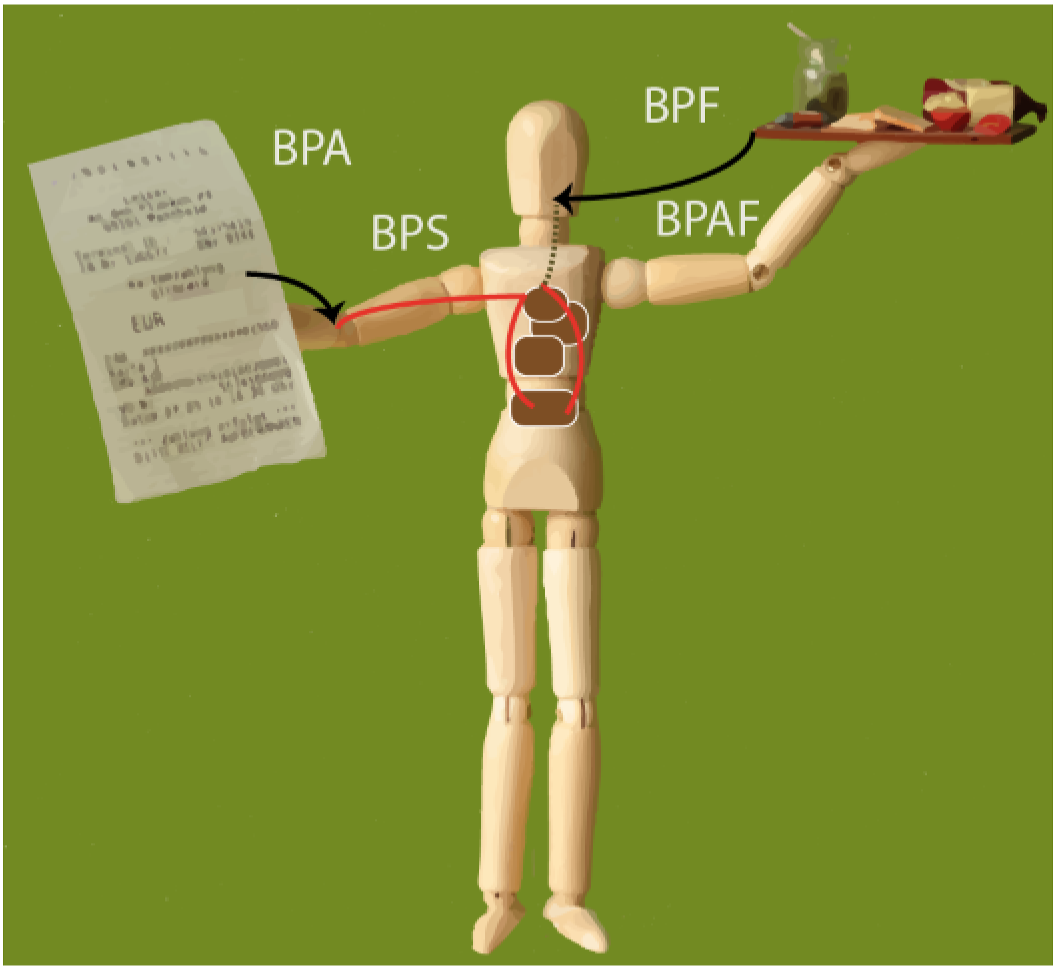


Figure 2.39: Graphical abstract 'Linking probabilistic exposure and pharmacokinetic modeling to assess the cumulative risk from the bisphenols BPA, BPS, BPF, and BPAF for Europeans.'

2.8 Risk modules

Exposures and *hazard characterisations* are compared in *risk estimates*.

2.8.1 Risks

Risks (health impacts) are quantified by comparing exposures and hazard characterisations at the chosen level (external or internal) via margins of exposure (MOE) or more generalised or integrated margins of exposure (IMOE). In addition, risks can be assessed from a plot of hazard characterisations vs. exposures.

This module has as primary entities: *Substances Effects Populations*

Risks calculation

'A method is proposed for integrated probabilistic risk assessment where exposure assessment and hazard characterization are both included in a probabilistic way. The aim is to specify the probability that a random individual from a defined (sub)population will have an exposure high enough to cause a particular health effect of a predefined magnitude, the critical effect size (*CES*). The exposure level that results in exactly that *CES* in a particular person is that person's individual critical effect dose (*ICED*). Individuals in a population typically show variation, both in their individual exposure (*IEXP*) and in their *ICED*. Both the variation in *IEXP* and the variation in *ICED* are quantified in the form of probability distributions. Assuming independence between both distributions, they are combined (by Monte Carlo) into a distribution of the individual margin of exposure (*IMoE*). The proportion of the *IMoE* distribution below unity is the probability of critical exposure (*PoCE*) in the particular (sub)population. Uncertainties involved in the overall risk assessment (i.e., both regarding exposure and effect assessment) are quantified using Monte Carlo and bootstrap methods. This results in an uncertainty distribution for any statistic of interest, such as the probability of critical exposure (*PoCE*). The method is illustrated based on data for the case of dietary exposure to the organophosphate acephate. We present plots that concisely summarize the probabilistic results, retaining the distinction between variability and uncertainty. We show how the relative contributions from the various sources of uncertainty involved may be quantified.' (abstract from [45]).

A statistical model is presented extending the integrated probabilistic risk assessment (IPRA) model of van der Voet and Slob (2007) The aim is to characterise the health impact due to one or more chemicals present in food causing one or more health effects. For chemicals with hardly any measurable safety problems we propose health impact characterisation by margins of exposure. In this probabilistic model not one margin of exposure is calculated, but rather a distribution of individual margins of exposure (*IMoE*) which allows quantifying the health impact for small parts of the population. A simple bar chart is proposed to represent the *IMoE* distribution and a lower bound (*IMoEL*) quantifies uncertainties in this distribution. It is described how *IMoE* distributions can be combined for dose-additive compounds and for different health effects. Health impact assessment critically depends on a subjective valuation of the health impact of a given health effect, and possibilities to implement this health impact valuation step are discussed. Examples show the possibilities of health impact characterisation and of integrating *IMoE* distributions. The paper also includes new proposals for modelling variable and uncertain factors describing food processing effects and intraspecies variation in sensitivity.' (abstract from: van der Voet et al, 2009 [46]).

Risks settings

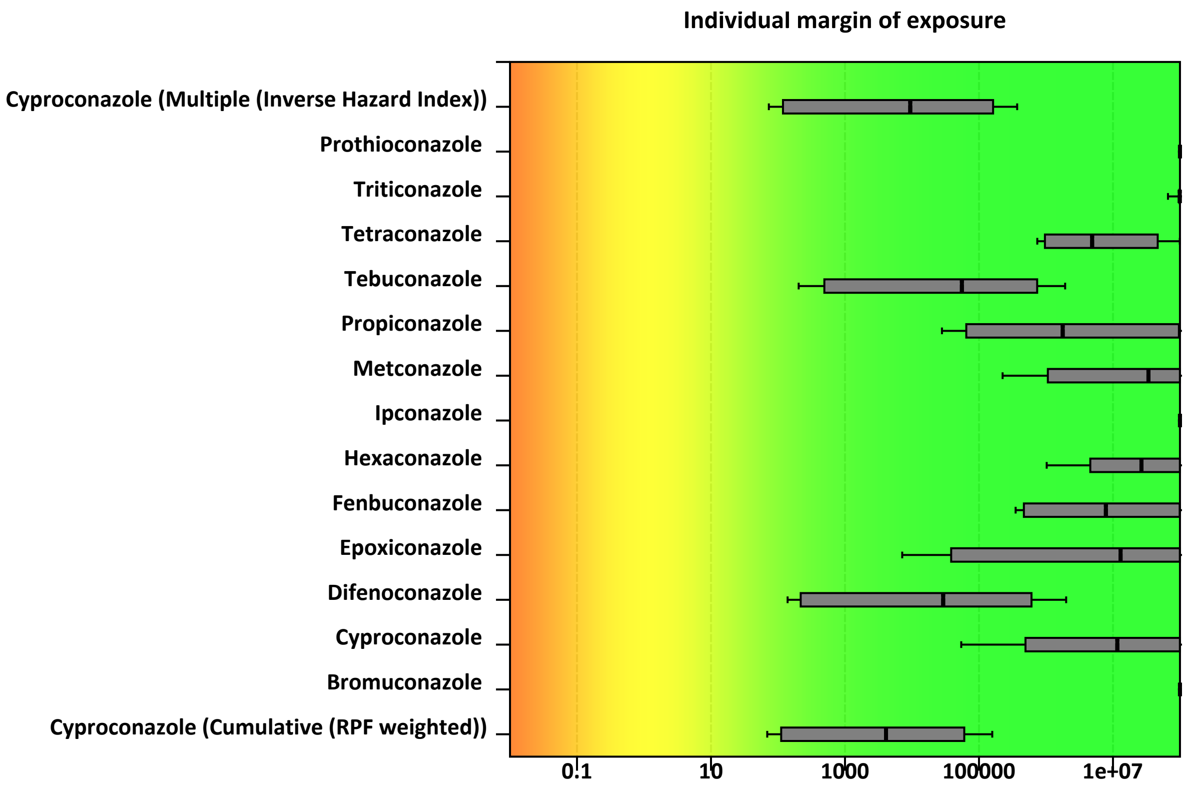


Figure 2.40: Individual Margin of Exposure (IMoE) plot for multiple chemicals.

Calculation settings

Table 2.171: Calculation settings for module Risks.

Name	Description
Health effect type	Specifies whether the health effect is a risk (negative) or benefit (positive).
Left margin safety plot	The left margin of the safety plot (Risks).
Right margin safety plot	The right margin of the safety plot (Risks).
Show equivalent animal dose output	Specifies whether equivalent animal doses should be reported in the output.
Threshold margin of exposure (MoE)	The threshold for the margin of exposure e.g., 1 (below unity) (Risks).
Inclusion percentage variability interval	The percentage of the variability distribution to include in intervals for exposure, hazard and IMoE estimates (e.g. 90) (Risks).
Number of plot labels	Number of labels to plot in hazard vs exposure plot.

Calculation of risks

Risks (health impacts) are quantified by comparing exposures and hazard characterisations at the chosen level (external or internal) via margins of exposure (MOE) or more generalised or integrated margins of exposure (IMOE).

- *Risks calculation*

Inputs used: *Exposures Hazard characterisations*

Settings used

- *Calculation Settings*

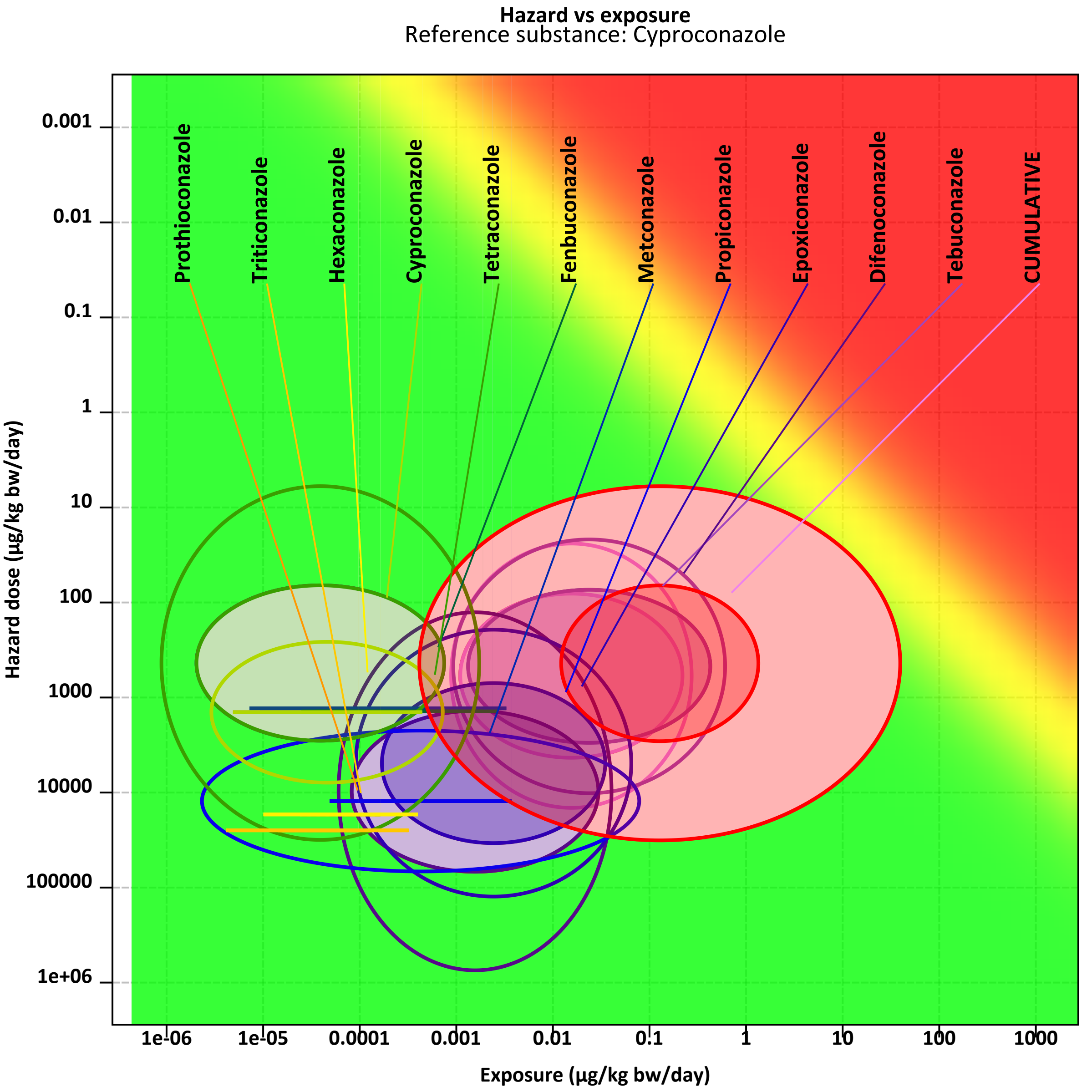


Figure 2.41: Example of MCRA Hazard vs. Exposure plot for multiple chemicals. 95% bivariate confidence areas for target hazard dose distribution *ICED* and exposure distribution *IEXP*. Inner ellipses express variability, outer ellipses uncertainty.

