MCRA Documentation

Release 9

Biometris, Wageningen University and Research

Apr 25, 2023

USER GUIDE

Reference and user manual for MCRA 9 (version 9.2.8).

Part I

User Guide

CHAPTER

ONE

INTRODUCTION TO MCRA

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. 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 MCRA 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 *Risks or 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 interest, 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. While running an action in MCRA, 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 *Modules*.

- The primary entity modules are data modules determining the scope of the assessments in MCRA. 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. Examples are,

the *relative potency factors* module: relative potency factors can be supplied as such (*Data*) or computed based on *hazard characterisations* (*Compute*); the *single value consumptions* module: Large Portions can be supplied as such (*Data*) or computed based on consumption distribution data of a population (*Compute*).

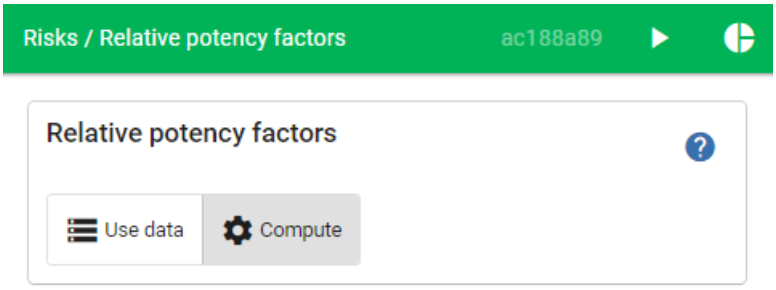


Figure 1.1: Relative potency factors supplied as data or computed based on hazard characterizations.

1.1.2 Nominal run and uncertainty analysis

Within MCRA two types of simulation runs are distinguished: the nominal run and the uncertainty analysis loop.

The nominal run represents a single simulation which is aimed to compute the most likely, unbiased estimates for the specified model. E.g., when a *dietary exposure assessment* is requested, in the nominal run a single exposure distribution is estimated using nominal values for all data and parameters.

In the *uncertainty analysis* loop, each simulation run is repeated a large number of times. Each run starts with a different scenario using data obtained with bootstrapping, parametric resampling and/or re-calculation of uncertain values. As a result, a large number of uncertain dietary exposure distributions is estimated which are used to estimate uncertainty limits (p5, p95).

Running a nominal run first has the advantage that the user may evaluate these modelling results before doing the final analysis. The model specification of rather complex simulation models and the corresponding output results are evaluated to detect any errors or misspecifications. Possible errors in the data and/or model settings are identified and corrected. In the final exposure assessment, the uncertainty analysis is included and the uncertainty of estimates is assessed.

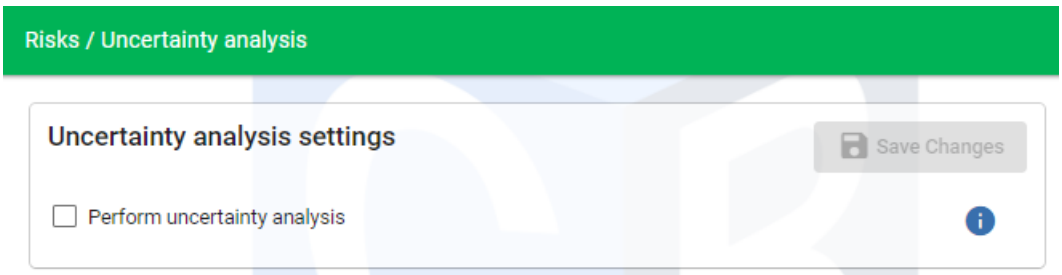


Figure 1.2: Uncertainty analysis settings.

1.1.3 Variability diagnostics

In MCRA, the nominal run might be followed by an uncertainty analysis to assess the uncertainty limits (e.g. 2.5 and 97.5%) of the nominal percentiles (e.g. p50, p95, p99, p99.9, p99.99) of the exposure or risk distribution. For these percentiles, the nominal run of an acute assessment consists of preferably 100.000 iterations, the uncertainty analysis preferably of 100 runs with 10.000 iterations each. Note that for the percentile p99.99 the minimal number of iterations should be 10.000. Likewise, to estimate uncertainty limits of 2.5 and 97.5%, the minimal number of bootstrap runs should be 100.

In general, the number of iterations and bootstrap runs will be restricted due to limited computational resources or simulation runs that are time consuming. MCRA offers a diagnostic tool to visualize whether the estimated percentiles and uncertainty limits are stable or vary due to small simulation runs with limited number of iterations and uncertainty runs.

The diagnostic tool focus on the stability of the percentiles or, re-frasing, quantify 1) the amount of Monte Carlo variability and 2) the amount of variability due to resampling e.g. consumption and monitoring data or others sources of uncertainty. By quantifying both quantities, the influence of both sources of variability on the estimated value of the percentiles is assessed.

The diagnostics are displayed in a number of graphs (as many as the number of specified percentiles). For each percentile, the graph is used to draw inference about the optimal number of MC-iterations, the number of uncertainty runs and the number of iterations in each uncertainty run.

In the section below it is assumed that the nominal run consists of 100.000 iterations and uncertainty (95% confidence) is assessed with 100 uncertainty runs of 10.000 iterations each.

To make inference, the set of nominal (Monte Carlo) values is split in 2 samples of 50.000 iterations each, 4 samples of 25.000 each, 8 samples of 12.500 each etc. By doing so, we get n partitions of samples and in each partition we have 2^n samples of size $100.000/2^n$. In each partion, the percentiles of the available samples are estimated and the standard deviation of the percentiles. So in partition n = 1, the estimate of the standard deviation is based on 2 percentiles derived from samples of size 50.000; in partition n = 2, the estimate of the standard deviation is based on 4 percentiles derived from samples of size 25.000, etc. The estimated standard deviations are plotted against the number of MC-iterations per sample of each partition. It is expected that the standard deviation decreases as a function of sample size, so for larger sample sizes MC-variability decreases. For each standard deviation the 90% confidence limits are calculated.

A similar procedure is applied to the 100 uncertainty runs (of size 10.000). In each uncertainty run the percentiles are estimated. Then, in partition n = 1, percentiles are estimated on the first 10.000/2 = 5000 iterations of each sample; in partition n = 2 percentiles are estimated on the first 10.000/4 = 2500 iterations of each sample, etc. Then standard deviations of the percentiles of the partitions of 100 x 10.000, 100 x 5000, 100 x 2500, etc are estimated and plotted against the number of iterations in each partition. It is expected that the standard deviation decreases as a function of sample size.

In Figure 1.4, the estimates and uncertainty limits are displayed. In Figure 1.3, the diagnostics are dlotted

- Red dots: standard deviation (sd) of percentile estimates between subsets of simulations. Error bars indicate parametric 90% confidence interval for the sd.
- Blue dots: standard deviation of percentile estimates between uncertainty iterations of subsets of simulations. The red square indicates the specified number of iterations (10000) of an uncertainty run.
- Each blue dot represents the standard deviation of 100 percentiles

For percentages, p50, p90, p95 and p99 the curves of the nominal estimates are smooth. For the extreme percentiles, p99.9 and p99.99, patterns do not monotonically decrease and vary, indicating that percentile estimates are variable. Also the 90% confidence interval around the sd's are much larger than for the smaller percentiles. An general conclusion is that percentiles and uncertainty limits for p50, p99, p95 and p99 are stable. The percentile of the p99.99 is unstable meaning that running the simulation with a different initialisation seed estimates will vary. The percentile p99.9 is an intermediate case.

The boxplots for uncertainty show the p25 and p75 as edges of the box, and p2.5 and p97.5 as edges of the whiskers. The reference value is indicated with the dashed black line, the median with the solid black line within the box. Outliers are displayed as dots outside the wiskers.

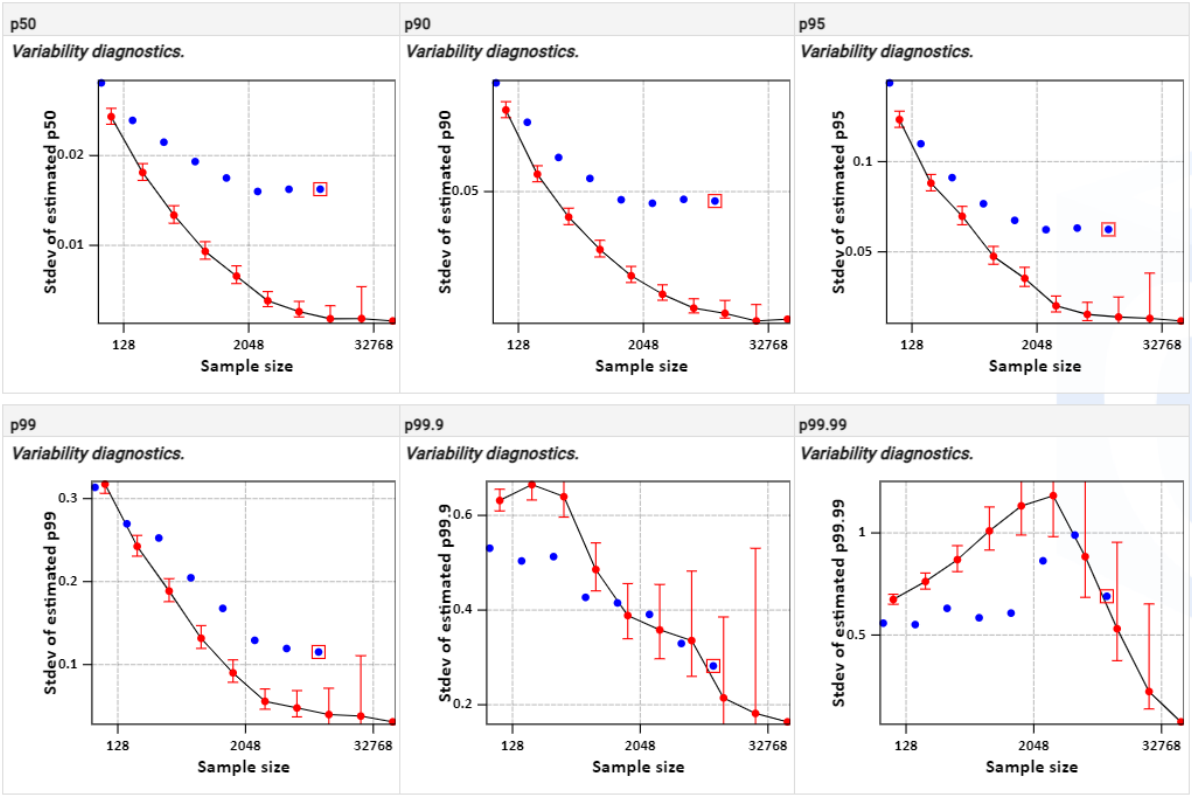


Figure 1.3: Variability diagnostics for percentiles p50, p90, p95, p99, p99.9 and p99.99.

Percentage	Exposure (µg/kg bw/day)	Median (p50)	Uncertainty lower bound (p2.5).	Uncertainty upper bound (p97.5).
50.00	0.2278	0.2026	0.1837	0.2395
90.00	0.6236	0.56	0.4896	0.653
95.00	0.82	0.7339	0.6417	0.8664
99.00	1.339	1.217	1.035	1.438
99.90	2.439	2.128	1.693	2.743
99.99	3.752	3.13	2.411	4.777

Figure 1.4: Exposure estimates for percentiles p50, p90, p95, p99, p99.9 and p99.99 with 95% confidence limits

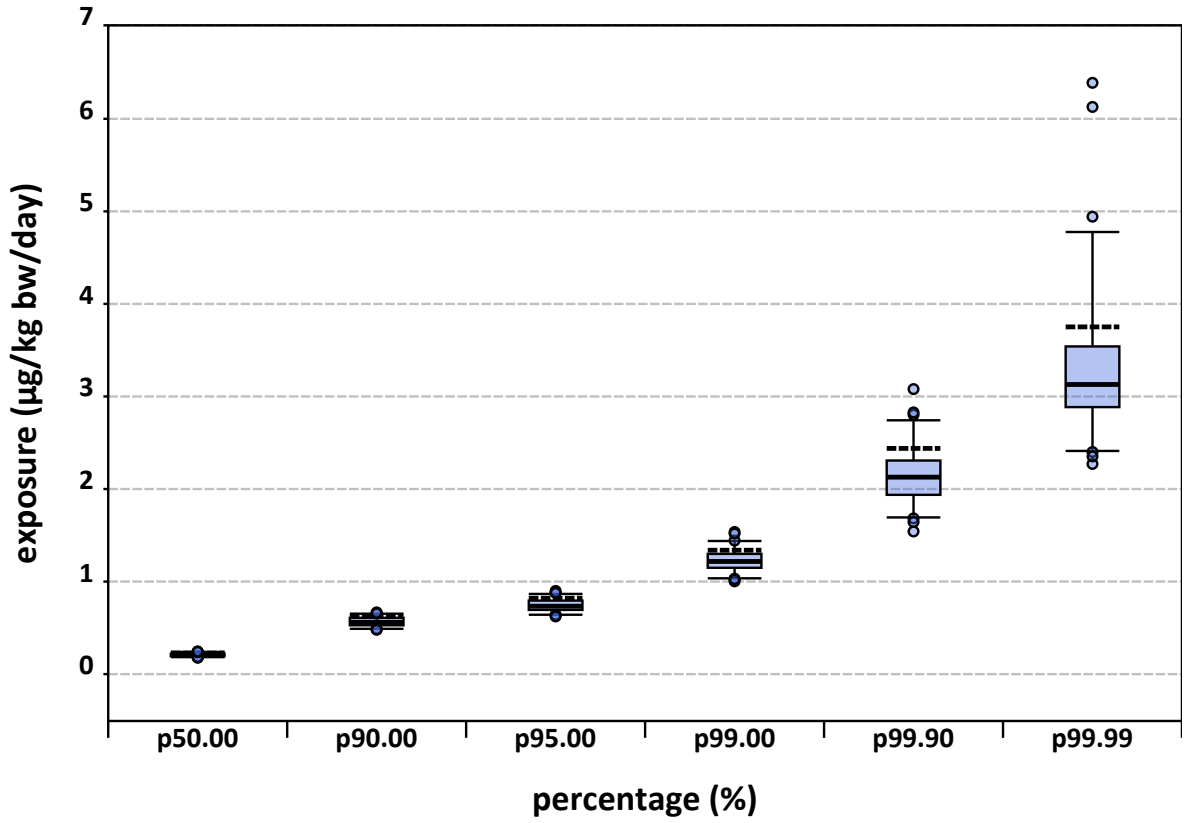


Figure 1.5: Boxplots for summary statistics for percentiles p50, p90, p95, p99, p99.9 and p99.99 with 95% confidence limits

1.1.4 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 approaches as tiers.

In the modular design, a tier is defined as a specific set of settings for a module or a group of modules. Tiers differ in many aspects, 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 MCRA, 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 substances may be assessed using a tier 'TTCx100' or 'sample from general NOAEL distribution x100' (which only requires knowledge of the Cramer class of the substance). As another example, in dietary exposure assessment some food-substance 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-substance combinations. So in this case the tiered entity is 'food-substance'. 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.

1.1.5 Uncertainty

Uncertainties arise in different forms in many of the models and data of MCRA.

- 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),
- uncertainty of the models, (e.g., due to a lack of detail).

MCRA offers the following options to handle uncertainty:

- for many types of data, the possibility to provide data including quantifications of uncertainty,
- imputation methods for filling in missing data in various types of models, and
- 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 is possible to not only provide nominal estimates of the data values, but also to provide quantified estimates of the uncertainties of these values. Occasionally, quantifications of the uncertainties of these estimates are not available. MCRA provides the possibility to work with both quantified and unquantified uncertainties: include uncertainties in a quantitative uncertainty analysis when available, or, when not available, use nominal estimates, followed by an offline qualitative uncertainty analysis.

Uncertainties of data values are available in different forms. 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*).

For each data sub-module that has quantified uncertainties, it is optional to include the source of uncertainty in the uncertainty analysis of the main module. Then, when specified, data values are resampled in each *uncertainty analysis cycle*.

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 (Efron (1979) and Efron and Tibshirani (1993)).

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)), 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 the censored values 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, data are 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

Model uncertainty or uncertainty of model outcomes arise by applying different modelling approaches or applying alternative model assumptions.

Note: TODO

1.2 Data repository

The data used for the *modelling actions* of MCRA is organised in the data repository. All users have their own (personal) repository folder in which they can *upload* their own data files and organise these in folders and sub-folders to their own preference. In addition, there are shared repository folders that are specifically created for sharing data sources among user groups. Users may be granted access to one or more shared repositories: shared, maintained, and used by multiple users. Shared repositories and their contents are free to use by granted users in their own calculations.

Each data source in the data repository contains data of one or more data groups, indicating the module(s) for which the data set can be used as data. Each primary entity module and data module of the *modular design* is linked to a data group. If a data source is recognized as a data source belonging to a module, then it can be used as a data source of this module in a modelling action. The data groups of a data source are automatically recognized when the data is uploaded, based on whether the data matches with the specified data format(s) of the data group.

Figure 1.6 shows the data repository browser. The repository browsers allows users to browse through the data repository, upload and organise their own datasets and share these with other users. 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 selected data source or folder. If the 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 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.

Users with read-write access (or higher) may upload new data source files by pressing the add button + on the bottom right and selecting the *upload new file(s)* item. 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 Creating and uploading data files

Users with an access level of read-write (or higher) are allowed to upload new data source files to a repository folder. This can be done by pressing the add button + on the bottom right and selecting the *upload new file(s)* item.

Accepted upload file types are:

- **Microsoft Excel files (.xlsx):** An Excel file contains one or more sheets, each sheet containing tabular data. The sheet names and the fields of the data tables must comply with the conventions as specified in the data format section of the module(s) for which the data is presented.

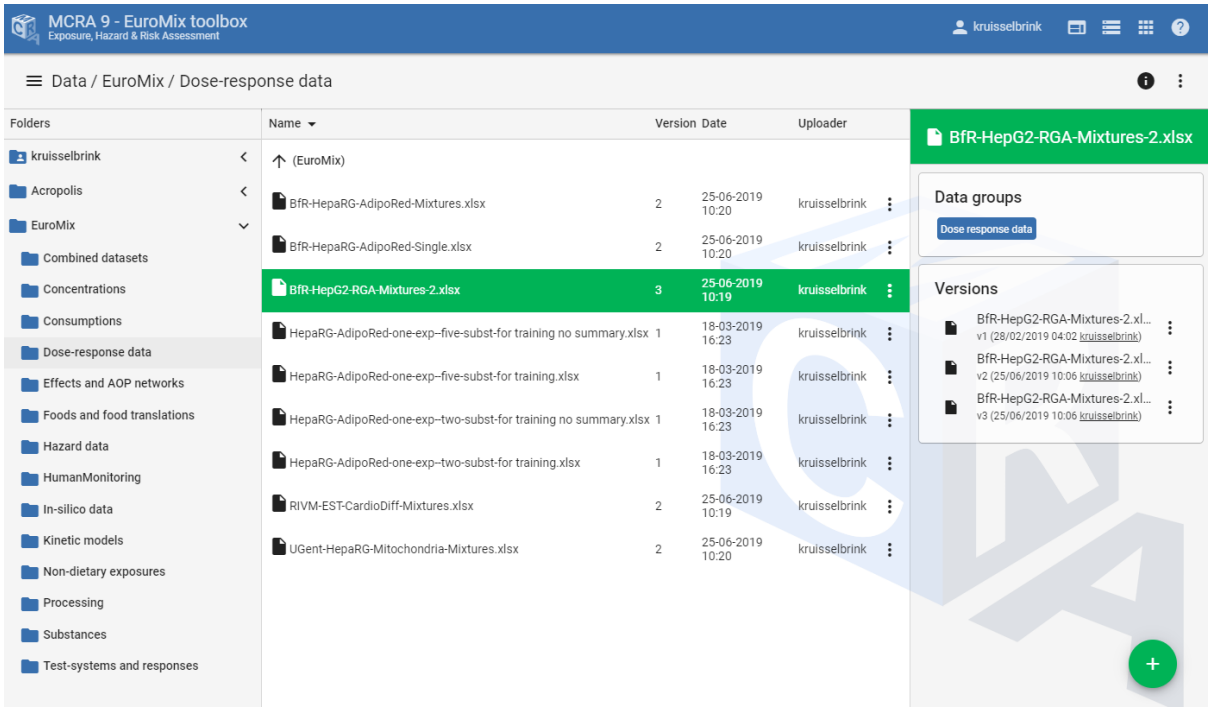


Figure 1.6: The data repository browser.

- **Microsoft Access files (.mdb):** An Access file contains one or more tables. The table names and the fields of the data tables must comply with the conventions as specified in the data format section of the module(s) for which the data is presented.
- **Zip archives with CSV files (.zip):** A CSV file (.csv) is a comma-separated values file containing data in tabular format. One or more CSV files are archived in a zipped file format (.zip) to facilitate the upload of collections of multiple CSV files. The names of the CSV files in the zip archive must follow the accepted table names of the module(s) for which the data is presented and the tables in the CSV files must follow the data format of that/those module(s). Note, that it is not allowed to upload single CSV files.

1.2.2 Moving repositories and data sources

The data repository browser (Figure 1.6) supports moving files (data sources) and folders (sub-repositories) via drag-and-drop functionality.

Click on a folder or file item in the browser and drag it to another folder while holding the primary mouse button. The mouse icon will change and the item will be visibly moving while dragging. Release the mouse button over the desired destination folder. MCRA will ask for confirmation before the move is executed.

The following rules apply:

- You **can** move items within the tree-view, list-view and between both, vice versa.
- You **cannot** move a root repository folder, a root repository folder is not draggable.
- You **cannot** move a repository to one of it's descendants.
- If you have insufficient privileges for the source or destination folder, MCRA will show an error message.

Note that moving a data source or any of it's parent repositories to a different folder does **not** affect any actions which use this data source (or any of it's versions). The action's data source will reflect the new location correctly after the data source or any of it's parent repositories has been moved.

1.2.3 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.
- **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.

Users with administrator or owner rights on a repository/folder are allowed to add/remove user and group access using the *edit shares dialog* (Figure 1.7) that can be opened by pressing the *edit shares* button .

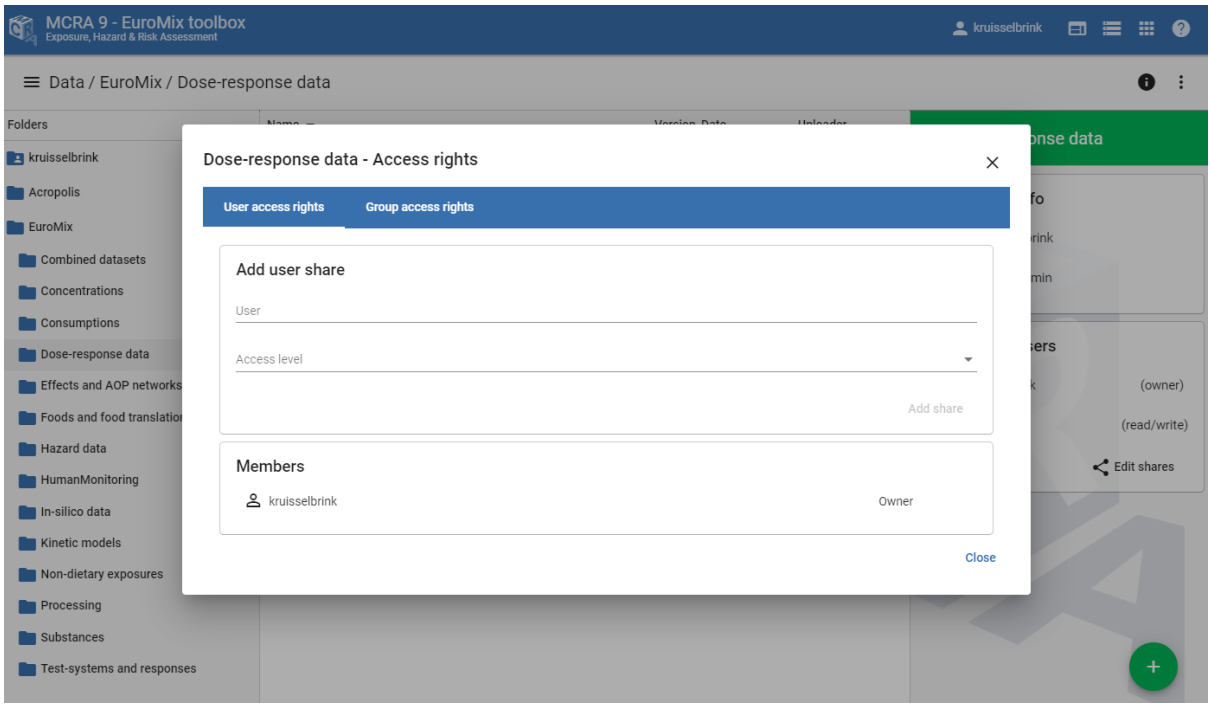


Figure 1.7: The edit-shares dialog of the data repository browser: user and group access rights are added and removed by repository owners and administrators.

1.2.4 Linking remote data repositories

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

Figure 1.8 shows how PROAST outputs of a PROASTweb user are linked to an external repository in MCRA. Data sources of remote repositories have to be explicitly imported in MCRA before they can be used in analyses. Initially, all data sources in a remote repository have a status of not-imported. Pressing the import button, MCRA 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 is created by pressing the add button + on the bottom right and selecting the *Create Proast link* option. A dialog (Figure 1.9) opens 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.

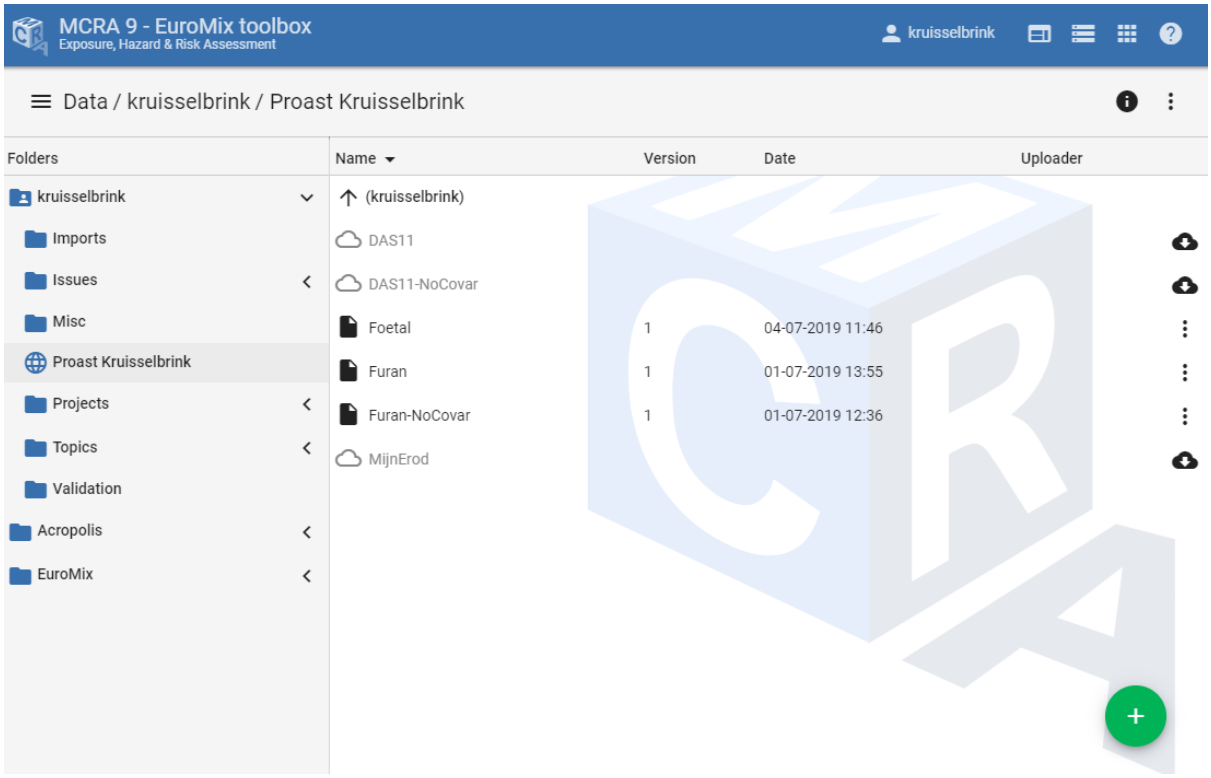


Figure 1.8: The remote (PROASTweb) repository in the data repository browser.

1.3 Workspaces and actions

In MCRA, user work is organised in workspaces. A workspace is collection of work items (actions) and data that are logically grouped together. Each workspace has a name and can be given a description and tags. The *workspace browser* page provides users with an overview of all available workspaces and allows for creation of new workspaces. The *workspace overview page* page is shown when opening a workspace and provides an overview of the actions, data and tasks of the workspace.

The modelling tasks of MCRA are specified through actions. Each action is of a certain action type, which is the *main module* for which this action specifies the modelling task. Depending on the type of action, the user can configure the settings and the data sources of the main module and all relevant/linked sub-modules in the *action area* or, if it is a so-called standard action, in the *standard action page*. When all required settings and data sources of an action are configured and the action is in a valid state, then the modelling task described by the action can be started by pressing

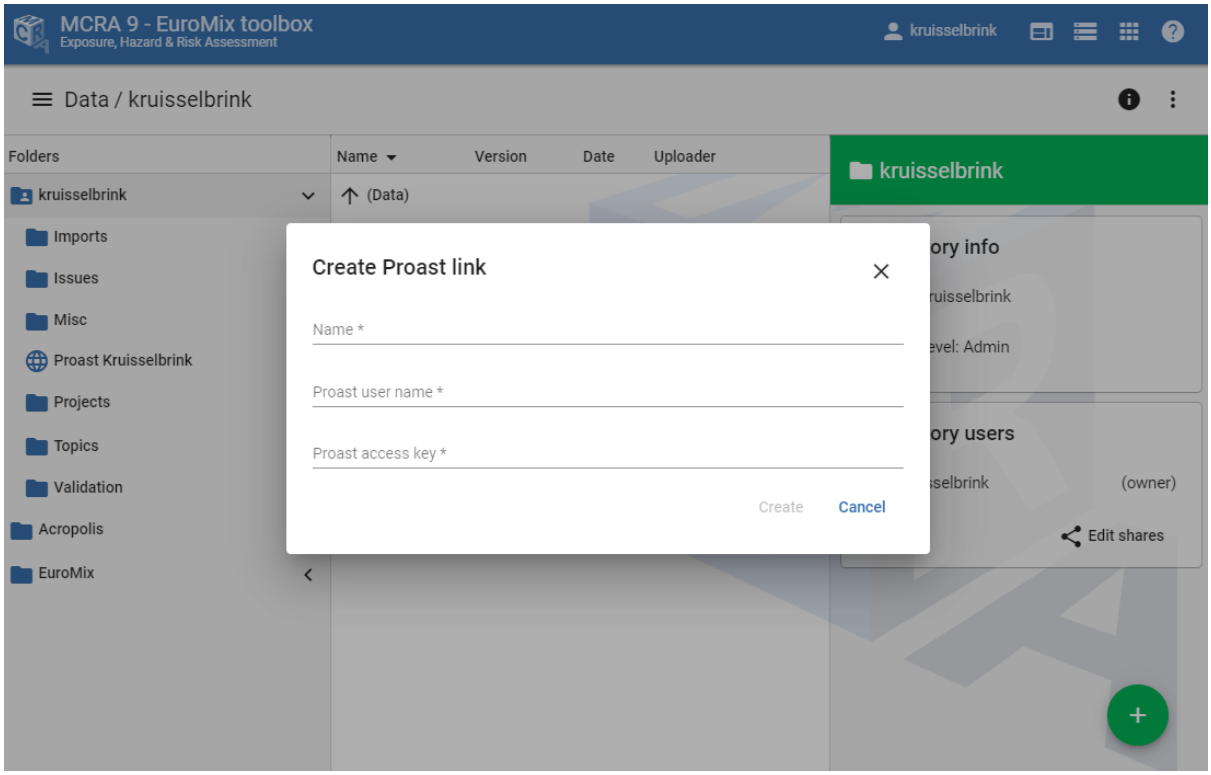


Figure 1.9: The dialog for creating a new PROASTweb remote repository link.

the run button. The status of this modelling task (which, depending on the complexity, may take some time to run) is shown in the results page and when the task has completed, output is available in the form of reports or in the form of data that can be used as input in other actions. Note that users are free to change the settings of an action at any time. Actions can therefore have multiple outputs from multiple tasks created with different settings.

A special kind of action is a so-called standard action. In a full action, the user has to link up all data and configure all settings to run the action. To get an action up and running can be quite cumbersome and the result is an output report which can be quite detailed. Although output sections follow the modular design, specific sections are not instantly available. To encourage users to explore MCRA, *standard actions* are implemented to facilitate an easy introduction to the use of MCRA. The actions can be specified with only a few settings that are important within the scope of the standard action, and the output report contains a limited selection of main sections only.

A new action can be added to a workspace by clicking the + button at the bottom right of the *workspace overview page*. This will present an action menu with three options (see Figure 1.10):

- **Create a new action:** will open a wizard to create a (full) *action* by selecting the desired action type, specifying name, description and tags, and, depending on the action type, some main action settings.
- **Create a new standard action:** will open a dialog to create a *standard action* by selecting one of the available standard actions.
- **Import an action from zip file:** will open a local file browser that lets the user *import* a specifically formatted zip-file containing an action definition.

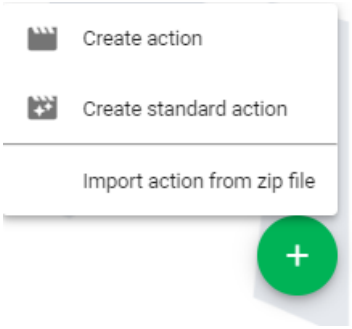


Figure 1.10: Three options to add an action to the workspace: create new action, create new standard action, and import action from zip file.

1.3.1 Workspace browser

Figure 1.11 shows the workspace browser. Users scroll through their workspaces and select the workspace which they want to work with. Detailed information about the 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* is used to quickly find/filter workspaces by name or tag. A workspace is opened by clicking on the workspace name or selecting the *open workspace* option of the *action menu* of the workspace. Opening a workspace will redirect you to the *workspace overview page*.

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

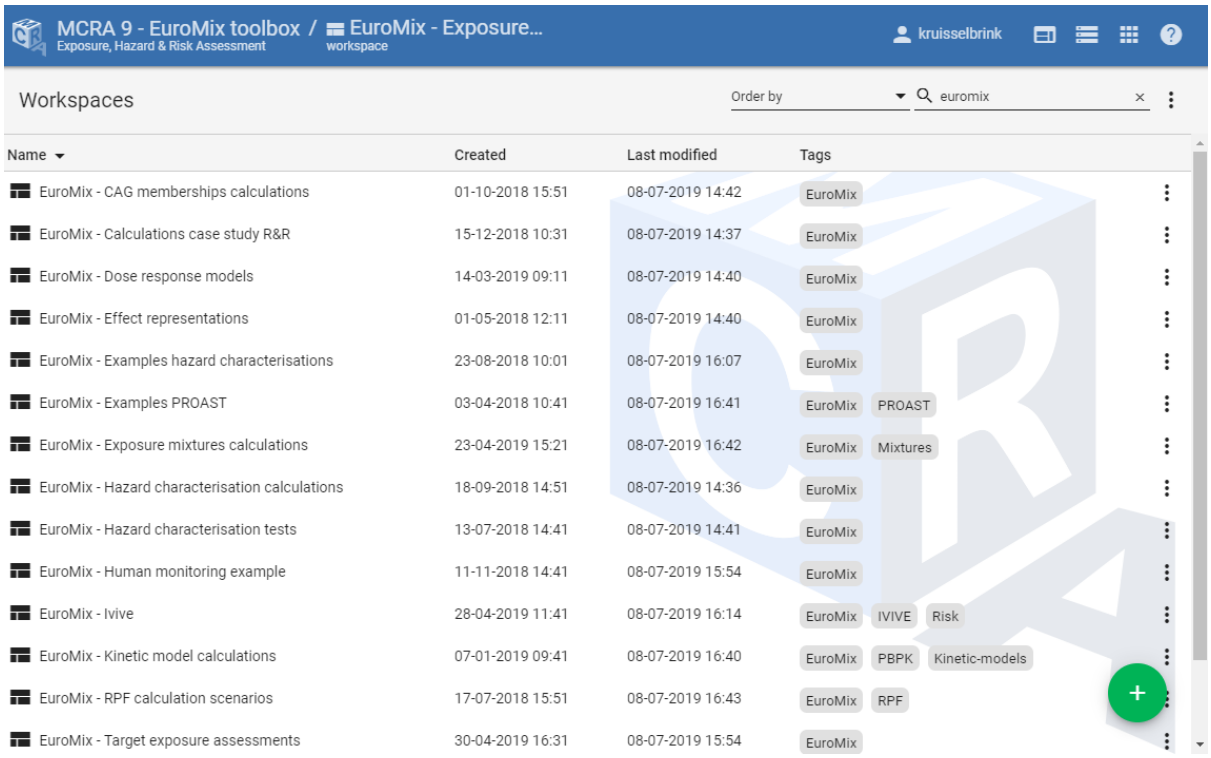


Figure 1.11: The workspace browser.

1.3.2 Workspace overview page

Figure 1.12 shows the workspace overview page. This page provides an overview of the actions, data, tasks, and results of a workspace, shown as four tabs at the top of the page. The actions tab shows all actions of the workspace, and from this tab, actions are 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 edit functionality.

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 is 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 redirect you to the *action area*. A new action is added to the workspace by pressing the *add button* + at the bottom right of the page.

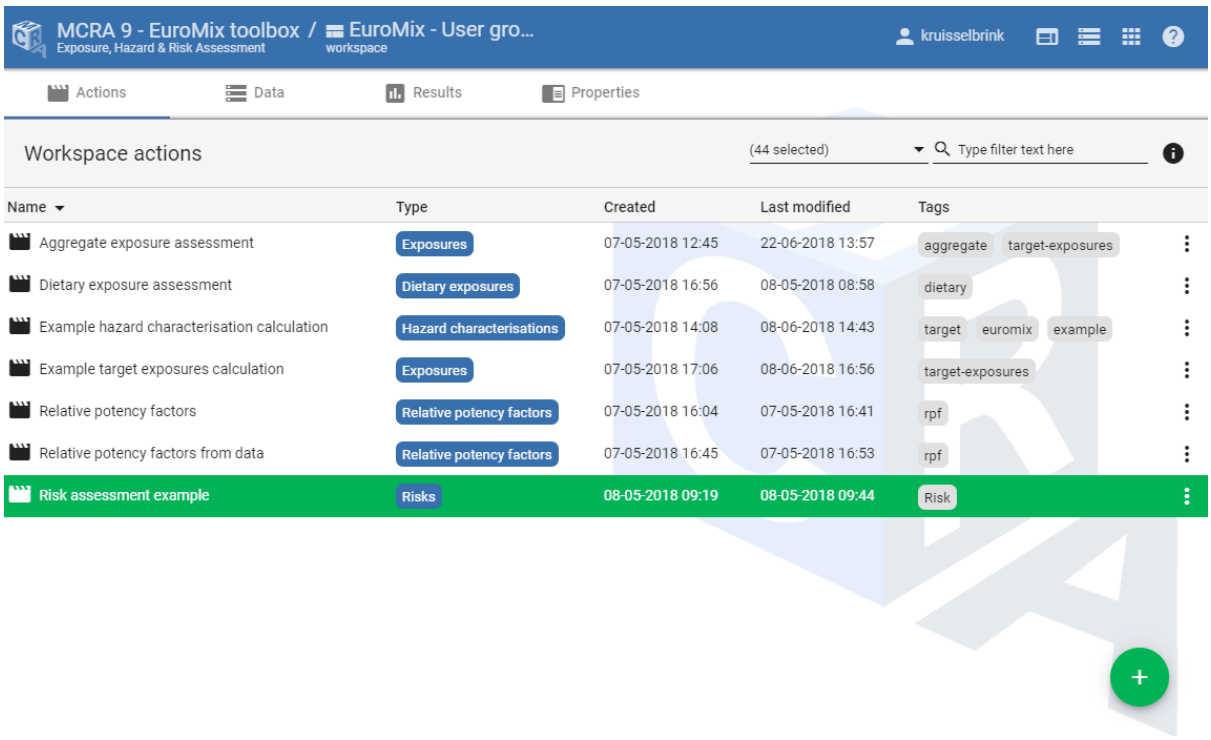


Figure 1.12: The workspace overview page.

1.3.3 Action area

After opening an action, the user is directed to the main panel of the action. Each action has its own specific panel. In the main action page and sub-action pages, an action is configured, simulation jobs started, and output results are evaluated. The panel in Figure 1.13 shows the following sections:

- **Scope:** Links to the scope-panels in which the scope entities of the action are set (e.g., foods or substances).
- **Inputs:** Links are shown for panels in which the calculation inputs or selection inputs are 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).
- **Settings:** A form is shown in which the calculation and/or selection settings of the action are set/changed (e.g., specify the exposure type, chronic/acute, of an exposure assessment).

All modules of MCRA have equally structured panels. In each panel, data sources and settings for the action are specified and the scope and input sub-module links that are relevant are shown. This presentation reflects the modular design and allows the user to select the data and settings required for running the action. In the summary panel

the main settings and data of the action are summarized. The output settings panel is used to specify general output settings. In the uncertainty settings panel the number of uncertainty runs and uncertainty sources is specified. In the results panel running tasks and output results of the actions are shown. 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 the menu button on the top left in the Action bar. In this menu, all required modules for the action are shown in one list, allowing a linear way of navigation.

An action is valid and ready to run when all scopes and inputs are valid and all required data and settings are configured. For each sub-action, the check symbol ✓ indicates that it has been configured correctly and is ready to run. In case a sub-action has a warning symbol ⚠, some user action is required. When the main action is ready to run, a simulation job is started by clicking the run button ▶ in the grey action bar on the top right. Optionally, sub-actions can be started by clicking the run button ▶ in the green (sub)action bar on the top right. 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 . After completing the task, output is available in the form of a screen report, download as pdf, or as an html report with tables (csv) and charts (svg) in a downloadable zip file.

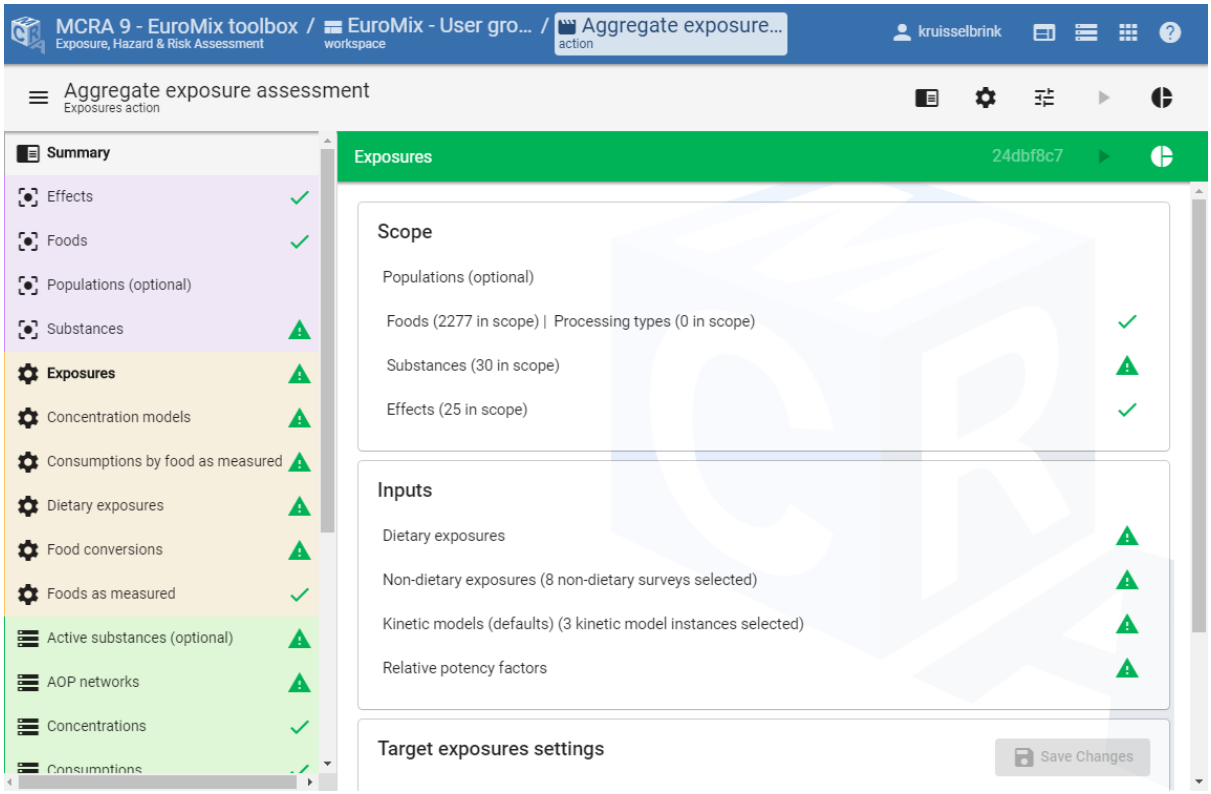


Figure 1.13: The main page of an action.

Scoping: entity selection

Each action starts with the selection of the relevant primary entities. In this context, entity selection or scoping plays an important role. Scoping of the action is defining the members for its primary entities, and, occasionally, also for other entities.

As an example, Figure 1.14 shows the substances module panel. At the top, the data source file with substances is selected containing the primary entity data of substance codes. In the selection card, a selection is made of the entities in the dataset that are relevant for the current action (3 in scope). Note that if no explicit selection is made, the scope is set to all entities by default. In the settings form, additional (selection) settings are shown, e.g., selection of the index substance (relevant for a cumulative assessment). In this way, the scope of the action is 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.

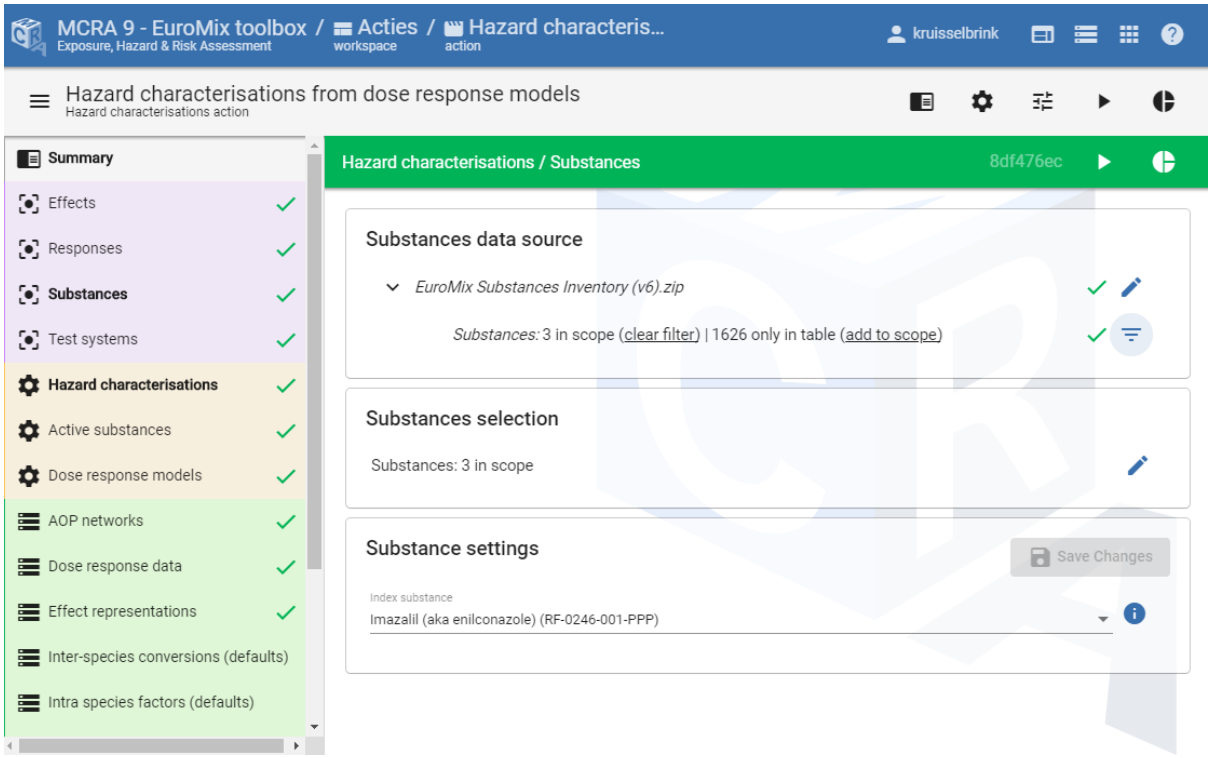


Figure 1.14: The substances module panel as an example of a primary entity module panel.

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 are linked to the action. For instance, the substance scope will contain all substance codes found. That is, not only substances as specified in the substance data source, but also all other substances found in data sources that link to substances like concentration sample data or points of departure data. These are implicit selections. Explicit selections are 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 1.14 is defined explicitly, having three substances in the scope, and excluding 1626 substances also present provided through substances data source and/or other linked data sources like concentration samples. By pressing the *clear filter* button, the explicit scope is cleared and is made implicit again. Then, the scope contains all substances found as primary entities and found in all linked data sources, in total 1629 (1626 + 3) substances.

Comparing new data to set scopes

After linking a data source to an action, MCRA performs a check whether the new data links well to the current scope (selected entities) of the action and reports the results. For instance, after linking new substance concentration data to an action which already has an implicit or explicit substance scope, it should be checked whether the substance codes used in the concentration data match with the current substances in scope. Note that this check is also performed after linking a primary entity substances data source to an action which already has a set of substances in scope, i.c. substances already specified in other selected data sources.

After linking a data table from a new data source to an action which already has 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, no further action is required. 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, for some substances no concentrations are available, and therefore MCRA allows missing concentration data for part of the substances in the scope: a green warning symbol is shown. The concentration data source may equally well contain codes that are not in the scope (e.g., concentrations for substances that are not specified in the primary entity data for substances). It may be desirable to extend the scope with these substances found in the concentration data. Also this situation is flagged with a green warning symbol.

Figure 1.15 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 is a non-critical error, normally leading to a green warning symbol, but in this case, it is overruled by the red warning symbol.

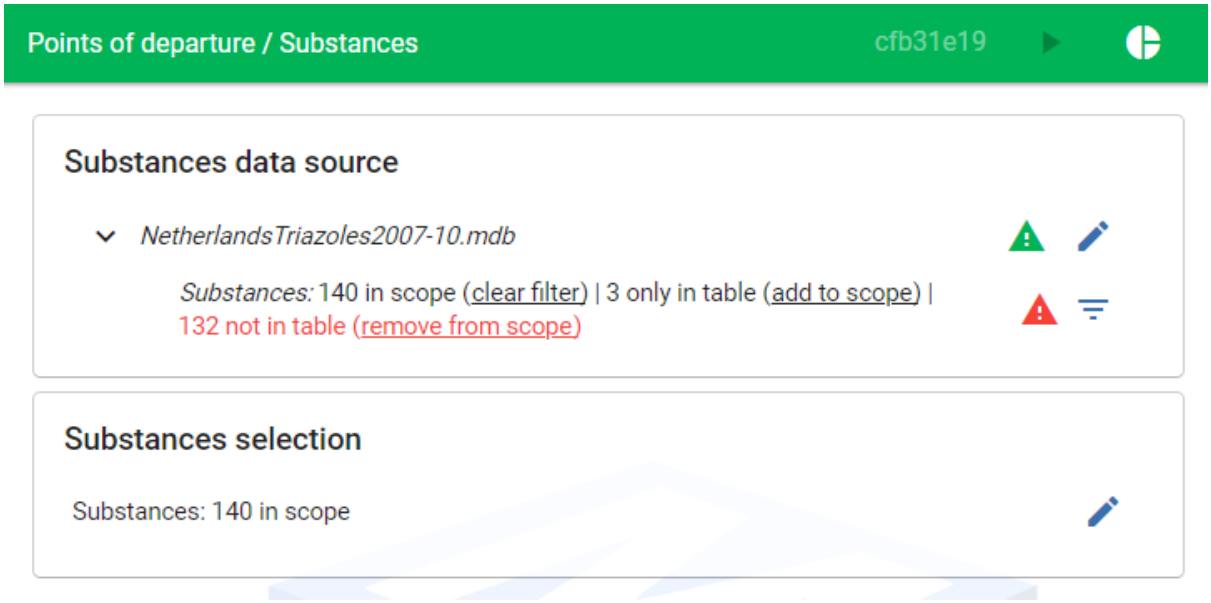


Figure 1.15: Checking substances data in a substances data source against an already set substances scope.

Another example is shown in Figure 1.16. The primary entities effects and substances are selected and in the scope. Then, a points of departure data source is selected containing effect and substance codes. For effects, no linking errors are observed, hence the new data source matches perfectly with the effects already in scope. For substances, we see that there are 7 substances that are in the points of departure data source but not in the substances scope (*new*) and for 3 substances in the scope no points of departure are available (*not in table*). The former is fine, but it might be needed to extend the scope with these 7 substances (*add to scope*). The latter, in general, is not a problem but just a point of consideration. These substances might be removed from the scope (*remove from scope*) or not.

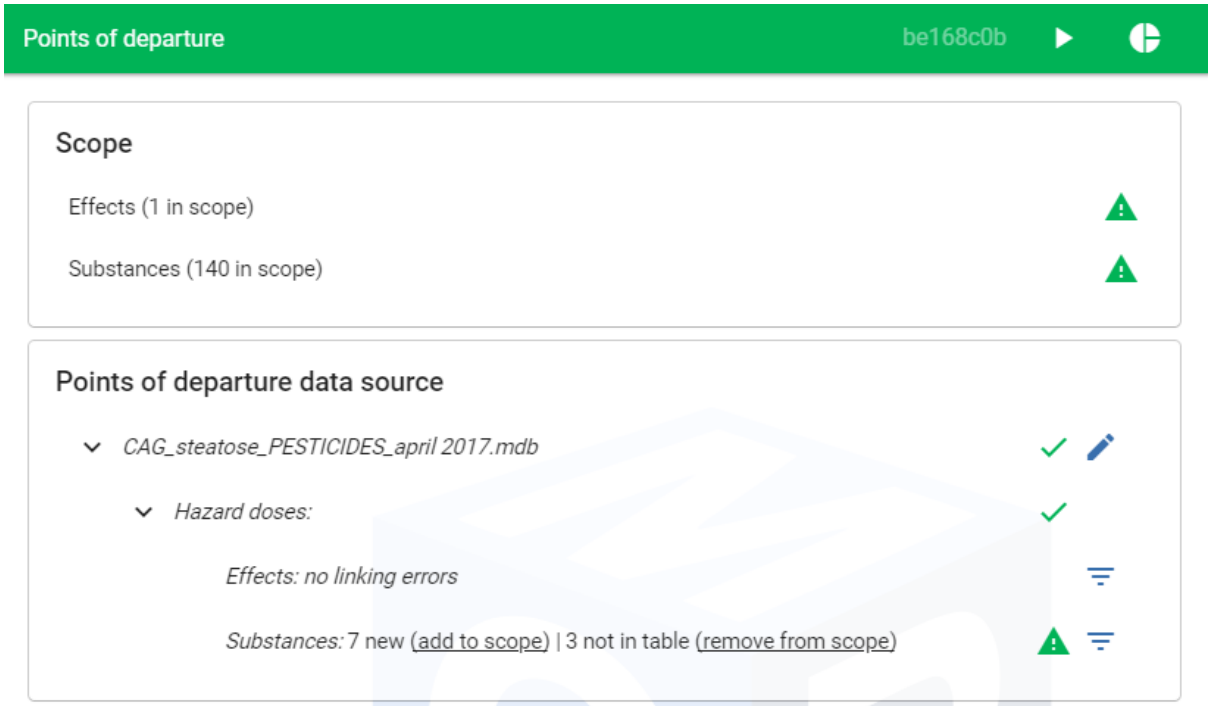


Figure 1.16: Checking substances data in a POD data source against an already set substances scope.

1.3.4 Standard action area

Standard actions facilitates MCRA users in organizing the data and configuring all setting to run an action. Some standard actions are for demonstration purposes only, other standard actions are realistic examples of risk assessment or dietary exposure assessment using real data.

Creating a standard action

A new standard action is added to the workspace by pressing the *add button* + at the bottom right of the workspace page, see *workspace overview page*. Select the *Create standard action* option.

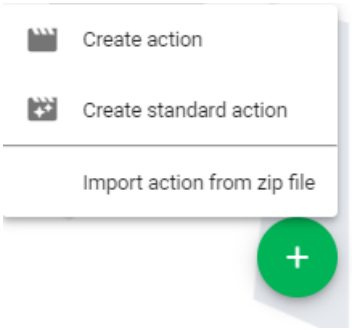


Figure 1.17: Add standard action.

Then a pop-up appears, see Figure 1.18 and after clicking one of the available panes the standard action is created. Currently, a limited number of standard actions is available for authorized users. Please contact the MCRA system administrator at https://rivm.nl for more information. New standard actions will be developed in the near future.

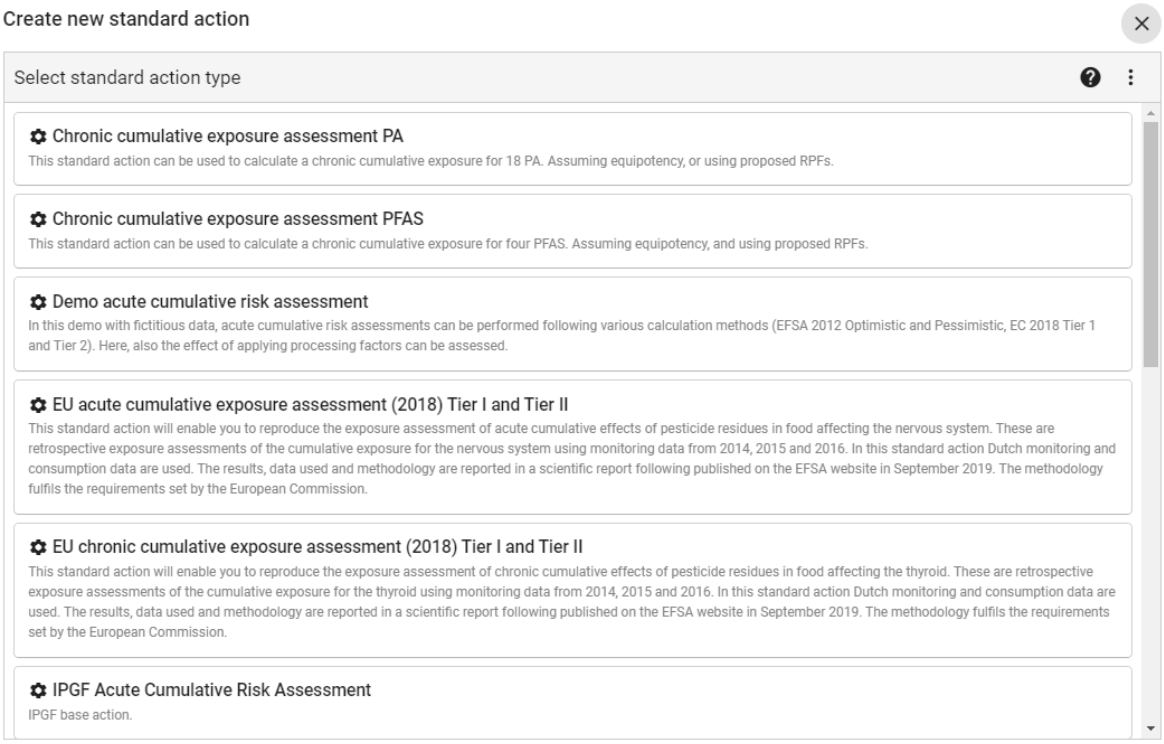


Figure 1.18: Create standard action.

Standard action reports

Although a standard action produces a short output report, see Figure 1.19, by clicking the *Show detailed report* the full report becomes available. By clicking *Show short report*, the short output report is returned.

Converting a standard action to a full action

A standard action is easily converted to a full action by opening the *action menu* ⋮ in the white bar of the standard workspace in your browser and selecting the 🎬 *Convert to full action* or 🎬 *Clone to full action* option. The first option replaces the standard action by a full action, the second option makes a clone to a full action and the standard action is still available.

1.3.5 Action zip files

MCRA also provides the functionality to import and export actions (with or without data) as specially formatted zip files (.zip) containing 'ready to run' actions. This special file archive contains two XML files (.xml) named *_ActionSettings.xml* and *_ActionData.xml* that respectively describe the action's settings and linked data sources. In addition, this zip archive may include the action data itself, either in their original form, located in a sub-folder (data) or in the form of (.csv) files.

These action zip files can be imported into and workspace by pressing the *add button* + at the bottom right of the workspace page, see *workspace overview page* and then selecting the *Import action from zip file* option. Once uploaded, the action doesn't require further user interaction and is ready to run.

An action can be exported as a zip archive by by clicking the *action menu* ⋮ located in the white bar on top of the panel. Currently, this allows export in three different formats (see Figure 1.21):

- The **download action (no data)** option will create a zip archive containing only the two xml files defining the action, but will not include the data.

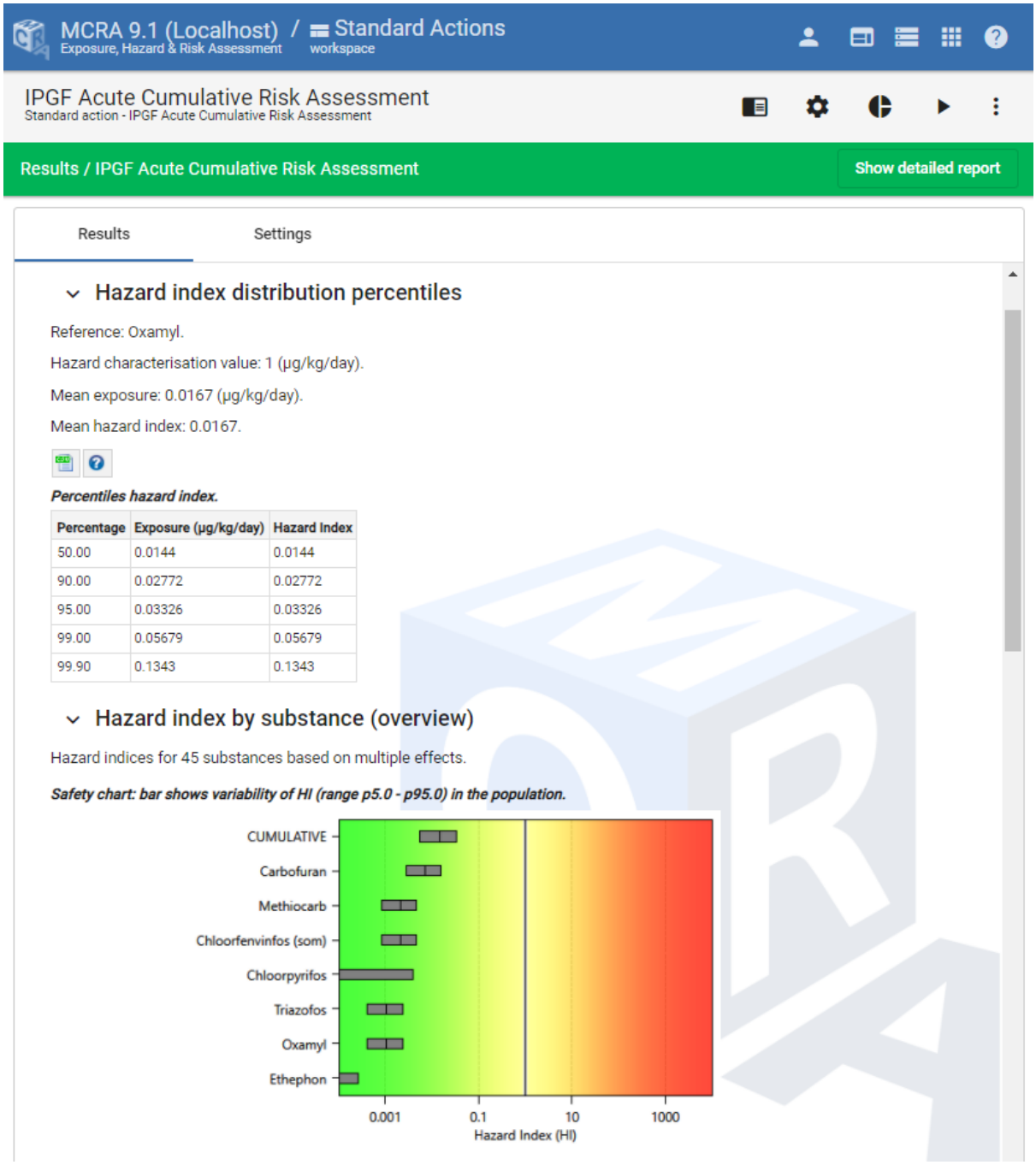


Figure 1.19: Short output report.