Risks are expressed as (individual) margins of exposure and as a probability to exceed a reference value (e.g. 1 or 100), comparing the exposures and the hazard characterisation for individuals or individual-days in a population. Exposures, hazard characterisations and risks can be acute or chronic. The default unit for exposures and hazard characterisations is $\mu g/kgBW/day$, but this can be changed by choosing non-default units for consumptions, concentrations and/or body weight.

The basic calculation is a graphical representation of hazard characterisations versus exposures.

In a low tier, the calculated ratio is equal to the traditional Margin Of Exposure (MOE). By including assessment factors in the hazard characterisations, the MOE can be generalised to account internally for e.g. interspecies and intraspecies uncertainty, making 1 the relevant limit for risk assessment. By using probabilistic tiers for exposure and hazard characterisation, the MOE is further generalised to a distribution of Integrated Margins Of Exposure (IMOEs), as described in [45] and [46].

Table 2.172: Overview of MCRA modules.

Category	Module	Inputs	Used by	Description
Primary entities	*Foods*		*Consumptions*, *Market shares*, *Food recipes*, *Concentrations*, *Processing factors*, *Unit variability factors*, *Occurrence patterns*, *Substance authorisations*, *Concentration limits*, *Concentration models*, *Foods as measured*, *Total diet study sample compositions*, *Food extrapolations*, *Food conversions*, *Consumptions by food as measured*, *Dietary exposures with screening*, *Dietary exposures*, *Exposures*, *Exposure mixtures*.	Foods are uniquely defined sources of dietary exposure to chemical substances. Foods may refer to 1) foods-as-eaten: foods as coded in food consumption data (e.g. pizza); 2) foods-as-measured: foods as coded in concentration data (e.g. wheat); 3) any other type of food (e.g. ingredients, e.g. flour).

Continued on next page

Table 2.172 – continued from previous page

Category	Module	Inputs	Used by	Description
	Substances		*Concentrations*, *Processing factors*, *Unit variability factors*, *Occurrence patterns*, *Substance authorisations*, *Substance conversions*, *Concentration limits*, *Concentration models*, *Foods as measured*, *Food conversions*, *Consumptions by food as measured*, *Dietary exposures with screening*, *Dietary exposures*, *Non-dietary exposures*, *Exposures*, *Exposure mixtures*, *Human monitoring data*, *Human monitoring analysis*, *QSAR membership models*, *Molecular docking models*, *Kinetic models*, *Active substances*, *Relative potency factors*, *Hazard characterisations*, *Points of departure*, *Dose response models*,	Substances are chemical entities that can refer to: 1) active substances such as investigated in toxicology; 2) measured substances such as defined in specific analytical methods. MCRA assessments can have one or more substances as the scope. When more than one substance is specified, there is an option to perform a cumulative assessment. In that case one of the substances has to be indicated as the index/reference substance, and results will be expressed in equivalents of the index substance.

Table 2.172 – continued from previous page

Category	Module	Inputs	Used by	Description
	Effects		*Concentration models*, *Dietary exposures with screening*, *Dietary exposures*, *Exposure mixtures*, *QSAR membership models*, *Molecular docking models*, *Active substances*, *Relative potency factors*, *Hazard characterisations*, *Points of departure*, *Effect representations*, *Inter-species conversions*, *Intra species factors*, *AOP networks*, *Risks*.	Effects are biological or toxicological consequences for human health, that may result from chemical exposure and are the focus of hazard or risk assessment.
	Populations		*Consumptions*, *Consumptions by food as measured*, *Dietary exposures*, *Non-dietary exposures*, *Exposures*, *Human monitoring analysis*, *Risks*.	Populations are groups of human individuals that are the scope of exposure or risk assessments. Optional descriptors of populations are location (e.g. a country), time period (start date, end date), age range and gender. Example: the French population in 2005-2007 of women of child-bearing age (18-45 yr).

Continued on next page

Table 2.172 – continued from previous page

Category	Module	Inputs	Used by	Description
	Test systems		*Responses*, *Dose response models*, *Dose response data*.	Test systems are biological or artificial systems used for assessing hazard in relation to chemical exposure from substances in varying doses. Test systems may refer to 1) in-vivo test systems (e.g. a rat 90-day study, a human biomonitoring study); 2) in-vitro test systems (e.g. HepaRG cells).
	Responses	*Test systems*.	*Dose response models*, *Dose response data*, *Effect representations*.	Responses are measurable entities in test systems. Responses are used to represent effects (see effect representations) and their measured values are collected in dose response data.
Consumption	*Consumptions*	*Populations*, *Foods*.	*Food conversions*, *Consumptions by food as measured*.	Consumptions data are the amounts of foods consumed on specific days by individuals in a food consumption survey. For acute exposure assessments, the interest is in a population of person-days, so one day per individual may be sufficient. For chronic exposure assessments, the interest is in a population of person, so preferably two or more days per individual are needed.
	Market shares	*Foods*.	*Food conversions*.	Market shares data specify for a given food, percentages of more specific foods (subfoods, e.g. brands) representing their share in a market. Market shares are used when consumption data are available at a more generalised level than concentration data.
	Food recipes	*Foods*.	*Food conversions*.	Food recipes data specify the composition of specific foods (typically: foods-as-eaten) in terms of other foods (intermediate foods or foods-as-measured) by specifying proportions in the form of a percentage.

Continued on next page

Table 2.172 – continued from previous page

Category	Module	Inputs	Used by	Description
Occurrence	*Concentrations*	*Foods*, *Substances*, *Focal food concentrations*, *Food extrapolations*, *Substance conversions*, *Relative potency factors*, *Substance authorisations*, *Active substances*, *Concentration limits*.	*Occurrence patterns*, *Concentration models*, *Foods as measured*.	Concentrations data are analytical measurements of chemical substances occurring in food samples. In their simplest form, concentration data can just be used as provided by datasets. Optionally, concentrations data can be manipulated for active substances, extrapolated to other foods, and/or default values can be added for water.
	Processing factors	*Foods*, *Substances*.	*Food conversions*, *Dietary exposures*.	Processing factors are multiplication factors to derive the concentration in a processed food from the concentration in an unprocessed food and can be specified for identified processing types (e.g., cooking, washing, drying). Processing factors are primarily used in dietary exposure assessments to correct for the effect of processing on substance concentrations in dietary exposure calculations.
	Unit variability factors	*Foods*, *Substances*.	*Dietary exposures*.	Unit variability factors specify the variation in concentrations between single units of the same food, which have been put together in a mixture sample on which the concentration measurements have been made. Unit variability factors are used to account for the fact that concentration data often relate to composite samples, whereas an acute risk may result from single food units.

Continued on next page

Table 2.172 – continued from previous page

Category	Module	Inputs	Used by	Description
	Occurrence patterns	*Foods*, *Substances*, *Concentrations*, *Substance authorisations*, *Active substances*.	*Concentration models*.	Occurrence patterns (OPs) are the combinations (or mixtures) of substances that occur together on foods and the frequencies of these mixtures occurring per food, expressed in percentages. In the context of pesticides, occurrence patterns can be associated with agricultural use percentages. Occurrence patterns are relevant to account for co-occurrence of active substances in exposed individuals. Occurrence patterns may be specified as data or modelled based on observed patterns of positive concentrations.
	Substance authorisations	*Foods*, *Substances*.	*Concentrations*, *Occurrence patterns*.	Substance authorisations specify which food/substance combinations are authorised for (agricultural) use. If substance authorisations are used, then only the food/substance combinations that are specified in the data are assumed to be authorised and all other combinations are assumed to be not authorised. This information may, for instance, be used to determine whether concentration measurements below the LOR could be assumed true zeros. I.e., if a food/substance combinations is assumed to be unauthorised, then the LOR may be assumed to be a zero.
	Substance conversions	*Substances*, *Active substances*.	*Concentrations*.	Substance conversions specify how measured substances are converted to active substances, which are the substances assumed to cause health effects. In the pesticide legislation such measured substances and the substance conversion rules are known as residue definitions.

Continued on next page

Table 2.172 – continued from previous page

Category	Module	Inputs	Used by	Description
	Concentration limits	*Foods*, *Substances*.	*Concentrations*, *Concentration models*, *Foods as measured*.	Concentration limits specify (legal) limit values for substance concentrations on foods and are sometimes used as conservative values for concentration data. In the framework of pesticides the legal Maximum Residue Limit (MRL) is the best known example.
	Concentration models	*Concentrations*, *Concentration limits*, *Foods as measured*, *Occurrence patterns*, *Relative potency factors*.	*Dietary exposures with screening*, *Dietary exposures*.	Concentration models are distributional models of substance concentrations on foods. They describe both the substance presence (yes/no, with no representing an absolute zero concentration) and the substance concentrations. Concentration models are specified per food/substance combination.
	Foods as measured	*Concentrations*, *Concentration limits*.	*Concentration models*, *Food conversions*.	Foods as measured are foods within the foods scope for which concentration data of substances are available (or expected).
	Focal food concentrations		*Concentrations*.	In some cases the attention in an assessment is on a specific food (focal food), against the background of other foods. Focal food concentrations are separate concentration data for one or more focal food commodities, that will take the place of any other concentration data for the focal food in the ordinary concentration data.
	Total diet study sample compositions	*Foods*.	*Food conversions*.	Total diet study sample compositions specify the composition of mixed food samples, such as used in a total diet study (TDS), in terms of their constituting foods.
	Food extrapolations	*Foods*.	*Concentrations*, *Food conversions*.	Food extrapolations data specify which foods (data rich foods) can be used to impute concentration data for other foods with insufficient data (data poor foods).

Continued on next page

Table 2.172 – continued from previous page

Category	Module	Inputs	Used by	Description
Exposure	*Food conversions*	*Consumptions*, *Foods as measured*, *Processing factors*, *Food recipes*, *Market shares*, *Food extrapolations*, *Total diet study sample compositions*, *Active substances*.	*Consumptions by food as measured*.	Food conversions relate foods-as-eaten, as found in the consumption data, to foods-as-measured, which are the foods for which concentration data are available. A food-as-eaten can be linked to one, or multiple food-as-measured using various conversion steps (e.g., using food recipes to translate a composite food to its ingredients, or using processing information to relate a processed food to its unprocessed form). There are several types of conversion steps, and a conversion path may comprise multiple conversion steps between a food-as-eaten and a food-as-measured.
	Consumptions by food as measured	*Consumptions*, *Food conversions*.	*Dietary exposures with screening*, *Dietary exposures*.	Consumptions by food as measured are consumptions of individuals expressed on the level of the foods for which concentration data are available (i.e., the foods-as-measured). These are calculated from consumptions of foods-as-eaten and food conversions that link the foods-as-eaten amounts to foods-as-measured amounts.
	Dietary exposures with screening	*Consumptions by food as measured*, *Concentration models*, *Active substances*, *Relative potency factors*.	*Dietary exposures*.	Dietary exposures with screening are just dietary exposures, but the calculation includes a prior screening step to identify the main risk drivers (food-substance combinations). This allows computations with more substances by suppressing some details for less important food-substance combinations.

Continued on next page

Table 2.172 – continued from previous page

Category	Module	Inputs	Used by	Description
	Dietary exposures	*Consumptions by food as measured*, *Concentration models*, *Processing factors*, *Unit variability factors*, *Dietary exposures with screening*, *Active substances*, *Relative potency factors*.	*Exposures*.	Dietary exposures are the amounts of substances, expressed per kg bodyweight or per individual, to which individuals in a population are exposed from their diet per day. Depending on the exposure type, dietary exposures can be short-term/acute exposures and then contain exposures for individual-days, or they can be long-term/chronic exposures, in which case they represent the average exposure per day over an unspecified longer time period.
	Non-dietary exposures	*Populations*, *Substances*, *Active substances*.	*Exposures*.	Non-dietary exposures are the amounts of substances to which individuals in a population are exposed via any of three non-dietary routes: dermal, inhalation or oral, per day.
	Exposures	*Dietary exposures*, *Non-dietary exposures*, *Active substances*, *Relative potency factors*, *Kinetic models*.	*Exposure mixtures*, *Human monitoring analysis*, *Risks*.	Exposures are amounts of substances, typically expressed per mass unit and per day, to which individuals in a population are exposed at a chosen target level. This target level may be external exposure (dietary exposure, expressed per unit body weight, or per person) or internal exposure (expressed per unit organ weight). Internal exposures may be aggregated from dietary and non-dietary exposures using either absorption factors or kinetic models to translate the external exposures to internal exposures. Exposures can be short-term/acute exposures and then contain exposures for individual-days, or they can be long-term/chronic exposures, in which case they represent the average exposure per day over an unspecified longer time period.

Continued on next page

Table 2.172 – continued from previous page

Category	Module	Inputs	Used by	Description
	Exposure mixtures	*Exposures*.		Exposure mixtures are mixtures of substances that contribute relatively much to the overall cumulative exposure (potential risk drivers).
	Human monitoring data	*Substances*.	*Human monitoring analysis*.	Human monitoring data quantify substance concentrations found in humans collected in human monitoring surveys.
	Human monitoring analysis	*Human monitoring data*, *Exposures*.		Human monitoring analysis compares observed human monitoring data with predictions made for the same population of individuals from dietary survey data, concentration data and (optionally) non-dietary exposure data.
In-silico	*QSAR membership models*	*Substances*, *Effects*, *AOP networks*.	*Active substances*.	QSAR membership models specify assessment group memberships for active substances related to a specific health effect (adverse outcome). Memberships should be derived externally from Quantitative Structure-Activity Relationship (QSAR) models.
	Molecular docking models	*Substances*, *Effects*, *AOP networks*.	*Active substances*.	Molecular docking models specify binding energies for substances in specific molecular docking models related to a specific health effect (adverse outcome).
Kinetic	*Kinetic models*	*Substances*, *Active substances*.	*Exposures*, *Hazard characterisations*.	Kinetic models convert exposures or hazard characterisations from one or more external routes or compartments to an internal (target) compartment. The reverse conversion from internal to external can also be made (reverse dosimetry).