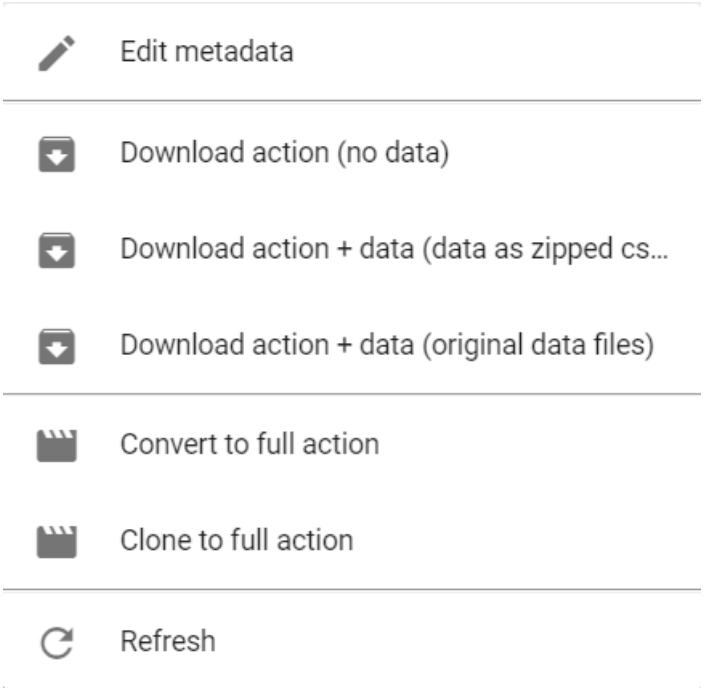


Figure 1.20: Convert to full action.

- The **download action + data (data as zipped cvs)** option will create a zip archive containing the action definition xml files and also the action data in the form of csv files of the internally used data table formats.
- The **download action + data (data as zipped cvs)** option will create a zip archive containing the action definition xml files and also the original action data files (i.e., the files that were originally uploaded in the repository).

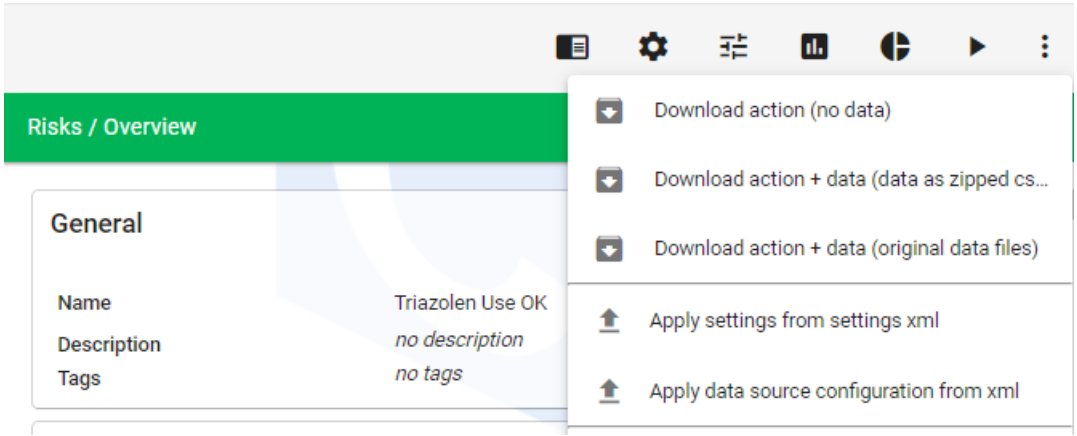


Figure 1.21: Export an action zip archive

1.4 MCRA web and core

The Monte Carlo Risk Assessment (MCRA) platform, developed by Wageningen University & Research (WUR, Biometris) for the Dutch National Institute for Public Health and the Environment (RIVM), is a web-based system for risk assessment of chemicals, which brings together statistical models, shared data and data uploaded by the users.

EFSA and RIVM previously agreed on the use of this MCRA system for the cumulative risk assessment of pesticides. This agreement was formalised in two framework partnership agreements for the periods 2015-2016 and 2017-2020. As part of these partnership agreements, the MCRA system was continuously improved in terms of capacity and functionality, in accordance with requirements defined by EFSA. In response to needs on transparency and data accessibility, the MCRA system was further developed into a transparent, collaborative, EU harmonised, interoperable, open-source accessible platform.

1.4.1 MCRA core

Source codes of the statistical models are open source (i.e., freely accessible and re-usable) and made more flexible to allow for cooperation and co-creation. The source code of MCRA core is available at a repository located at https://github.com/rivm-syso/mcra-core/. RIVM is the owner of the MCRA core, with co-ownerships of specific parts by WUR Biometris, FERA, INERIS and possibly others in the future. RIVM will manage MCRA core, such that it is available as open source.

1.4.2 MCRA web

The MCRA web portal provides tailored access to the MCRA core actions. In the web portal, each module of MCRA can be accessed as a starting point to perform the corresponding action. RIVM is the owner and manager of the MCRA web platform. The source code of the MCRA web platform is maintained at a dedicated RIVM github site. Access (both for reading and contributing) is restricted to the MCRA development and operations team. It is not anticipated that co-developers will contribute to the code of the web platform.

1.4.3 Running MCRA core using the command line interface

The MCRA core library is supplied with a *command line interface* (CLI) utility to run MCRA actions using input files and producing output files.

1.5 Results panel

The output of all runs is shown in the results panel. By clicking the hyperlink of an output section the output is opened in the browser. Most output tables have three icons. Click the left icon to download the table in csv format, click the middle icon to sort table columns or multiple columns (sort column, press/hold shift and sort next column) or to filter (search box). Press the question mark for additional information about the table headers.

All plots and charts can be saved by right clicking the picture.

An easy way to compare multiple outputs is to select a few outputs as shown in Figure 1.22, then click om the three dots and click Compare selected.

A new panel pops up and each output section can be reached by clicking the output name, e.g., *PARC Exposure mixtures training 0.8*. Expand the tree on the left side and select the sections you want to compare. In Figure 1.23, *Concentrations by substance* is selected in the Human monotoring analysis sub-action. Navigate to the output by clicking name of the output sections e.g., *PARC Exposure mixtures training 0.8*.

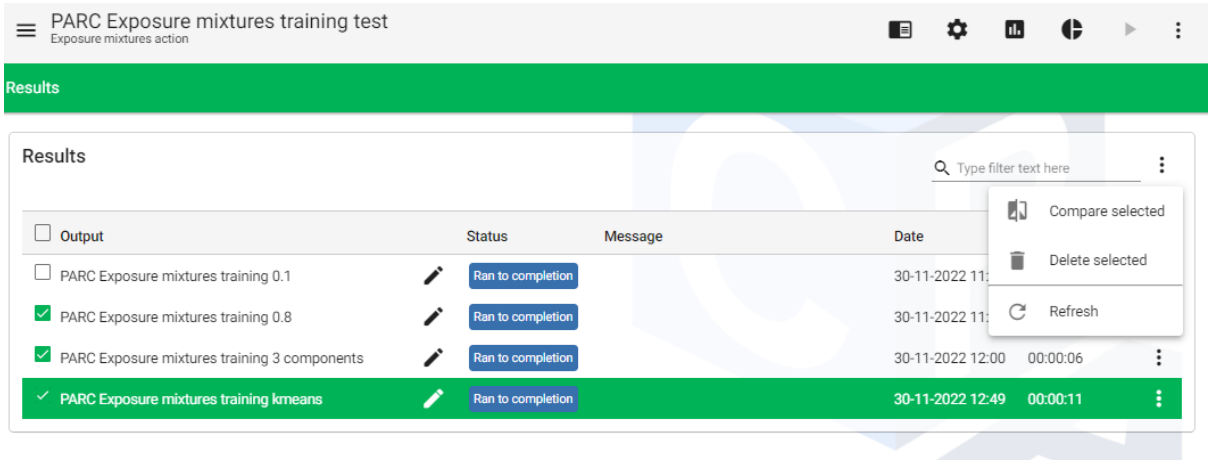


Figure 1.22: Compare selected outputs

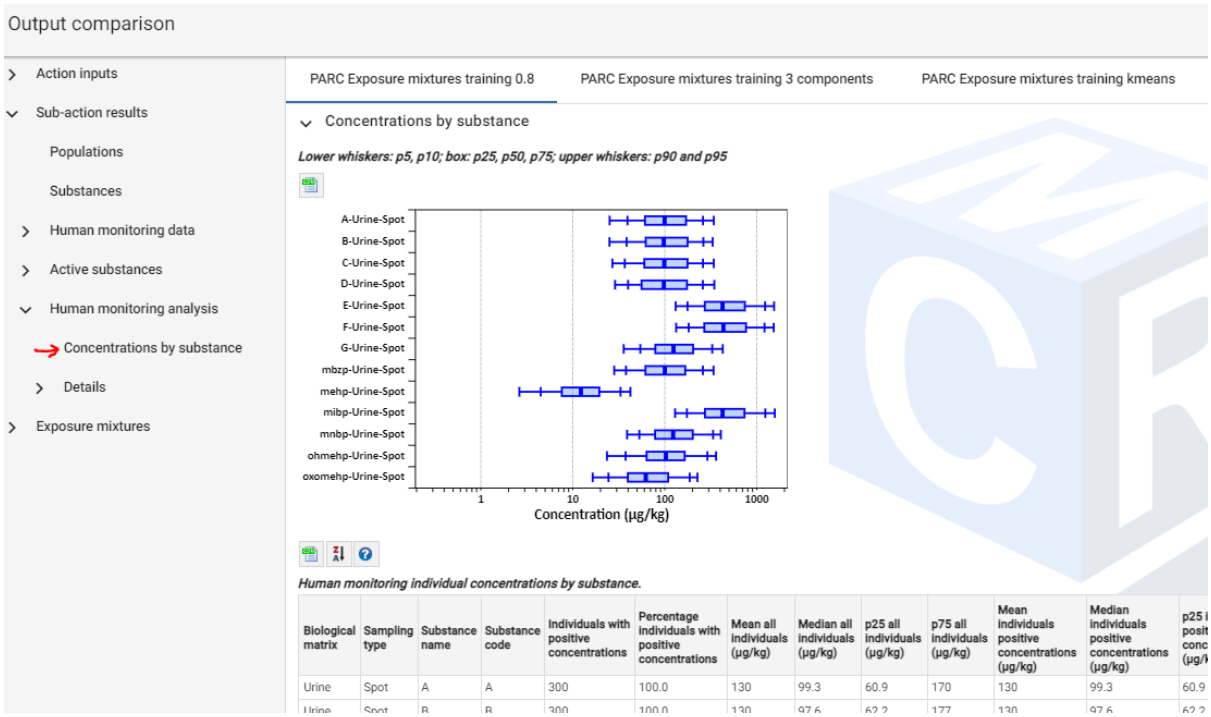


Figure 1.23: Output comparison panel

CHAPTER

TWO

EXAMPLES

Note: This section is under construction.

Guidance notes used in PARC training sessions:

- Guidance how to use PARC HBM data in MCRA
- Guidance how to perform mixture analysis on PARC HBM data in MCRA
- Example HBM data

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.

2.1 Cumulative dietary exposure assessment

2.1.1 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.

2.1.2 Preparation

In the workspace browser (icon), create a new workspace *Examples*, using the + button in the bottom right corner.

2.1.3 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 similar 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 data files 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 modelled foodd* (path: *Dietary exposures / Consumptions by modelled food*)
  - Click *Consumptions* (path: *Dietary exposures / Consumptions by modelled food / 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 modelled food* to go up one level
  - Click *Food conversions* (path: *Dietary exposures / Consumptions by modelled food / Food conversions*)

- Click *Modelled foods* (path: *Dietary exposures / Consumptions by modelled food / Food conversions / Modelled foods*)
  - Click *Concentrations* (path: *Dietary exposures / Consumptions by modelled food / Food conversions / Modelled foods / 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 modelled food / Food conversions / Food recipes*)
    - At *Food recipes data source*, click ✎ 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*, click ✎ 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 exercise, 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

2.1.4 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 modelled food / 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, which should be the same as for Example 1.

2.1.5 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

2.2 TDS-based exposure and risk assessment

Total Diet Studies (TDS) monitor chemical levels in representative consumed foods as eaten (e.g. bread, pizza), so after processing steps (Lee et al. (2015)). TDSs contrast to the common chemical monitoring studies on raw primary commodities (e.g. wheat, tomato). In some TDSs, samples are analysed directly, in other TDSs samples of similar food products are pooled before chemical analysis. Pooling of samples means that average concentrations can be well estimated, but that less or no infromation is available about variability. Therefore, in the context of risk assessment, TDS data are used for the assessment of chronic risks from long-term exposure.

This section contains three demonstrators (*MCRA standard actions*) on how to use TDS data for exposure and risk assessment. The demonstrators were developed in the FNS-Cloud project. Two examples illustrate simple TDS data on consumed foods. The first demonstrator shows the exposure to methyl-mercury and the risk for German children. The second demonstrator shows the exposure to nickel and the risk for several Belgian age groups. DON for Dutch children. The third demonstrator contains the previous two as specific cases, but also includes a case where foods are pooled before analysis, for the exposure to the mycotoxin DON and the risk for Dutch children. In this last case additional data are included to describe the TDS sample compositions in terms of the modelled foods.

After trying out the standard actions, prospective users can convert or clone these standard actions to full actions for use with other data or using other exposure or risk models.

2.2.1 Standard actions

The following standard actions are available as TDS exposure and risk assessment demonstrators:

- *TDS-based long-term exposure and risk assessment of methylmercury for German children*
- *Long-term dietary exposure and risk of nickel for the Belgian population*
- *TDS-based long term dietary exposure and risk assessment*

2.3 Aggregate exposure assessment

2.3.1 Introduction

The goal of this exercise is to assess aggregate exposure assessment.

2.3.2 Preparation

If you haven't done so, in the workspace browser (use the icon), create a new workspace named *Examples*, using the +.

The data files used in the example(s) in this section, are located in the data folder *Documentation-Examples / Exercise Aggregate Exposure Assessment*.

2.3.3 Example 1

- In the *Examples* workspace, create a new action (using +)
  - Then select ✓ *Show all action types*, select *Exposures*
  - Name it *exposures*
  - At *Exposure settings* choose:
    - As *Risk type Chronic*
    - Check ✓ *Include dietary and non-dietary routes of exposure*

- Press *Create*

- Then go to the Actions settings ⚙ of this action (path: *Exposures*)
  - At *Scope*, click *Effects* (path: *Exposures / Effects*)
    - At *Effects data source* with ✎ browse to the file *Effect - Steatosis.xlsx* and *Select*
    - At *Effect settings* for *Focal effect* select *Steatosis-liver* and press 🖫 *Save Changes*
    - In the green navigation bar, click *Exposures* to go up one level
  - At *Scope*, click *Foods* (path: *Exposures / Foods*)
    - At *Foods data source* with ✎ browse to the file *Foods.xlsx* and *Select*
    - In the green navigation bar, click *Exposures* to go up one level
  - At *Scope*, click *Substances* (path: *Exposures / Substances*)
    - At *Substances data source* with ✎ browse to the file *Substances.xlsx* and *Select*
    - At *Substance settings*
      - for *Index substance* select *Cyproconazole* and press 🖫 *Save Changes*
    - In the green navigation bar, click *Exposures* to go up one level
  - At *Inputs*, click *Dietary exposures* (path: *Exposures / Dietary exposures*)
    - At *Inputs*, click *Consumptions by modelled food* (path: *Exposures / Dietary exposures / Consumptions by modelled food*)
      - At *Inputs*, click *Consumptions* (path: *Exposures / Dietary exposures / Consumptions by modelled food / Consumptions*)
        - At *Consumptions data source* with ✎ browse to the file *Consumptions.xlsx* and *Select*
        - At *Consumptions data selection* with ✎ open the food consumption surveys selection.
          - The file contains two surveys, but only one is allowed. So deselect everything by clicking ✓ on the first line, next to the word *Code*
          - Now select *DNFCS_2003* and press *Save* (the red warning ⚠ should now be gone)
        - In the green navigation bar, click *Consumptions by modelled food* to go up one level
      - At *Inputs*, click *Food conversions* (path: *Exposures / Dietary exposures / Consumptions by modelled food / Food conversions*)
        - At *Inputs*, click *Modelled foods* (path: *Exposures / Dietary exposures / Consumptions by modelled food / Food conversions / Modelled foods*)
          - At *Inputs*, click *Concentrations* (path: *Exposures / Dietary exposures / Consumptions by modelled food / Food conversions / Modelled foods / Concentrations*)
            - At *Concentration data source* with ✎ browse to the file *ConcentrationData.xlsx* and *Select*
            - In the green navigation bar, click *Food conversions* to go up two levels
        - At *Inputs*, click *Food recipes* (path: *Exposures / Dietary exposures / Consumptions by modelled food / Food conversions / Food recipes*)
          - At *Food recipes data source*, with ✎ browse to the file *FoodRecipes.xlsx* and *Select*
          - In the green navigation bar, click *Dietary exposures* to go up three levels
      - At *Inputs*, click *Concentration models* (path: *Exposures / Dietary exposures / Concentration models*)
        - At *Inputs*, click *Relative potency factors* (path: *Exposures / Dietary exposures / Concentration models / Relative potency factors*)
          - At *Relative potency factors data source* with ✎ browse to the file *RelativePotencyFactors.xlsx* and *Select*

        - In the green navigation bar, click *Dietary exposures* to go up two levels
    - At *Inputs*, click *Processing factors* (path: *Exposures / Dietary exposures / Processing factors*)
        - At *Processing factors data source* with ✎ browse to the file *ProcessingFactors.xlsx* and *Select*
        - In the green navigation bar, click *Dietary exposures* to go up one level
    - At *Inputs*, click *Active substances (optional)* (path: *Exposures / Dietary exposures / Active substances*)
        - At *Inputs*, click *Points of departure* (path: *Exposures / Dietary exposures / Active substances / Points of departure*)
            - At *Points of departure data source* with ✎ browse to the file *HazardDoses - Triazoles.xlsx*
            - In the green navigation bar, click *Dietary exposures* to go up two levels
    - At *Dietary exposure settings*, for *Dietary exposure calculation tier* select *EFSA 2012 Optimistic*, and press 🖫 *Save Changes*
    - In the green navigation bar, click *Exposures* to go up one level
- At *Inputs*, click *Non-dietary exposures* (path: *Exposures / Non-dietary exposures*)
    - At *Non-dietary exposures data source* with ✎ browse to the file *NonDietaryExposures.xlsx* and *Select*
- Now run the model, by pressing the ▶ run icon in the grey bar.

Try to find the following results:

1. Exposure percentiles daily intakes with uncertainty bounds
2. Substance with highest contribution to the total exposure distribution
3. The modelled food measured with the highest contribution to the upper tail of the exposure distribution

2.3.4 Example 2

In this example we will elaborate on the previous one with kinetic models.

- Go to the Actions settings ⚙ of this action (path: *Exposures*)
    - At *Inputs*, click *Kinetic models (default)* (path: *Exposures / Kinetic models*)
        - At *Kinetic models data source* with ✎ browse to the file *UserGroupDemo-KineticModelsArtificial.xlsx* and *Select*
        - At *Kinetic model settings* for *Kinetic model* select *Cosmos Version 5*
- Now run the model, by pressing the ▶ run icon in the green bar.

2.4 Hazard characterisations from PoDs

2.4.1 Introduction

The goal of this exercise is to try to establish hazard characterisations from PoDs (NOAELs).

2.4.2 Preparation

If you haven't done so, in the workspace browser (use the icon), create a new workspace named *Examples*, using the +.

The data files used in the example(s) in this section, are located in the data folder *Documentation-Examples / Exercise Hazard characterisations*.

2.4.3 Example 1

In this example, Imazalil target dose from NOAEL will be calculated.

- In the *Examples* workspace, create a new action (using +)
  - Then select ✓ *Show all action types*, and select *Hazard characterisations*
  - Name it *TargetDoseImazalil*
  - Use as Hazard characterization settings
    - Risk type: *Chronic*
    - Target level: *External*
  - Press Create
- Then go to the Actions settings of this action.
  - At *Scope*, click *Effects* (path: *Hazard characterisations / Effects*)
    - At *Effects data source* with ✎ browse to the file *Effects and AOP Network Steatosis.xlsx* and *Select*
    - At *Effects selection* with ✎
      - Deselect everything by clicking ✓ on the first line, next to the word *Code*
      - On the second page, select only *Steatosis-liver*, and Save
    - At *Effect Settings* for *focal effect* select *Steatosis-liver* and press Save Changes.
    - In the green navigation bar, click *Hazard characterisations* to go up one level
  - At *Scope*, click *Substances* (path: *Hazard characterisations / Substances*)
    - At *Substances data source* with ✎ browse to the file *TargetDosescalculation-Substances.xlsx* and *Select*
    - At *Substances selection* with ✎
      - Deselect everything, by clicking the ✓ on the first line, next to the word *Code*
      - Select only *Imazalil*, and Save
      - In the green navigation bar, click *Hazard characterisations* to go up one level
  - At *Inputs*, click *Points of departure* (path: *Hazard characterisations / Points of departure*)
    - At *Points of departure data source* with ✎ browse to the file *TargetDosesCalculation-HazardDoses.xlsx* and *Select*
    - In the green navigation bar, click *Hazard characterisations* to go up one level
  - At *Hazard characterisations settings*, for *Expression type* select *NOAEL (convert all hazard characterisations as NOAELs)*
  - At *Hazard characterisations settings*, Select ✓ Use inter-species conversions
  - At *Hazard characterisations settings*, Select ✓ Use intra-species factors, and press Save Changes
- Now run the model, by pressing the ▶ run icon in the grey bar.

Try to find the following results:

1. The NOAEL for Imazalil used as point of departure.
2. The target hazard dose based on the default assessment factors 1/10 and 1/10 for inter-species and within-species conversion.

Answers:

- In the grey bar, browse to the results panel by clicking the icon and click on the latest output (path: *Results / TargetDoseImazalil*)
  - In the *Hazard characterisations* tab, browse in the tree (unfold by clicking › where necessary) to ˅ *Available hazard characterisations*
    - The NOAEL for Imazalil is 40 µg/kg bw/day.

2.5 Health impact estimates

2.5.1 Introduction

The goal of this exercise is to assess a health impact estimate.

2.5.2 Preparation

If you haven't done so, in the workspace browser (use the icon), create a new workspace named *Examples*, using the +.

The data files used in the example(s) in this section, are located in the data folder *Documentation-Examples / Exercise Health Impact*.

2.5.3 Example 1

- In the *Examples* workspace, create a new action (using +)
  - Then select ✓ *Show all action types*,select *Risks*
  - Name it *Risks*
  - Press *Create*
- Then go to the Actions settings of this action (path: *Risks*)
  - At *Scope*, click *Effects* (path: *Risks / Effects*)
    - At *Effects data source* with browse to the file *Effects and AOP Network Steatosis.xlsx* and *Select*
    - At * Effect settings*, for *focal effect* select *Steatosis-liver* and press *Save Changes*
    - In the green navigation bar, click *Risks* to go up one level
  - At *Scope*, click *Foods* (path: *Risks / Foods*)
    - At *Foods data source* with browse to the file *Foods.xlsx* and *Select*
    - In the green navigation bar, click *Risks* to go up one level
  - At *Scope*, click *Substances* (path: *Risks / Substances*)
    - At *Substances data source* with browse to the file *Substances.xlsx* and *Select*
    - At *Substance settings*, for *index substance* select *Cyproconazole*, and press *Save Changes*
    - In the green navigation bar, click *Risks* to go up one level
  - At *Inputs*, click *Exposures* (path: *Risks / Exposures*)

- At *Inputs*, click *Dietary exposures* (path: *Risks /Exposures / Dietary exposures*)
  - At *Inputs*, click *Consumptions by food measured* (path: *Risks /Exposures / Dietary exposures / Consumptions by modelled food*)
    - At *Inputs*, click *Consumptions* (path: *Risks /Exposures / Dietary exposures / Consumptions by modelled food / Consumptions*)
      - At *Consumptions data source* with ✎ browse to the file *Consumptions.xlsx* and *Select*
      - At *Consumptions data selection* with ✎ open the food consumption surveys selection.
        - The file contains two surveys, but only one is allowed. So deselect everything by clicking ✓ on the first line, next to the word *Code*
        - Now select *DNFCS_2003* and press *Save* (the red warning ⚠ should now be gone)
      - At *Consumptions settings* for *Food survey* select *DNFCS_2003* and press 🖫 *Save Changes*
      - In the green navigation bar, click *Consumptions by modelled food* to go up one level
    - At *Inputs*, click *Food conversions* (path: *Risks /Exposures / Dietary exposures / Consumptions by modelled food / Food conversions*)
      - At *Inputs* click *Food as modelled* (path: *Risks /Exposures / Dietary exposures / Consumptions by modelled food / Food conversions / Food as modelled*)
        - At *Inputs*, click *Concentrations* (path: *Risks /Exposures / Dietary exposures / Consumptions by modelled food / Food conversions / Food as modelled / Concentrations*)
          - At *Concentrations data source* with ✎ browse to the file *UserGroupDemo-ConcentrationData.xlsx* and *Select*
          - In the green navigation bar, click *Food conversions* to go up two levels
      - At *Inputs*, click *Food recipes* (path: *Risks /Exposures / Dietary exposures / Consumptions by modelled food / Food conversions / Food recipes*)
        - At *Food recipes data source* with ✎ browse to the file *UserGroupDemo-FoodRecipes.xlsx* and *Select*
        - In the green navigation bar, click *Dietary exposures* to go up three levels
    - At *Inputs*, click *Processing factors* (path: *Risks /Exposures / Dietary exposures / Processing factors*)
      - At *Processing factors data source* with ✎ browse to the file *UserGroupDemo-ProcessingFactors.xlsx* and *Select*
      - In the green navigation bar, click *Risks* to go up three levels
- At *Inputs*, click *Hazard chararcterisations* (path: *Risks / Hazard chararcterisations*)
  - At *Inputs*, click *Active substances* (path: *Risks / Hazard chararcterisations / Active substances*)
    - At *Inputs*, click *Points of departure* (path: *Risks / Hazard chararcterisations / Active substances / Points of departure*)
      - At *Points of departure data source* with ✎ browse to the file *UserGroupDemo-HazardDoses.xlsx* and *Select*
      - In the green navigation bar, click *Active substances* to go up one level
    - At 'Active substances' click ⚙ Compute

2.6 Assessment group membership probabilities

2.6.1 Introduction

The goal of this exercise is to assess group membership probabilities.

2.6.2 Preparation

If you haven't done so, in the workspace browser (use the icon), create a new workspace named *Examples*, using the +.

The data files used in the example(s) in this section, are located in the data folder *Documentation-Examples / Exercise Dietary Exposure Assessment*.

2.6.3 Example 1

- In the *Examples* workspace, create a new action (using +)
  - Then select *Dietary exposures*
  - Name it *Dietary exposures*
  - Use as Dietary exposures settings
    - Tier: *EFSA Guidance Optimistic*
    - Risk type *Chronic*
    - Select ✓ *Cumulative*
  - Press *Create*
- Then go to the actions settings of this action (path: *Dietary exposures*)
  - At *Scope*, click *Foods* (path: *Dietary exposures / Foods*)
    - At *Foods data source* with browse to the file *UserGroupDemo-Foods.xlsx* and *Select*
    - In the green navigation bar, click *Dietary exposures* to go up one level
  - At *Scope*, click *Substances* (path: *Dietary exposures / Substances*)
    - At *Substances data source* with browse to the file *UserGroupDemo-Substances.xlsx* and *Select*
    - At *Substance settings* for *Index substance* select *Cyproconazole* and press *Save Changes*
    - In the green navigation bar, click *Dietary exposures* to go up one level
  - At *Scope*, click *Effects* (path: *Dietary exposures / Effects*)
    - At *Effects data source* with browse to the file *Effect - Steatosis.xlsx* and *Select*
    - At *Effect Settings* for *focal effect* select *Steatosis-liver* and press *Save Changes*
    - In the green navigation bar, click *Dietary exposures* to go up one level.
  - At *Inputs*, click *Consumptions by modelled food* (path: *Dietary exposures / Consumptions by modelled food*)
    - At *Inputs*, click *Consumptions* (path: *Dietary exposures / Consumptions by modelled food / Consumptions*)
      - At *Consumptions data source* with browse to the file *UserGroupDemo-Consumptions.xlsx* and *Select*
      - At *Consumption settings* for *Food survey* select *DNFCS_2003* and press *Save Changes*

- In the green navigation bar, click *Consumptions by modelled food* to go up one level
- At *Inputs*, click *Food conversions* (path: *Dietary exposures / Consumptions by modelled food / Food conversions*)
  - At *Inputs*, click *Modelled foods* (path: *Dietary exposures / Consumptions by modelled food / Food conversions / Modelled foods*)
    - At *Inputs*, click *Concentrations* (path: *Dietary exposures / Consumptions by modelled food / Food conversions / Modelled foods / Concentrations*)
      - At *Concentrations data source* with ✎ browse to the file *UserGroupDemo-ConcentrationData.xlsx* and *Select*
      - In the green navigation bar, click *Food conversions* to go up two levels
    - At *Inputs*, click *Food recipes* (path: *Dietary exposures / Consumptions by modelled food / Food Food recipes*)
      - At *Food recipes data source*, with ✎ browse to the file *UserGroupDemo-FoodRecipes.xlsx*
      - In the green navigation bar, click *Dietary exposures* to go up three levels
- At *Inputs*, click *Concentration models* (path: *Dietary exposures / Concentration models*)
  - At *Inputs*, click *Relative potency factors* (path: *Dietary exposures / Concentration models / Relative potency factors*)
    - At *Relative potency data source* with ✎ browse to the file *UserGroupDemo-RelativePotencyFactors.xlsx* and *Select*
    - In the green navigation bar, click *Dietary exposures* to go up two levels
- At *Inputs*, click *Processing factors* (path: *Dietary exposures / Processing factors*)
  - At *Processing factors data source* with ✎ browse to the file *UserGroupDemo-ProcessingFactors.xlsx* and *Select*
  - In the green navigation bar, click *Dietary exposures* to go up one level
- At *Inputs*, click *Active substances (optional)* (path: *Dietary exposures / Active substances*)
  - At *Inputs*, click *Points of departure* (path: *Dietary exposures / Active substances / Points of departure*)
    - At *Points of departure data source*, with ✎ browse to the file *HazardDoses - Triazoles.xlsx*
    - In the green navigation bar, click *Dietary exposures* to go up two levels
- Now run the model, by pressing the ▶ run icon in the grey bar.

Try to find the following results:

1. Exposure percentiles daily intakes
2. Substance with highest contribution to the total exposure distribution
3. The food-as-measured with the highest contribution to the upper tail of the exposure distribution

2.6.4 Example 2

Repeat the run of the previous task, but instead of the nominal run, now do an uncertainty analysis loop.

- Click on the ⚙ icon (in the grey bar) to open the uncertainty settings panel, and 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.

Compare with the previous results, to find:

1. Exposure percentiles daily intakes with uncertainty bounds

2. Substance with highest contribution to the total exposure distribution
3. The food-as-measured with the highest contribution to the upper tail of the exposure distribution

Part II

Reference Manual

CHAPTER

THREE

MODULES

MCRA is a modular system. The diagram of Figure 3.1 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.

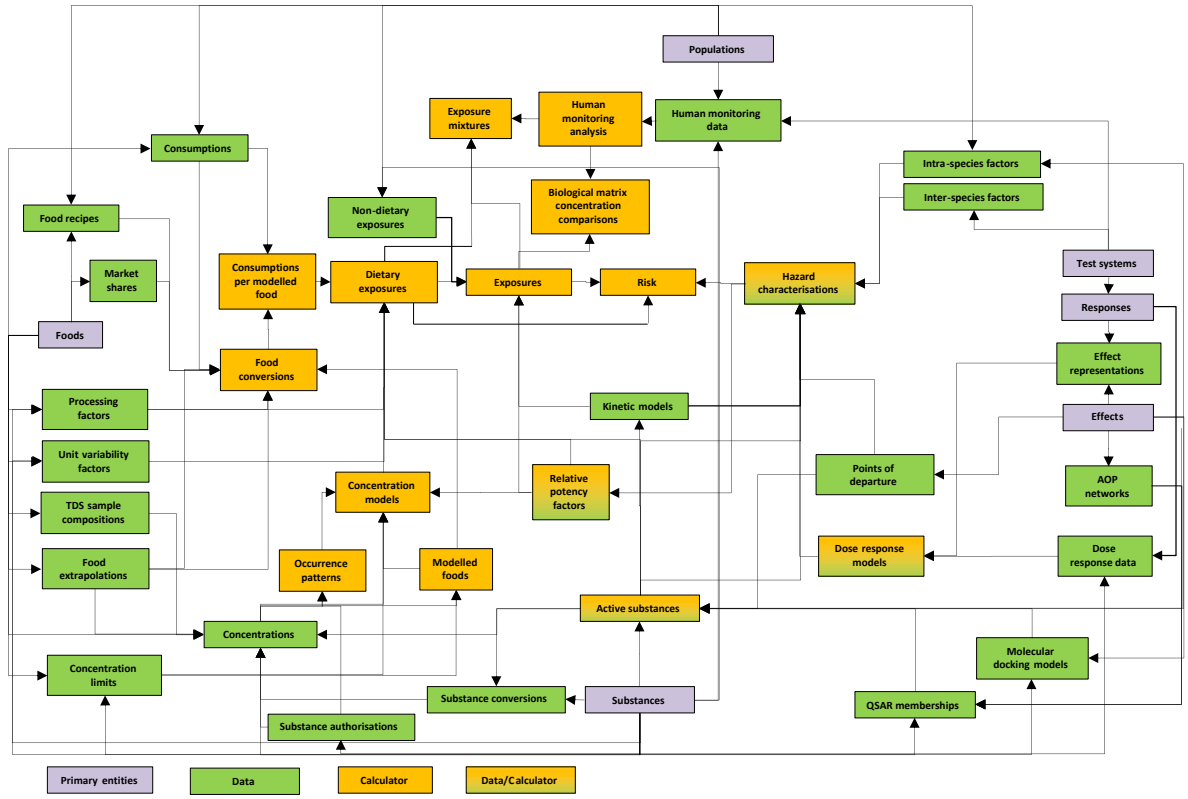


Figure 3.1: Modular design of MCRA.

3.1 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*).

3.1.1 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 High exposure food-substance combinations 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 Single value risks*

Effects data formats

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

Download empty dataset template: Zipped CSV Excel

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 3.1: Table definition for Effects.

Name	Type	Description	Aliases	Required
idEffect	AlphaNumeric (50)	Unique identification code of the effect.	idEffect, EffectId, CodeFocalEffect, 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
BiologicalOrganisation	*BiologicalOrganisationType*	Biological organisation of the effect: Molecular, Cellular, Tissue, Organ, Individual, Population. This is in line with AOP wiki terminology and can be used for grouping.	BiologicalOrganisation	No
KeyEventProcess	AlphaNumeric (100)	Description of AOP Key event component process. E.g., receptor signalling.	Process	No
KeyEventObject	AlphaNumeric (100)	Description of AOP Key event component object. E.g., PPAR-alpha.	Object	No
KeyEventAction	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

Accepted table names: Effects, Effect, KeyEvents, KeyEvent.

Effects calculation

Option *Multiple effects analysis* selects multiple selects. Effects are selected using **Effects selection**. Press **Clear filter** and/or **change selection** and select multiple effects in the scroll down box. If both this option and *Include related effects of AOP network* is unchecked, it is obligatory to select one and only one effect. If *Include related effects of AOP network* is checked, a **Focal effect** is specified and all related effects in the *AOP network* are selected as well.

Effects settings

Selection settings

Table 3.2: Selection settings for module Effects.

Name	Type	Description
Multiple effects analysis	Boolean	Specifies whether the analysis should consider multiple effects. Otherwise, a single focal effect should be selected.
Focal effect	AlphaNumeric	The main (health) effect of interest.
Include related effects of AOP network	Boolean	Include all related key events of the AOP network.

Effects as data

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

- *Effects data formats*
- *Effects calculation*

3.1.2 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) modelled foods, foods as coded in concentration data (e.g. wheat, tomato); 3) any other type of food (e.g. ingredients like flour, tomato sauce).

Output of this module is used by: *Consumptions Single value consumptions Market shares Food recipes Concentrations Concentration distributions Single value concentrations Processing factors Unit variability factors Occurrence patterns Occurrence frequencies Substance authorisations Deterministic substance conversion factors Concentration limits Concentration models Modelled foods Focal food concentrations Total diet study sample compositions Food extrapolations Food conversions Consumptions by modelled food High exposure food-substance combinations Dietary exposures Single value dietary exposures Exposures*

Foods data formats

Foods are of interest in (dietary) consumption assessments and the sources of exposure within exposure 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 and brand loyalty), facets, and hierarchies, can be described in the food properties, food hierarchies, and faces and facet descriptors tables.

Download empty dataset template: Zipped CSV Excel

Foods

Each food is identified by a unique code (idFood) in a code system of choice, a name, and a description. Food codes can have a hierarchical structure (as in the FoodEx1 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 3.3: 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	No
Description	AlphaNumeric (200)	Food description.	Description	No

Accepted table names: Foods, Food.

Food properties

Additional food properties. This table is deprecated. See table FoodUnitWeights which partly replaces this table for data on food unit weights.

Table 3.4: Table definition for Food properties.

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

Accepted table names: FoodProperties, FoodProperty.

Food unit weights

Food unit weights as specified for a food, and possibly a location.

Table 3.5: Table definition for Food unit weights.

Name	Type	Description	Aliases	Required
idFood	AlphaNumeric (50)	The unique identification code of the food.	idFood, Code, FoodId, FoodCode, Food, Id	Yes
Location	AlphaNumeric (50)	The location for which this food unit weight is defined. If not specified, then the value is considered a default unit weight that can be used when there is no location specific unit weight.	Location	No
ValueType	*UnitWeightValue-Type*	The value type of the unit weight value (i.e., raw agricultural commodity or edible portion). Controlled terminology.	ValueType, UnitWeight-ValueType	No
Qualifier	*ValueQualifier*	Qualifier of the unit weight value, e.g. equal-to (=) or smaller-than (<). If omitted, = is assumed.	Qualifier, QualifierType	Yes
Value	Numeric	Unit weight value in grams.	Value, Unit-WeightValue, UnitWeight	Yes
Reference	AlphaNumeric (200)	External reference(s) to source of the unit weight value.	Reference, References	No

Accepted table names: FoodUnitWeights, UnitWeights.

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 3.6: Table definition for Food hierarchies.

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

Accepted table names: FoodHierarchies, FoodHierarchy, FoodsHierarchy.

Facets

Food codes can be linked to facets, as e.g. in FoodEx2. FoodEx2 is a comprehensive food classification and description system aimed at covering the need to describe food in data collections across different food safety domains. See the EFSA catalogue browser for facets currently used e.g. the facet to describe processing technology is: F28 process.

Table 3.7: Table definition for Facets.

Name	Type	Description	Aliases	Required
idFacet	AlphaNumeric (10)	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	No
Description	AlphaNumeric (200)	Additional description of the facet.	Description	No

Accepted table names: Facets, Facet, FoodFacets, FoodFacet.

Facet descriptors

Facet descriptors are elements of additional information on a facet such as processing.

Table 3.8: Table definition for Facet descriptors.

Name	Type	Description	Aliases	Required
idFacet-Descriptor	AlphaNumeric (10)	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	No
Description	AlphaNumeric (200)	Additional description of the facet descriptor.	Description	No

Accepted table names: FacetDescriptors, FacetDescriptor, FoodFacetDescriptors, FoodFacetDescriptor.

Processing types

Table 3.9: Table definition for Processing types.

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	*Processing-DistributionType*	The distribution type. Simulated processing factors are restricted to the interval (0,1) using a logistic-normal distribution (default) or simulated processing factors are restricted to positive values using a log-normal distribution.	Distribution-Type, DistType	No
Bulking-Blending	Boolean	For types of processing applied on large batches, e.g., juicing, sauce/puree. Default is no bulking blending.	Bulking-Blending, BulkBlending, IsBulkBlending	No

Accepted table names: ProcessingTypes, ProcessingType.

Food consumption quantifications

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

Table 3.10: Table definition for Food consumption quantifications.

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

Accepted table names: FoodConsumptionQuantifications, FoodConsumptionQuantification.

Foods as data

Food definitions are provided as lists/catalogues of food definitions, optionally with encompassing processing type definitions, facet definitions, hierarchy 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 **modelled food**. MCRA implements a *recursive search algorithm* to link foods-as-eaten to modelled foods. This means that there can be intermediate steps, e.g. if unpeeled *apple* and *grapes* are the modelled foods, the food-as-eaten *apple pie* contains *peeled apple* and *raisins*, *peeled apple* is linked to unpeeled *apple*, and *raisins* are dried *grapes*. *Peeled* and *dried* are the *processing types*.

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 the food processing type (e.g. peeling). Any code can be followed by a hyphen ('-') plus a processing type code (e.g. FP0226-2). More than one level of processing code is allowed (e.g. FP0226-2-13). Subtype codes should precede processing codes (e.g. NL005$123$456-2).

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 (EFSA (2011a), EFSA (2011b)), 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 modelled foods 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 (e.g. FP0226-2) will be converted to the food code of the unprocessed food (FP0226) and a processing type (2). In an exposure assessment, the concentration in the unprocessed food will then be multiplied by the processing factor. Special attention is needed when food processing also includes changes of 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* (modelled food) 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 modelled food *grape* is needed, see also *processing correction*.

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 modelled food (TDS-food). A conversion from *apple-pie* (food-as-eaten) to *FruitMix* (modelled food) 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

The conversion algorithm links food as eaten codes to modelled food codes 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 3.11: 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 see EFSA (2011b). 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 3.12: 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 (EFSA (2011a)) '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 3.13 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 3.13: 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 Facet descriptors*.

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 modelled foods, 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 3.14: Example of a MCRA processing factors table using FoodEx 2 foods and facets codes.

FacetCode	Substance	FoodCode	ProcNom	ProcUpp	ProcNomUncUpp	ProcUppUncUpp
A07LC	SubstanceX	A01CE	0.5	0.6	0.05	0.06
F28.A07GV	SubstanceX	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 3.15: 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 substance 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 substance-specific, and non-substance-specific. Processing factors that are defined substance-specific always precede non-substance 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 data formats* is added. In this group, a new *table for Food hierarchies* 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/modelled foods and the exposure by food-as-eaten/modelled food 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

modelled food. 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 3.2: 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 3.2 for an example (*Citrus Fruits* versus *Citrus Fruits (unspecified)*). In MCRA, foods-with-facets are added as child nodes of the foods-without-facets.

Food unit weights

Food unit weights specify the standard weights of food units. E.g., the standard weight of an apple. This unit weight may be specified as the weight of the whole food (raw agricultural commodity/RAC) or the weight of the edible portion (EP), e.g., without peel. Unit weights are specified in the table *table for Food unit weights* and used in combination with *unit variability factors* to account for unit-to-unit variation in concentrations between single units of the same food in *single value dietary exposures assessments* and *(individual) dietary exposures assessments*.

Food unit weights can be location specific or specified as overall (default) unit weights. For some models, e.g., the *IESTI model*, location specific unit weights are preferred over overall unit weights. The overall unit weights are then used when no location specific uses are available. For other methods, only overall unit weights are used. If, for a food, an overall unit weight is not available, but there are location specific unit weights available, then the overall unit weight is computed as the average weight of the location specific unit weights (similar to EFSA PRIMo revision 3 EFSA (2018)).

Note: Note that in earlier versions of the software, food unit weights were specified in the *table for Food properties*. Although this is still possible, the recommended way of specifying unit weights is in the *table for Food unit weights*. If, for a food, unit weights are specified in both tables, then the unit weights specified in the *table for Food unit weights* have priority. The unit weights specified in the *table for Food properties* are then only used as fallbacks for the overall unit weight when no overall unit weight is specified in the *table for Food unit weights*.

3.1.3 Non-dietary exposure sources

Non-dietary exposure sources are the sources containing chemical substances to which individuals in a population are exposed via any of three non-dietary routes: dermal, inhalation or oral, per day.

Non-dietary exposure sources data formats

Non-dietary exposure sources are defined in the non-dietary sources table.

Download empty dataset template: Zipped CSV Excel

Non-dietary exposure sources

Each non-dietary exposure source is identified by a unique code in a code system of choice, a name, and a description.

Table 3.16: Table definition for Non-dietary exposure sources.

Name	Type	Description	Aliases	Required
idNonDietary-ExposureSource	AlphaNumeric (50)	The unique identification code of the non-dietary exposure source.	idNonDietary-ExposureSource, idSource, Code, Id	Yes
Name	AlphaNumeric (100)	The name of the non-dietary exposure source.	Name	No
Description	AlphaNumeric (200)	Description of the non-dietary exposure source.	Description	No

Accepted table names: NonDietaryExposureSources.

Non-dietary exposure sources as data

Non-dietary exposure sources are lists/catalogues of definitions of non-dietary sources through which people can be exposed to chemical substances (e.g., personal care products and plant-protection products).

- *Non-dietary exposure sources data formats*

3.1.4 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 (with a start and end date), age range (with a minimum and maximum age) and gender. Example: the French population in 2005-2007 (= time period) of women (= gender) of child-bearing age 18-45 yr (= age range).

Output of this module is used by: *Consumptions Single value consumptions Concentrations Consumptions by modelled food Dietary exposures Single value dietary exposures Non-dietary exposures Exposures Human monitoring data Human monitoring analysis Biological matrix concentration comparisons Hazard characterisations Risks Single value risks*

Populations data formats

Populations are primary entities of the data model.

Download empty dataset template: Zipped CSV Excel

Populations

Populations identify human groups in e.g. dietary, nondietary and human monitoring surveys. Optionally, a name and description can be added. Specify a standard bodyweight (optional) as a descriptor of the average or nominal bodyweight in the population. Use table PopulationIndividualPropertyValues to specify descriptors/properties that characterise the population. These population individual property values will be used to restrict the population to e.g. a certain time period (through specifying a start and end date, both dates are inclusive), age (through specifying a minimum and maximum age, both limits are inclusive) or gender (male or female). To facilitate the user, dynamic properties may be added to the table. Three kind of additional properties are available: alphanumeric properties (property name and level), numeric properties (a range through specifying a minimum and maximum using suffixes Min and Max, both bounds are inclusive) and datetime properties (a range through specifying a start and end date using prefixes Start and End, both dates are inclusive). Dynamic or additional properties are ignored when table PopulationIndividualPropertyValues is present in the upload. In table IndividualProperties each property used in the Populations table is described.

Table 3.17: 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 of of the population.	Description	No
Location	AlphaNumeric (50)	Location.	Location, Country	No
StartDate	DateTime	Starting date of the specific time window marking this population.	StartDate	No
EndDate	DateTime	End date of the specific time window marking this population.	EndDate	No
NominalBody-Weight	Numeric	Nominal body weight (in kg) of the individuals of this population.	NominalBody-Weight, BodyWeight	No
Additional individual properties, type = Alpha-Numerical		AlphaNumerical population properties specifying a level or levels (comma separated), [property name]. E.g. for individual property [Region] specify a region like [North] or [South]. Note that table IndividualProperties should contain the property Region with PropertyLevel = Individual and Type = Categorical. For other type of properties use Type = Boolean or Gender. For properties specifying the sampling date use PropertyLevel = IndividualDay and Type = Month. See also Type and Unit definitions for accepted Individual property types (controlled terminology).		No
Additional individual properties, type = Numerical		Numerical population properties specifying a range. Specify an individual property name followed by a suffix 'Min' or 'Max': [property name]Min or [property name]Max. E.g. for individual property [Height] specify the range as [HeightMin] and [HeightMax]. Note that table IndividualProperties should contain the property Height with PropertyLevel = Individual and Type = Numeric, Nonnegative, Integer or NonnegativeInteger. See also Type and Unit definitions for accepted Individual property types (controlled		No

Accepted table names: Populations, Population.

Individual properties

This table is used to describe the properties used in the Populations or PopulationIndividualPropertyValues table characterising the population (table Populations) and/or the properties used in the Individuals table characterising an individual. Properties like Age, Gender, Region are describing an individual (PropertyLevel = Individual). Properties like Period (for populations) or Month (sampling date for an individual day) are describing an individual day (PropertyLevel = IndividualDay).

Table 3.18: Table definition for Individual properties.

Name	Type	Description	Aliases	Required
idIndividual-Property	AlphaNumeric (50)	The code of the property.	idIndividual-Property, Individual-PropertyId, Individual-Property	Yes
Name	AlphaNumeric (100)	The name of the property.	Name	No
PropertyLevel	*PropertyLevelType*	The level of the property. This type follows a controlled terminology, with possible values: Individual or IndividualDay.	PropertyLevel, LevelProperty	No
Description	AlphaNumeric (200)	Description of the property.	Description	No
Type	*IndividualProperty-Type*	This field specifies the type of the values of this individual property. This type follows a controlled terminology, with possible values: Boolean, Categorical (default), Numeric, Nonnegative, Integer, NonnegativeInteger, Month, Datetime, Gender.	Type	No

Accepted table names: IndividualProperties, IndividualProperty.

Population individual property values

This table describes population individual properties, such as Age, Gender, Period, Region or Breastfeeding. Population individual property value are used to restrict the population to e.g. a range of ages, a gender, a certain time period, a geographical location or women giving breast feeding. For numerical properties use MinValue and MaxValue to specify a range. For Gender, Region and Breastfeeding use Value to specify a gender level, a categorical level or a boolean, respectively. In table IndividualProperties each property used in the PopulationIndividualPropertyValues table is described.

Table 3.19: Table definition for Population individual property values.

Name	Type	Description	Aliases	Required
idPopulation	AlphaNumeric (50)	The code of the population to which the property is attached. The provided population code should match with a code of the populations table.	idPopulation, PopulationId, Population, PopulationCode, Code	Yes
idIndividual-Property	AlphaNumeric (50)	The name or reference of the individual property.	idIndividual-Property, Individual-PropertyId, Individual-Property	Yes
Value	AlphaNumeric (50)	The value of the property.	Value	No
MinValue	Numeric	Minimum value of the value of the property.	MinValue, ValueMin	No
MaxValue	Numeric	Maximum value of the value of the property.	MaxValue, ValueMax	No

Accepted table names: PopulationIndividualPropertyValues, PopulationPropertyValues.

Populations calculation

In an exposure or risk assessment, the population of interest is implicitly or explicitly defined. An implicit definition is made by selecting one of the available food consumption surveys in the **Compute** option. For example, by selecting the food survey 'NL-Toddlers' with corresponding consumption data, the Dutch population of toddlers is implicitly considered as the population of interest.

After selecting a survey, the population is further defined by checking *Define population based on specified individual properties* in the *Population definition from dietary surveys panel*. Include implicitly defined individual(day) properties in the population definition. E.g., the Dutch population of toddlers is restricted to females by including the property gender with level *'Female'* in the population definition. The implicitly defined population becomes the Dutch female toddlers.

Although implicit definition of the population of interest works fine, there is a need to make more explicit that the focus of an assessment is on the population. The user should be aware that assessments are about assessing the exposure or risk in a specific population of interest. This becomes even more urgent when dietary exposures are combined with non-dietary exposures. Without explicitly specifying the population of interest, the exposure of Dutch toddlers may unintentionally be combined with non-dietary exposures of e.g. adult operators. The explicit definition of a population is made by selecting a population datasource in the **Use data** option. By specifying population properties like age, gender or any other property in the data, the population may be further restricted.

For the **Compute** option:

- Setting *Population definition from dietary surveys* Check *Define populations based on specified individual properties* for including individual properties.

When the population of interest is explicitly defined, (**Use data** option), the population definitions are based on the data in the *IndividualProperties table* referring to populations in the *Populations table.*.

Population definition from dietary surveys

This panel is only available when the population of interest is implicitly defined, use the *Compute* option. Currently, population definitions are based on dietary surveys. It is foreseen that the population definition will be extended as soon as Human Biomonitoring (HBM) data and/or non-dietary survey data become available.

Specify the *nominal population bodyweight*, needed for single value calculations, e.g. 70 kg (default).

After checking *Define populations based on specified individual properties* in Figure 3.3, include individual properties for the population definition.

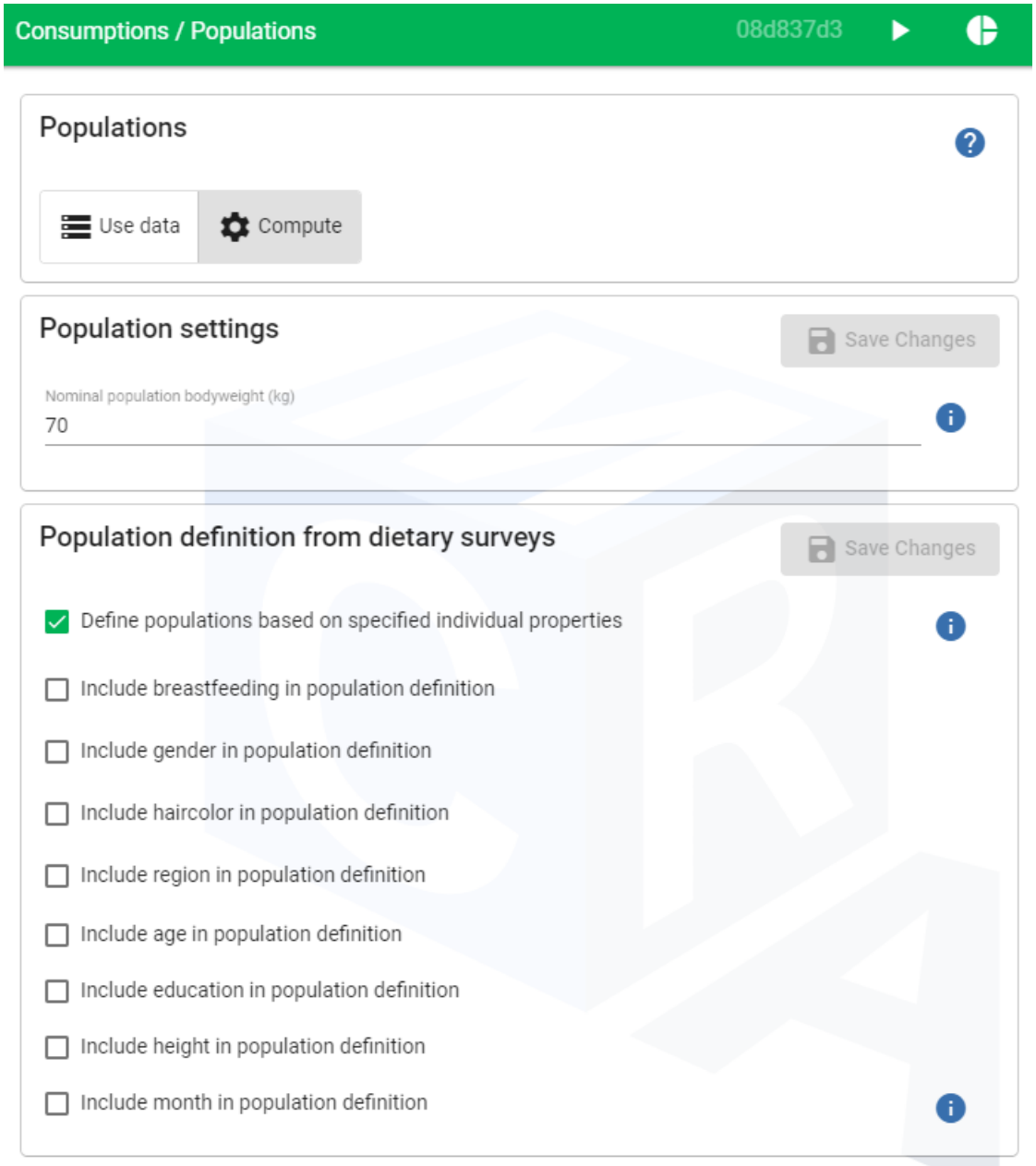


Figure 3.3: Define populations by including individual properties.

Select the requested levels for each property.

See also the *consumption panel documentation* for further information on population definition based on a selection of individual properties.

For individual day properties, an extra option is available to include the records without date data. Currently, the

Figure 3.4: Check gender, region and age and select the requested levels Female, East, North and South and age min = 0, age max = 5, resp.

defualt value is to include records without dates.

- *Include individual day records with missing dates:*. If checked then individual day records with missing dates are included in the individual day subset.

Populations data selection

This option is only available when the population of interest is explicitly defined, use the *Use data* option and select a data source. Select one of the available populations in the *Populations selection* pane.

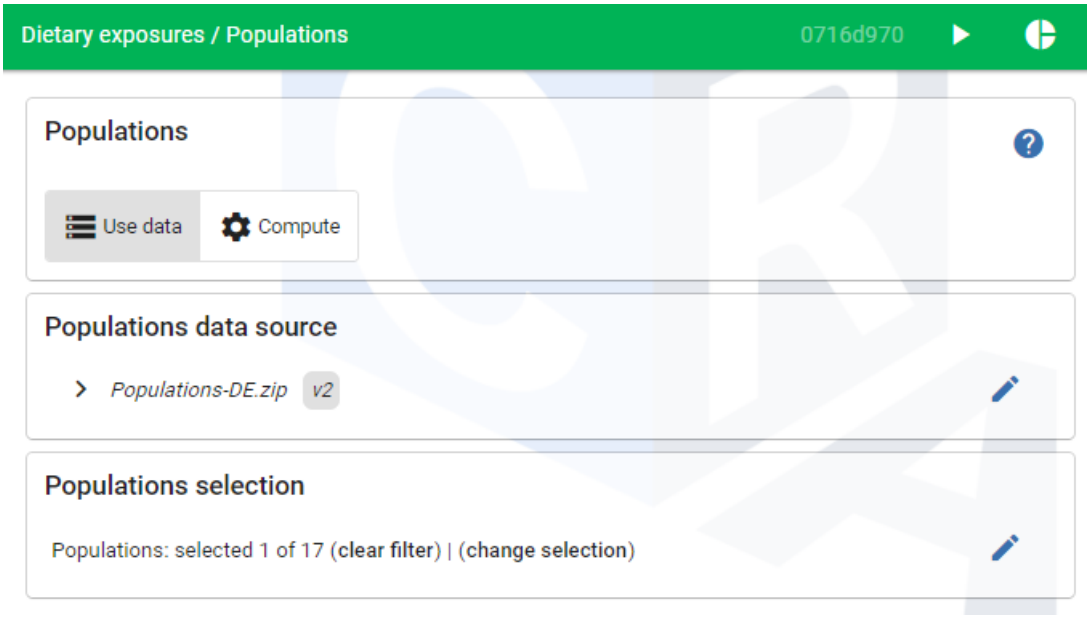


Figure 3.5: Population data selection (*Use Data*).

Selecting two or more populations, initiates a loop over multiple populations. Check the *Loop over multiple populations :* checkbox. See also Figure 3.8

Figure 3.6: Available population data: **change selection**.

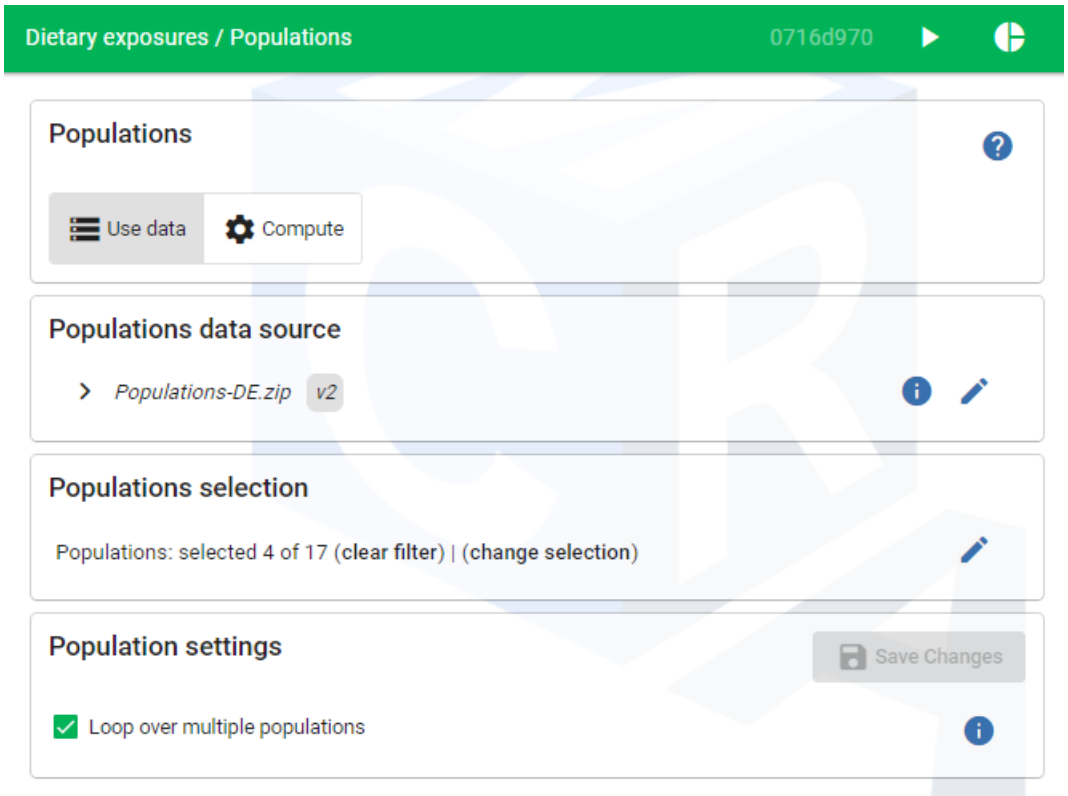


Figure 3.7: Loop over multiple populations (*Use Data*).