Continued on next page

Table 2.172 – continued from previous page

Category	Module	Inputs	Used by	Description
Hazard	*Active substances*	*AOP networks*, *Points of departure*, *Molecular docking models*, *QSAR membership models*.	*Concentrations*, *Occurrence patterns*, *Substance conversions*, *Non-dietary exposures*, *Kinetic models*, *Food conversions*, *Dietary exposures with screening*, *Dietary exposures*, *Exposures*, *Hazard characterisations*.	Active substances are substances that may lead (P>0) to a specific health effect (adverse outcome). Active substances can be specified directly as data or calculated from POD presence, QSAR models or Molecular docking models. Active substances can have an assessment group membership 1 (crisp), or values in the range (0,1] (probabilistic).
	Relative potency factors	*AOP networks*, *Hazard characterisations*.	*Concentrations*, *Concentration models*, *Dietary exposures with screening*, *Dietary exposures*, *Exposures*.	Relative potency factors (RPFs) quantify potencies of substances with respect to a defined effect, relative to the potency of a chosen index substance. RPFs can be used to express combined exposures of multiple substances in terms of a the exposure value of the chosen index substance (i.e., in index substance equivalents). In MCRA, hazard characterisations, and therefore also RPFs are based on mass units (e.g., µg), and not on mol units. RPFs can be different for different levels of the human organism (external, internal, specific compartment). RPFs can be given as data or computed from hazard characterisations. RPFs can be specified with uncertainty. Computation from uncertain hazard characterisations allows to include correlations between uncertain RPFs which originate from using the same index substance.

Continued on next page

Table 2.172 – continued from previous page

Category	Module	Inputs	Used by	Description
	Hazard characterisations	*AOP networks*, *Active substances*, *Dose response models*, *Effect representations*, *Points of departure*, *Inter-species conversions*, *Intra species factors*, *Kinetic models*.	*Relative potency factors*, *Risks*.	Hazard characterisations are benchmark doses for active substances and for the chosen effect at the chosen target level (external or internal) of the hazard assessment. Hazard characterisations are based on points of departure, such as BMDs from dose-reponse models or externally specified points of departure (MDSs, NOAELs or LOAELs). The computation may involve inter-species conversion, intra-species factors and the use of kinetic models or absorption factors to convert external doses to internal doses.
	Points of departure	*Substances*, *Effects*, *AOP networks*.	*Active substances*, *Hazard characterisations*.	Externally specified points of departure can be used as an alternative to calculated BMDs from dose response models. Points of departure can be of various types, such as NOAEL, LOAEL or BMD. They can be used to construct the list of active substances, to derive relative potency factors, and to perform health impact assessments.
	Dose response models	*Dose response data*, *Effect representations*.	*Hazard characterisations*.	Dose response models are models fitted to dose response data and can be provided as data or calculated using a local or remote version of PROAST. The main results for hazard and risk assessment are benchmark doses (BMDs), related to a specified substance, response, optionally covariate value, and the benchmark response (BMR).
	Dose response data	*Substances*, *Test systems*, *Responses*.	*Dose response models*.	Dose response data are data on response values of test systems at specified doses of substances (or mixtures of substances) from dose response experiments.
	Effect representations	*Effects*, *Responses*, *AOP networks*.	*Hazard characterisations*, *Dose response models*.	Effect representations specify the responses that can be used to measure specified effects and which response levels, the benchmark response (BMR), define the hazard limits for the effects.

Continued on next page

Table 2.172 – continued from previous page

Category	Module	Inputs	Used by	Description
	Inter-species conversions	*Substances*, *Effects*.	*Hazard characterisations*.	Inter-species conversions specify how to convert a hazard characterisation for a given species to a hazard characterisation for humans. In the simplest approach, this specifies a fixed inter-species factor. In a higher tier, this specifies a geometric mean (GM) and geometric standard deviation (GSD) for a lognormal uncertainty distribution of the interspecies factor. Inter-species conversion are specified per effect and can be general or substance-specific.
	Intra species factors	*Substances*, *Effects*.	*Hazard characterisations*.	Intra-species factors specify how to convert a hazard characterisation from the average to a sensitive human individual.
	AOP networks	*Effects*.	*QSAR membership models*, *Molecular docking models*, *Active substances*, *Relative potency factors*, *Hazard characterisations*, *Points of departure*, *Effect representations*.	Effects can be related to each other using the toxicological concept of adverse outcome pathways (AOPs) and adverse outcome pathway networks (see https://aopwiki.org). Adverse Outcome Pathway (AOP) Networks specify how biological events (effects) can lead to an adverse outcome (AO) in a qualitative way through relations of upstream and downstream key events (KEs), starting from molecular initiating events (MIEs). Using AOPs, the adverse outcome (AO), e.g., liver steatosis, is linked to key events (KEs), e.g., triglyceride accumulation in the liver, and to molecular initialing events (MIEs), e.g., PPAR-alpha receptor antagonism. In general, multiple AOPs may lead to the same AO, and therefore AOP networks can be identified.

Continued on next page

Table 2.172 – continued from previous page

Category	Module	Inputs	Used by	Description
Risks	*Risks*	*Exposures*, *Hazard characterisations*.		Risks (health impacts) are quantified by comparing exposures and hazard characterisations at the chosen level (external or internal) via margins of exposure (MOE) or more generalised or integrated margins of exposure (IMOE). In addition, risks can be assessed from a plot of hazard characterisations vs. exposures.

CHAPTER

THREE

EXAMPLES

Note: This section is under construction. Please contribute!

Training materials used in EuroMix training sessions:

- EuroMix dietary exposure
- RPF-exercise 1-for training-draft

There are a few exercises prepared that you could follow to get started.

3.1 Cumulative dietary exposure assessment

3.1.1 Introduction

The goal of this exercise is to perform a probabilistic cumulative dietary exposure assessment, illustrating all data needed. In Example 1 we will upload and use nine different files containing the data. In Example 2 we will upload and use a single data file for the same purpose. In the example the exposure will be characterised by upper tail percentiles, and the risk driving substances and foods can be examined. In Example 3 an uncertainty analysis is added.

3.1.2 Preparation

In the workspace browser (icon), create a new workspace *Examples*, using the + button in the bottom right corner.

3.1.3 Example 1

Calculate a cumulative chronic dietary exposure for Dutch young adults in 2003 regarding a group of eight triazole substances according to the basic optimistic model of the EFSA 2012 guidance document. Use liver steatosis as a focal effect and Cyproconazole as an index substance. The data files are already available in the data folder *Documentation-Examples / Exercise Dietary Exposure Assessment*.

Detailed steps are as follows.

- In the *Examples* workspace, create a new action using the + button in the bottom right corner.
  - Select action type *Dietary exposures*
    - Name it, e.g. *Triazoles exposures*
    - (Optional) You can also add tags (e.g. triazoles, NL, steatosis) as labels that can be used later to find simiar actions
    - (Optional) You can add a description for further information
    - Click Next

- Specify Dietary exposures settings
  - Tier: *EFSA 2012 Optimistic*
  - Risk type *Chronic*
  - Click *Create*

You are now directed to the main page of the new action. You can always return to this main page by clicking Action settings or the action type name (*Dietary exposures*) in the green bar.

The main page contains at least three blocks of information: Scope, Inputs and Settings. We will now first link all nine datafiles needed for this cumulative assessment. For most settings we will use default values in accordance with the chosen tier (*EFSA 2012 Optimistic*).

Scope of the assessment:

- Click *Effects* (path in the green bar changes Total *Dietary exposures / Effects*)
  - At *Effects data source*, click and browse to the file *Effect - Steatosis.xlsx*, then click *Select*
  - At *Effect Settings* for *focal effect* select *Steatosis-liver* and click *Save Changes*
  - In the green navigation bar, click *Dietary exposures* to go up one level.
- Click *Foods* (path: *Dietary exposures / Foods*)
  - At *Foods data source*, click and browse to the file *Foods.xlsx*, then click *Select*
  - In the green navigation bar, click *Dietary exposures* to go up one level
- Click *Populations (optional)* (path: *Dietary exposures / Populations*)
  - At *Populations data source*, click and browse to the file *Populations.xlsx*, then click *Select*
  - This file contains two populations, only one is allowed. Click under Populations selection, this opens a pop-up window. Deselect *NL_2006*, then click *Save*. The red warning signs should now be gone. (Note: green warning signs point at details and can usually be ignored)
  - In the green navigation bar, click *Dietary exposures* to go up one level.
- Click *Substances* (path: *Dietary exposures / Substances*)
  - At *Substances data source*, click and browse to the file *Substances - Triazoles.xlsx*, then click *Select*
  - At *Substance settings* for *Index substance* select *Cyproconazole* and click *Save Changes*
  - In the green navigation bar, click *Dietary exposures* to go up one level

Next we choose the other input data:

- Click *Consumptions by food as measured* (path: *Dietary exposures / Consumptions by food as measured*)
  - Click *Consumptions* (path: *Dietary exposures / Consumptions by food as measured / Consumptions*)
    - At *Consumptions data source*, click and browse to the file *FoodConsumptions.xlsx* and *Select*
    - At *Consumptions data selection*, with open the food consumption surveys selection.
      - The file contains two surveys, but only one is allowed. Click under Consumptions data selection, this opens a pop-up window. Deselect *VCP-kids*, then click *Save* (the red warning should now be gone)
    - In the green navigation bar, click *Consumptions by food as measured* to go up one level
  - Click *Food conversions* (path: *Dietary exposures / Consumptions by food as measured / Food conversions*)
    - Click *Foods as measured* (path: *Dietary exposures / Consumptions by food as measured / Food conversions / Foods as measured*)
      - Click *Concentrations* (path: *Dietary exposures / Consumptions by food as measured / Food conversions / Foods as measured / Concentrations*)

- At *Concentrations data source*, click and browse to the file *ConcentrationData.xlsx*, then click *Select*
- In the green navigation bar, click *Food conversions* to go up two levels

- Click *Food recipes* (path: *Dietary exposures / Consumptions by food as measured / Food conversions / Food recipes*)
  - At *Food recipes data source*, clikc and browse to the file *FoodTranslations.xlsx*. then click *Select*
  - In the green navigation bar, click *Dietary exposures* to go up three levels

- Click *Concentration models* (path: *Dietary exposures / Concentration models*)
  - Click *Relative potency factors* (path: *Dietary exposures / Concentration models / Relative potency factors*)
    - At *Relative potency data source*, click and browse to the file *RPFs.xlsx*, then click *Select*
    - In the green navigation bar, click *Dietary exposures* to go up two levels
- Click *Processing factors* (path: *Dietary exposures / Processing factors*)
  - At *Processing factors data source*, clikc and browse to the file *ProcessingFactors.xlsx*, then click *Select*
  - In the green navigation bar, click *Dietary exposures* to go up one level
- Click *Active substances (optional)* (path: *Dietary exposures / Active substances*)
  - In this example we have a fixed list of relative potency factors for the eight substances, and don't need point of departure (POD) data to decide which substances are active with respect to the health effect and therefore belong to the cumulative assessment group. Deselect the setting "Derive memberships from POD presence", then click *Save Changes*
  - In the green navigation bar, click *Dietary exposures* to go up one level

Now run the model, either by clicking the run icon in the grey bar, or by clicking the run icon in the green bar (Note: in the green bar can also be used to run subactions on their own).

The icon is replaced by the text "Running". When the run has finished, the interface automatically changes to the Results screen. You can also click the Results icon to go there.

As an exerciose, try find the following results:

1. The 99th percentile of cumulative exposure
2. The substance(s) with highest contribution to the total exposure
3. The food(s)-as-measured with the highest contribution to the upper tail of the exposure distribution

Answers:

- In the grey bar, browse to the results panel by clicking the icon and click on the latest output (path: *Results / Dietary exposures*)
  - In the *Dietary exposures* tab, browse in the tree (unfold by clicking where necessary) to *Dietary exposures* *Distribution (OIM)* *Percentiles*
    - In the table it states that the 99% exposure percentile is at an exposure of 0.02127 µg/kg bw/day.
  - In the *Dietary exposures* tab, browse in the tree (unfold by clicking where necessary) to *Dietary exposures* *Details* *Exposures by substance* *Total distribution*
    - From the pie chart it is clear that Tebuconazole contributes the most to the total exposure distribution with 32.7%. In the table below the graph more details can be found.
  - In the *Dietary exposures* tab, browse in the tree (unfold by clicking where necessary) to *Dietary exposures* *Details* *Exposures by food and substance* *Risk drivers upper tail*
    - From the pie chart it is clear that Flusilazole in grapefruit contributes the most (16.7%) to the upper tail exposure distribution

3.1.4 Example 2

We will create a new action to demonstrate uploading all the data at once. All data is now contained within one file, *MCRA-Documentation Example Dietary exposures.xlsx*.

Detailed steps are as follows.

- In the *Examples* workspace, create a new action (using +)
  - Select action type *Dietary exposures*
  - Name it, e.g. *Triazoles exposures from one data file*
  - Click Next
- Specify Dietary exposures settings
  - Tier: *EFSA 2012 Optimistic*
  - Risk type *Chronic*
  - Click *Create*
- Then go to the actions settings of this action (path: *Dietary exposures*)
  - Click *Effects* (path: *Dietary exposures / Effects*)
    - At *Effects data source*, click and browse to the file *MCRA-Documentation Example Dietary exposures.xlsx*. Click *Toggle all*, then *Select*. This will load all available data tables for all subactions of *Dietary exposures*.

You still need to specify the focal effect (under *Effects*), index substance (under *Substances*), and food surveys (under *Consumptions by food as measured / Consumptions*). You also need to deselect the "Derive memberships from POD presence" setting under *Active substances*. Navigate to the subaction where these changes have to be made using the green bar.

You now have achieved the same as in Example 1, only with the upload of one single file. You can now run the model, and inspect the results, whcih shouldbe the same as for Example 1.

3.1.5 Example 3

Repeat the run of the previous task, but in addition to the nominal run, perform an uncertainty analysis as well.