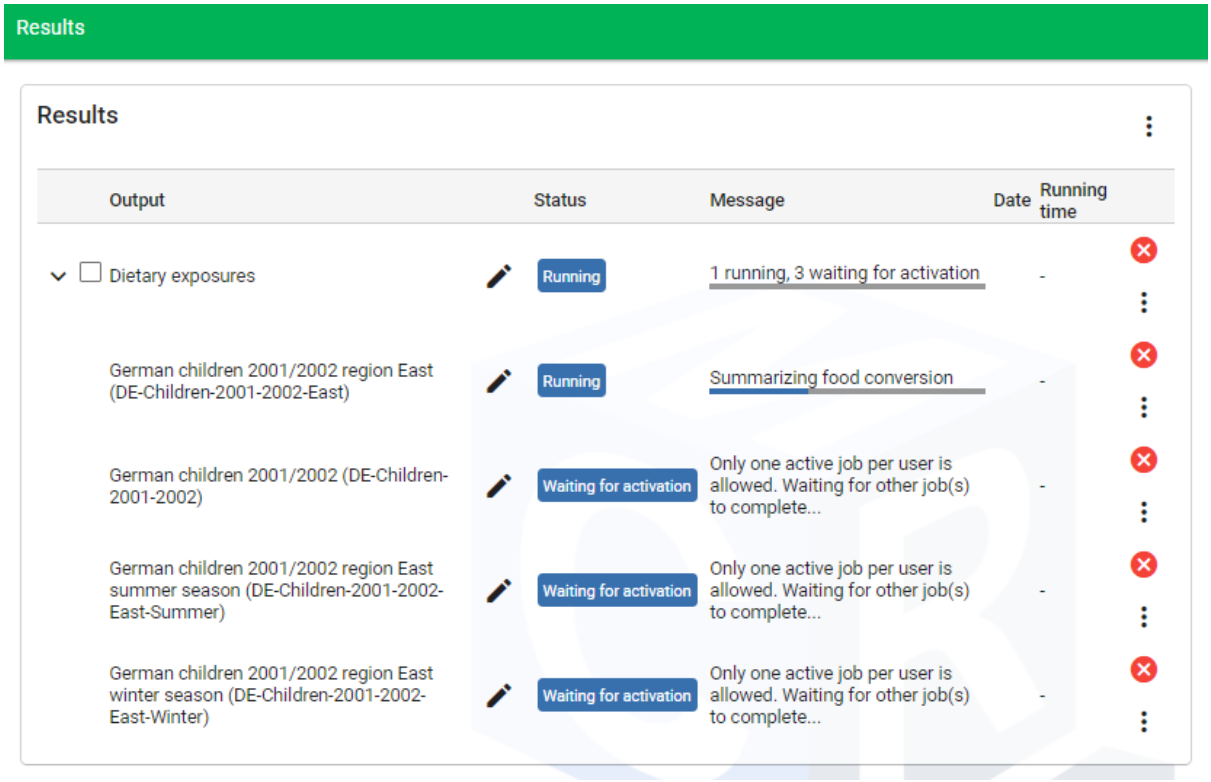


Figure 3.8: Results pane with loop over multiple populations (*Use Data*).

Populations settings

Selection settings

Table 3.20: Selection settings for module Populations.

Name	Type	Description
Population	AlphaNumeric	Specifies which population is selected.
Define populations based on specified individual properties	Boolean	Define a population by selecting specific ranges/values of individual properties. E.g., the female population between ages and 45 is composed of the properties gender (female) and age (between 18 and 45).
Include individual day records with missing dates	Boolean	If checked, then individual day records with missing dates are included in the individual day subset.
Loop over multiple populations	Boolean	Loop over the selected populations.
Nominal population bodyweight (kg)	Numeric	Nominal population bodyweight in kg (needed for single value calculations).

Populations as data

Populations are provided as data. When the population of interest is explicitly defined, subset selections are based on the data in the IndividualProperties and PopulationIndividualPropertyValues table.

- *Populations data formats*

Calculation of populations

Populations are implicitly defined by the individuals as used in the dietary consumption module and/or the non-dietary module and/or the human based monitoring module.

- *Populations calculation*

3.1.5 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

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.

Download empty dataset template: Zipped CSV Excel

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., standardised assay kit may also be specified (GuidelineMethod).

Table 3.21: 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., standardised assay kit.	Guideline-Method	No
ResponseType	*ResponseType*	The data type of the response measurements (e.g., continuous multiplicative, continuous additive, binary, quantal, count, ordinal, categorical). Controlled terminology.	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

Accepted table names: Responses, Response.

Responses settings

Selection settings

Table 3.22: Selection settings for module Responses.

Name	Type	Description
Response(s)	AlphaNumeric	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*

3.1.6 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 Concentration distributions Single value concentrations Processing factors Unit variability factors Occurrence patterns Occurrence frequencies Substance authorisations Substance approvals Substance conversions Deterministic substance conversion factors Concentration limits Concentration models Modelled foods Focal food concentrations Food conversions Consumptions by modelled food High exposure food-substance combinations Dietary exposures Single value dietary exposures Non-dietary exposures Exposures Exposure mixtures Human monitoring data Human monitoring analysis Biological matrix concentration comparisons 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 Single value risks*

Substances data formats

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.

Download empty dataset template: Zipped CSV Excel

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 3.23: Table definition for Substances.

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	Yes
Name	AlphaNumeric (100)	The substance name.	Name, SubstanceName, PesticideName	No
Description	AlphaNumeric (200)	Substance description.	Description	No
Concentration-Unit	*ConcentrationUnit*	Contains a coding to determine the default unit in which concentrations for this substance are expressed.	Concentration-Unit, Unit, Reference-Concentration-Unit	No
CramerClass	Integer	The Cramer class of the substance.	CramerClass	No
MolecularMass	Numeric	The molecular (molar) mass.	MolecularMass, Mass, MolarMass, Molecular-Weight, MolarWeight	No
IsLipidSoluble	Boolean	States whether the substance is soluble in lipid (0 = no, 1 = yes).	IsLipidSoluble, IsSoluble	No

Accepted table names: Substances, Substance.

Substances settings

Selection settings

Table 3.24: Selection settings for module Substances.

Name	Type	Description
Multiple substances analysis	Boolean	Specifies whether the assessment involves multiple substances.
Index substance	AlphaNumeric	The substance of interest or index substance.

Substances as data

Substances are provided as data (code, name).

- *Substances data formats*

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

Download empty dataset template: Zipped CSV Excel

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 3.25: Table definition for Test Systems.

Name	Type	Description	Aliases	Required
idTestSystem	AlphaNumeric (50)	Unique identification code of the test system.	idTestSystem, idSystem, Id, Code	Yes
CodeSystem	AlphaNumeric (100)	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	*TestSystemType*	The type of the test system, i.e., in-vivo, cell-line, primary cells, tissue, organ. Controlled terminology.	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	*ExposureRouteType*	If applicable, the route of exposure associated with the in vivo test-system, oral, dermal, inhalation, s.c., i.v. Controlled terminology.	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

Accepted table names: TestSystems, TestSystem, Systems, System.

Test systems as data

Test systems are provided as data.

- *Test systems data formats*

3.2 Consumption modules

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

3.2.1 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 persons, 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 modelled food*

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 targeted 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 Souverein et al. (2011).

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.

Download empty dataset template: Zipped CSV Excel

Food consumption surveys

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

Table 3.26: Table definition for Food consumption surveys.

Name	Type	Description	Aliases	Required
idSurvey	AlphaNumeric (50)	Unique identification code of the food consumption survey.	idSurvey, idFoodSurvey, Survey, FoodSurvey, SurveyId, FoodSurveyId, Code, Id	Yes
Name	AlphaNumeric (100)	The name of the food consumption survey.	Name, SurveyName	No
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	*BodyWeightUnit*	The unit of bodyweight of the individuals of the survey: kg (default) or g.	BodyWeight-Unit, UnitBody-Weight, WeightIn	No
Consumption-Unit	*ConsumptionUnit*	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

Accepted table names: FoodConsumptionSurveys, ConsumptionSurveys, FoodSurveys, Surveys.

Individuals

The individuals of a survey are recorded in the Individuals table. Add additional properties like Region, Breastfeeding to further describe an individual. In table IndividualProperties, each property in the Individuals table is described (recommended way). Note that only those properties that are available in the Individuals table are used in module Populations, table Populations or PopulationIndividualPropertyValues to subset the individuals. This is only relevant when the UseData option in the population module is used.

Table 3.27: 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
Name	AlphaNumeric (100)	Name or label of the individual.	Name	No
Description	AlphaNumeric (200)	Additional description of the individual.	Description	No
Individual properties		Other individual properties can be added like the fields Age, Gender, Region etc. These properties are automatically parsed as co-factors or co-variables.		No

Accepted table names: Individuals, SurveyIndividuals, ConsumptionSurveyIndividuals, FoodConsumptionSurveyIndividuals.

IndividualDays

The individuals and days in the survey are recorded in the individualDays table.

Table 3.28: Table definition for IndividualDays.

Name	Type	Description	Aliases	Required
idIndividual	AlphaNumeric (50)	The identification code of the individual.	idIndividual, IndividualId, Individual, Id	Yes
idDay	AlphaNumeric (50)	Identification code of the day of consumption, sequential number	idDay, DayId, Day, DayOfSurvey	Yes
SamplingDate	DateTime	The date of the consumption.	Date, SamplingDate, SurveyDate	No

Accepted table names: IndividualDays, SurveyIndividualDays, ConsumptionSurveyIndividualDays, FoodConsumptionSurveyIndividualDays.

Individual properties

This table is used to describe the properties used in the Populations or PopulationIndividualPropertyValues table characterising the population (table Populations) and/or the properties used in the Individuals table characterising an individual. Properties like Age, Gender, Region are describing an individual (PropertyLevel = Individual). Properties like Period (for populations) or Month (sampling date for an individual day) are describing an individual day (PropertyLevel = IndividualDay).

Table 3.29: Table definition for Individual properties.

Name	Type	Description	Aliases	Required
idIndividual-Property	AlphaNumeric (50)	The code of the property.	idIndividual-Property, Individual-PropertyId, Individual-Property	Yes
Name	AlphaNumeric (100)	The name of the property.	Name	No
PropertyLevel	*PropertyLevelType*	The level of the property. This type follows a controlled terminology, with possible values: Individual or IndividualDay.	PropertyLevel, LevelProperty	No
Description	AlphaNumeric (200)	Description of the property.	Description	No
Type	*IndividualProperty-Type*	This field specifies the type of the values of this individual property. This type follows a controlled terminology, with possible values: Boolean, Categorical (default), Numeric, Nonnegative, Integer, NonnegativeInteger, Month, Datetime, Gender.	Type	No

Accepted table names: IndividualProperties, IndividualProperty.

Individual property values

Not recommended. This table describes individual property values. Property values are describing an individual for properties like e.g. Region, Breastfeeding. The recommended way is to add these columns as additional columns in the Individuals table. In table IndividualProperties, each property in the IndividualPropertyValues table is described.

Table 3.30: Table definition for Individual property values.

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

Accepted table names: IndividualPropertyValues, IndividualPropertyValue.

Consumptions

The individual consumptions are recorded in the consumptions table.

Table 3.31: 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
Facets	AlphaNumeric	The codes of the facets/treatments recorded for this consumption. Multiple treatments are separated by a '$'.	Treatments, Treatment, Facets	No
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 (25)	Identification code of the day of consumption, sequential number	idDay, DayId, Day, DayOfSurvey	Yes
idMeal	AlphaNumeric (25)	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

Accepted table names: FoodConsumptions, FoodConsumption, Consumptions, Consumption.

Consumptions calculation

The consumptions module offers a number of options and filters that specify how the consumption data should be included in the assessment.

- After including individual properties for defining a population in the *populations panel* an option becomes available *Match consumption data to population definition options*. Use all populations definitions (default), ignore all populations definitions or use a selection of properties to define the population.
- When individual sampling weights have been recorded in the *data*, the option *ignore sampling weights* specifies whether sampling weights are ignored or not in all calculations.
- It is possible to *restrict to consumptions of specific foods*.
- Depending on the *exposure type* (acute or chronic), the individuals (chronic) or individual-days (acute) can be filtered in several ways. It is possible to restrict the consumers or consumer days to consumers that consume or consumer days with a consumption. Use option *consumers or consumer days only*, and only consumers or consumer days with consumption of *specific (focal) products* are selected.
- For a chronic dietary or risk assessment (OIM, LNN or BBN), the individuals with less than N surveys days may be filtered out. Check setting *Exclude individuals with less than N days*, then./ option *N (number of days in survey)* appears. Specify a number, all individuals with a number of survey days less than N are filtered out.
- In general, one consumption data source is selected. However, the menu allows selection of multiple data-sources through the use of the *add another data source* option. Press the add button + on the right side of the

panel and select a datasource. Then, in the settings panel a new checkbox appears *Loop over multiple surveys* and after checking it all red triangles in the left panel disappear. See also *multiple surveys*.

In Figure 3.9 the consumption panel is shown. The setting *Match consumption data to population definition* is selected as default. All properties that are included in the *populations definition* are used. See *other options* for more info.

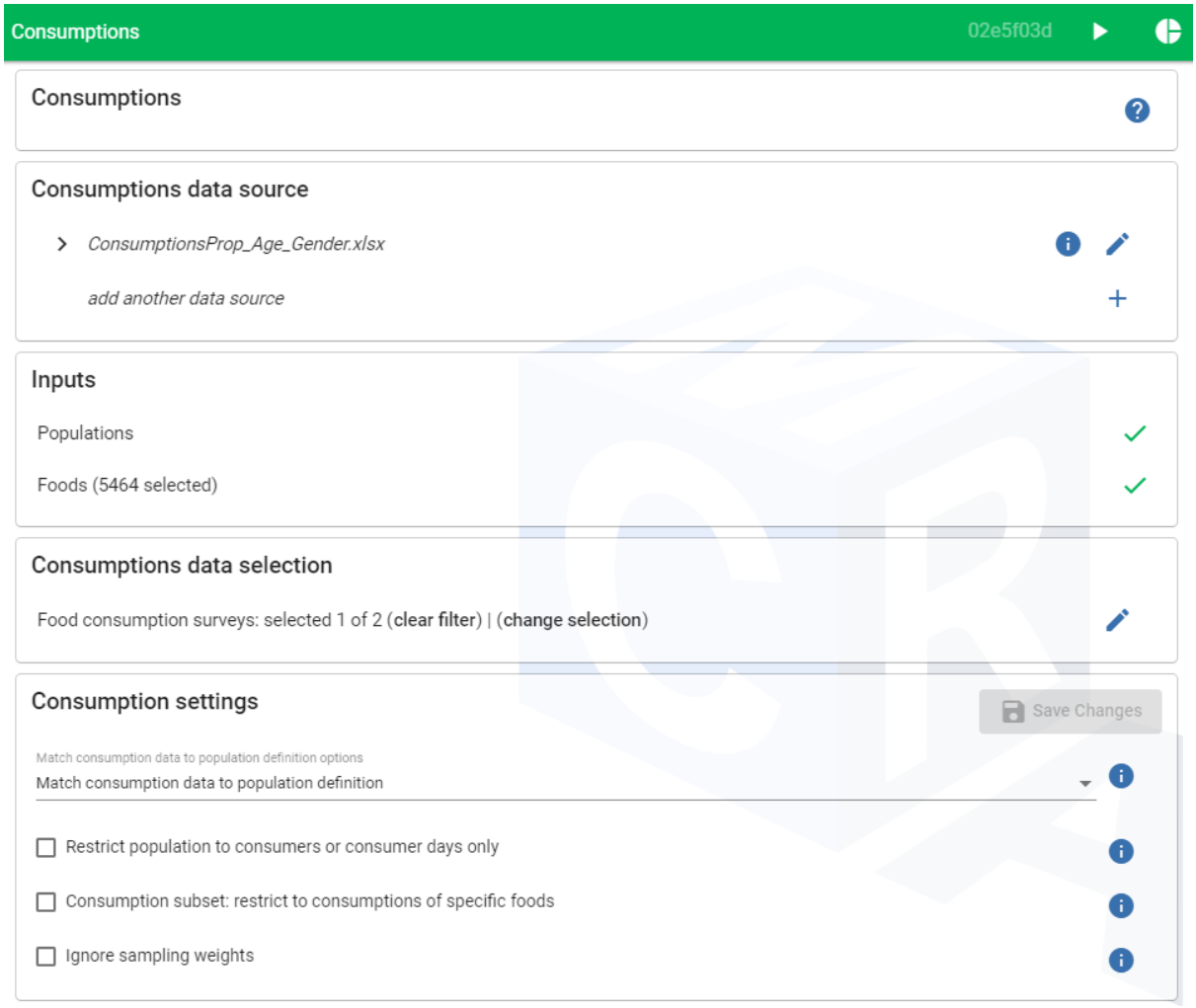


Figure 3.9: Consumption and consumption settings.

Consumption population definition

In Figure 3.10 all options of setting *Match consumption data to population definition options* are shown.

- Match consumption data to population definition: all selected properties in the population definition are used (default),
- Ignore population definition (use all individuals in survey): all selected properties in the population definition are ignored,
- Match consumption data to population definition using selected properties only: a selection of individual properties is used to define the population.

In the *populations panel*, the population was defined by including individual properties. These properties are all used (default), ignored or *a subset is selected* for further restricting the population.

In Figure 3.12 the selected individual properties that are available are shown. Check one or more for restricting the population definition to the specified set.

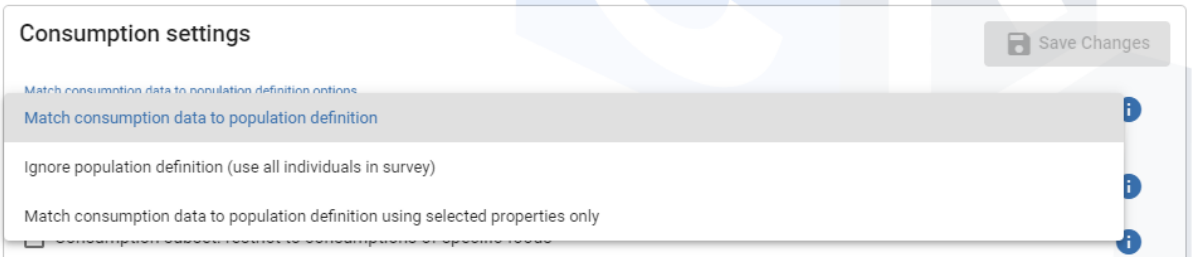


Figure 3.10: Consumption settings panel: Match consumption data to population definition options.

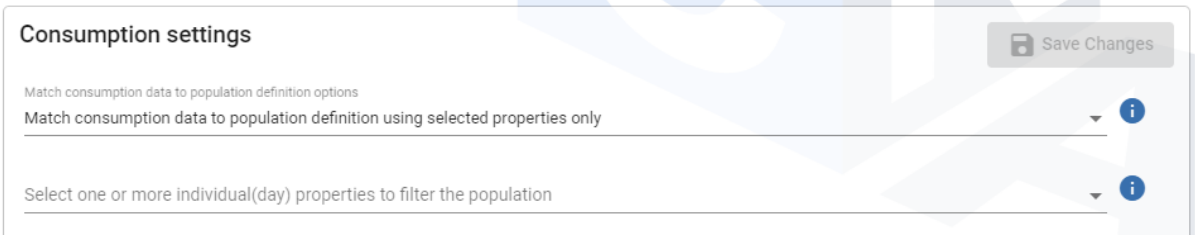


Figure 3.11: Match consumption data to population definition using selected properties only.

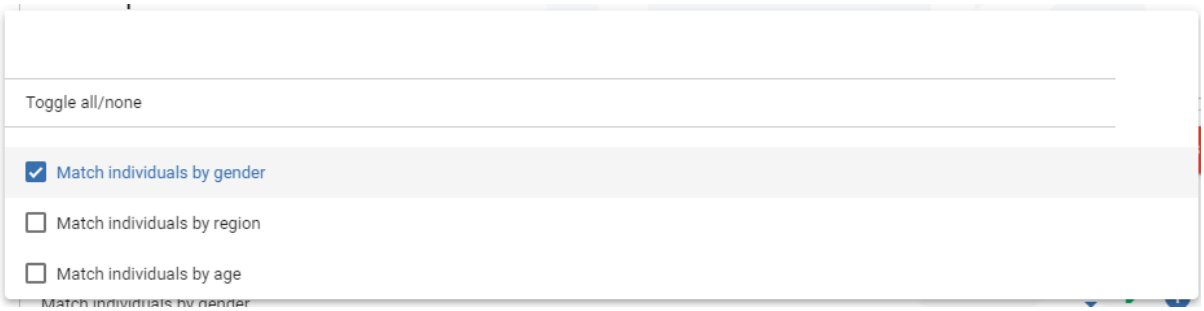


Figure 3.12: Check one or more of the available properties after selecting

Multiple surveys

In the Consumption panel, consumption data sources are selected. After selecting two or more data sources (*add another data source*), MCRA performs a loop over the selected surveys, see Figure 3.13. Check *Loop over multiple surveys*.

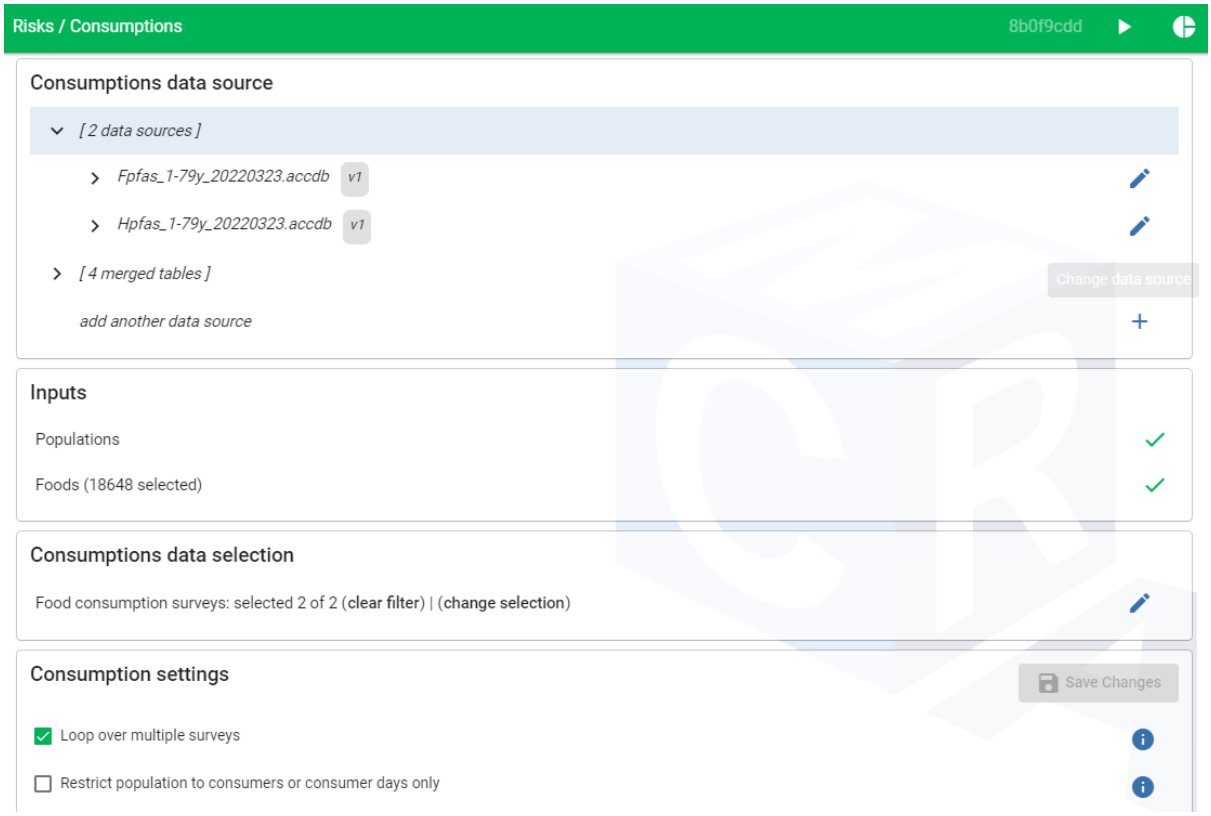


Figure 3.13: Loop over multiple surveys.

The advantage of running a multiple survey loop instead of running separate actions for each surveyn is that outputs are directly compared. In the Results panel, Figure 3.14, a toc appears containing links to the output of the risk or dietary exposure assessment (main action) for the selected surveys, but also a link to the combined output panel.

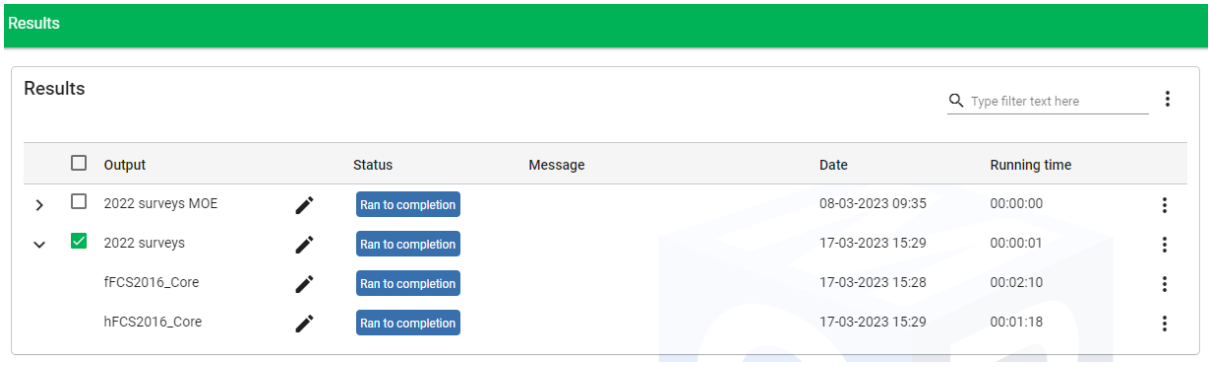


Figure 3.14: Results panel multiple surveys.

After clicking the link *2022 surveys*, the combined output is represented in violin plots and a table with percentiles for the specified percentages. Here a combined risks action is run and the p99 of the hazard index distribution is selected. For a dietary exposure action similar results are available.

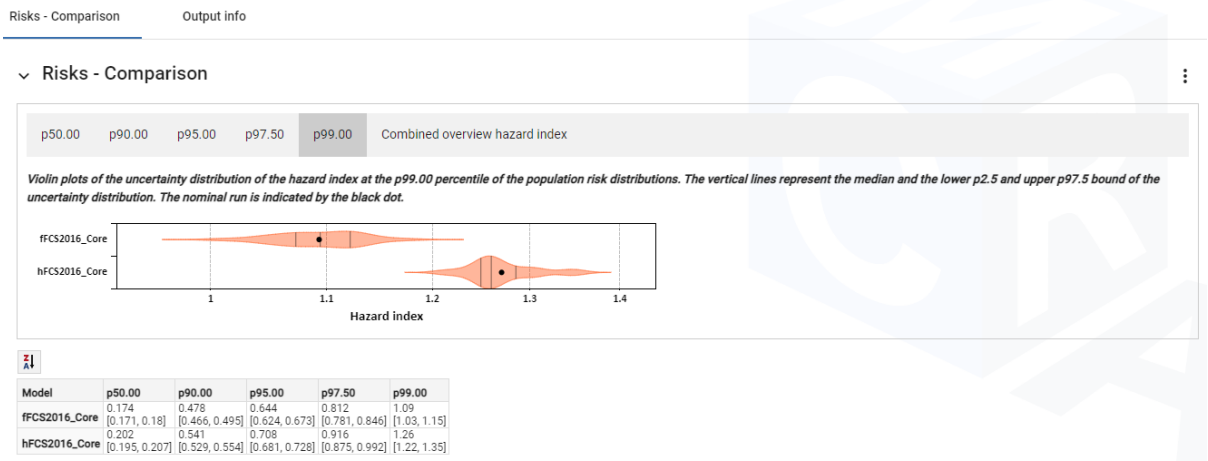


Figure 3.15: Risks comparison multiple surveys.

Consumptions settings

Selection settings

Table 3.32: Selection settings for module Consumptions.

Name	Type	Description
Risk type	*ExposureType*	The type of exposure considered in the assessment; acute (short term) or chronic (long-term).
Food survey	AlphaNumeric	The food consumption survey representative for the population interest.
Restrict population to consumers or consumer days only	Boolean	Specifies whether the population should be restricted to the individuals (chronic) or individual days (acute) that have non-ze consumption.
Restrict population to consumers or consumer days with consumptions of specific foods	Boolean	Specifies whether the population should be restricted to the individuals (chronic) or individual days (acute) consuming any the foods of the specified subset.
Selected foods-as-eaten	AlphaNumeric	Set of consumed foods that are of particular interest for restrict the consumers / consumption days.
Consumption subset: restrict to consumptions of specific foods	Boolean	If checked, then the consumptions are restricted to those of the specified food-as-eaten subset.
Selected foods-as-eaten	AlphaNumeric	Set of consumed foods that are of particular interest.
Ignore sampling weights	Boolean	If checked, individual sampling weights are not used (sampling weight = 1). If unchecked, the specified sampling weights are used.
Match consumer selection to population definition options	*IndividualSubsetType*	Match consumption data to population definition. Use populati definitions (default), ignore all population definitions or use a selection of properties.
Exclude individuals with less than N days	Boolean	Filter out all individuals with less than N survey days.
N (number of days in survey)	Numeric	Specify the nominal number of days in the survey to filter out a individuals with less than N survey days.
Select one or more individual(day) properties to filter the population	AlphaNumeric	Select one or more individual(day) properties to filter the individuals(days) in the population.

Uncertainty settings

Table 3.33: Uncertainty settings for module Consumptions.

Name	Type	Description
Resample individuals	Boolean	Individual data are resampled from the original database using t bootstrap methodology (Efron 1979, Efron & Tibshirani 1993)
Resample portion sizes	Boolean	Specifies whether portion sizes should be resampled based on food consumption quantification data, see (Souverein et al. 201

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*
- *Consumptions calculation*

3.2.2 Food recipes

Food recipes data specify the composition of specific foods (typically: foods-as-eaten) in terms of other foods (intermediate foods or modelled foods) 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

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).

Download empty dataset template: Zipped CSV Excel

Recipes

Table 3.34: Table definition for Recipes.

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

Accepted table names: FoodTranslations, FoodTranslation, FoodCompositions, FoodComposition.

Food recipes as data

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

- *Food recipes data formats*

3.2.3 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

Describes the shares (proportions) in a market.

Download empty dataset template: Zipped CSV Excel

Market shares

Market shares main table.

Table 3.35: Table definition for Market shares.

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

Accepted table names: MarketShares, MarketShare, FoodMarketShares, FoodMarketShare.

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 (Goodhardt et al. (1984)). 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.

3.2.4 Single value consumptions

Single value consumption data are the single value amounts (Large Portion, Mean Consumption, p97.5Consumption) of modelled foods (foods-as-measured) consumed in a population.

This module has as primary entities: *Populations Foods*

Output of this module is used by: *Single value dietary exposures*

Single value consumptions data formats

Single value consumptions data provides a single per-individual-day and per-food consumption amount for a population. Also the bodyweight should be specified, and optionally also age, sex, and other properties may be recorded. 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 Souverein et al. (2011).

Single value consumptions are described using one table: PopulationConsumptionSingleValues.

Download empty dataset template: Zipped CSV Excel

Population consumption single values

Population consumption single values describe population food consumptions in the form of single value statistics.

Table 3.36: Table definition for Population consumption single values.

Name	Type	Description	Aliases	Required
idPopulation	AlphaNumeric (50)	Unique identification code of the population.	IdPopulation, PopulationId	Yes
idFood	AlphaNumeric (50)	The unique identification code of the consumed food.	idFood, FoodCode, Food	Yes
Value type of the single value consumption amount.	*ConsumptionValue-Type*	The value type of this consumption value.	Consumption-Type, ValueType, Consumption-ValueType, Consumption-SingleValue-Type	Yes
Percentile	Numeric	The percentile (if consumption value type is a percentile).	Percentile	No
Consumption-Amount	Numeric	The consumed amount.	Amount, Consumption, Consumption-Amount, Amount-Consumed	Yes
Consumption-Unit	*ConsumptionIntake-Unit*	The unit of the consumption amount.	AmountUnit, UnitAmount, Consumption-Unit	No
Reference	AlphaNumeric (200)	Reference to the source from which this value is obtained.	Reference, References, Source, Sources	No

Accepted table names: ConsumptionSingleValues, SingleValueConsumptions, PopulationConsumptionSingleValues, PopulationConsumptionValues.

Single value consumptions calculation

Single value consumptions can be supplied *as data* or computed. When single value consumptions are computed from *consumptions by modelled food*, then the mean, median and large portion (p97.5 percentile) are computed for all modelled food consumption distributions. Besides these statistics, also the mean bodyweight of the population is computed. The following options are relevant in this calculation:

- Set the *risk type* option to *acute* if the consumptions should be based on the individual-day distributions. Otherwise, choose *chronic* to base them on the distributions aggregated by individual.
- Checking the *apply processing factors* option will compute the single value consumptions for the processed foods. When using this option, the output will also show a reverse yield factor, that is the ratio of the quantity of the raw commodity required to to obtain the processed commodity. Note, when no processing factors are available, the single-value consumption amounts of processed foods are expressed in terms of the processed commodities. The yield factor, i.c. the factor for translating the processed amount to the unprocessed amount, is not applied. In the IESTI calculations (and also chronic single-value calculations), calculations are done using the processed amounts, this is on the level of raw processed foods.
- Check the *restrict population to consumers or consumer days only (modelled-food)* option to compute the single value consumption statistics for each food based on the food consumers only. Note that checking this option will also affect the computed bodyweight, which is then computed by food based on the food-consumers only and can be different for each food.
- There is also an option to *ignore sampling weights* in the calculation.

- Check the *standardise consumption with body weight before calculation of single values or afterwards (with mean bodyweight)* option to compute the single value consumptions from the per bodyweight distribution. If unchecked, the per-person distribution will be used for computing the statistics. Note that although the results are reported per-day, the statistics are established by multiplying the statistics obtained from the per bw distribution by the bodyweight.

Single value consumptions settings

Calculation settings

Table 3.37: Calculation settings for module Single value consumptions.

Name	Type	Description
Risk type	*ExposureType*	The type of exposure considered in the assessment; acute (short term) or chronic (long-term).
Restrict population to consumers or consumer days only	Boolean	Specifies whether the population should be restricted to the individuals (chronic) or individual days (acute) that have non-ze consumption.
Ignore sampling weights	Boolean	If checked, individual sampling weights are not used (sampling weight = 1). If unchecked, the specified sampling weights are used.
Use standardised consumption distributions before calculation of single values	Boolean	Specifies whether single values are calculated on individual consumptions standardised with body weight and then multiplie by the mean body weight. Otherwise, single values are calculate on the original consumptions (per day). Note that both methods lead to different estimates for the single value.
Apply processing factors	Boolean	Specified in table ProcessingFactor. If checked, processing fact are applied. Concentrations in the consumed food may be different from concentrations in the modelled food in monitorin programs (typically raw food) due to processing, such as peelin washing, cooking etc. If unchecked, no processing information used. This is in most (though not all) cases a worst-case assumption
Restrict population to consumers or consumer days only (food-as-measured)	Boolean	Specifies whether the population should be restricted to the individuals (chronic) or individual days (acute) with consumptio containing any of the modelled foods.

Single value consumptions as data

Single value consumption data are the single value amounts of modelled foods (foods-as-measured) consumed in a population.

- *Single value consumptions data formats*

Calculation of single value consumptions

Single value consumptions are calculated as a percentile (p97.5 or p99) or mean of the modelled food consumption distribution. For an acute single value dietary exposure assessment, this is the individual day consumption distribution, for chronic single value dietary exposure assessment, the individual consumption distribution is used.

- *Single value consumptions calculation*

Inputs used: *Consumptions by modelled food*

Settings used

- *Calculation Settings*

3.3 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 *modelled-foods*. 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 *modelled-foods* is derived.

Active substance concentrations in *modelled-foods* 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*.

3.3.1 Concentration distributions

Concentration distributions describe substance concentrations on foods in the form of summary statistics.

This module has as primary entities: *Foods Substances*

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

Concentration distributions data formats

Concentration distributions describe substance concentrations on foods in the form of summary statistics. These distributions can be characterised by a mean and a dispersion factor, the standard deviation or, preferably, a percentile point e.g. p95.

Download empty dataset template: Zipped CSV Excel

Concentration distributions

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

Table 3.38: Table definition for Concentration distributions.

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	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	*ConcentrationUnit*	The unit of the limit value (default mg/kg).	Concentration-Unit, Unit	No

Accepted table names: ConcentrationDistributions, ConcentrationDistribution.

Concentration distributions as data

Concentration distributions describe substance concentrations on foods in the form of summary statistics.

- *Concentration distributions data formats*

3.3.2 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 Single value concentrations Concentration models Modelled foods*

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

Download empty dataset template: Zipped CSV Excel

Concentration limits

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

Table 3.39: Table definition for Concentration limits.

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	Yes
Value	Numeric	Residue limit value.	Value, Limit, MaximumResidueLimit, MaximumResidueLimits, 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
ConcentrationUnit	*ConcentrationUnit*	The unit of the limit value (default mg/kg).	ConcentrationUnit, Unit	No
ValueType	*ConcentrationLimitValueType*	Value type of the concentration value.	ValueType, ConcentrationLimitValueType, ConcentrationSingleValueType	No
Reference	AlphaNumeric (200)	Reference to the source from which this concentration single value is obtained.	Reference, References, Source, Sources	No

Accepted table names: ResidueLimits, ResidueLimit, MaximumResidueLimits, MaximumResidueLimit, MRLs, MRL.

Concentration limits as data

Maximum Residue Limits (MRL) are provided as data.

- *Concentration limits data formats*

3.3.3 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: *High exposure food-substance combinations 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 data from *Total Diet Studies* does not include variability information. In order to include variability in the concentration models created for TDS samples, the variance of the TDS concentrations may be *estimated using variances of the foods making up the composite TDS food*.

Concentration model types

Empirical model

Data points are sampled at random from the available set. Censored values (non-detect, i.c. < LOD, non-quantifications, i.c. < LOQ) are handled by imputation. If *occurrence patterns* are used, a proportion p_0/p_{ND} of censored values is set as 0. See also *concentration models*.

Censored 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 censored value with probability $p_{ND} = 1 - p$ or a value sampled from the fitted lognormal distribution with probability p. Censored values (non-detects or non-quantifications) are handled by imputation. If occurrence patterns are used, a proportion p_0/p_{ND} of censored values is set as 0. Minimum requirements: at least two positive concentration values. See also *concentration models*.

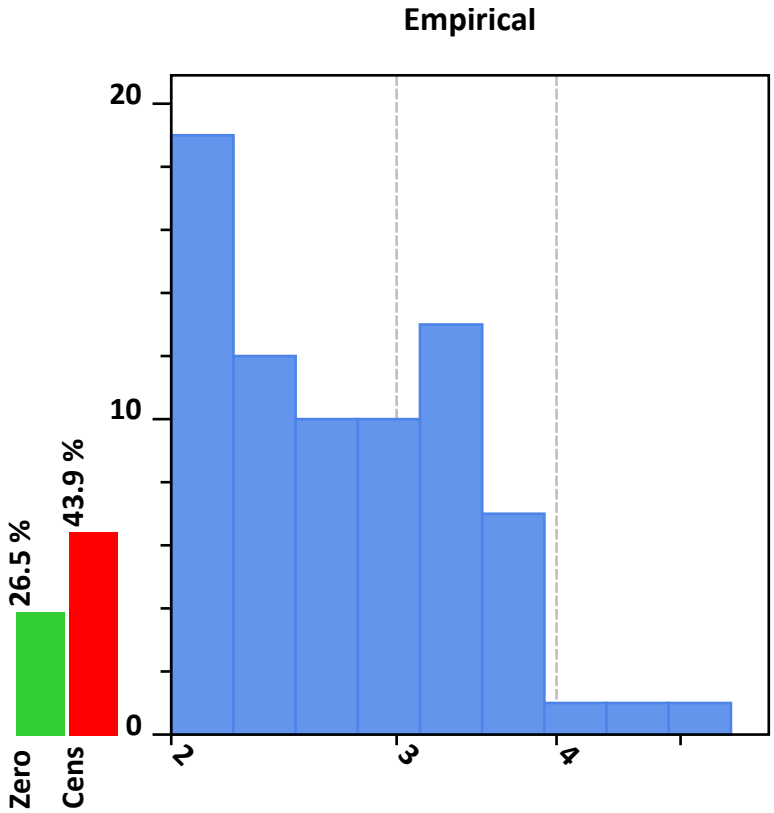


Figure 3.16: Empirical distribution

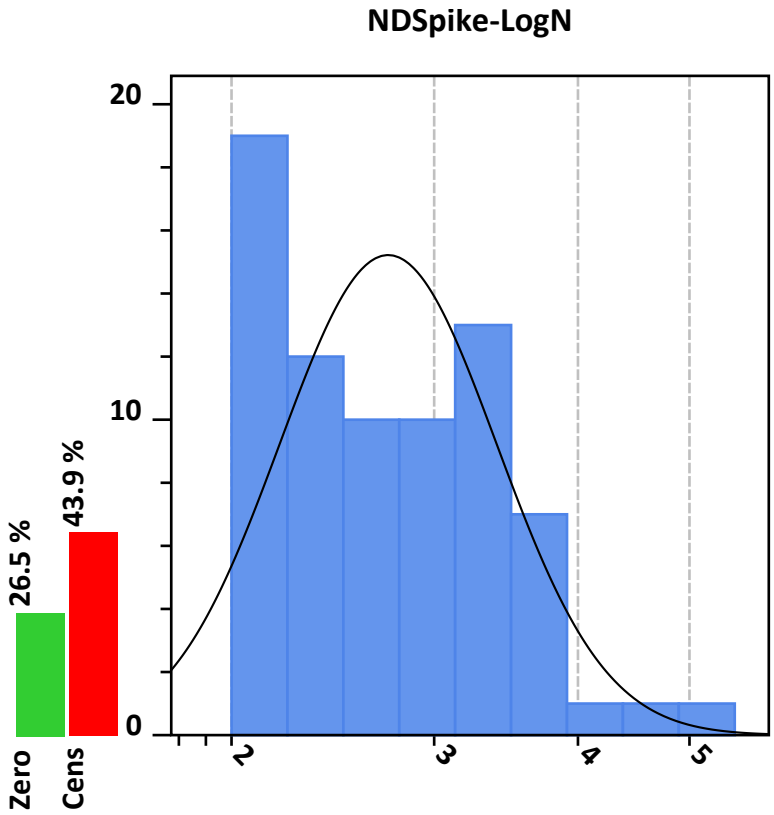


Figure 3.17: Censored Spike Lognormal distribution

Censored-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 censored with probability $p_{ND} = 1 - p$ or a value sampled from the fitted lognormal distribution with probability p. Censored values (non-detects or non-quantifications) are handled by imputation. If occurrence patterns are used, a proportion p_0/p_{ND} of censored values is set as 0. Minimum requirements: at least two positive concentration values, all censored values must have one LOR value. See also *concentration models*.

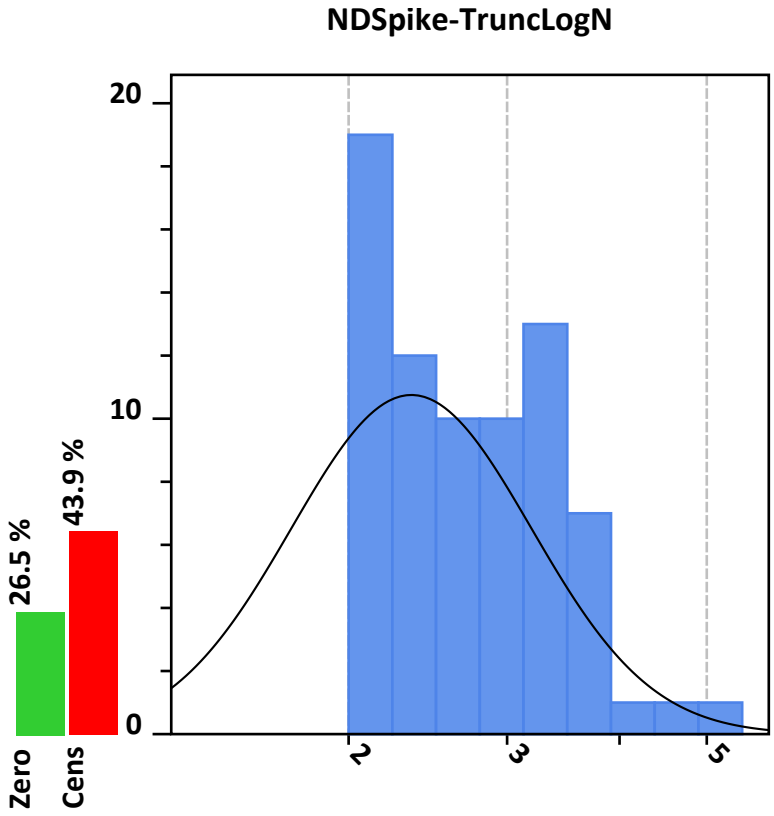


Figure 3.18: Censored Spike Truncated Lognormal distribution

Censored Lognormal model

A censored lognormal model, with LOR as the censoring limit, is fitted to the data, both positives and censored values. 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 censored values 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 *concentration models*.

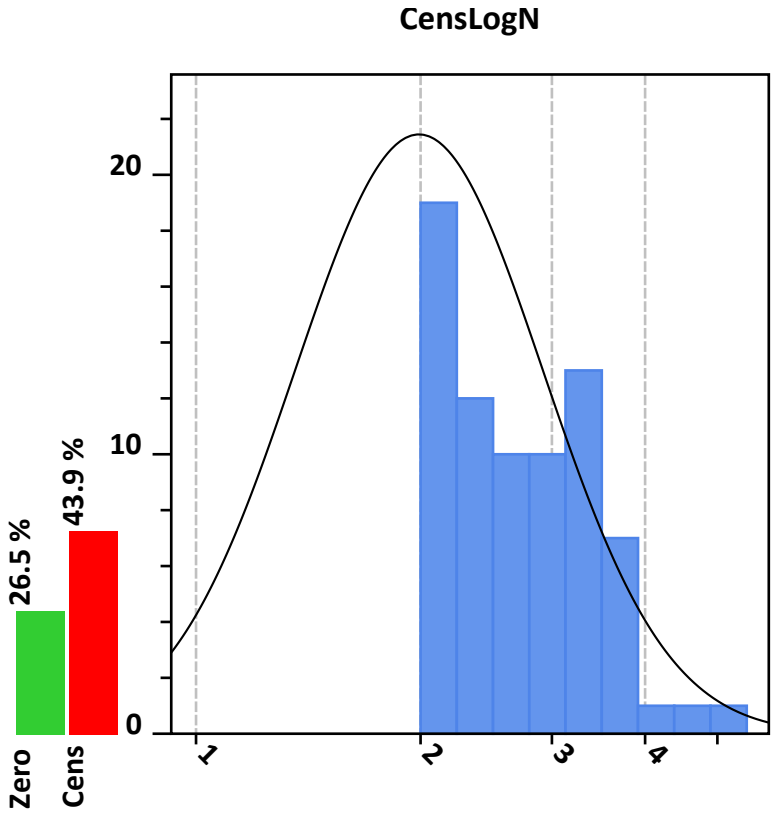


Figure 3.19: Censored Lognormal distribution

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 (censored values 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-substance combination (which directly specify p_0, therefore it should not be estimated from the data). See also *concentration models*.

Censored spike MRL model

This model simply takes values specified in an input table as Maximum Residue Level (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.

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 Study* assessments. 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*.

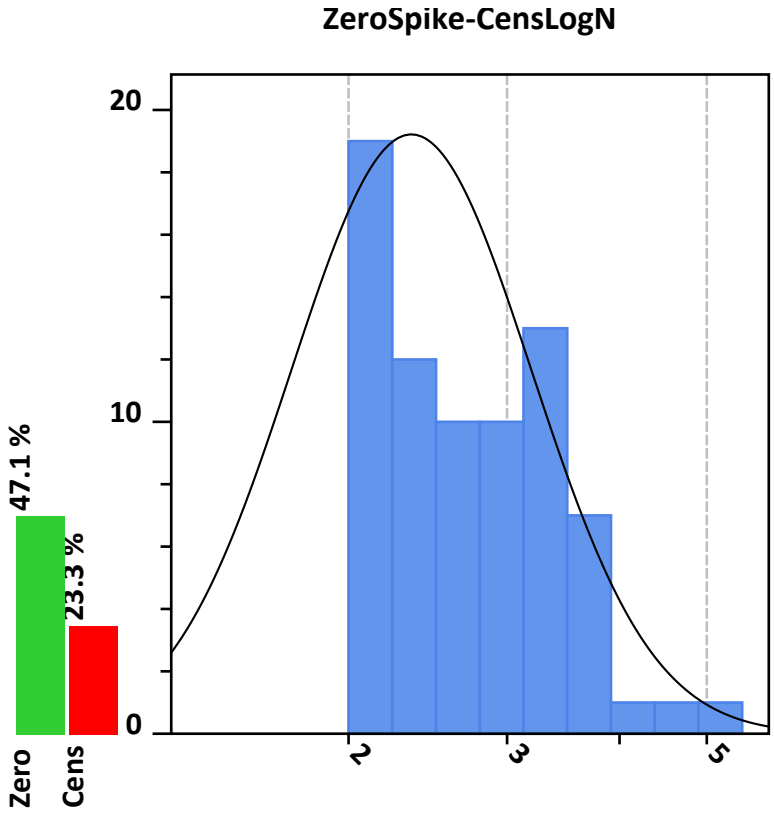


Figure 3.20: Zero Spike Censored Lognormal distribution

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.

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.

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.

Mixture censored 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.

Mixture censored 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.

Imputation of non-detect measurements

A complication in concentration modelling occurs when concentration measurements are reported as below a certain limit. Different names may be used for such a limit, e.g. **limit of detection** (LOD), **limit of quantification** (LOQ) or the more general term **limit of reporting** (LOR). Results that are only reported to be below the LOD (non-detects) or below the LOQ (non-quantifications) are generally referred to as censored values or non-reports. Censored values are a very common phenomenon for some classes of substances like pesticides. When modelling the substance concentrations, they can be handled by incorporating them in a parametric model, or by replacing them with a given value (**imputation**). For imputation, different models are available:

- **Replace by zero**: All left censored measurement values are replaced by zero. This is an optimistic modelling choice.
- **Replace by factor x LOR**: Measurements reported below LOD or LOQ are replaced by a factor times LOQ. When the LOQ is unspecified, then the LOD is used. When the factor is equal to one, this method can be assumed to be a conservative approach.
- **Replace LOD by factor x LOD and LOQ by factor x (LOQ - LOD)**: Measurements reported below LOD are replaced by a factor times LOD and measurements reported below LOQ are replaced by a factor times LOQ-LOD (i.e., a value between LOD and LOQ).

An additional option of the imputation methods that replace the left censored values with some positive value is to *use occurrence frequencies for imputation*. When this option is used, a part of the left-censored values are replaced by a positve value and another part is replaced by zero, based on the *occurrence frequencies* of the modelled foods and substances. Another option is to restrict imputation with positive values to only the authorised substances.

In Figure 3.21 to Figure 3.24, the various scenarios are displayed. Two substances, Fenamidine and Hexythiazox are indicated with a brown box, these substances are authorised.

No imputation

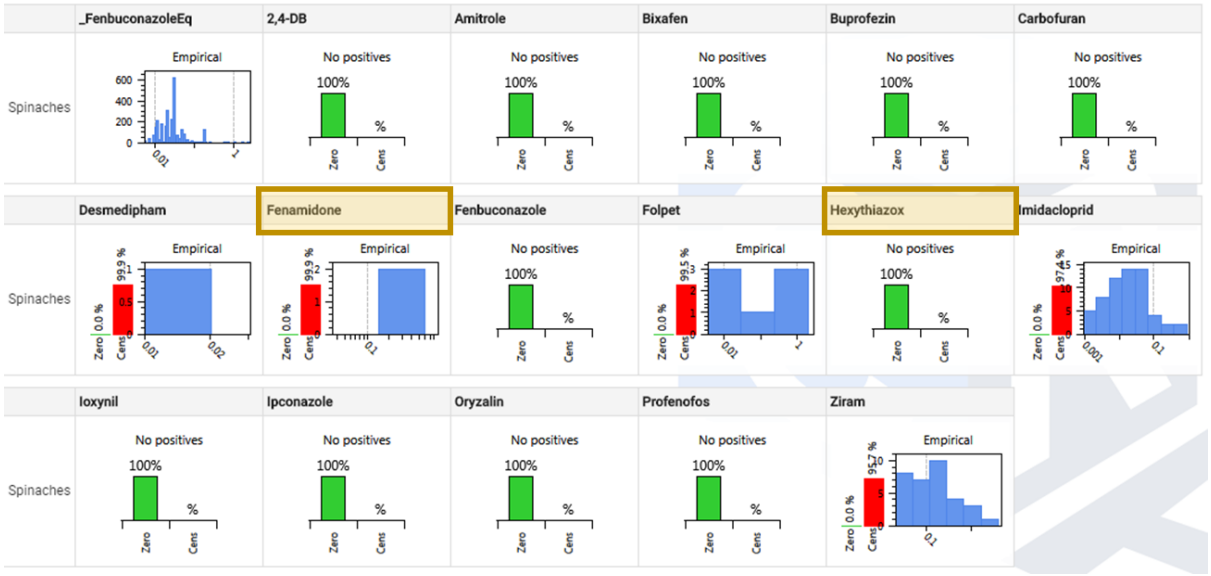


Figure 3.21: Tier 1: Censored values are replaced by zero. For Fenamidine and Hexythiazox (brown boxes) authorized use is assumed.

Impute all censored values by factor times LOQ

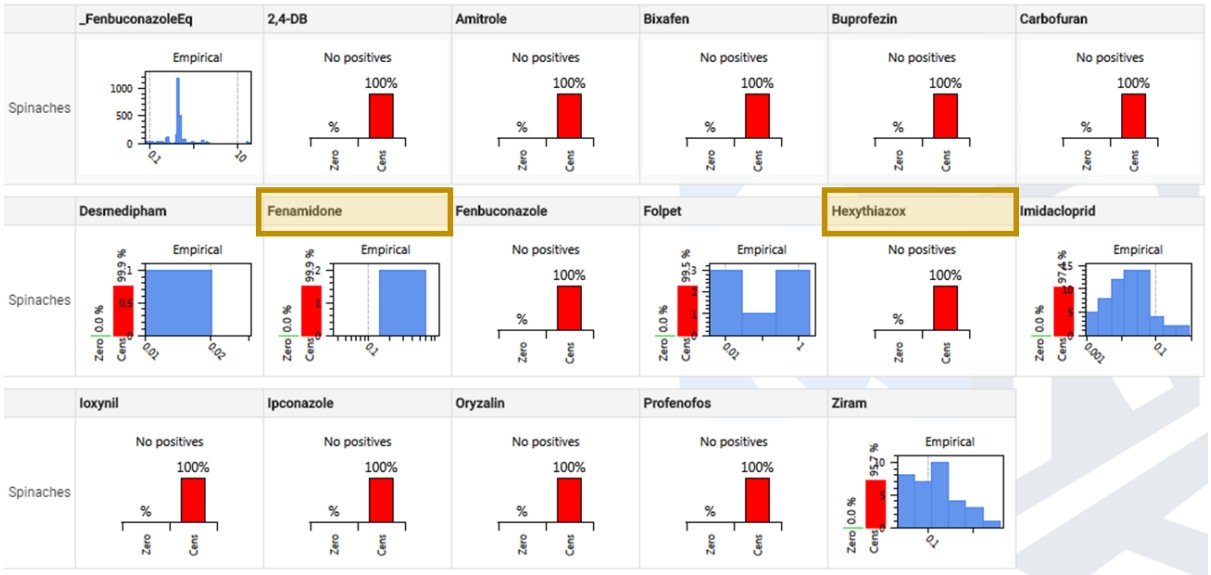


Figure 3.22: All censored values are replaced by a **constant** x LOR. For Fenamidine and Hexythiazox (brown boxes) authorized use is assumed.

Impute censored values based on authorized uses

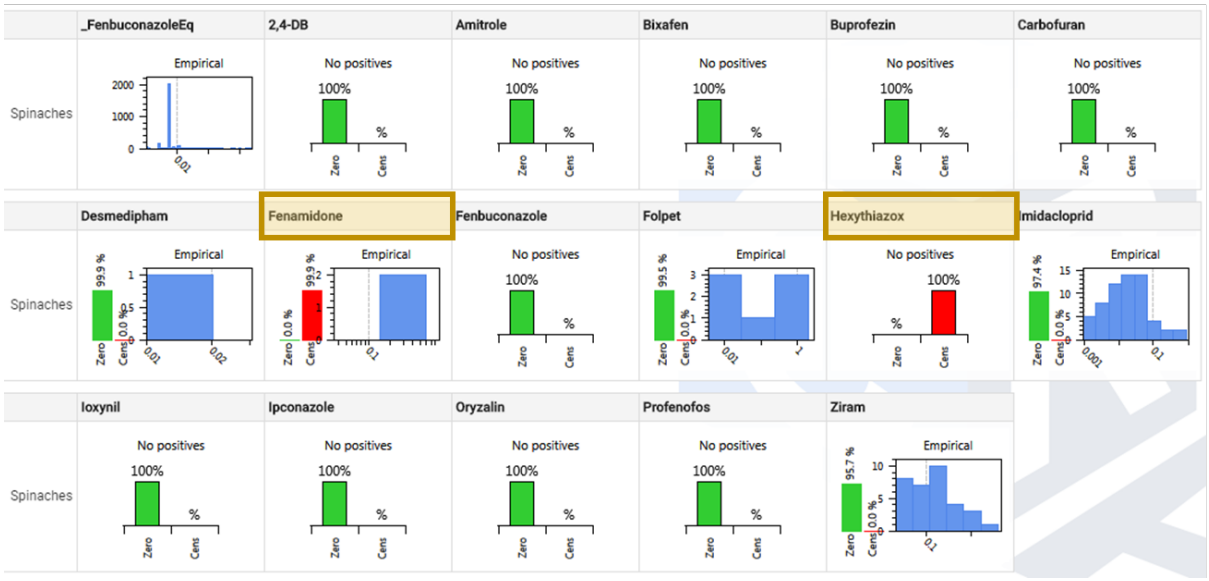


Figure 3.23: Censored values are replaced by a **constant** x LOR for authorized uses. For Fenamidine and Hexythiazox (brown boxes) authorized use is assumed.

No imputation except for authorized uses

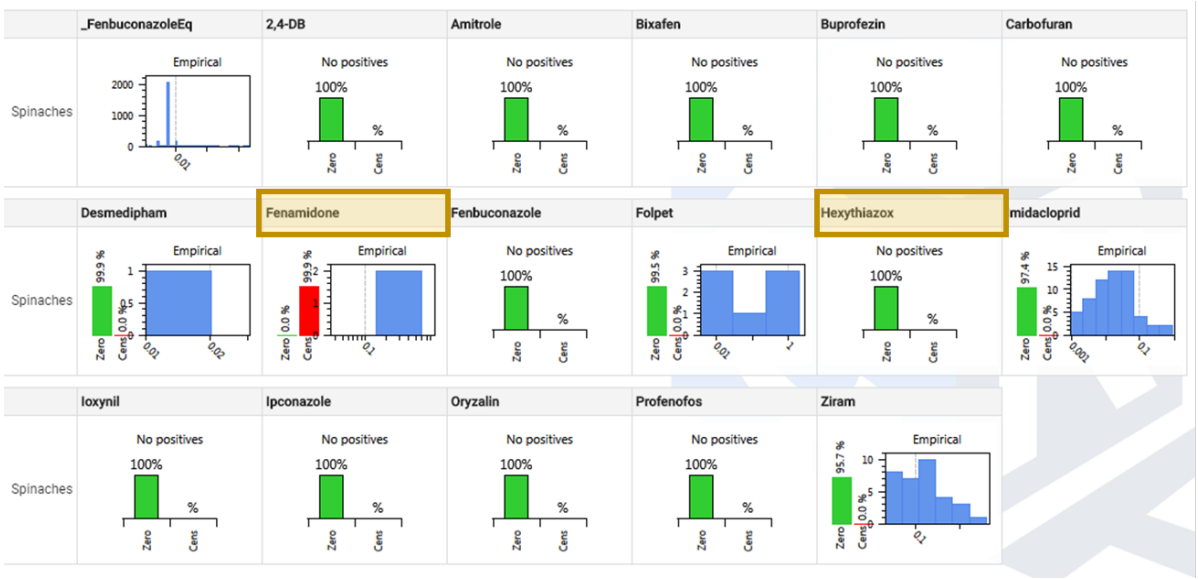


Figure 3.24: Tier 1: Censored values are not replaced except for authorized uses (replaced by a **constant** x LOR). For Fenamidine and Hexythiazox (brown boxes) authorized use is assumed.

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 (defined by the *TDS compositions*). 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 $i = 1...k, w_i$ 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. $CV_i = \sqrt{\exp(var(c + i) - 1)}$

Concentration models settings

Calculation settings

Table 3.40: Calculation settings for module Concentration models.

Name	Type	Description
Concentration model tier	*ConcentrationModelChoice*	Custom model, or set according to EFSA Guidance 2012. Note you may need to set the tier separately in sub-modules.
Default concentration model	*ConcentrationModelType*	The concentration model type that will be used as default for al food/substance combinations. If this model type cannot be fitte e.g., due to a lack of data, a simpler model will be chosen automatically as a fall-back.
Include MRL fallback model	Boolean	Use the MRL as fallback model in case the occurrence data is insufficient for other concentration modelling options.
Restrict LOR imputation to authorised uses	Boolean	Specifies whether imputation of factor x LOR should be limited authorised uses only.
Censored values replacement	*NonDetectsHandlingMethod*	How to replace censored values (when not co-modelled, as in censored models).
Factor f (f x LOR or f x LOD or LOD + f x (LOQ - LOD)	Numeric	Replace censored values by Limit of reporting (LOR), Non-detects (LOD) or Non-quantifications (LOQ) times this factor. Constant (f), e.g. 0.5.
MRL Factor (f x MRL)	Numeric	Use f x MRL as concentration estimate of the MRL models.
Sample based	Boolean	Include co-occurrence of substances in samples in simulations. checked, substance residue concentrations are sampled using th correlations between values on the same sample. If unchecked, any correlation between substances is ignored, substance residu concentrations are sampled ignoring the correlations between values on the same sample.
Impute missing values from available values (if unchecked, missing values are imputed with 0)	Boolean	If checked, in procedure of EFSA Guidance 2012, Appendix 1 impute missing values using substance based concentration models. If unchecked, missing values are imputed by 0.
Correlate imputed values with sample potency	Boolean	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 frequencies for imputation	Boolean	Use of occurrence frequencies (e.g., agricultural use frequencie is relevant for imputation of censored values in the concentratio data. Part of the observed censored values and missing values m be imputed with zero when the occurrence frequency is smaller than 100%. If checked, occurrence frequencies are expected as input of this action, otherwise 100% potential presence is assumed for all substances on all foods.

Uncertainty settings

Table 3.41: Uncertainty settings for module Concentration models.

Name	Type	Description
Parametric uncertainty	Boolean	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* (EFSA (2012)).
- 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 (EC (2018)).

Overview

Table 3.42: Tier overview for module Concentration models.