- Click on the icon (in the grey bar) to open the uncertainty settings panel
  - At *Uncertainty settings*, check ✓ *Perform uncertainty analysis*
    - For *Monte Carlo iterations per uncertainty run* choose *100*, and press *Save Changes*
- Now run the model, by pressing the ▶ run icon in the grey bar. Note that the run will take much more time.

Compare with the previous results, to find:

1. 95% uncertainty bounds for the 99% exposure percentile
2. 95% uncertainty bounds for the highest contribution from a substance to the total exposure distribution
3. 95% uncertainty bounds for the highest contribution from a food to the total exposure distribution

CHAPTER

FOUR

APPENDICES

4.1 Munro collection

This collection can be downloaded here.

4.2 Unit definitions

4.2.1 Benchmark response types

Accepted benchmark response types.

Table 4.1: Unit definition for Benchmark response types.

Name	Short name	Aliases	Description
Fraction change	Fraction change	FractionChange, FactorChange	The benchmark response is defined as a fraction change of the background response (i.e., defined for both increase and decrease). E.g., for a factor of 0.1, the benchmark response is at +/- 10% of background response.
Percentage change	Percentage change	PercentageChange	The benchmark response is defined as a percentage change of the background response (i.e., defined for both increase and decrease). E.g., for a percentage of 10, the benchmark response is at +/- 10% of background response.
Fraction of background response	Fraction of background	Factor, FactorOfBackground	The benchmark response is defined as a fraction of the background response. E.g., for a factor of 0.9, the benchmark response is at 0.9 times the background response (i.e., a decrease).
Percentage of background response	Percentage of background	Percentage, PercentageOfBackground	The benchmark response is defined as a percentage of the background response. E.g., for a percentage of 90, the benchmark response is at 90% of the background response (i.e., a decrease).
Extra risk	ER	ExtraRisk	For quantal response types. The benchmark dose is defined as the dose that corresponding with an extra risk of a factor times the background risk. A factor of 0.05 corresponds with 5% extra risk.
Additional risk	AR	AdditionalRisk	For quantal response types. The benchmark dose is defined as the dose that corresponding with an additional risk of a factor times the background risk. A factor of 0.05 corresponds with 5% additional risk.
ED50	ED50	ED50	For quantal response types. The benchmark dose is defined as the dose that corresponds with an estimated risk of 50% (ED50).
Absolute threshold value	Threshold value	Absolute	The benchmark dose is defined as an absolute threshold value.
Absolute difference	Absolute difference	Difference	The benchmark dose is defined an absolute difference with the background risk.

4.2.2 Body weight units

Units for describing person body weights.

Table 4.2: Unit definition for Body weight units.

Name	Short name	Aliases
Kilogram	kg	kg, kilograms, kilogr, 3, G167A
Gram	g	g, grams, gr, 0, G148A

4.2.3 Concentration units

Units for describing substance concentrations.

Table 4.3: Unit definition for Concentration units.

Name	Short name	Aliases
kilogram/kilogram	kg/kg	kg/kg, kilogram/kilogram, kilogram/kg, 0, G063A
gram/kilogram	g/kg	g/kg, gram/kilogram, gram/kg, gr/kg, -3, G015A, G060A, G191A
milligram/kilogram	mg/kg	mg/kg, milligram/kilogram, milligram/kg, milligr/kg, -6, G049A, G061A
micro-gram/kilogram	µg/kg	µg/kg, ug/kg, microgram/kilogram, microgram/kg, microgr/kg, -9, G050A, G076A
nanogram/kilogram	ng/kg	ng/kg, nanogram/kilogram, nanogram/kg, nanogr/kg, -12, G077A, G080A
picogram/kilogram	pg/kg	pg/kg, picogram/kilogram, picogram/kg, picogr/kg, -15, G081A
kilogram/liter	kg/L	kg/l, kg/L, kilogram/liter, kilogram/litre, G017A
gram/liter	g/L	g/l, g/L, gram/liter, gram/litre, gr/l, gr/L, G016A
milligram/liter	mg/L	mg/l, mg/L, milligram/liter, milligram/litre, milligr/l, milligr/L, G052A, G062A
microgram/liter	µg/L	µg/l, ug/L, microgram/liter, microgram/litre, microgr/l, microgr/L, G051A, G079A
nanogram/liter	ng/L	ng/l, ng/L, nanogram/liter, nanogram/litre, nanogr/l, nanogr/L, G078A
picogram/liter	pg/L	pg/l, pg/L, picogram/liter, picogram/litre, picogr/l, picogr/L
micro-gram/milliliter	µg/mL	µg/ml, ug/mL, microgram/milliliter, microgram/millilitre, microgr/ml, microgr/mL
nanogram/milliliter	ng/mL	ng/ml, ng/mL, nanogram/milliliter, nanogram/millilitre, nanogr/ml, nanogr/mL

4.2.4 Consumption units

Units for consumption amounts.

Table 4.4: Unit definition for Consumption units.

Name	Short name	Aliases
kilogram	kg	kg, kilograms, kilogr, 3, G167A
Gram	g	g, grams, gr, 0, G148A

4.2.5 Dose response model types

Known dose response model types.

Table 4.5: Unit definition for Dose response model types.

Name	Short name	Aliases	Description
Exp-m1	Exp-m1	Expm1	
Exp-m2	Exp-m2	Expm2	
Exp-m3	Exp-m3	Expm3	
Exp-m4	Exp-m4	Expm4	
Exp-m5	Exp-m5	Expm5	
Hill-m1	Hill-m1	Hillm1	
Hill-m2	Hill-m2	Hillm2	
Hill-m3	Hill-m3	Hillm3	
Hill-m4	Hill-m4	Hillm4	
Hill-m5	Hill-m5	Hillm5	
TwoStage	TwoStage	TwoStage	
LogLogist	LogLogist	LogLogist	
Weibull	Weibull	Weibull	
LogProb	LogProb	LogProb	
Gamma	Gamma	Gamma	
Logistic	Logistic	Logistic	
Probit	Probit	Probit	
LVM Exp m2	LVM Exp m2	LVM Exp m2	
LVM Exp m3	LVM Exp m3	LVM_Exp_M3	
LVM Exp m4	LVM Exp m4	LVM_Exp_M4	
LVM Exp m5	LVM Exp m5	LVM_Exp_M5	
LVM Hill m2	LVM Hill m2	LVM Hill m2	
LVM Hill m3	LVM Hill m3	LVM_Hill_M3	
LVM Hill m4	LVM Hill m4	LVM_Hill_M4	
LVM Hill m5	LVM Hill m5	LVM Hill m5	

4.2.6 Dose units

Units for describing substance doses.

Table 4.6: Unit definition for Dose units.

Name	Short name	Aliases
gram/kilogram bodyweight/day	g/kg bw/day	g/kg bw/day, gram/kg bw/day, gr/kg bw/day
milligram/kilogram bodyweight/day	mg/kg bw/day	mg/kg bw/day, milligram/kg bw/day, milligr/kg bw/day
micro-gram/kilogram bodyweight/day	µg/kg bw/day	µg/kg bw/day, microgram/kg bw/day, microgr/kg bw/day
nanogram/kilogram bodyweight/day	ng/kg bw/day	ng/kg bw/day, nanogram/kg bw/day, nanogr/kg bw/day
picogram/kilogram bodyweight/day	pg/kg bw/day	pg/kg bw/day, picogram/kg bw/day, picogr/kg bw/day
fem-togram/kilogram bodyweight/day	fg/kg bw/day	fg/kg bw/day, femtogram/kg bw/day, femtogr/kg bw/day
gram/gram bodyweight/day	g/g bw/day	g/g bw/day, gram/g bw/day, gr/g bw/day
milligram/gram bodyweight/day	mg/g bw/day	mg/g bw/day, milligram/g bw/day, milligr/g bw/day

Continued on next page

Table 4.6 – continued from previous page

Name	Short name	Aliases
microgram/gram bodyweight/day	µg/g bw/day	µg/g bw/day, microgram/g bw/day, microgr/g bw/day
nanogram/gram bodyweight/day	ng/g bw/day	ng/g bw/day, nanogram/g bw/day, nanogr/g bw/day
picogram/gram bodyweight/day	pg/g bw/day	pg/g bw/day, picogram/g bw/day, picogr/g bw/day
femtogram/gram bodyweight/day	fg/g bw/day	fg/g bw/day, femtogram/g bw/day, femtogr/g bw/day
kilogram/day	kg/day	kg/day, kilogram/day, kilogr/day
gram/day	g/day	g/day, gram/day, gr/day
milligram/day	mg/day	mg/day, milligram/day, milligr/day
microgram/day	µg/day	µg/day, microgram/day, microgr/day
nanogram/day	ng/day	ng/day, nanogram/day, nanogr/day
picogram/day	pg/day	pg/day, picogram/day, picogr/day
femtogram/day	fg/day	fg/day, femtogram/day, femtogr/day
kilogram/kilogram	kg/kg	kg/kg, kilogram/kilogram, kilogram/kg, kg/kg bw
gram/kilogram	g/kg	g/kg, gram/kilogram, gram/kg, gr/kg, g/kg bw
milligram/kilogram	mg/kg	mg/kg, milligram/kilogram, milligram/kg, milligr/kg, mg/kg bw
micro-gram/kilogram	µg/kg	µg/kg, microgram/kilogram, microgram/kg, microgr/kg, µg/kg bw
nanogram/kilogram	ng/kg	ng/kg, nanogram/kilogram, nanogram/kg, nanogr/kg, ng/kg bw
picogram/kilogram	pg/kg	pg/kg, picogram/kilogram, picogram/kg, picogr/kg, pg/kg bw
Molar	M	M, mol/L
millimolar	mM	mM, mmol/L
micromolar	µM	uM, µM, umol/L
nanomolar	nM	nM, nmol/L
moles	moles	moles, Moles
millimoles	mmoles	mmoles, mMoles
micromoles	µmoles	umoles, uMoles
nanomoles	nmoles	nmoles, nMoles

4.2.7 Exposure route types

The different routes in which an individual is exposed to substance concentrations.

Table 4.7: Unit definition for Exposure route types.

Name	Short name	Aliases	Description
Dietary exposure	Dietary	Dietary	Dietary exposure.
Non-dietary oral exposure	Oral	Oral	Non-dietary oral exposure.
Non-dietary dermal exposure	Dermal	Dermal	Non-dietary dermal exposure.
Non-dietary inhalation exposure	Inhalation	Inhalation	Non-dietary inhalation exposure.
At target	At target	AtTarget	Exposures directly at the target (organ).

4.2.8 Exposure units

Units for describing substance exposures.

Table 4.8: Unit definition for Exposure units.

Name	Short name	Aliases
gram/kilogram bodyweight/day	g/kg bw/day	g/kg bw/day, gram/kg bw/day, gr/kg bw/day, G212A
milligram/kilogram bodyweight/day	mg/kg bw/day	mg/kg bw/day, milligram/kg bw/day, milligr/kg bw/day, G211A
micro-gram/kilogram bodyweight/day	µg/kg bw/day	µg/kg bw/day, microgram/kg bw/day, microgr/kg bw/day, G210A
nanogram/kilogram bodyweight/day	ng/kg bw/day	ng/kg bw/day, nanogram/kg bw/day, nanogr/kg bw/day, G214A
picogram/kilogram bodyweight/day	pg/kg bw/day	pg/kg bw/day, picogram/kg bw/day, picogr/kg bw/day
fem-togram/kilogram bodyweight/day	fg/kg bw/day	fg/kg bw/day, femtogram/kg bw/day, femtogr/kg bw/day
gram/gram bodyweight/day	g/g bw/day	g/g bw/day, gram/g bw/day, gr/g bw/day
milligram/gram bodyweight/day	mg/g bw/day	mg/g bw/day, milligram/g bw/day, milligr/g bw/day
microgram/gram bodyweight/day	µg/g bw/day	µg/g bw/day, microgram/g bw/day, microgr/g bw/day
nanogram/gram bodyweight/day	ng/g bw/day	ng/g bw/day, nanogram/g bw/day, nanogr/g bw/day
picogram/gram bodyweight/day	pg/g bw/day	pg/g bw/day, picogram/g bw/day, picogr/g bw/day
femtogram/gram bodyweight/day	fg/g bw/day	fg/g bw/day, femtogram/g bw/day, femtogr/g bw/day
kilogram/day	kg/day	kg/day, kilogram/day, kilogr/day
gram/day	g/day	g/day, gram/day, gr/day
milligram/day	mg/day	mg/day, milligram/day, milligr/day
microgram/day	µg/day	µg/day, microgram/day, microgr/day
nanogram/day	ng/day	ng/day, nanogram/day, nanogr/day
picogram/day	pg/day	pg/day, picogram/day, picogr/day
femtogram/day	fg/day	fg/day, femtogram/day, femtogr/day

4.2.9 Point of departure types

Known point of departure types.

Table 4.9: Unit definition for Point of departure types.

Name	Short name	Aliases
Benchmark dose	BMD	BMD
No observed adverse effect level	NOAEL	NOAEL
Lowest observed adverse effect level	LOAEL	LOAEL

4.2.10 Response types

Available response types.

Table 4.10: Unit definition for Response types.

Name	Short name	Aliases	Description
Continuous multiplicative	CM	Continuous-Multiplicative	Response values are positive real numbers, e.g., weight, size.
Continuous additive	CA	ContinuousAdditive	Response values are real numbers, e.g., weight change, temperature.
Binary	B	Binary	Response values have binary outcomes (yes/no, true/false, success/failure, 0/1, etc.).
Quantal	Q	Quantal, Binomial	Response is measured in terms of number of successes out of N possible.
Quantal group	QG	QuantalGroup	Individual responses are measured as binary values, which may be grouped to form a quantal response.
Count	C	Count	Number of items (cells, molecules, deaths, etc.) in given interval/area/volume.
Ordinal	O	Ordinal	Relative scores (or graded scores) useable only for ranking.

4.2.11 Test system types

Available test system types.

Table 4.11: Unit definition for Test system types.