Name	EFSA 2012 Optimistic	EFSA 2012 Pessimistic - Acute	EFSA 2012 Pessimistic - Chronic	EC 2018 Tier 1	EC 2018 Tier 2
Default concentration model	Empirical	NonDetectSpikeLogNormal	NonDetectSpikeLogNormal	Empirical	Empirical
Include MRL fallback model	false	true	true	false	false
Censored values replacement	ReplaceByZero	ReplaceByLOR	ReplaceByLOR	ReplaceByLOR	ReplaceByLOR
Sample based	true	true	true	true	true
Impute missing values from available values (if unchecked, missing values are imputed with 0)	false	true	true	true	true
Correlate imputed values with sample potency	false	true	true	true	false
Use occurrence frequencies for imputation	false			true	true
Parametric uncertainty	false	true	false	false	false
Restrict LOR imputation to authorised uses		false	false	false	false
Factor f (f x LOR or f x LOD or LOD + f x (LOQ - LOD)		1	1	0.5	0.5
MRL Factor (f x MRL)		1	1		
Apply occurrence pattern percentages				false	true
Substance conversion method				UseMostToxic	DrawRandom
Retain all allocated substances after active substance allocation				true	true
Account for substance authorisations in substance conversions				false	true
Fix duplicate substance allocation inconsistencies				false	false
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				true	true

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. Censored values and missing values are replaced by zero.

Table 3.43: Tier definition for EFSA 2012 Optimistic.

Name	Setting
Default concentration model	Empirical
Include MRL fallback model	false
Censored values replacement	ReplaceByZero
Sample based	true
Impute missing values from available values (if unchecked, missing values are imputed with 0)	false
Correlate imputed values with sample potency	false
Use occurrence frequencies for imputation	false
Parametric uncertainty	false

EFSA 2012 Pessimistic - Acute

Concentration model settings for acute pessimistic dietary exposure assessments according to the EFSA Guidance 2012. A censored value spike lognormal model is fitted to the positive residue values and censored values 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 3.44: Tier definition for EFSA 2012 Pessimistic - Acute.

Name	Setting
Default concentration model	NonDetectSpikeLogNormal
Include MRL fallback model	true
Restrict LOR imputation to authorised uses	false
Censored values replacement	ReplaceByLOR
Factor f (f x LOR or f x LOD or LOD + f x (LOQ - LOD)	1
MRL Factor (f x MRL)	1
Sample based	true
Impute missing values from available values (if unchecked, missing values are imputed with 0)	true
Correlate imputed values with sample potency	true
Parametric uncertainty	true

EFSA 2012 Pessimistic - Chronic

Concentration model settings for acute pessimistic dietary exposure assessments according to the EFSA Guidance 2012. A censored value spike lognormal model is fitted to the positive residue values and censored values 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 3.45: Tier definition for EFSA 2012 Pessimistic - Chronic.

Name	Setting
Default concentration model	NonDetectSpikeLogNormal
Include MRL fallback model	true
Restrict LOR imputation to authorised uses	false
Censored values replacement	ReplaceByLOR
Factor f (f x LOR or f x LOD or LOD + f x (LOQ - LOD)	1
MRL Factor (f x MRL)	1
Sample based	true
Impute missing values from available values (if unchecked, missing values are imputed with 0)	true
Correlate imputed values with sample potency	true
Parametric uncertainty	false

EC 2018 Tier 1

Table 3.46: Tier definition for EC 2018 Tier 1.

Name	Setting	From input tier	In module
Default concentration model	Empirical		
Include MRL fallback model	false		
Restrict LOR imputation to authorised uses	false		
Censored values replacement	Replace-ByLOR		
Factor f (f x LOR or f x LOD or LOD + f x (LOQ - LOD)	0.5		
Sample based	true		
Impute missing values from available values (if unchecked, missing values are imputed with 0)	true		
Correlate imputed values with sample potency	true		
Use occurrence frequencies for imputation	true		
Parametric uncertainty	false		
Apply occurrence pattern percentages	false	*EC 2018 Tier 1*	*Occurrence patterns*
Substance conversion method	UseMostToxic	*EC 2018 Tier 1*	*Concentrations*
Retain all allocated substances after active substance allocation	true	*EC 2018 Tier 1*	*Concentrations*
Account for substance authorisations in substance conversions	false	*EC 2018 Tier 1*	*Concentrations*
Fix duplicate substance allocation inconsistencies	false	*EC 2018 Tier 1*	*Concentrations*
Use extrapolation rules	true	*EC 2018 Tier 1*	*Concentrations*
Threshold for extrapolation	10	*EC 2018 Tier 1*	*Concentrations*
Restrict extrapolations to equal MRLs	true	*EC 2018 Tier 1*	*Concentrations*
Restrict extrapolations to authorised uses	true	*EC 2018 Tier 1*	*Concentrations*
Impute water concentrations	true	*EC 2018 Tier 1*	*Concentrations*
Water concentration value (µg/kg)	0.1	*EC 2018 Tier 1*	*Concentrations*
Restrict water imputation to the five most toxic substances	true	*EC 2018 Tier 1*	*Concentrations*
Restrict water imputation to authorised uses	false	*EC 2018 Tier 1*	*Concentrations*
Restrict water imputation to approved substances	false	*EC 2018 Tier 1*	*Concentrations*

EC 2018 Tier 2

Table 3.47: Tier definition for EC 2018 Tier 2.

Name	Setting	From input tier	In module
Default concentration model	Empirical		
Include MRL fallback model	false		
Restrict LOR imputation to authorised uses	false		
Censored values replacement	Replace-ByLOR		
Factor f (f x LOR or f x LOD or LOD + f x (LOQ - LOD)	0.5		
Sample based	true		
Impute missing values from available values (if unchecked, missing values are imputed with 0)	true		
Correlate imputed values with sample potency	false		
Use occurrence frequencies for imputation	true		
Parametric uncertainty	false		
Apply occurrence pattern percentages	true	*EC 2018 Tier 2*	*Occur-rence patterns*
Scale up use frequency to 100%	true	*EC 2018 Tier 2*	*Occur-rence patterns*
Restrict use percentage up-scaling to authorised uses	true	*EC 2018 Tier 2*	*Occur-rence patterns*
Substance conversion method	DrawRan-dom	*EC 2018 Tier 2*	*Concen-trations*
Retain all allocated substances after active substance allocation	true	*EC 2018 Tier 2*	*Concen-trations*
Account for substance authorisations in substance conversions	true	*EC 2018 Tier 2*	*Concen-trations*
Fix duplicate substance allocation inconsistencies	false	*EC 2018 Tier 2*	*Concen-trations*
Use extrapolation rules	true	*EC 2018 Tier 2*	*Concen-trations*
Threshold for extrapolation	10	*EC 2018 Tier 2*	*Concen-trations*
Restrict extrapolations to equal MRLs	true	*EC 2018 Tier 2*	*Concen-trations*
Restrict extrapolations to authorised uses	true	*EC 2018 Tier 2*	*Concen-trations*
Impute water concentrations	true	*EC 2018 Tier 2*	*Concen-trations*
Water concentration value (µg/kg)	0.05	*EC 2018 Tier 2*	*Concen-trations*
Restrict water imputation to the five most toxic substances	true	*EC 2018 Tier 2*	*Concen-trations*
Restrict water imputation to authorised uses	false	*EC 2018 Tier 2*	*Concen-trations*
Restrict water imputation to approved substances	false	*EC 2018 Tier 2*	*Concen-trations*

EFSA 2012 Pessimistic

Note: This tier is deprecated and has been replaced by separate acute/chronic tiers.

Concentration model settings for pessimistic dietary exposure assessments according to the EFSA Guidance 2012. A censored value spike lognormal model is fitted to the positive residue values and censored values 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 3.48: Tier definition for EFSA 2012 Pessimistic.

Name	Setting
Default concentration model	NonDetectSpikeLogNormal
Include MRL fallback model	true
Restrict LOR imputation to authorised uses	false
Censored values replacement	ReplaceByLOR
Factor f (f x LOR or f x LOD or LOD + f x (LOQ - LOD)	1
MRL Factor (f x MRL)	1
Sample based	true
Impute missing values from available values (if unchecked, missing values are imputed with 0)	true
Correlate imputed values with sample potency	true
Parametric uncertainty	true

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 necessary steps of missing value imputation and imputation of censored values).

When parametric uncertainty is preferred over empirical 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 Active substances Modelled foods Substance authorisations Occurrence frequencies Relative potency factors Concentration distributions Total diet study sample compositions*

Settings used

- *Calculation Settings*

3.3.4 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 Populations*

Output of this module is used by: *Single value concentrations Occurrence patterns Concentration models Modelled foods*

Concentrations data formats

Concentrations data are analytical measurements of chemical substances occurring in food samples. The following data formats are supported/available for providing concentration data in MCRA:

1. **Relational concentration data format:** Concentration data is provided using a number of relational tables describing (food) samples (e.g., sampling date and location), the analytical methods with their properties for substances (e.g., LOQ and LOD), sample analyses describing the analyses performed on the samples (e.g., analysis date and the used analytical method), and concentration measurements (e.g., the measurement type and value). This relational table structure is also used internally by MCRA.
2. **EFSA Standard Sample Description (SSD) format:** Concentration data according to the EFSA Standard Sample Description (SSD) format. During upload, SSD data are converted automatically to the relational data structure used internally by MCRA.
3. **Tabulated data scheme:** This is a simplified data format where samples and analytical methods are not explicitly specified. Tabulated concentration data are converted automatically to the MCRA scheme.

Relational concentration data format

The relational data format is the data format that is used internally in MCRA.

Download empty dataset template: Zipped CSV Excel

Analytical methods

The analytical methods used for analysing 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 3.49: Table definition for Analytical methods.

Name	Type	Description	Aliases	Required
idAnalytical-Method	AlphaNumeric (50)	The code for the method of analysis.	idAnalytical-Method, Analytical-MethodId, Analytical-MethodName, Id	Yes
Name	AlphaNumeric (100)	Name of the analytical method.	Name	No
Description	AlphaNumeric (255)	Additional description of method of analysis.	Description	No

Accepted table names: AnalyticalMethod, AnalyticalMethods.

Analytical method properties for substances

This table describes the substances analysed by an analytical method. For each substance analysed by an analytical method a record should be included that describes the unit of measurement and the reporting limits (LOQ/LOD).

Table 3.50: Table definition for Analytical method properties for substances.

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	Yes
LOD	Numeric	The limit of detection (LOD) is the lowest concentration of an substance in a sample that can be consistently detected.	LOD	No
LOQ	Numeric	The limit of quantification (LOQ) is the lowest concentration of a substance that can be quantified. The LOQ should be larger than the LOD.	LOQ, LOR	No
Concentration-Unit	*ConcentrationUnit*	The unit used for reporting the LOD, LOQ, and the substance concentrations. When not specified, then a default unit of mg/kg is assumed.	Concentration-Unit, Units, Unit	No

Accepted table names: AnalyticalMethodSubstances, AnalyticalMethodSubstance, AnalyticalMethodCompounds, AnalyticalMethodCompound.

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 3.51: Table definition for Food samples.

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
Region	AlphaNumeric (50)	The area or region within the sampling location.	Region, Area, Sampling-Region, SamplingArea	No
DateSampling	DateTime	The date of sampling.	DateSampling, SamplingDate	No
Production-Method	AlphaNumeric (50)	Additional information on the type/method of production of the sampled food.	Production-Method, ProductionType	No
Name	AlphaNumeric (100)	Name of the food sample.	Name	No
Description	AlphaNumeric (200)	Additional description of the food sample.	Description	No

Accepted table names: FoodSamples, FoodSample, Samples, Sample, PrimarySample, PrimarySamples.

Sample properties

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

Table 3.52: Table definition for Sample properties.

Name	Type	Description	Aliases	Required
Name	AlphaNumeric (50)	The name of the property.	Id	Yes
Description	AlphaNumeric (200)	Additional description of the sample property.	Description	No

Accepted table names: SampleProperties, SampleProperty.

Sample property values

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

Table 3.53: Table definition for Sample property values.

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

Accepted table names: SamplePropertyValues, SamplePropertyValue.

Sample Analyses

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 3.54: Table definition for Sample Analyses.

Name	Type	Description	Aliases	Required
idSample-Analysis	AlphaNumeric (50)	The identification number of the analysed sample.	id, idSample-Analysis, SampleAnalysis, idAnalysis-Sample, AnalysisSample-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
Name	AlphaNumeric (100)	Name of the analysis sample.	Name	No
Description	AlphaNumeric (200)	Additional description of the the analysis sample.	Description	No

Accepted table names: AnalysisSamples, AnalysisSample, SampleAnalysis, SampleAnalyses.

Sample concentrations

This table contains substance concentration values specified in the unit defined by the analytical method. The analytical method contains the list of all substances which have been analyzed in the analysis sample. This ConcentrationsPerSample table contains the analysis results where substances with positive concentrations are included. Censored values (i.e. results 'less than LOQ or LOD') are reported as follows: 1) Substances for which an LOD (Limit of detection) is reported are included with ResType 'LOD', without a concentration value. 2) Substances for which only an LOQ (Limit of quantification) is reported are EXCLUDED, because the LOQ value from the analytical method substances table (AnalyticalMethodCompounds) is used by default. Explicitly missing concentration values are specified with ResType 'MV' (obligatory).

Table 3.55: Table definition for Sample concentrations.

Name	Type	Description	Aliases	Required
idSample-Analysis	AlphaNumeric (50)	The identification number of the analysed sample.	idSample-Analysis, SampleAnalysis, idAnalysis-Sample, AnalysisSample-Id	Yes
idSubstance	AlphaNumeric (50)	The substance code.	idSubstance, SubstanceId, Substance	Yes
Concentration	Numeric	The measured concentration.	Concentration	No
ResType	*ResType*	The type of residue. Should be VAL (= default), LOQ, LOD or MV.	ResType	No

Accepted table names: SampleConcentrations, ConcentrationsPerSample, ConcentrationPerSample.

SSD concentration data format

The Standard Sample Description (SSD) concentration is the standard data format proposed by EFSA. Optionally, additional sample properties may be specified of fields that are not part of the SSD format (e.g., season of sampling). For this, the sample properties table and the sample property values table can be used.

Download empty dataset template: Zipped CSV Excel

SSD concentrations

MCRA uses the concept of samples analysed by analytical methods, where the analytical method contains the substances analysed and the LOQs and LODs for these substances. However, the SSD data do not provide information on the analytical methods at this level of detail. Therefore, the provided SSD sample records are used to generate analytical methods which are linked to the samples. All SSD records with the same labSampCode and labSubSampCode compose one MCRA analysis sample. All SSD samples that contain the same substance, LOQ/LOD values and resUnit combinations are linked to the corresponding generated analytical method. If both LOQ and LOD are provided, the LOQ is used as the LOR of the generated analytical method. It is highly recommended to supply LOQ/LOD values, even for positive measurements, because this reduces the number of gererated analytical methods.

Table 3.56: Table definition for SSD concentrations.

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	No
sampArea	AlphaNumeric (5)	Area where the sample was collected.	sampArea	No
prodCode	AlphaNumeric (50)	Code identifying the modelled food. Should be equal to a code idFood in the Foods table.	prodCode	Yes
prodProdMeth	AlphaNumeric (50)	Code providing additional information on the type of production for the food under analysis.	prodProdMeth	No
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 (50)	Code identifying the substance.	paramCode	Yes
resUnit	*ConcentrationUnit*	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	*ResType*	Type of residue data. Should be VAL, LOQ, LOD or MV.	resType	Yes

Accepted table names: ConcentrationsSSD, SSDConcentrations.

Sample properties

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

Table 3.57: Table definition for Sample properties.

Name	Type	Description	Aliases	Required
Name	AlphaNumeric (50)	The name of the property.	Id	Yes
Description	AlphaNumeric (200)	Additional description of the sample property.	Description	No

Accepted table names: SampleProperties, SampleProperty.

Sample property values

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

Table 3.58: Table definition for Sample property values.

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

Accepted table names: SamplePropertyValues, SamplePropertyValue.

Tabulated concentration data format

The tabulated concentration data format is an old data format for entering concentration data.

Download empty dataset template: Zipped CSV Excel

Tabulated concentrations

In the tabulated concentrations data table, each record represents one or multiple samples, and each sample contains a concentration value for a food/substance combination. Censored values (i.e. concentrations less than LOR) are specified as negative values, i.e. 'less than LOR' should be specified as minus the LOR value. 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 tabulated data do not provide this information explicitly. Samples are reconstructed from the tabulated records using the NumberOfSamples field to create that number of single substance samples. Analytical methods are reconstructed from the data, with each analytical method having only one analysed substance with a LOR and concentration unit. When a negative concentration value is given (i.e., it is a censored measurement), this value is recorded as the LOR (negated). All censored measurements of the same substance with the same LOR and concentration unit are linked to the same analytical method. When a positive concentration value is given, this value is recorded as the measured concentration of the sample. All positive measurements of the same substance are linked to the same analytical method that has an artificial LOR that is smaller

than the lowest positive concentration. When a concentration of 0 (zero) is given, the measurement is considered to be a censored measurement and the LOR is set to the default value 1E-08.

Table 3.59: Table definition for Tabulated concentrations.

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	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	*ConcentrationUnit*	The unit of the specified concentrations/LORs (default mg/kg).	Concentration-Unit, Unit	No

Accepted table names: ConcentrationTabulated, ConcentrationValues, TabulatedConcentrations, TabulatedConcentration.

Concentrations calculation

Occasionally, concentrations of substances measured in food samples are exceeding a specified concentration limit e.g. the *Maximum Residue Limits* (MRL). An MRL is the highest level of a substance that is legally tolerated in or on food or feed when substances are applied correctly. *Filter samples* exceeding the concentration limits filter out all samples where one of the substances measured is exceeding the *MRL*.

Substance conversions rules 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 (provided as data) may be applicable, for example, when a measured substance represents multiple substances and these measurements 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. *points of departure* or *relative potency factors* might be needed when the substance conversion should select the most toxic candidate in case a measured substance translates to multiple active substances.

If there are only a few measurements in the concentration data, then *extrapolation of concentration data* may be desired. In that 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 concentration data, but it may be desirable to include them 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.

In some scenarios it may be desired to perform a prospective analysis in which anticipated (or foreground) *focal commodity concentration data* for a particular focal commodity food (and substance) is added to, or replaces part of the background concentration data that is used for the null-scenario. The concentrations module offers various options to perform such *focal commodity scenario analyses*.

It is also possible to *filter (or subset) samples by specific sample properties* (e.g., year, location). This can be done by checking the option to *filter samples by specific property values (subset selection)*.

Filter samples exceeding the concentration limits

If the option **Filter samples exceeding the concentration limits** is checked, all samples with one or more substance concentrations exceeding the *MRL* are filtered out. Then a **concentration limit filter exceedance factor** ($factor$) is specified, which filters out samples with at least one substance concentration higher than $factor \cdot MRL$.

If the option **Filter samples exceeding the concentration limits** remains unchecked all samples are retained in the analysis.

Substance conversion

When concentration data at the level of measured substances have to be converted into concentration data at the level of *active substances* (or perhaps also inactive substances), then *substance conversion rules* are specified to provide the rules, the so-called residue definitions. This section describes the basic substance conversion, and then the refinements using available *substance authorisations*.

For each measured substance in the concentration data, there may be zero or more conversion rules, each one linking to an active or inactive substance. Each rule is represented by a record in the substance conversion data source. Substance conversion rules may specify a link to an exclusive substance or not. For an exclusive conversion it is assumed that only one substance is present in the sample, therefore the measured substance is considered to be just one of the linked substances. Another possibility is that measured substances link to one or more exclusive substances, plus one (non-exclusive) substance that is considered a metabolite of the other exclusive substances. 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 exactly one rule is marked as exclusive and the others are marked as not exclusive (case 2). If this does not apply for any set of rules linked to a measured substance, data are regarded as erroneous.

Four methods for substance conversion are implemented:

1. Allocate most potent (EC 2018 Tier 1): 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 LOQ/LOD is multiplied by the molecular weight correction factor. See *EC2018 Tier 1*.

2. Random allocation (EC 2018 Tier 2): one of the conversion rules is drawn randomly (with equal probability), including the rules of both active and other substances. Then, the rule is used as follows to generate active substance concentrations:

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

All assigned concentrations are multiplied by the molecular weight correction factor. All unselected candidate substances are assigned a zero concentration. See *EC2018 Tier 2*.

3. Nominal estimate: the substances specified through the conversion rules are allocated with 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 allocated 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 allocated 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.

4. Allocate all: the concentration of a measured substance is allocated to each active substance associated with the measured substance as if it were the most potent substance. I.e., the same measured substance is allocated to all associated active substances simultaneously. This method is not sensible when using it in a cumulative assessment, but it is of use in substance screening assessments, where in a combined analysis of multiple substances all active substances are considered independently.

Use of substance authorisations in substance conversion

When *substance authorisations* are available, 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:

1. Allocate 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.

Table 3.60: Most toxic allocation of a measured substance to active substances based on RPFs and authorised use

Active substance	RPF	Authorised use	Active substance	RPF	Authorised use	Measured substance allocated to
AS1	1	true	AS2	2	true	AS1
AS1	1	false	AS2	2	true	AS2
AS1	1	true	AS2	2	false	AS1
AS1	1	false	AS2	2	false	AS1

In Table 3.60, results of most toxic allocation.

2. Random allocation: 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.

Table 3.61: Random allocation of a measured substance to active substances based on authorised use

Active substance	RPF	Authorised use	Active substance	RPF	Authorised use	Measured substance allocated to
AS1	1	true	AS2	2	true	AS1 or AS2
AS1	1	false	AS2	2	true	AS2
AS1	1	true	AS2	2	false	AS1
AS1	1	false	AS2	2	false	AS1 or AS2

In Table 3.61, results of random allocation.

3. Nominal estimate: 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.

4. Allocate all: For this method, the same rules apply as for *allocate most potent*. The set of candidate active substances that are to be allocated is reduced to only the substances with authorised uses. Hence, a substance is not allocated when it is not authorised and there is at least one other candidate active substance that is authorised. However, if none of the candidate active substances is authorised, then the most potent of the unauthorised substances is selected for active substance allocation.

Multiple allocations of the same active substance in one sample

In some datasets substance conversion can cause the same active substance to be allocated multiple times in one sample. For example, when an active substance is measured directly, but also a measurement is recorded for a (measured) substance that converts to the active substance. By default, MCRA does not accept such cases, because often it is associated with errors in the data. Therefore, an error will be reported with a message

"Unexpected substance translation in sample xxx: substance X is translated from multiple measured substances."

However, if such cases are known to exist and accepted in the data, then a method is available to *fix duplicate substance allocation inconsistencies*. If active substance allocation leads to multiple allocated measurements then the following procedure is implemented for resolving these inconsistencies:

- If any measurement is positive or zero then: clone one of these records to create the "aggregate measurement record"; to make this selection deterministic, prefer records that have measured-substance equal to active-substance over other records, and take the record with the highest residue. Update the measured value of the clone with the mean of all positive/zero measurements.
- If all allocated active substance measurements are censored, then clone one of the censored values. Here also, records that have measured-substance equal to active-substance are preferred over other records, and then the measurement with the smallest LOQ/LOD is used.

Note that these rules are quite generic. They work quite well even when there are many measurements for the same active substance. In practice, only a few (two) are expected.

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 to-food are imputed in a random order by active substance concentrations (positive, censored or zero) from a randomised list obtained from the fromfood(s). By matching the randomised lists, each from-food 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.

1. 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 fulfil 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.

Imputation of water concentrations

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*.

Focal commodity scenario analysis

There are different methods for modifying the (background) concentration data for specific (prospective) focal commodity scenario analyses. In the front end, these focal commodity scenario analysis method are accessible through the option *include focal commodity concentrations*. Checking this option will open the focal commodity scenario analysis form (see Figure 3.25) where the method and the focal commodity food/substance can be selected, and accompanying other settings can be configured.

Replace samples with focal commodity samples

This method will replace all samples for the selected focal commodity food by samples from the *focal commodity concentration data*. It works substance independent, and will therefore replace all substance concentrations of the focal commodity food in the background concentration data.

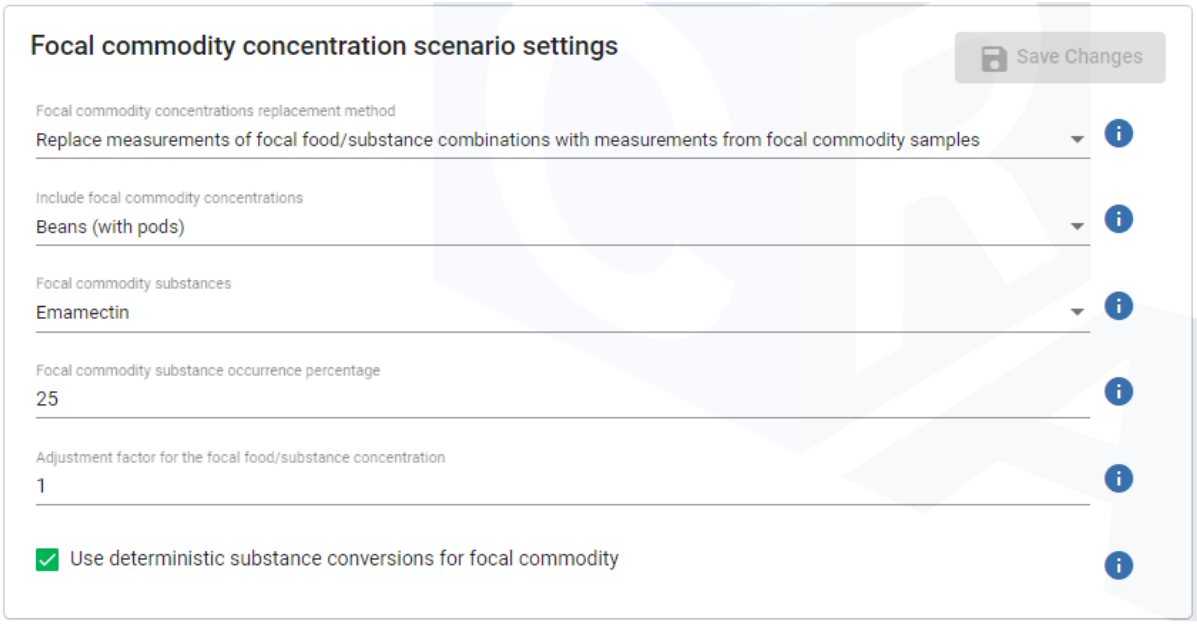


Figure 3.25: Focal commodity scenario analysis form of the front end. This form is a sub-form of the concentrations module panel.

Append focal commodity samples

This method adds the *focal commodity samples* of the selected focal commodity food to the background concentration data. This method is also substance-independent and may be a useful approach when the substances measured in the field trial do not overlap with the substances of the (background) concentration data. In this case, the focal commodity substance concentrations will be missing for the background concentration data and (also the other way around) the substance concentrations of all other substances will be considered missing for the focal commodity samples. These missing values may be imputed at a later stage following the "normal procedures".

Replace measurements of focal food/substance combinations

This method replaces, for the selected (focal) combination of food and substance, all substance concentrations with focal concentrations. This method knows two variants:

- **Replace by focal commodity samples:** The focal food/substance measurements are obtained from *focal commodity samples*. Here, substance measurements of the focal commodity food in the background concentration data set are replaced by randomly assigned substance measurements of the focal commodity samples.
- **Replace by concentration limits:** The focal food/substance measurements are obtained from *focal commodity samples*. Here, substance measurements of the focal commodity food in the background concentration data set are replaced by the concentration limit value (e.g., an MRL) obtained from the provided *concentration limits data*.

Using the *focal commodity substance occurrence percentage*, it is possible to specify an occurrence percentage for the combination of focal food and substance. When this percentage is less than 100%, this will partly (i.e., for the selected percentage) replace the concentrations of the focal commodity food and substance with the focal concentrations, and for the other part replace the concentrations with zero concentrations. E.g., when aiming to replace background concentrations of the substance fluopyram on potatoes with an MRL value, then specifying a focal commodity substance occurrence percentage of 40% will replace 40% of the measurements with the MRL, and 60% of the measurements with zero concentrations. Note that, because the allocation is random (i.e., each substance measurement has a probability of being assigned a focal concentration or a zero defined by the percentage), the realized replacement percentage may differ from the specified percentage. This option can, for example, be used to simulate a percentage of agricultural use.

Using the *adjustment factor for the focal food/substance concentration*, it is possible to adjust the (positive) concentrations of the focal food and substance measurements. This factor can be used when the focal commodity concentrations (e.g., from field trials) are assumed to be higher than what may be reasonably expected in practice. In this case, this factor could be set for instance, to the expected ratio of mean monitoring concentration and mean field trial concentration. Note that for replacement by focal commodity measurements, this factor will only adjust the positive concentrations and not the LORs.

By default, the focal commodity substance measurements are replaced before the optional step of *converting the concentrations from measured to active substance concentrations*. This also means that for these replaced measurements, the same rules apply, and the measurements may be converted to active substance measurements after replacement. Alternatively, it is possible to replace substance measurements after having done the allocation, and to use *deterministic substance conversions factors* for the focal commodity food and substance to convert these measurements to the level of *active substances*.

Note that when also using using *substance authorisations*, the focal food and substance combination will be treated as authorised, even if there is no authorisation supplied for the combination. The approved authorisation status is considered to be part of this scenario analysis.

Remove measurements of focal food/substance combinations

This method will simply remove all background concentrations for the selected focal commodity food and substance combination, and will not replace them with other values. This method may be useful when a separate analysis is desired for the background and foreground concentrations.

Filter samples by specific sample properties

When the option to *filter samples by specific property values (subset selection)* is checked in the *main panel of the concentrations module*, a new form will appear in which various sample property subsets can be specified (see Figure 3.26). The visibility/availability of the sample property filters depends on the availability of these properties in the data. The following filters are available:

- The options to *filter samples by year* and *filter samples by month* allow the user to filter the samples by sampling date. The additional option to *include samples with missing sampling date* determines whether samples for which the sampling date is unknown/missing should be included or not.
- The options to *filter samples by location* and *filter samples by region* allow for subsets by location (country) and/or region within a location. Also for these options there is the possibility to include/exclude samples for which the location/region is unknown/missing.
- The option to *filter samples by production method* allows for subset selection on production method of the sampled product (e.g., organic or conventional).
- The option to filter samples by additional sample properties available in the data (see *data format*).

Concentrations settings

Selection settings

Table 3.62: Selection settings for module Concentrations.