Name	Short name	Aliases	Description
In vivo	In vivo	InVivo	In vivo
Cell line	Cell line	CellLine	CellLine
Primary cells	Primary cells	PrimaryCells	PrimaryCells
Tissue	Tissue	Tissue	Tissue
Organ	Organ	Organ	Organ

4.3 Transformations

4.3.1 Box Cox power transformation

The Box-Cox power transformation is a data transformation to achieve a better normality and to stabilize the variance. In MCRA, the transformation parameter p in $(y^p - 1)/p$ is determined by maximizing the log-likelihood function

$$l(p) = -\frac{n}{s} \log \left[\frac{1}{n} \sum_{i=1}^{n} (y_i^{(p)} - \overline{y^{(p)}})^2 \right] + (p-1) \sum_{i=1}^{n} \log y_i$$

where i indexes the n observations and

$$\overline{y^{(p)}} = \frac{1}{n} \sum_{i=1}^{n} y_i^{(p)}$$

is the average of the $y_i^{(p)}$ (Box & Cox, 1964) [9].

4.4 Gauss-Hermite

4.4.1 Gauss-Hermite integration

4.4.2 One-dimensional Gauss-Hermite integration

Gauss-Hermite integration approximates a specific integral as follows

$$\int_{-\infty}^{\infty} f(x)\exp(-x^2)\mathrm{d}x \approx \sum_{j=1}^{N} w_j f(x_j$$

in which w_j and x_j are weights and abscissas for N-point Gauss-Hermite integration, see Abramowitz and Stegun (1972) [7]. N-point integration is exact for all polynomials $f(x)$ of degree 2N-1, see Dahlquist and Björck (1974) [14]. This can for instance be used to approximate the mean of a function $F(Y)$ of a normally distributed random variable Y with mean μ and variance σ^2:

$$\int_{-\infty}^{\infty} F(x)\frac{1}{\sqrt{2\pi\sigma}}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)\mathrm{d}y$$

$$= \int_{-\infty}^{\infty} F(\mu+\sqrt{2}\sigma x)\frac{1}{\sqrt{\pi}}\exp(-x^2)\mathrm{d}x$$

$$= \frac{1}{\sqrt{\pi}}\sum_{j=1}^{N} w_j F(\mu+\sqrt{2}\sigma x_j)$$

4.4.3 Two-dimensional Gauss-Hermite integration

One-dimensional Gauss-Hermite integration can readily be extended to two dimensions. The following principal result in two dimensions is more or less given in Jäckel (2005) [24] for the standard bivariate normal distribution $\phi(x, y; \rho)$ with correlation parameter ρ :

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(x,y)\phi(x,y;\rho)\mathrm{d}x\mathrm{d}y \approx \frac{1}{\pi}\sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j F(\sqrt{2}[ax_i+bx_j], \sqrt{2}[bx_i+ax_j])$$

in which

$$a = \frac{\sqrt{1+\rho}+\sqrt{1-\rho}}{2}$$

and

$$b = \frac{\sqrt{1+\rho}-\sqrt{1-\rho}}{2}$$

as given in Jäckel (2005) [24] .

Jäckel (2005) discusses other Gauss-Hermite approximations to the two-dimensional integral, but found that the approximation given above generally gives the most accurate results. For the general bivariate normal distribution with means (μ_x, μ_y) and variances (σ_x^2, σ_y^2) the integral can be approximated by means of

$$\frac{1}{\pi}\sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j F(\mu_x+\sigma_x\sqrt{2}[ax_i+bx_j], \mu_y+\sigma_y\sqrt{2}[bx_i+ax_j])$$

The product $w_i w_j$ can be very small, especially when many quadrature points are used, thus wasting possibly precious calculation time. This can be remedied by pruning, i.e. by dropping combinations of (i, j) with very small values of the product $w_i w_j$.

4.4.4 Maximum likelihood for the LNN model with two-dimensional Gauss-Hermite integration

Denote non-consumption on day j for individual i as $Y_{ij} = 0$. The conditional likelihood, i.e. given random effects b_i and v_i, of a non-consumption on day j equals, with $H()$ the inverse of the logit function

$P(Y_{ij} = 0|b_i, v_i) = 1–H(\lambda + v_i)$.

The conditional likelihood of a positive intake $Y_{ij} > 0$ equals, with ϕ the density of the normal distribution

$$f(Y_{ij} = y_{ij}|y_{ij} > 0, b_i, v_i) = H(\lambda + v_i)\phi(y_{ij} - \mu - b_i; 0, \sigma_w^2)$$

The conditional likelihood contribution for individual i is the product of the individual contributions for each day. The marginal likelihood contribution for individual i is obtained by integrating over the possible values of b_i and v_i. Since the pair (b_i, v_i) follows a bivariate normal distribution, the likelihood contribution for individual i can be approximated by means of two-dimensional Gauss-Hermite integration. Individually based covariabels, such as sexe or age, imply that μ_i and λ_i must be used instead of μ and λ. The likelihood must be optimized by means of some general optimization routine.

4.5 Concentration models

Let x denote a random variable from a lognormal distribution. Then, the log transformed variable $y = ln(x)$ is normally distributed with μ and variance σ. The probability density function (p.d.f.) of y may be expressed as:

$$f_y(y, p_0, \mu_y, \sigma_y^2) = p_0 I(y; 0) + (1 - p_0)(1 - I(y; 0)) \cdot \frac{1}{\sqrt{2\pi\sigma_y}} \exp \frac{(y - \mu_y)^2}{2\sigma_y^2}$$

where $p_0 = Pr(y < log(X_{lor}))$, x_{lor} is the limit of reporting and $I(y; 0)$ is an indicator function for $y < log(X_{lor})$. For $p_0 = 0$ the p.d.f. of y reduces to the usual lognormal density. The left truncated density for $y \geq \log(X_{lor})$ may be expressed as:

$$f_y(y; \mu_y, \sigma_y^2) = \frac{1}{\sqrt{2\pi\sigma_y}} \exp \frac{(y - \mu_y)^2}{2\sigma_y^2} / (1 - \Phi(z))$$

with $\Phi(\cdot)$ the standard normal c.d.f. and $z = (\log(x_{lor}) - \mu_y)/\sigma_z$. Model parameters are estimated using maximum likelihood estimation based on the loglikelihood functions specified below. The loglikelihood functions are evaluated in R, using the **optim** algorithm to find estimates for μ_y, σ_y^2 and p_0.

4.5.1 Mixture zero spike and censored lognormal

The loglikelihood may be expressed as:

$$\log L(p_0, \mu_y, \sigma_y^2) = \sum_{i=1}^{n_0} \log(p_0 + (1 - p_0)\Phi(z_i)) + n_1 \log(\frac{1 - p_0}{\sqrt{2\pi\sigma_y}}) - \sum_{i=n_0+1}^{n} \frac{(y_i - \mu_y)^2}{2\sigma_y^2}$$

where $y_i = \log(x_i)$, $\Phi(\cdot)$ is the standard normal c.d.f., $z = (\log(x_{i,lor}) - \mu_y)/\sigma_y$, $z_{lor} = (\log(lor) - \mu_y)/\sigma_y$ with n_0 number of censored values $(x_i < x_{i,lor})$, n_1 number of uncensored values $(x_i \geq x_{i,lor})$ and x_i,$i = 1 \cdots n$.

Multiple values for LOR are allowed.

4.5.2 Censored lognormal

When $p_0 = 0$ the loglikelihood reduces to:

$$\log L(\mu_y, \sigma_y^2) = \sum_{i=1}^{n_0} \log(\Phi(z)) + n_1 \log(\frac{1}{\sqrt{2\pi\sigma_y}}) - \sum_{i=n_0+1}^{n} \frac{(y_i - \mu_y)^2}{2\sigma_y^2}$$

Multiple values for LOR are allowed.

4.5.3 Mixture non-detect spike and truncated lognormal

Ignoring the n_0 values below x_{lor}, the loglikelihood may be expressed as:

$$\log L(\mu_y, \sigma_y^2) = -n_1 \log(1 - \Phi(z)) + n_1 \log(\frac{1}{\sqrt{2\pi\sigma_y}}) - \sum_{i=n_0+1}^{n} \frac{(y_i - \mu_y)^2}{2\sigma_y^2}$$

Only one value for LOR is allowed.

4.5.4 Mixture non-detect spike and lognormal

Ignoring the n_0 values below x_{lor}, the loglikelihood may be expressed as:

$$\log L(\mu_y, \sigma_y^2) = n_1 \log(\frac{1}{\sqrt{2\pi\sigma_y}}) - \sum_{i=n_0+1}^{n} \frac{(y_i - \mu_y)^2}{2\sigma_y^2}$$

Only one value for LOR is allowed.

4.6 Chronic exposure assessment, daily consumed foods

4.6.1 Model based usual intake

Foods are consumed on a daily basis.

For individual i on day j let Y_{ij} denote the 24 hour recall of a food ($i = 1...n; j = 1...n_i$). In most cases within-individual random variation is dependent on the individual mean and has a skewed distribution. It is therefore customary to define a one-way random effects model for Y_{ij} on some transformed scale

$$Y_{ij}^* = g(Y_{ij}) = \mu_i + b_i + w_{ij}$$

with $b_i \sim N(0, \sigma_b^2)$ and $w_{ij} \sim N(0, \sigma_w^2)$

Note that b_i represents variation between individuals and w_{ij} represents variation within individuals between days.

The mean μ_i may depend on a set of covariate $Z_i = (Z_{i1}, ..., Z_{ip})$:

$$\mu_i = \beta_0 + \beta_1^t Z_i$$

where β_0 and β_1 are regression coefficients.

The usual intake T_i for an individual i is defined as the mean consumption over many many days. This assumes that the untransformed intakes Y_{ij} are unbiased for true usual intake rather than the transformed intakes Y_{ij}^*. In mathematical terms T_i is the expectation of the intake for this individual where the expectation is taken over the random day effect:

$$T_i = E_w[g^{-1}(\mu_i + b_i + w_{ij})|b_i] = F(b_i)$$

4.6.2 Model based usual intake on the transformed scale

For the model based usual intake first note that the conditional distribution

$$(\mu_i + b_i + w_{ij}|b_i) \sim N(\mu_i + b_i, \sigma_w^2)$$

It follows that the usual intake T_i is given by

$$T_i = E_w[g^{-1}(\mu_i + b_i + w_{ij}|b_i)] = \int_{-\infty}^{\infty} g^{-1}(\mu_i + b_i + w_{ij}) \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left(-\frac{w^2}{2\sigma_w^2}\right) \mathrm{d}w$$

4.6.3 Model based using a logarithmic transformation

For the logarithmic transform the usual intake T_i can be written in closed form using the formula for the mean of the lognormal distribution:

$$T_i = \exp(\mu_i + b_i + \sigma_w^2/2)$$

In this case T_i follows a log-normal distribution with mean $\mu_i + \sigma_w^2/2$ and variance σ_b^2. This fully specifies the usual intake distribution, e.g. the mean and variance of the usual intake are given by

$$\mu_{iT} = E[T_i] = \exp(\mu_i + \sigma_w^2/2 + \sigma_b^2/2)$$

$$\sigma_{iT}^2 = Var[T_i] = [\exp(\sigma_b^2) - 1]\exp(2\mu_i + \sigma_w^2 + \sigma_b^2)$$

4.6.4 Model based using a power transformation

For the *power transformation* the integral can be approximated by means of N-point Gauss-Hermite integration. This results in the following usual intake

$$T_i \approx \frac{1}{\sqrt{\pi}} \sum_{j=1}^{N} w_j(\mu_i + b_i + \sqrt{2}\sigma_w x_j)^p$$

with p the inverse of the power transformation. A similar approximation can be used for the Box-Cox transformation. There can be a small problem with Gauss-Hermite integration. The summation term $(\mu_i + b_i + \sqrt{2}\sigma_w x_j)^p$ can not be calculated when the factor between round brackets is negative and the power p is not an integer. This can happen when $(\mu_i + b_i)$ is small relative to the between day standard error σ_w. In that case the corresponding term is set to zero. This is not a flaw in the numerical method but in the statistical model since the model allows negative intakes on the transformed scale which cannot be transformed back to the natural scale. The mean and variance of T_i can be approximated again by using Gauss-Hermite integration:

$$\mu_{iT} = E[T_i] = \frac{1}{\sqrt{\pi}} \sum_{k=1}^{N} w_k \frac{1}{\sqrt{\pi}} \sum_{j=1}^{N} w_j(\mu_i + \sqrt{2}\sigma_w x_j + \sqrt{2}\sigma_b x_k)$$

$$\sigma_{iT} = Var[T_i] = \frac{1}{\sqrt{\pi}} \sum_{k=1}^{N} w_k \left[\frac{1}{\sqrt{\pi}} \sum_{j=1}^{N} w_j(\mu_i + \sqrt{2}\sigma_w x_j + \sqrt{2}\sigma_b x_k)\right]^2 - \mu_T^2$$

An alternative method for obtaining model based usual intakes for the power transformation employs a Taylor series expansion for the power, see e.g. Kipnis (2009) [31]. This is however less accurate than Gauss-Hermite integration. For the power transformation simulation is required to derive the usual intake distribution: simulate a random effect b_i for many individuals and then approximate T_i for these individuals. The T_i values then form a sample form the usual intake distribution.

4.6.5 Model assisted usual intake on the transformed scale

The model assisted approach employs a prediction for the usual intakes of every individual in the study. This requires a prediction of the individual random effect b_i for every individual.

In the one-way random effects model the Best Linear Unbiased Prediction for $(\mu_i + b_i)$ is given by

$$BLUP_i = \mu_i + (\bar{Y}_i^* - \mu_i)\left(\frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2/n_i}\right)$$

in which $\bar{Y}_i^*$ is the mean of the transformed intakes for individual i. BLUPs have optimal properties for some purposes, but not for the purpose of representing the variation σ_b^2 between individuals. This can be seen by noting that

$$Var(\bar{Y}_i^*) = \sigma_b^2 + \sigma_w^2/n_i$$

and thus

$$Var(BLUP_i) = \left(\frac{\sigma_b^4}{\sigma_b^2 + \sigma_w^2/n_i}\right)$$

which is smaller than the between individual variance σ_b^2. As an alternative a modified BLUP can be defined by means of

$$modifiedBLUP_i = \mu_i + (\bar{Y}_i^* - \mu_i)\sqrt{\left(\frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2/n_i}\right)}$$

which has the correct variance σ_b^2 and also the correct mean μ_i. However these optimal properties disappear when modified BLUPs are directly backtransformed to the original scale.

4.6.6 Model assisted using a logarithmic transformation

For the logarithmic transformation the usual intake T_i follows a log-normal distribution with mean $\mu_i + \sigma_w^2/2$ and variance σ_b^2. If we can construct a BLUP like stochastic variable with the same mean and variance, then this variable be an unbiased predictor with the correct variance. It is easy to see that the following variable has the same distribution as T_i

$$modelassistedBLUP_i = \mu_i + \frac{\sigma_w^2}{2} + (\bar{Y}_i^* - \mu_i)\sqrt{\left(\frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2/n_i}\right)}$$

So the model assisted individual intake $\exp(modelassistedBLUP_i)$ has the same distribution as the usual intake and is thus the best predictor for usual intake.

Kipnis et al. (2009) [31] employs the conditional distribution of b_i given the observations $Y_{i1}, \cdots, Y_{in_i}$ to obtain a prediction. First note that

$$(b_i|Y_{i1}, \cdots, Y_{in_i}) = (b_i|Y_{i1}^*, \cdots, Y_{in_i}^*) = (b_i|\bar{Y}_i^*)$$

Since all distributions in the one-way random effects model are normal it follows that:

$$(b_i, \bar{Y}_i^*) \sim BivariateNormal(0, \mu_i, \sigma_b^2, \sigma_b^2 + \sigma_w^2/n_i, \sigma_b^2)$$

where the last parameter represents the covariance between b_i and $\bar{Y}_i^*$. It follows that the conditional distribution

$$(b_i|\bar{Y}_i^*) \sim N(\mu_c, \sigma_c^2)$$

with

$$\mu_c = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2/n_i}(\bar{Y}_i^* - \mu_i)$$

and

$$\sigma_c^2 = \frac{\sigma_b^2\sigma_w^2/n_i}{\sigma_b^2 + \sigma_w^2/n_i}$$

A prediction for the usual intake $T_i = F(b_i)$ is then obtained by the expectation

$$E[F(b_i)|\bar{Y}_i^*] = \int F(b)\phi(b; \mu_c, \sigma_c^2)\mathrm{d}b$$

For the logarithmic transform $F(b_i) = \exp(\mu_i + b_i + \sigma_w^2/2)$ and the expectation reduces to

$$E[F(b_i)|\bar{Y}_i^*] = \exp(\mu_i + \mu_c + \sigma_c^2/2 + \sigma_w^2/2)$$

which is a function of $\bar{Y}_i^*$ through μ_c. To obtain the mean and variance of the prediction note that

$$\mu_i + \mu_c + \sigma_c^2/2 + \sigma_w^2/2 \sim N\left(\mu_i + \frac{\sigma_b^2\sigma_w^2/n_i}{2(\sigma_b^2 + \sigma_w^2/n_i)} + \frac{\sigma_w^2}{2}, \frac{\sigma_b^4}{\sigma_b^2 + \sigma_w^2/n_i}\right)$$

It follows that the expectation of the prediction equals

$$E[E[F(b_i)|\bar{Y}_i^*]] = \exp\left(\mu_i + \frac{\sigma_b^2\sigma_w^2/n_i}{2(\sigma_b^2+\sigma_w^2/n_i)} + \frac{\sigma_w^2}{2} + \frac{\sigma_b^4}{2(\sigma_b^2+\sigma_w^2/n_i)}\right)$$
$$= \exp\left(\mu_i + \frac{\sigma_b^2}{2} + \frac{\sigma_w^2}{2}\right)$$

which equals the mean of the usual intake. However the variance of the prediction equals

$$Var[E[F(b_i|\bar{Y}_i^*]] = \left[\exp\left(\frac{\sigma_b^4}{(\sigma_b^2+\sigma_w^2/n_i)}\right) - 1\right]\exp(2\mu_i + \sigma_b^2 + \sigma_w^2)$$

Which is less than the variance of the usual intake. The approach of Kipnis et al (2009) [31] will therefor result in too much shrinkage of the model assisted usual intake.

4.6.7 Model assisted using a power transformation

For the power transformation a model assisted BLUP with optimal properties, as derived above, cannot be constructed. The approach of Kipnis et al. (2009) [31] can however be used to obtain a prediction in the following way. First approximate $T_i = F(b_i)$ by Gauss-Hermite integration:

$$F(b_i) = T_i \approx \frac{1}{\sqrt{\pi}}\sum_{j=1}^{N} w_i(\mu_i + b_i + \sqrt{2}\sigma_w x_i)^p$$

Secondly again use Gauss-Hermite to approximate the expectation of the conditional distribution giving the prediction P_i.

$$P_i = E[F(b_i)|\bar{Y}_i^*] = \int F(b_i)\phi(b;\mu_c,\sigma_c^2)\mathrm{d}b \approx \frac{1}{\pi}\sum_{k=1}^{N} w_k \sum_{j=1}^{N} w_j(\mu_i + \mu_c + \sqrt{2}\sigma_w x_j + \sqrt{2}\sigma_c x_k)^p$$

which is a function of $\bar{Y}_i^*$ through μ_c. It is likely that the thus obtained predictions P_i have a variance that is too small. If we would know the mean μ_{iP} and variance σ_{iP}^2 of the predictions, the predictions could be linearly rescaled to have the correct mean μ_{iT} and variance ${}^2_{iT}$. The mean and variance of the prediction can be calculated using *Gauss-Hermite integration*.

$$\mu_{iP} = \frac{1}{\sqrt{\pi}}\sum_{l=1}^{N} w_l \frac{1}{\pi}\sum_{k=1}^{N} w_k \sum_{j=1}^{N} w_j(\mu_i + \sqrt{2}\frac{\sigma_b^2}{\sigma_b^2+\sigma_w^2/n_i}x_l + \sqrt{2}\sigma_w x_j + \sqrt{2}\sigma_c x_k)^p$$

$$\sigma_{iP}^2 = \frac{1}{\sqrt{\pi}}\sum_{l=1}^{N} w_l\left[\frac{1}{\pi}\sum_{k=1}^{N} w_k \sum_{j=1}^{N} w_j(\mu_i + \sqrt{2}\frac{\sigma_b^2}{\sigma_b^2+\sigma_w^2/n_i}x_l + \sqrt{2}\sigma_w x_j + \sqrt{2}\sigma_c x_k)^p\right]^2 - \mu_{iP}^2$$

The proposed prediction then equals

$$P_i^* = \mu_{iT} + \frac{\sigma_{iT}}{\sigma_{iP}}(P_i - \mu_{iP})$$

4.7 Chronic exposure assessment, episodically consumed foods

For episodically consumed foods we need to take the probability of consumption into account. Define p_i as the probability that individual i consumes the food on any given day. The usual intake for this individual is then given by the product of p_i and T_i which is now defined as the usual amount on consumption days. Since individuals will vary in their probability pi, besides modelling the amounts as for daily consumed foods, it is also necessary to model the frequency of consumption. A three stage analysis of 24-hour recall data is the necessary:

1. A model for the frequency of consumption
2. A model for the intakes on consumption days

3. Integration of both models in order to obtain a usual intake distribution.

Step 2 uses the analysis outlined in the previous section for the positive intakes only. For step 1 two popular models which describe between-individual variation for the probability of consumption are the beta-binomial model and the logistic-normal model.

4.7.1 Beta-Binomial model for frequencies (BBN)

Let n_i be the total number of recall days for individual i and X_i the number of days with a positive intake. The distribution of X_i, with p_i the probability of consumption for individual i, is given by

$$X_i = \mathit{Binomial}(n_i, p_i)$$

In this model the probability p_i varies among individuals according to the Beta distribution:

$$f(p) = B^{-1}(\alpha, \beta)p^{\alpha-1}(1-p)^{\beta-1}$$

with

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Combining the binomial and the Beta distribution results in the betabinomial distribution:

$$P(X_i = x) = \begin{pmatrix} n_i \\ r \end{pmatrix} \frac{B(\alpha + x, n_i + \beta - x)}{B(\alpha, \beta)}$$

The mean and variance of the betabinomial distribution are given by

$$E[X_i] = n_i \frac{\alpha}{\alpha + \beta}$$

and

$$\mathit{Var}[X_i] = n_i \frac{\alpha\beta(\alpha + \beta + n_i)}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Using the reparameterization $\pi = \alpha/(\alpha + \beta)$ and $\phi = 1/(\alpha + \beta + 1)$, it follows that

$$E[X_i] = n_i\pi$$

and

$$\mathit{Var}[X_i] = n_i\pi(1-\pi)[1 + (n_i - 1)\phi]$$

This reparameterization enables to model the probability π_i of consumption for individual i directly as a logistic regression:

$$\mathit{logit}(\pi_i) = \gamma_0 + \gamma_1^t Z_i$$

Note that the dispersion parameter ϕ: is assumed to be equal for all individuals. The betabinomial logistic regression model can be fitted by means of maximum likelihood.

4.7.2 Model based frequencies for usual intake

For the model based usual intake distribution the estimated parameters π_i and ϕ are backtransformed using $\alpha_i = \pi_i\phi/(1-\phi)$ and $\beta_i = (1-\pi_i)\phi/(1-\phi)$. These can then be used to draw from the Beta distribution.

4.7.3 Model assisted frequencies for usual intake

For the model assisted usual intake distribution a prediction of the consumption probability is required for every individual. Simple predictions are

1. the observed frequencies for every individual or
2. the fitted probability for evey individual. When there are no covariables the fitted probability is the same for every individual.
3. Alternatively one can use the approach outlined in Kipnis et al (2009) employing the conditional expectation of the probability given the observed frequency:

$$
\begin{aligned}
E(p_i|X_i = x) &= \int_p p f(p|X_i = x)\mathrm{d}p \\
&= \int_p p \frac{f(X_i = x|p)f(p)}{\int f(X_i = x|p)f(p)dp}\mathrm{d}p \\
&= \frac{1}{P(x_i = x)} \int_p p \begin{pmatrix} n_i \\ r \end{pmatrix} p^x (1-p)^{n_i - x} B^{-1}(\alpha_i, \beta_i) p^{\alpha_i - 1}(1-p)^{\beta_i - 1}\mathrm{d}p \\
&= \frac{B^{-1}(\alpha_i, \beta_i)}{P(x_i = x)} \begin{pmatrix} n_i \\ r \end{pmatrix} \int_p p^{\alpha_i + x}(1-p)^{n_i + \beta_i - x - 1}\mathrm{d}p \\
&= \frac{B(\alpha_i + x + 1, n_i + \beta_i - x)}{B(\alpha_i + x, n_i + \beta_i - x)} \\
&= \frac{\alpha_i + x}{\alpha_i + \beta_i - x}
\end{aligned}
$$

For individual with zero intakes on all recall days a prediction for the random individual amount effect b_i is not available. There seem to be two option for predicting the usual intake for such individuals:

- Set the individual intake to zero
- Simulate a model based prediction for the amount and combine this with the conditional expected probability given above to obtain an individual usual intake.

4.7.4 Logistic-Normal model for frequencies (LNN0)

In this model the distribution of X_i is again binomial:

$$X_i = \mathit{Binomial}(n_i, p_i)$$

The probability p_i is now given by a logistic regression with a random effect in the linear predictor which represents the between-individual variation in the probability p_i

$\mathit{logit}(p_i) = \lambda_i + v_i$ with $v_i \sim N(0, \sigma_v^2)$ and the regression equation $\lambda_i = \gamma_0 + \gamma_1^t Z_i$

The marginal probability π_i is obtained by integrating over the random effect v_i, i.e. using *Gauss-Hermite integration*

$$\pi_i = \int H(\lambda_i + v)f(v)dv \approx \frac{1}{\sqrt{\pi}} \sum_{j=1}^{N} w_j H(\lambda_i + \sqrt{2}\sigma_v x_j)$$

in which $H()$ is the inverse of the logit transformation. Note that this is different from $\mathit{logit}^{-1}(\lambda_i)$ which is the median probability. The model can be fitted by maximum likelihood using Gauss-Hermite integration. An (approximate) maximum likelihood procedure is implemented in routine glmer of the lme4 package in R. For a new vector of covariates Z_i^* the linear predictor λ_i^* can be calculated along with its standard error $\mathit{Se}(\lambda_i^*)$. The marginal predicted probability π_i^* can be calculated by means of Gauss-Hermite integration and the standard error of the predicted probability can be calculated by means of the usual Taylor series expansion:

$$\mathit{Se}(\pi_i^*) \approx \frac{\mathit{Se}(\lambda_i^*)}{\sqrt{\pi}} \sum_{j=1}^{N} w_j \frac{d}{d\lambda_i^*} H(\lambda_i^* + \sqrt{2}\sigma_v x_j)$$

$$= \frac{Se(\lambda_i^*)}{\sqrt{\pi}} \sum_{j=1}^{N} w_j H(\lambda_i^* + \sqrt{2}\sigma_v x_j)[1 - H(\lambda_i^* + \sqrt{2}\sigma_v x_j)]$$

4.7.5 Model based frequencies for usual intake

For the model based usual intake distribution the estimated parameters λ_i and σ_v^2 can be used to generate individual probabilities.

4.7.6 Model assisted frequencies for usual intake

For the model assisted usual intake distribution simple predictors are (a) the observed frequencies and (b) the marginal probability π_i. The conditional expectation (c) is given by

$$E(p_i|X_i = x) = \int_v H(\lambda_i + v) f(v|X_i = x)\mathrm{d}v$$

$$= \int_v H(\lambda_i + v) \frac{f(X_i = x_i|v) f(v)}{\int f(X_i = x_i|v) f(v)\mathrm{d}v} \mathrm{d}v$$

$$= \frac{\int_v H(\lambda_i + v)[H(\lambda_i + v)]^{x_i}[1 - H(\lambda_i + v)]^{n_i - x_i} f(v)\mathrm{d}v}{\int_v [H(\lambda_i + v)]^{x_i}[1 - H(\lambda_i + v)]^{n_i - x_i} f(v)\mathrm{d}v}$$

and both nominator and denominator can be approximated by means of the *Gauss-Hermite integration*. For individual with zero intakes on all recall days see above for the two options.

4.7.7 Logistic-Normal model for frequencies correlated with amounts (LNN)

This model is extends the LNN0 model with a correlation between the individual random effect b_i for amounts and the individual random effect v_i for frequencies. This model is also known as the NCI model and is introduced by Tooze et al (2006) [43] with further mathematical details in Kipnis et al (2009) [31]. The model can be written as

$$\mathit{logit}(P(Y_{ij} > 0)) = \lambda_i + v_i$$

$$g(Y_{ij}) = \mu_i + b_i + w_{ij}$$

and $(v_i, b_i) \sim \mathit{BivariateNormal}(0, 0, \sigma_v^2, \sigma_b^2, \rho)$ and $w_{ij} \sim N(0, \sigma_w^2)$

The model can be fitted by maximum likelihood employing *two-dimensional Gauss-Hermite integration*.

4.7.8 Model based usual intake.

Model based usual intake requires generation of the pair (v_i, b_i) for many hypothetical individual. The usual intake U_i for such a hypothetical individual is then given by

$$U_i = H(\lambda_i + \nu_i) T_i$$

$$= H(\lambda_i + \nu_i) E_w[g^{-1}(\mu_i + b_i + w_{ij})|b_i]$$

$$= H(\lambda_i + \nu_i) F(b_i)$$

The second term can be calculated using the method outlined for daily intakes.

4.7.9 Model assisted usual intake.

This requires simulatenous prediction of the random effect for frequency and for amount as outlined in Kipnis et al (2009) [31]. We have for individual i in the study $(U_i|Y_{i1}, \cdots, Y_{in_i}) = (U_i|Y^*_{i1}, \cdots, Y^*_{in_i}) = (U_i|x_i, \bar{Y}^*_i)$ where x_i is the number of positive intakes and $\bar{Y}^*_i$ is the mean of the transformed **positive** intakes. It follows that the required conditional expectation P_i equals

$$P_i = E[U_i|x_i, \bar{Y}^*_i]$$

$$= E_{v_i,b_i}[H(\lambda_i + v_i)F(b_i)|x_i, \bar{Y}^*_i]$$

$$= \frac{\int\int H(\lambda_i + v_i)F(b_i)f(x_i, \bar{Y}^*_i|v_i, b_i)\phi(v_i, b_i)dv_i db_i}{\int\int f(x_i, \bar{Y}^*_i|v_i, b_i)\phi(v_i, b_i)dv_i db_i}$$

where

$$f(x_i, \bar{Y}^*_i|v_i, b_i) = [H(\lambda_i + v_i)]^{x_i}[1 - H(\lambda_i + v_i)]^{n_i - x_i}\phi(\bar{Y}^*_i - \mu_i - b_i; 0, \sigma^2_w/x_i)$$

Both nominator and denominator can be approximated by two-dimensional Gauss-Hermite integration. Note that for the log-transform $F(b_i) = T_i$ = $\exp(\mu_i + b_i + \sigma^2_w)/2)$ can be calculated exactly; for the power transformation an approximation must be used. It can be expected that the predicted usual intake will not have the correct variance. This can possibly be remedied by equating the mean and variance of U_i and P_i. These are however rather involved to calculate.

For individual with zero intakes on all recall days the model assisted usual intake can be set to zero, or can be simulated as follows

1. Calculate the Model assisted frequency P_0 for usual intake (see LNN0)
2. Transform P_0 back to the logistic scale, i.e. $L_0 = logit(P_0)$. Get the conditional distribution of

$$(b|v = L_0 - \lambda_i) \sim N\left(\frac{\sigma_b}{\sigma_v}\rho(L_0 - \lambda_i), (1 - \rho^2)\sigma^2_b\right)$$

3. Simulate a draw b_0 from this conditional distribution and obtain the usual intake as $P_0 \exp(\mu_i + b_0 + \sigma^2_w)$

Note that the backtransformation from P_0 to L_0 is according to the median of the distribution rather than the mean.

4.8 Unit variability

A composite sample for food k is composed of nu_k units with nominal unit weight wu_k. The weight of a composite sample is $wm_k = nu_k \cdot wu_k$ with mean residue value cm_k.

4.8.1 Beta distribution

Under the beta model simulated unit values are drawn from a bounded distribution on the interval (0, c_{max}) with $c_{max} = nu_k \cdot cm_k$. The standard beta distribution is defined on the interval (0, 1) and is usually characterised by two parameters a and b, with $a > 0$, $b > 0$ (see e.g. Mood et al. 1974) [33]. Alternatively, it can be parameterised by the mean

$$\mu = a/(a + b)$$

and the variance

$$\sigma^2 = ab/(a + b + 1)^{-1}(a + b)^{-2}$$

or, as applied in MCRA, by the mean μ and the squared coefficient of variation

$$cv^2 = ba^{-1}(a + b + 1)^{-1}$$

For the simulated unit values in *each iteration of the program* we require an expected value cm_k. This scales down to a mean value $\mu = cm_k/c_{max} = 1/nu_k$ in the (standard) beta distribution. From this value for μ and an externally specified value for cv_k the parameters a and b of the beta distribution are calculated as:

$$a = b(nu_k - 1)^{-1}$$

and

$$b = \frac{(nu_k - 1)(nu_k - 1 - cv_k^2)}{nu_k cv_k^2}$$

From the second formula it can be seen that cv_k should not be larger than $\sqrt{nu_k - 1}$ in order to avoid negative values for b. When the unit variability is specified by a variability factor

$$v_k = \frac{p97.5_k}{cm_k}$$

instead of a coefficient of variation cv_k then MCRA applies a bisection algorithm to find a such that the cumulative probability

$$P[Beta(a,b)] = 0.975$$

for $b = a(nu_k - 1)$.

Sampled values from the beta distribution are rescaled by multiplication with cm_{max} to unit concentrations c_{ijk} on the interval $(0, cm_{max})$.

4.8.2 Lognormal distribution

The lognormal distribution is characterised by μ and σ, which are the mean and standard deviation of the log-transformed concentrations. The unit log-concentrations are drawn from a normal distribution with mean $\mu = ln(cm_{ik}) - 1/2\sigma^2$. The coefficient of variation cv is turned into the standard deviation σ on the log-transformed scale with:

$$\sigma = \sqrt{ln(cv^2 + 1)}$$

The variability factor is defined as the 97.5th percentile of the concentration in the individual measurements divided by the corresponding mean concentration seen in the composite sample. A variability factor v is converted into the standard deviation σ as follows:

$$v = \frac{p97.5}{mean} = \frac{e^{\mu+1.96\sigma}}{e^{\mu+1/2\sigma^2}} = e^{1.96\sigma-1/2\sigma^2}$$

with µ and σ representing the mean and standard deviation of the log-transformed concentrations. So

$$ln(v) = 1.96\sigma – 1/2\sigma^2$$

Solving for σ gives:

$$\sigma^2 – 2 \cdot 1.96\sigma + 2log(v) = 0$$

with roots for σ according to:

$$\sigma = 1.96 \pm \sqrt{(1.96^2 – 2log(v))}$$

The smallest positive root is taken as an estimate for σ.

4.8.3 Bernoulli distribution

The bernoulli model is a limiting case of the beta model, which can be used if no information on unit variability is available, but only the number of units in a composite sample is known (see van der Voet et al. 2001). As a worst case approach we may take the coefficient of variation cv as large as possible. When cv is equal to the maximum possible value $\sqrt{nu_k - 1}$, the (unstandardised) beta distribution simplifies to a bernoulli distribution with probability

$$(nu_k–1)\,/nu_k$$

or

$$(v_k - 1)/v_k$$

for the value 0 and probability

$$1/nu_k$$

or

$$1/v_k$$

for the value $c_{max} = nu_k \cdot cm_k$.

In MCRA values 0 are actually replaced by cm_k, to keep all values on the conservative side. For example, with nu_k = 5, there will be 80% probability at $c_{ijk} = cm_k$ and 20% probability at $c_{ijk} = c_{max}$. When the number of units nu_k in the composite sample is missing, the nominal unit weight wu_k is used to calculate the parameter for unit variability.

4.9 Screening calculation for large Cumulative Assessment Groups

4.9.1 Statistical model for the screening step (acute exposure)

The screening step implements a simple model that is applied to each SCC. Assume independent NonDetectSpike-LogNormal (NDS-LN) models for both the consumptions of food-as-measured in source S and the concentrations of compound C in source S. A non-detect consumption is assumed to be a zero consumption. A non-detect concentration will be imputed by a user-specified fraction f of the Limit of Reporting. Then the model for consumption has 3 parameters and the model for concentration has four parameters, as specified in Table 4.12. Note that the parameters of the consumption distribution are estimated from the consumption data using sampling weights if these have been provided in the consumption data set.

Table 4.12: Parameters for screening models (per source/compound)

parameter	consumptions	concentrations
probability of a positive	π_x	π_c
mean positives (ln scale)	μ_x	μ_c
standard deviation positives (ln scale)	σ_x	σ_c
value to use for NonDetects (ln scale)		$f \cdot L_c$

Exposure is consumption times concentration, so on logarithmic scale they can be added

$$e = x + c.$$

The assessment will focus on a chosen percentile of exposure, e.g. p95. The relevant fraction will be denoted by p, for example $p = 0.95$ for the 95th percentile. The two NDS-LN models combine to three possibilities, depending on whether there is consumption and if so, whether the concentration is non-detect or positive. In the screening model the two possibilities that lead to potential exposure are modelled with a mixture of two lognormal distribution. For the non-detect case the positive exposure distribution equals the positive consumption distribution modified by the multiplication of a user-chosen factor times an estimate of the average worst-case limit value for concentration (LOR):

$$\pi_1 = \pi_x(1-\pi_c); \mu_1 = \mu_x + f \cdot L_c; \sigma_1 = \sigma_x$$

where L_c is the logarithm of the LOR, or, if there are multiple analytical methods with different LOR, a weighted average of these different LORs.

For the detect case the positive exposure distribution is easily combined from the positive consumption distribution and the positive concentration distribution:

$$\pi_2 = \pi_x \pi_c; \mu_2 = \mu_x + \mu_c; \sigma_{12} = \sqrt{\sigma_x^2 + \sigma_c^2}$$

p can be corrected for the non-consumptions to the appropriate fraction needed in the mixture of the two positive distributions:

$$p' = \frac{p - (1-\pi_x)}{\pi_x}$$

If $p' \leq 0$ then all positive exposures are beyond the requested fraction, and the estimated exposure is just 0.

If $p' > 0$ then the relevant log exposure e_p satifies

$$(1-\pi_c) \cdot \Phi\left(\frac{e_p - \mu_1}{\sigma_1}\right) + \pi_c \cdot \Phi\left(\frac{e_p - \mu_{12}}{\sigma_2}\right) = p'$$

where $\Phi(\cdot)$ represents the cumulative standard normal distribution function. The value of e_p can easily be found in a bisection search within the interval

$$[\mu_{min} - 4\sigma_{max}, \mu_{max} + max(0, z_{p'}\sigma_{max})].$$

The final exposure percentile estimate then is exp(e_p).

Denote by $e_{(p,max)}$ the highest estimate (for the SCC denoted by $SSC_{highest}$). Then evaluate for each SCC the probability to exceed $e_{(p,max)}$.

$$P_i = Pr(e > e_{p,max}) = \pi_x \cdot \left[(1-\pi_c) * \Phi\left(\frac{e_{p,max} - \mu_1}{\sigma_1}\right) + \pi_c \cdot \Phi\left(\frac{e_{p,max} - \mu_2}{\sigma_1}\right)\right]$$

P_i is a tentative measure for the 'probability of a high exposure'. For $SSC_{highest}$ $P_i = 1 - p$, for all other SCCs it will be lower. The sum of all these probabilities is not a meaningful probability in itself. However, this sum is used to scale the individual P_i values to measures of relative importance for the SCCs

$$Imp_i = P_i / \sum P_i$$

Rank all SCCs according to Imp_i and calculate cumulative importance. The relative importance of the two mixture components at e_p can be estimated as

$$w_{1,2} = \frac{\pi_{1,2} \cdot \phi\left(\frac{e_p - \mu_{1,2}}{\sigma_{1,2}}\right)/\sigma_{1,2}}{\pi_1 \cdot \phi\left(\frac{e_p - \mu_1}{\sigma_1}\right)/\sigma_1 + \pi_2 \cdot \phi\left(\frac{e_p - \mu_2}{\sigma_2}\right)/\sigma_2}$$

where $\phi(.)$ represent the standard normal probability density function. The user interface should allow to select the top-N SCCs from the list, based on a chosen percentage (e.g. 95%) of cumulative importance included. The full analysis will calculate exactly the same exposure distribution as a full analysis without screening. However, less information is retained in the output. This concerns tables with information on foods-as-eaten, which is only shown for the selected risk driver components (SCCs). Risk drivers are groupings of SCCs (risk driver components) at the level of measured-source-compound combinations (MSCCs). Note that output for an MSSC (e.g. APPLE/captan) only covers the selected SCCs (e.g. APPLE from apple juice/captan and APPLE from apple pie/captan), but not unselected SCCs (e.g. APPLE from fruit yoghurt/captan).

4.9.2 Statistical model for the screening step (chronic exposure)

In chronic exposure assessments, the mean concentration of chemicals is calculated first, and combined with the consumption distribution. For this reason a chronic calculation uses less memory, and therefore larger datasets can be handled. The model described under Acute can be simplified for a chronic screening. The concentration distribution is only used to estimate a mean exposure, incorporating any effect from the imputation of non-detects. The exposure distribution is therefore only a scaled version of the consumption distribution.

$$\pi_2 = \pi_x \pi_c; \mu_2 = \mu_x + \mu_c; \sigma_2 = \sigma_x$$

The parameters of the consumption distribution (π_x, μ_x, σ_x) are calculated from the observed individual means (OIMs), i.e. the mean daily consumptions over the survey days of each person in the data (allowing for sampling weights). The percentiles are calculated as $e_p = \mu_2 + z_p$ where z is a percentile of the standard normal distribution. The exceedances of the maximum percentile are calculated as

$$P_i = Pr(e > e_{p,max}) = \pi_x \cdot \Phi\left(\frac{e_{p,max} - \mu_2}{\sigma_2}\right)$$

CHAPTER

FIVE

COLOPHON

WUR/Biometris, Wageningen University & Research
FERA, Food and Environmental Research Agency
RIVM, National Institute for Public Health and the Environment

This manual was created for MCRA 9.0 (build 9.0.35.0) for Prod-RIVM on Wed Jan 8 09:43:20 2020.

5.1 Contributors to MCRA

Main programmers of MCRA are:
Waldo de Boer, Johannes Kruisselbrink, Marco van Lenthe

Many people contributed to the MCRA code over the years:
Frits van Evert, Jack van Galen, Paul Goedhart, Gerie van der Heijden, Hans van den Heuvel, Paul Keizer, Marcel Koenders, Jaap Kokorian, Sanne Korzec, Helen Owen, Gerrit Polder, Pim Reijersen, Willem Roelofs, Gert-Jan Swinkels, Jac Thissen, Hilko van der Voet.

MCRA has been tested in practice by many people. Feed-back over many years from the team at RIVM is gratefully acknowledged:
Jan Dirk te Biesebeek, Polly Boon, Gerda van Donkersgoed, Jacob van Klaveren, Corinne Sprong.

Contributors to the MCRA Reference Manual:
Waldo de Boer, Paul Goedhart, Andy Hart, Marc Kennedy, Johannes Kruisselbrink, Helen Owen, Willem Roelofs, Hans van den Heuvel, Hilko van der Voet.

WUR/Biometris is the unit for **Mathematical and Statistical Methods of Wageningen University & Research**
P.O. Box 16, 6700 AA Wageningen, Netherlands
WUR Campus, Building 107 (Radix), Droevendaalsesteeg 1, 6708 PB Wageningen
Telephone: +31 (0)317 476925
https://wur.nl
http://www.biometris.nl

Fera Food and Environmental Research Agency
Sand Hutton, York, YO41 1LZ, United Kingdom
Telephone: +44 (0)1904 462000
https://www.fera.co.uk

RIVM National Institute for Public Health and the Environment
P.O. Box 1, 3729 BA Bilthoven, Netherlands
Antonie van Leeuwenhoeklaan 9, 3721 MA Bilthoven
Telephone: +31 30 2749111
https://rivm.nl

BIBLIOGRAPHY

[1] European Food Safety Authority (EFSA). Report on the development of a food classification and description system for exposure assessment and guidance on its implementation and use. *EFSA Journal*, 9(12):84, 2011. doi:doi:10.2903/j.efsa.2011.2489.

[2] European Food Safety Authority (EFSA). The food classification and description system foodex 2 (draft-revision 1). *EFSA Journal*, pages 438, 2011.

[3] European Food Safety Authority (EFSA). Guidance on the use of probabilistic methodology for modelling dietary exposure to pesticide residues. *EFSA Journal*, 10(10):2839, 2012. doi:doi:10.2903/j.efsa.2012.2839.

[4] European Food Safety Authority (EFSA). Guidance on the assessment of exposure of operators, workers, residents and bystanders in risk assessment for plant protection products. *EFSA Journal*, 12(10):3874, 2014. doi:10.2903/j.efsa.2014.3874.

[5] European Food Safety Authority (EFSA). Guidance on dermal absorption. *EFSA Journal*, 6 2017. doi:10.2903/j.efsa.2017.4873.

[6] World Health Organization (WHO). *Guidance document on evaluating and expressing uncertainty in hazard characterization*. World Health Organization, 2018.

[7] MILTON Abramowitz. L. a. stegun, 1972: handbook of mathematical functions. *National Bureau of Standards Applied Mathematics Series*, 55:589–626, 85.

[8] Stephanie Bopp and BERGGREN ELISABET; KIENZLER AUDE; VAN DER LINDEN SANDER; WORTH Andrew. Scientific methodologies for the assessment of combined effects of chemicals - a survey and literature review. *EUR - Scientific and Technical Research Reports*, 2015. doi:10.2788/093511.

[9] George EP Box and David R Cox. An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2):211–243, 1964.

[10] M.C. Butler Ellis, Marc C. Kennedy, C.J. Kuster, R. Alanis, and C.R. Tuck. Improvements in modelling bystander and resident exposure to pesticide spray drift: investigations into new approaches for characterizing the 'collection efficiency'of the human body. *Annals of work exposures and health*, 62(5):622–632, 2018. doi:10.1093/annweh/wxy017.

[11] Camille Béchaux, Mélanie Zetlaoui, Jessica Tressou, Jean-Charles Leblanc, Fanny Héraud, and Amélie Crépet. Identification of pesticide mixtures and connection between combined exposure and diet. *Food and chemical toxicology*, 59:191–198, 2013.

[12] Tebby Cleo, Hilko van der Voet, Georges de Sousa, Emiel Rorije, Vikas Kumar, Waldo de Boer, Johannes H. Kruiselbrink, Frédéric Y. Bois, Moosa Faniband, Angelo Moretto, and Céline Brochot. In prep: integration of in silico and in vitro data in PBPK modeling for risk assessment of food- and non-food-borne chemicals using the Euromix toolbox. *xxx*, 2019.

[13] G.M. Cramer, R.A. Ford, and R.L. Hall. Estimation of toxic hazard—a decision tree approach. *Food and cosmetics toxicology*, 16(3):255–276, 1976. doi:10.1016/S0015-6264(76)80522-6.

[14] G Dahlquist and A Bjorck. Numerical methods (transl. by n. anderson). 1974.

[15] Waldo J de Boer and van Der Voet. Mcra 7. a web-based program for monte carlo risk assessment. reference manual 2011-12-19, documenting mcra release 7.1. Technical Report, Biometris, Wageningen UR and National

Institute for Public Health and the Environment (RIVM), Bilthoven, Wageningen., 2011. URL: https//mcra.rivm.nl.

[16] Waldo J de Boer, Hilko van Der Voet, Bas GH Bokkers, Martine I Bakker, and Polly E Boon. Comparison of two models for the estimation of usual intake addressing zero consumption and non-normality. *Food Additives and Contaminants*, 26(11):1433–1449, 2009.

[17] KW Dodd. A technical guide to c-side. *Ames, Iowa: Department of Statistics and Center for Agricultural and Rural Development, Iowa State University*, 1996.

[18] B Efron. Bootstrap methods: another look at the jackknife annals of statistics 7: 1–26. *View Article PubMed/NCBI Google Scholar*, 1979.

[19] Bradley Efron and Robert J Tibshirani. An introduction to the bootstrap chapman & hall. *New York*, 1993.

[20] Nicolas Gillis and Robert J Plemmons. Sparse nonnegative matrix underapproximation and its application to hyperspectral image analysis. *Linear Algebra and its Applications*, 438(10):3991–4007, 2013.

[21] Paul W. Goedhart, Hilko. van der Voet, S. Knüppel, Arnold L.M. Dekkers, Kevin W. Dodd, Hermann. Boeing, and Jacob D. van Klaveren. A comparison by simulation of different methods to estimate the usual intake distribution for episodically consumed foods. Technical Report, Report: Supporting Publications 2012:EN-299, 2012. URL: http//www.efsa.europa.eu/publications.

[22] Gerald Joseph Goodhardt, Andrew SC Ehrenberg, and Christopher Chatfield. The dirichlet: a comprehensive model of buying behaviour. *Journal of the Royal Statistical Society. Series A (General)*, pages 621–655, 1984.

[23] Patrik O Hoyer. Non-negative matrix factorization with sparseness constraints. *Journal of machine learning research*, 5(Nov):1457–1469, 2004.

[24] Peter Jäckel. A note on multivariate gauss-hermite quadrature. *London: ABN-Amro. Re*, 2005.

[25] Karrer, Cecile, Waldo de Boer, Christiaan Delmaar, Yaping Cai, Amélie Crépet, Konrad Hungerbühler, and Natalie van Goetz. In prep: linking probabilistic exposure and pharmacokinetic modeling to assess the cumulative risk from the bisphenols BPA, BPS, BPF, and BPAF for europeans. *xxx*, 2019.

[26] Marc C Kennedy, C Richard Glass, Bas Bokkers, Andy DM Hart, Paul Y Hamey, Johannes W Kruisselbrink, Waldo J de Boer, Hilko van der Voet, David G Garthwaite, and Jacob D van Klaveren. A european model and case studies for aggregate exposure assessment of pesticides. *Food and Chemical Toxicology*, 79:32–44, 2015.

[27] Marc C Kennedy, Hilko van der Voet, Victoria J. Roelofs, Willem Roelofs, C. Richard Glass, Waldo J de Boer, Johannes W. Kruisselbrink, and Andy D.M. Hart. New approaches to uncertainty analysis for use in aggregate and cumulative risk assessment of pesticides. *Food and Chemical Toxicology*, 79:54–64, 2015.

[28] Marc C. Kennedy, Clare M.J. Butler Ellis, and Paul C.H. Miller. Bream: a probabilistic bystander and resident exposure assessment model of spray drift from an agricultural boom sprayer. *Computers and electronics in agriculture*, 88:63–71, 2012. doi:10.1016/j.compag.2012.07.004.

[29] Marc C. Kennedy and M.C. Butler Ellis. Probabilistic modelling for bystander and resident exposure to pesticides using the browse software. *Biosystems engineering*, 154:105–121, 2017. doi:10.1016/j.biosystemseng.2016.08.012.

[30] Marc C. et al. Kennedy. In prep.: a retain and refine approach for grouping chemicals into cumulative exposure assessment. *xxx*, 2019.

[31] Victor Kipnis, Douglas Midthune, Dennis W Buckman, Kevin W Dodd, Patricia M Guenther, Susan M Krebs-Smith, Amy F Subar, Janet A Tooze, Raymond J Carroll, and Laurence S Freedman. Modeling data with excess zeros and measurement error: application to evaluating relationships between episodically consumed foods and health outcomes. *Biometrics*, 65(4):1003–1010, 2009.

[32] Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788, 1999.

[33] Alexander McFarlane Mood, Franklin A Graybill, and Duane C Boes. *Introduction to the Theory of Statistics 1974*. McGraw-Hill Kogakusha, 1974.

[34] Ian C. Munro, Richard A. Ford, Elke Kennepohl, and James G. Sprenger. Correlation of structural class with no-observed-effect levels: a proposal for establishing a threshold of concern. *Food and Chemical Toxicology*, 34(9):829–867, 1996. doi:10.1016/S0278-6915(96)00049-X.

[35] Sarah M Nusser, Alicia L Carriquiry, Kevin W Dodd, and Wayen A Fuller. A semiparametric transformation approach to estimating usual daily intake distributions. *Journal of the American Statistical Association*, 91(436):1440–1449, 1996.

[36] Sarah M Nusser, Wayne A Fuller, Patricia M Guenther, and others. Estimating usual dietary intake distributions: adjusting for measurement error and nonnormality in 24-hour food intake data. Technical Report, Center for Agricultural and Rural Development (CARD) at Iowa State University, 1997.

[37] European Commission Standing Committee on Plants Animals Food and Feed. European commission working document sante-2015-10216 rev. 7. 2018.

[38] Paul S Price and Xianglu Han. Maximum cumulative ratio (mcr) as a tool for assessing the value of performing a cumulative risk assessment. *International journal of environmental research and public health*, 8(6):2212–2225, 2011.

[39] Lawrence K Saul and Daniel D Lee. Multiplicative updates for classification by mixture models. In *Advances in Neural Information Processing Systems*, 897–904. 2002.

[40] Wout Slob. Probabilistic dietary exposure assessment taking into account variability in both amount and frequency of consumption. *Food and Chemical Toxicology*, 44(7):933–951, 2006.

[41] Wout Slob, Waldo J de Boer, and Hilko van der Voet. Can current dietary exposure models handle aggregated intake from different foods? a simulation study for the case of two foods. *Food and chemical toxicology*, 48(1):178–186, 2010.

[42] Olga W. Souverein, Waldo J. de Boer, Anouk Geelen, Hilko van Der Voet, Jeanne H. de Vries, Max Feinberg, and Pieter van't Veer. Uncertainty in intake due to portion size estimation in 24-hour recalls varies between food groups. *The Journal of nutrition*, 141(7):1396–1401, 2011. doi:10.3945/jn.111.139220.

[43] Janet A Tooze, Douglas Midthune, Kevin W Dodd, Laurence S Freedman, Susan M Krebs-Smith, Amy F Subar, Patricia M Guenther, Raymond J Carroll, and Victor Kipnis. A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution. *Journal of the American Dietetic Association*, 106(10):1575–1587, 2006.

[44] F. van den Berg, C.M.J. Jacobs, M.C. Butler Ellis, P. Spanoghe, K. Doan Ngoc, and G. Fragkoulis. Modelling exposure of workers, residents and bystanders to vapour of plant protection products after application to crops. *Science of the Total Environment*, 573:1010–1020, 2016. doi:10.1016/j.scitotenv.2016.08.180.

[45] Hilko van Der Voet and Wout Slob. Integration of probabilistic exposure assessment and probabilistic hazard characterization. *Risk Analysis: An International Journal*, 27(2):351–371, 2007. doi:10.1111/j.1539-6924.2007.00887.x.

[46] Hilko van der Voet, Gerie W.A.M. van der Heijden, Peter M.J. Bos, Sieto Bosgra, Polly E. Boon, Stefan D. Muri, and Beat J. Brüschweiler. A model for probabilistic health impact assessment of exposure to food chemicals. *Food and Chemical Toxicology*, 47(12):2926–2940, 2009. doi:10.1016/j.fct.2008.12.027.

[47] Janneke Verkaik-Kloosterman, Kevin W Dodd, Arnold LM Dekkers, Pieter van't Veer, and Marga C Ocké. A three-part, mixed-effects model to estimate the habitual total vitamin d intake distribution from food and dietary supplements in dutch young children. *The Journal of nutrition*, 141(11):2055–2063, 2011.

[48] Mélanie Zetlaoui, Max Feinberg, Philippe Verger, and Stephan Clémençon. Extraction of food consumption systems by nonnegative matrix factorization (nmf) for the assessment of food choices. *Biometrics*, 67(4):1647–1658, 2011.