Name	Type	Description
Concentrations tier	*ConcentrationsTier*	Specifies the concentration data should be treated according to pre-defined tier or custom.
Filter samples exceeding the concentration limits	Boolean	If checked, samples with at least one substance concentration higher than some factor (concentration limit filter exceedance factor) times the MRL are filtered out.

continues on next pa

Table 3.62 – continued from previous page

Name	Type	Description
Concentration limit filter exceedance factor	Numeric	The multiplication factor for the concentration limit exceedance filter.
Use substance conversion rules	Boolean	If checked, concentrations are modelled in terms of active substances (using substance conversion).
Substance conversion method	*SubstanceTranslationAllocationMethod*	Allocation method for assigning active substance concentrations from measured substance concentrations based on substance translations.
Retain all allocated substances after active substance allocation	Boolean	If checked, all allocated substances kept after substance conversion. Otherwise, the concentration data is restricted to th active substances of the assessment group.
Account for substance authorisations in substance conversions	Boolean	Account for substance authorisations when allocating measured substances to active substance using substance conversions.
Fix duplicate substance allocation inconsistencies	Boolean	Resolve inconsistencies when active substance allocation leads t multiple concentration value estimates for the same active substance. This method uses the mean of the positives or zero concentrations when available, or else the lowest of the censore values.
Use extrapolation rules	Boolean	Use extrapolation rules to extrapolate food samples for foods w a limited amount of samples (data poor foods) from other foods (data rich foods).
Threshold for extrapolation	Numeric	Threshold for extrapolation.
Restrict extrapolations to equal MRLs	Boolean	Restrict extrapolations to equal MRLs.
Restrict extrapolations to authorised uses	Boolean	Only extrapolate if substance use is authorised.
Impute water concentrations	Boolean	Impute constant concentration values on the selected (water) commodity.
Water commodity	AlphaNumeric	The commodity for which constant concentration values should added.
Water concentration value (µg/kg)	Numeric	Constant concentration value that should be used for water (in µg/kg).
Restrict water imputation to the five most toxic substances	Boolean	Restrict water imputation to the five most toxic substances.
Restrict water imputation to authorised uses	Boolean	Restrict water imputation to authorised uses.
Restrict water imputation to approved substances	Boolean	Specifies whether imputation of water should be limited to approved substances only.
Include focal commodity concentrations	Boolean	Specifies whether there is monitoring data that should replace p of the consumption data for the specified focal commodities.
Focal commodity foods	AlphaNumeric	The foods for which background concentration data are to be replaced by focal commodity concentrations.
Focal commodity substances	AlphaNumeric	The substances for which background concentration data are to replaced by focal commodity concentrations.
Focal commodity concentrations replacement method	*FocalCommodityReplacementMethod*	Replacement method to be used for replacing base concentratio data with concentration data of the focal commodity/commodit concentrations.
Focal commodity substance occurrence percentage	Numeric	Anticipated occurrence percentage / agricultural use percentage of the focal commodity.
Adjustment factor for the focal food/substance concentration	Numeric	Optional adjustment factor for the focal food/substance concentration. E.g., the expected ratio of mean monitoring concentration and mean field trial concentration.

continues on next pa

Table 3.62 – continued from previous page

Name	Type	Description
Use deterministic substance conversions for focal commodity	Boolean	Convert measured substance concentrations of focal commodity to active substance concentrations using deterministic substance conversion factors.
Filter samples by specific property values (subset selection)	Boolean	Specifies whether a subset selection on specific sample properti should be made (e.g., by country or by year).
Sample locations	AlphaNumeric	The locations for which samples are filtered.
Sample years	AlphaNumeric	The years for which samples are filtered.
Restrict to specific modelled foods (modelled foods subset)	Boolean	If checked, then the assessment is restricted to the specified modelled foods.
Align sampling location subset with population	Boolean	If checked, the samples are filtered based on the location of the selected population.
Include samples with unspecified location	Boolean	If checked, then samples for which the sample location is not specified are also included by the sample location filter.
Align sample date subset with population	Boolean	If checked, the samples are filtered based on the period of the selected population.
Align sampling month subset with population	Boolean	If checked, the samples are filtered based on the month/period the selected population.
Include samples with unspecified sampling date	Boolean	If checked, then samples for which the sample date is not specified are also included by the sample date filter.
Filter samples by location	Boolean	If checked, samples are filtered based on the selected locations.
Filter samples by year	Boolean	If checked, samples are filtered based on the selected years.

Uncertainty settings

Table 3.63: Uncertainty settings for module Concentrations.

Name	Type	Description
Resample concentrations	Boolean	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.

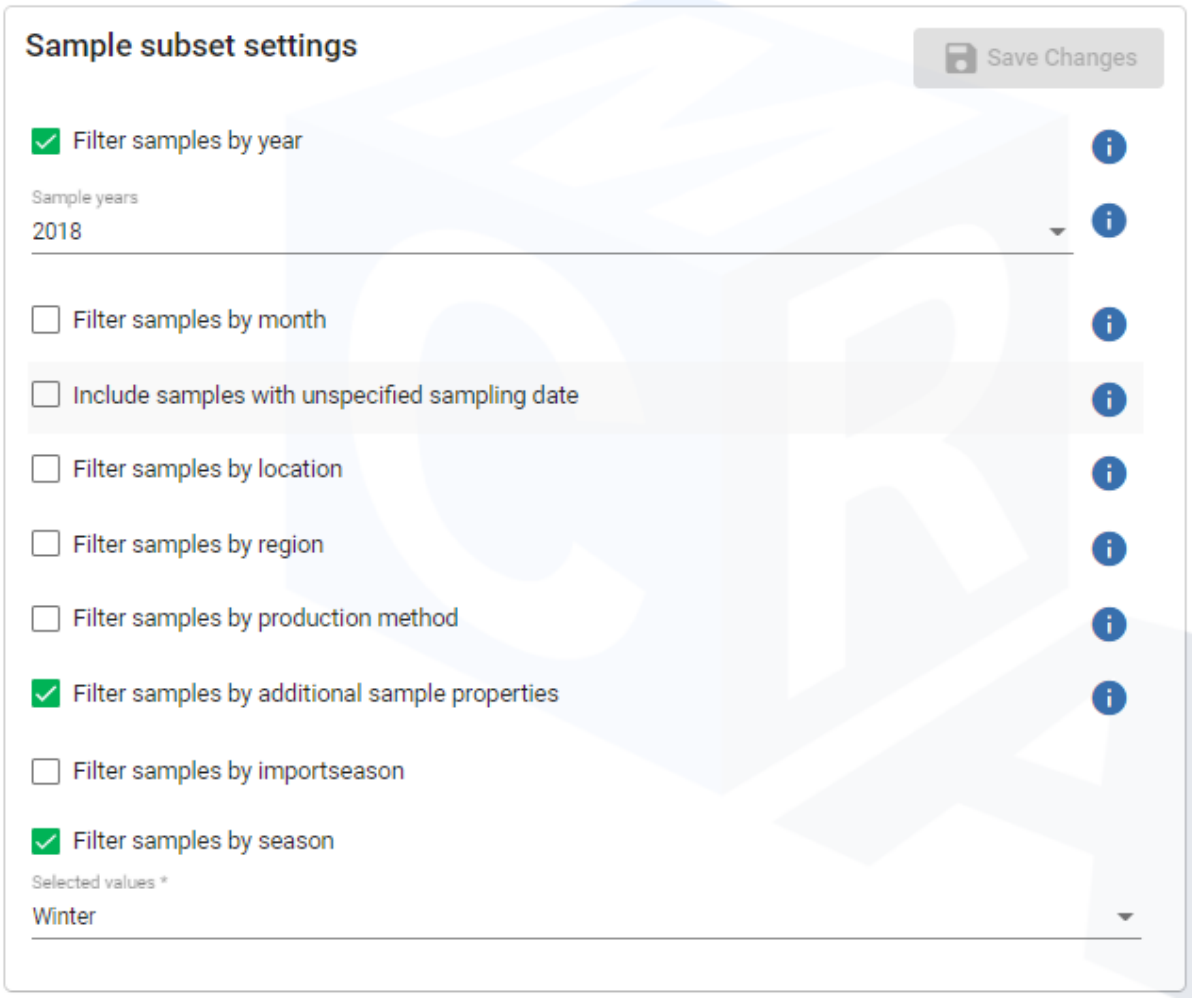


Figure 3.26: Sample subset selection form of the front end. This form is a sub-form of the concentrations module panel.

Overview

Table 3.64: Tier overview for module Concentrations.

Name	EC 2018 Tier 1	EC 2018 Tier 2
Substance conversion method	UseMostToxic	DrawRandom
Retain all allocated substances after active substance allocation	true	true
Account for substance authorisations in substance conversions	false	true
Fix duplicate substance allocation inconsistencies	false	false
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 the five most toxic substances	true	true
Restrict water imputation to authorised uses	false	false
Restrict water imputation to approved substances	false	false

EC 2018 Tier 1

Table 3.65: 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
Fix duplicate substance allocation inconsistencies	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 the five most toxic substances	true
Restrict water imputation to authorised uses	false
Restrict water imputation to approved substances	false

EC 2018 Tier 2

Table 3.66: 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
Fix duplicate substance allocation inconsistencies	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)	0.05
Restrict water imputation to the five most toxic substances	true
Restrict water imputation to authorised uses	false
Restrict water imputation to approved substances	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 Deterministic substance conversion factors Relative potency factors Substance authorisations Active substances Concentration limits Substance approvals*

3.3.5 Deterministic substance conversion factors

Deterministic substance conversion factors.

This module has as primary entities: *Substances Foods*

Output of this module is used by: *Concentrations Single value concentrations*

Deterministic substance conversion factors data formats

Deterministic substance conversion factors. Foods are optional.

Download empty dataset template: Zipped CSV Excel

Deterministic substance conversion factors

Deterministic substance conversion factors for translating measured substance concentrations to active substance concentrations.

Table 3.67: Table definition for Deterministic substance conversion factors.

Name	Type	Description	Aliases	Required
idMeasured-Substance	AlphaNumeric (50)	Substance code of the measured substance.	idMeasured-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
idFood	AlphaNumeric (50)	The unique identification code of the food.	idFood, Code, FoodId, FoodCode, Food	No
Conversion-Factor	Numeric	Specifies the conversion factor to translate concentrations of the measured substance to (equivalent) concentrations of the active substance according to e.g. the system used in PRIMo.	Factor, Conversion-Factor	Yes
Reference	AlphaNumeric (200)	Reference to the source from which this value is obtained.	Reference, References, Source, Sources	No

Accepted table names: SingleValueSubstanceConversionFactors, SingleValueConversionFactors, SingleValueConversions, SubstanceConversionsFixed, DeterministicSubstanceConversionFactors.

Deterministic substance conversion factors as data

Deterministic substance conversion factors.

- *Deterministic substance conversion factors data formats*

3.3.6 Focal food concentrations

In some cases the attention in an assessment is to evaluate concentrations (e.g., from specific field trials) for a specific food (and substance), in combination with a background of concentration data for other foods. Focal food concentrations can be included to provide these separate (foreground) concentration data for one or more focal food commodities that should replace measurements in the (background) *concentration data* in *focal commodity scenario analyses*.

This module has as primary entities: *Foods Substances*

Output of this module is used by: *Concentrations*

Focal food concentrations data formats

See *concentration data formats*.

Focal food concentrations settings

Selection settings

Table 3.68: Selection settings for module Focal food concentrations.

Name	Type	Description
Focal commodity foods	AlphaNumeric	The foods for which background concentration data are to be replaced by focal commodity concentrations.
Focal commodity substances	AlphaNumeric	The substances for which background concentration data are to replaced by focal commodity concentrations.

Calculation settings

Table 3.69: Calculation settings for module Focal food concentrations.

Name	Type	Description
Focal commodity concentrations replacement method	*FocalCommodityReplacementMethod*	Replacement method to be used for replacing base concentratio data with concentration data of the focal commodity/commodit concentrations.

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*

3.3.7 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 (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).

Download empty dataset template: Zipped CSV Excel

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 3.70: Table definition for Food extrapolations.

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.	IdFoodData-Poor, FoodDataPoor, idFromFood, FromFoodId, FromFood, FoodFrom, Food, IdFood	Yes
CodeDataRich-Food	AlphaNumeric (50)	The code of the read-across food (or data rich food). I.e., the food from which data is used for extrapolation.	IdFoodData-Rich, FoodDataRich, IdFoodRead-Across, FoodRead-Across, IdReadAcross-Food, ReadAcross-Food, idToFood, ToFoodId, ToFood, FoodTo	Yes

Accepted table names: ReadAcrossFoodTranslations, ReadAcrossFoodTranslation, ReadAcrossTranslations, ReadAcrossTranslation, FoodExtrapolations, FoodExtrapolation.

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*

3.3.8 Modelled foods

Modelled foods are foods within the foods scope for which concentration data or MRLs 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*

Modelled foods calculation

Modelled foods are the foods within the foods scope for which concentration data or MRLs of substances are available (or expected). Modelled foods are derived primarily from *concentration data*. That is, all foods for which food samples are available in the concentration data or MRL data are considered to be modelled foods. 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 modelled foods. The set of foods can also be restricted by omitting foods with only censored measurements (see *calculation settings*).

Modelled foods settings

Calculation settings

Table 3.71: Calculation settings for module Modelled foods.

Name	Type	Description
Restrict to specific modelled foods (modelled foods subset)	Boolean	If checked, then the assessment is restricted to the specified modelled foods.
Selected modelled foods	AlphaNumeric	Set of modelled foods that are of particular interest.
Derive modelled foods from concentrations	Boolean	Derive modelled foods from sample based concentration data.
Derive modelled foods from single value concentrations	Boolean	Derive modelled foods from single value concentrations.
Derive modelled foods from concentration limits	Boolean	Derive modelled foods from concentration limits.
If checked, include foods with only censored value measurements	Boolean	Specifies whether foods with only censored value measurement are part of the exposure assessment (default yes).
If checked, include substances with only censored value measurements	Boolean	Specifies whether substances with only censored value measurements are part of the exposure assessment (default yes)
If checked, include substances without measurements	Boolean	Specifies whether substances without any measurements should included.

Calculation of modelled foods

Modelled foods are computed from concentration data (which may also be in the form of single-value concentrations) and/or derived from available maximum residue limits.

- *Modelled foods calculation*

Inputs used: *Concentrations Single value concentrations Concentration limits*

Settings used

- *Calculation Settings*

3.3.9 Occurrence frequencies

Occurrence frequencies specify how often substances occur on foods. Frequencies are expressed as percentages.

This module has as primary entities: *Foods Substances*

Output of this module is used by: *Concentration models Single value dietary exposures*

Occurrence frequencies data formats

Occurrence frequencies are described by one simple table, specifying for pairs of food and substance, the associated occurrence frequencies as percentages.

Download empty dataset template: Zipped CSV Excel

Occurrence frequencies

Occurrence frequencies are specified as percentages for pairs of food and substance. Optionally, a reference can be included in each record to specify the source (e.g., from literature) from which the percentage was obtained.

Table 3.72: Table definition for Occurrence frequencies.

Name	Type	Description	Aliases	Required
idFood	AlphaNumeric (50)	The food code.	idFood, CodeFood, FoodId, FoodCode, Food	Yes
idSubstance	AlphaNumeric (50)	Code of the substance.	idSubstance, CodeSubstance, SubstanceId, SubstanceCode, Substance	Yes
Percentage	Numeric	The occurrence frequency percentage.	Percentage, Frequency-Percentage	Yes
Reference	AlphaNumeric (200)	Reference to the source from which this use frequency value is obtained.	Reference, References, Source, Sources	No

Accepted table names: OccurrenceFrequencies.

Occurrence frequencies calculation

Occurrence frequencies can be provided as data or computed from *occurrence patterns*. For a food and substance, they are computed by collecting all occurrence patterns of this food and summing up the frequencies of the occurrence patterns containing the substance. In the unlikely case that the total frequency of the occurrence patterns of a food exceeds 100%, then a rescaling is applied first. If the sum of the frequencies does not sum up to 100%, the interpretation of the remaining unspecified percentage can be designated as either "no use" or "all use". In the the first case it is assumed that none of the substances occur on this remaining percentage. In the latter it is assumed that all of the substances occur on this remaining percentage. This choice is available as the setting *associate the unspecified percentage with no-occurrence for foods with at least one specified occurrence pattern*.

Depending on the setting *apply occurrence pattern percentages*, occurrence frequencies can be computed in a crisp form in which the occurrence frequency is either 0% or 100% or as percentages ranging from 0% to 100%.

Occurrence frequencies Settings

Selection settings

Table 3.73: Selection settings for module Occurrence frequencies.

Name	Type	Description
Associate the unspecified percentage with no-occurrence for foods with at least one specified occurrence pattern	Boolean	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 i samples). Note that this setting cannot be used for foods that ha no specified AUs. These foods have 100% potential presence o all substances. To declare all AUs on such a food un-authorised include an empty AU with percentage 100% in the AU data tab (i.e., use an AU for this food, without specifying substances in t AU Substances table)
Apply occurrence pattern percentages	Boolean	If checked, use the percentages of potential presence as specifie by the occurrence patterns. If unchecked, 100% potential presence in samples is assumed for all substances identified by t occurrence patterns.

Occurrence frequencies as data

Occurrence frequencies are described by one table, specifying for a food and substance the associated occurrence frequency as percentage.

- *Occurrence frequencies data formats*

Inputs used: *Active substances*

Calculation of occurrence frequencies

Occurrence frequencies for a food and substance are computed according to the model that is part of the EC 2018 Tier II definition (see van Klaveren et al. 2019)

- *Occurrence frequencies calculation*

Inputs used: *Occurrence patterns*

3.3.10 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 are 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: *Occurrence frequencies Dietary exposures*

Occurrence patterns data formats

Agricultural use percentages for plant protection products (PPPs) may be of use for concentration modelling, as they provide information about what substance mixtures are expected to be present simultaneously on food samples. Especially for censored concentration measurements, this information may aid to determine whether the censored measurement originated from a true zero or may be a positive measurement below the limit of detection. Agricultural use percentages are 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%.

Download empty dataset template: Zipped CSV Excel

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 3.74: Table definition for Agricultural uses.

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
Name	AlphaNumeric (100)	Name of the agricultural use.	Name	No
Description	AlphaNumeric (200)	Additional description of the agricultural use.	Description	No

Accepted table names: AgriculturalUses, AgriculturalUse.

Agricultural use substances

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

Table 3.75: Table definition for Agricultural use substances.

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	Yes

Accepted table names: AgriculturalUseHasSubstances, AgriculturalUsesHasSubstances, AgriculturalUseSubstances, AgriculturalUseGroups, AgriculturalUseGroup.

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 censored 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 3.76: Selection settings for module Occurrence patterns.

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

Uncertainty settings

Table 3.77: Uncertainty settings for module Occurrence patterns.

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

Occurrence patterns tiers

Overview

Table 3.78: Tier overview for module Occurrence patterns.

Name	EC 2018 Tier 1	EC 2018 Tier 2
Apply occurrence pattern percentages	false	true
Substance conversion method	UseMostToxic	DrawRandom
Retain all allocated substances after active substance allocation	true	true
Account for substance authorisations in substance conversions	false	true
Fix duplicate substance allocation inconsistencies	false	false
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 the five most toxic substances	true	true
Restrict water imputation to authorised uses	false	false
Restrict water imputation to approved substances	false	false
Scale up use frequency to 100%		true
Restrict use percentage up-scaling to authorised uses		true

EC 2018 Tier 1

Table 3.79: Tier definition for EC 2018 Tier 1.

Name	Setting	From input tier	In module
Apply occurrence pattern percentages	false		
Substance conversion method	UseMostToxic	*EC 2018 Tier 1*	*Concentrations*
Retain all allocated substances after active substance allocation	true	*EC 2018 Tier 1*	*Concentrations*
Account for substance authorisations in substance conversions	false	*EC 2018 Tier 1*	*Concentrations*
Fix duplicate substance allocation inconsistencies	false	*EC 2018 Tier 1*	*Concentrations*
Use extrapolation rules	true	*EC 2018 Tier 1*	*Concentrations*
Threshold for extrapolation	10	*EC 2018 Tier 1*	*Concentrations*
Restrict extrapolations to equal MRLs	true	*EC 2018 Tier 1*	*Concentrations*
Restrict extrapolations to authorised uses	true	*EC 2018 Tier 1*	*Concentrations*
Impute water concentrations	true	*EC 2018 Tier 1*	*Concentrations*
Water concentration value (µg/kg)	0.1	*EC 2018 Tier 1*	*Concentrations*
Restrict water imputation to the five most toxic substances	true	*EC 2018 Tier 1*	*Concentrations*
Restrict water imputation to authorised uses	false	*EC 2018 Tier 1*	*Concentrations*
Restrict water imputation to approved substances	false	*EC 2018 Tier 1*	*Concentrations*

EC 2018 Tier 2

Table 3.80: Tier definition for EC 2018 Tier 2.

Name	Setting	From input tier	In module
Apply occurrence pattern percentages	true		
Scale up use frequency to 100%	true		
Restrict use percentage up-scaling to authorised uses	true		
Substance conversion method	DrawRandom	*EC 2018 Tier 2*	*Concentrations*
Retain all allocated substances after active substance allocation	true	*EC 2018 Tier 2*	*Concentrations*
Account for substance authorisations in substance conversions	true	*EC 2018 Tier 2*	*Concentrations*
Fix duplicate substance allocation inconsistencies	false	*EC 2018 Tier 2*	*Concentrations*
Use extrapolation rules	true	*EC 2018 Tier 2*	*Concentrations*
Threshold for extrapolation	10	*EC 2018 Tier 2*	*Concentrations*
Restrict extrapolations to equal MRLs	true	*EC 2018 Tier 2*	*Concentrations*
Restrict extrapolations to authorised uses	true	*EC 2018 Tier 2*	*Concentrations*
Impute water concentrations	true	*EC 2018 Tier 2*	*Concentrations*
Water concentration value (µg/kg)	0.05	*EC 2018 Tier 2*	*Concentrations*
Restrict water imputation to the five most toxic substances	true	*EC 2018 Tier 2*	*Concentrations*
Restrict water imputation to authorised uses	false	*EC 2018 Tier 2*	*Concentrations*
Restrict water imputation to approved substances	false	*EC 2018 Tier 2*	*Concentrations*

Occurrence patterns as data

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

- *Occurrence patterns data formats*

Inputs used: *Substance authorisations Active substances*

Calculation of occurrence patterns

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

- *Occurrence patterns calculation*

Inputs used: *Concentrations*

3.3.11 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 Single value dietary exposures*

Processing factors data formats

Processing factors connect to a food id and a processing type or (food) facet (FoodEx 2). The specification of a substance (id) is optional. Specify the unprocessed food code and the processing type or facet. The combination of idFood-idProcessingType represents a processed food.

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.

Download empty dataset template: Zipped CSV Excel

Processing factors

Processing factor records should be linked to processing types (or facets) using the processing type (or facet) 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 (of facets) definitions.

Table 3.81: Table definition for Processing factors.

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
idSubstance	AlphaNumeric (50)	The code of the substance.	idSubstance, SubstanceId, SubstanceCode, Substance	No
idFood-Unprocessed	AlphaNumeric (50)	The code of the unprocessed food.	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	Yes
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

Accepted table names: ProcessingFactors, ProcessingFactor, Processing.

Processing factors calculation

Processing factors fixed or distribution based

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

- 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 *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

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

See also *processing correction*

Processing factors settings

Uncertainty settings

Table 3.82: Uncertainty settings for module Processing factors.

Name	Type	Description
Resample processing factors	Boolean	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 Processing factors*).

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 Processing factors*).

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)). 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 little 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

Specify for a combination of processing type, food and substance the processing factor (nominal, upper).

- *Processing factors data formats*
- *Processing factors calculation*

3.3.12 Single value concentrations

Single value concentrations data are the single value estimates (High Residue, Maximum Residue Limit, Supervised Trials Median Residue) of residue concentrations on modelled foods.

This module has as primary entities: *Foods Substances*

Output of this module is used by: *Modelled foods Single value dietary exposures*

Single value concentrations data formats

Single value concentrations data provides a single value concentration for a substance.

Download empty dataset template: Zipped CSV Excel

Concentration single values

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

Table 3.83: Table definition for Concentration single values.

Name	Type	Description	Aliases	Required
idFood	AlphaNumeric (50)	Code of the food of this concentration single value.	idFood, FoodId, Food	Yes
idSubstance	AlphaNumeric (50)	Code of the substance of this concentration single value.	idSubstance, SubstanceId, Substance, idCompound, CompoundId, Compound	Yes
Value	Numeric	Concentration single value.	Value, Concentration, Concentration-Value	Yes
ValueType	*ConcentrationLimit-ValueType*	Value type of the concentration value.	Concentration-SingleValue-Type, Concentration-ValueType, SingleValue-Type, Concentration-Type, ValueType, Type	Yes
Percentile	Numeric	Percentile.	Percentile	No
Concentration-Unit	*ConcentrationUnit*	The unit of the concentration single value (default mg/kg).	Concentration-Unit, Unit	No
Reference	AlphaNumeric (200)	Reference to the source from which this concentration single value is obtained.	Reference, References, Source, Sources	No

Accepted table names: ConcentrationSingleValues, SingleValueConcentrations.

Single value concentrations calculation

Single value concentrations as data are supplied as mean concentrations, median concentrations, highest residues, percentiles, LOQs or maximum residue limits. Specify the 'Use data' option in the interface. In a retrospective context, the single values are computed based on the concentration distributions available for the modelled food as supplied in the *Concentrations module*. Specify option 'Compute' in the Single value concentrations action.

Single value concentrations settings

Selection settings

Table 3.84: Selection settings for module Single value concentrations.

Name	Type	Description
Use substance conversion factors	Boolean	Specifies whether to use substance conversion factors to convert measured substance concentrations to active substance concentrations.

Single value concentrations as data

Single value concentrations data are the single value concentrations of residues on modelled foods.

- *Single value concentrations data formats*

Inputs used: *Active substances*

Calculation of single value concentrations

Single value concentrations are calculated as a percentile (p50, p97.5 or maximum residue limit) of the modelled food concentration distribution.

- *Single value concentrations calculation*

Inputs used: *Concentrations Concentration limits Deterministic substance conversion factors*

3.3.13 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 LOQ or LOD could be assumed true zeros. I.e., if a food/substance combinations is assumed to be unauthorised, then the LOQ, LOD 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 Concentration models*

Substance authorisations data formats

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

Download empty dataset template: Zipped CSV Excel

Authorised uses

The authorised uses table

Table 3.85: Table definition for Authorised uses.

Name	Type	Description	Aliases	Required
idFood	AlphaNumeric (50)	The food code.	idFood, FoodId, Food	Yes
idSubstance	AlphaNumeric (50)	The substance 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

Accepted table names: AuthorisedUses, AuthorisedUse.

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 associated with no authorised use.

- *Substance authorisations data formats*

3.3.14 Substance approvals

Substance approvals specify which substances are approved within the definition under regulation (EC) No 1107/2009. This information may, for instance, be used to to restrict water imputation to approved substances only.

This module has as primary entities: *Substances*

Output of this module is used by: *Concentrations*

Substance approvals data formats

Substance approvals specify which substances are approved within the definition under regulation (EC) No 1107/2009. This information may, for instance, be used to restrict water imputation to approved substances only.

Download empty dataset template: Zipped CSV Excel

Substance approvals

Substance approval records are tuples consisting of an identifier of the substance and a boolean value specifying the approval status of the substance. Substances that are not included in the table are assumed to be NOT approved.

Table 3.86: Table definition for Substance approvals.

Name	Type	Description	Aliases	Required
idSubstance	AlphaNumeric (50)	The substance code.	idSubstance, Substance, SubstanceId	Yes
IsApproved	Boolean	Specifies whether the substance is approved or not. Substances not included in the table are assumed to be NOT approved.	IsApproved, Approved	Yes

Accepted table names: SubstanceApprovals, ApprovedSubstances.

Substance approvals as data

Substance approvals are specified as data in the form of a list of approved substances. Substances not on the list are assumed to be not approved.

- *Substance approvals data formats*

3.3.15 Substance conversions

Substance conversions specify how measured substances are converted into active substances, which are the substances assumed to cause health effects. In 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.

Note, it is not allowed to have conversion factors equal to 0.

Substance conversions are described by a single substance conversions table.

Download empty dataset template: Zipped CSV Excel

Substance conversion rules

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 3.87: Table definition for Substance conversion rules.

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	Only applicable for non-exclusive conversions. The proportion of the concentration that is assumed to be attributed to the active substance.	Proportion	No

Accepted table names: ResidueDefinitions, ResidueDefinition.

Substance conversions as data

Substance conversions are provided as data.

- *Substance conversions data formats*

Inputs used: *Active substances*

3.3.16 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: *Concentration models Food conversions*

Total diet study sample compositions data formats

Total diet studies (TDS) complement traditional monitoring of substance concentrations on raw commodities by measuring substance occurrence in main foods prepared as consumed and pooled into representative food groups. To include occurrence data from TDS for exposure assessment, the composition of the TDS samples is needed in order to link the composite samples to the consumed foods (either directly or indirectly). TDS composition data describes the composition of TDS samples by specifying the foods (and the amounts) of TDS samples.

Download empty dataset template: Zipped CSV Excel

TDS food sample compositions

The TDS food sample compositions table contains the descriptions of the TDS samples and specifications of the foods (with amounts) included in the TDS samples.

Table 3.88: Table definition for TDS food sample compositions.

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 (250)	Regionality information.	Regionality	No
Seasonality	AlphaNumeric (250)	Seasonality information.	Seasonality	No

Accepted table names: TDSFoodSampleCompositions, TDSFoodSampleComposition, CompositionTDSFoodSamples, CompositionTDSFoodSample.

Total diet study sample compositions as data

Total diet study sample compositions are provided as data.

- *Total diet study sample compositions data formats*

3.3.17 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 for *modelling unit variability* in acute *(individual) dietary exposures calculations* to account for the fact that concentration data often relate to composite samples, whereas an acute risk may result from consumption of single food units. For the same purpose, they are also used in the *IESTI model* for *single value dietary exposures calculations*.

This module has as primary entities: *Foods Substances*

Output of this module is used by: *Dietary exposures Single value dietary exposures*

Unit variability factors data formats

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.

Download empty dataset template: Zipped CSV Excel

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 3.89: Table definition for Unit variability factors.

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	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

Accepted table names: UnitVariabilityFactors, UnitVariabilityFactor, VariabilityFactor, VariabilityFactors, VariabilityProcCompProd, UnitVariability.

IESTI special cases

IESTI special cases for specified combinations of food, substance. The application type (post-harvest or pre-harvest) determines whether Case 1 or Case 3 should be used.

Table 3.90: Table definition for IESTI special cases.

Name	Type	Description	Aliases	Required
idFood	AlphaNumeric (50)	The unique identification code of the food.	idFood, Code, FoodId, FoodCode, Food, Id	Yes
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	Yes
Application-Type	*HarvestApplication-Type*	Harvest application type (pre-harvest or post-harvest).	Application-Type, Harvest-ApplicationType	Yes
Reference	AlphaNumeric (200)	External reference(s) to pre-harvest use.	Reference	No

Accepted table names: IestiSpecialCases.

Unit variability factors as data

Unit variability factors are provided as data.

- *Unit variability factors data formats*

3.4 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 *modelled foods* are made using *food conversions*, and the consumptions are expressed as *consumptions per modelled food*. 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*.

3.4.1 Consumptions by modelled food

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

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

Output of this module is used by: *Single value consumptions High exposure food-substance combinations Dietary exposures*

Consumptions by modelled food calculation

Consumptions by modelled food are calculated from *consumptions* of *modelled foods* and *food conversions* that link the foods-as-eaten amounts to modelled-foods amounts. Given that the food conversion is already available, the procedure for computing the consumptions by modelled-food is straightforward. For each consumption of each individual, a modelled-food consumption record is created for each modelled-food 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 in the *food conversion algorithm* one or more *processing types* are found, then these types are recorded in the consumption by modelled food record.

Consumptions by modelled food

Calculation settings

Table 3.91: Calculation settings for module Consumptions by modelled food.

Name	Type	Description
Restrict population to consumers or consumer days only (food-as-measured)	Boolean	Specifies whether the population should be restricted to the individuals (chronic) or individual days (acute) with consumptio containing any of the modelled foods.
Risk type	*ExposureType*	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 modelled foods only	Boolean	Specifies whether the population should be restricted to the individuals (chronic) or individual days (acute) with consumptio containing any of the specified modelled food subset.
Selected modelled foods	AlphaNumeric	Set of consumed modelled foods that are of particular interest f restricting the consumers / consumption days.

Calculation of consumptions by modelled food

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

- *Consumptions by modelled food calculation*

Inputs used: *Consumptions Food conversions*

Settings used

- *Calculation Settings*

3.4.2 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 Exposure mixtures Risks*

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 acute 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 are 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 modelled foods. After screening, contributions related to food-as-eaten are available for the risk drivers only.

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 substances are individuals exposed?

   Co-exposure of substances is a qualitative approach where the number of combinations of substances to which an individual is exposed is recorded. 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. For more information see *co-exposure of substances*.

2. quantitative approach: *maximum cumulative ratio (MCR)*. To what degree are mixtures more important than single substances?

For a quantitative approach, see also the *exposure mixtures module*.

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 w_i when the number of MC iterations is > 0 (see *table for Individuals*, field *SamplingWeight*).

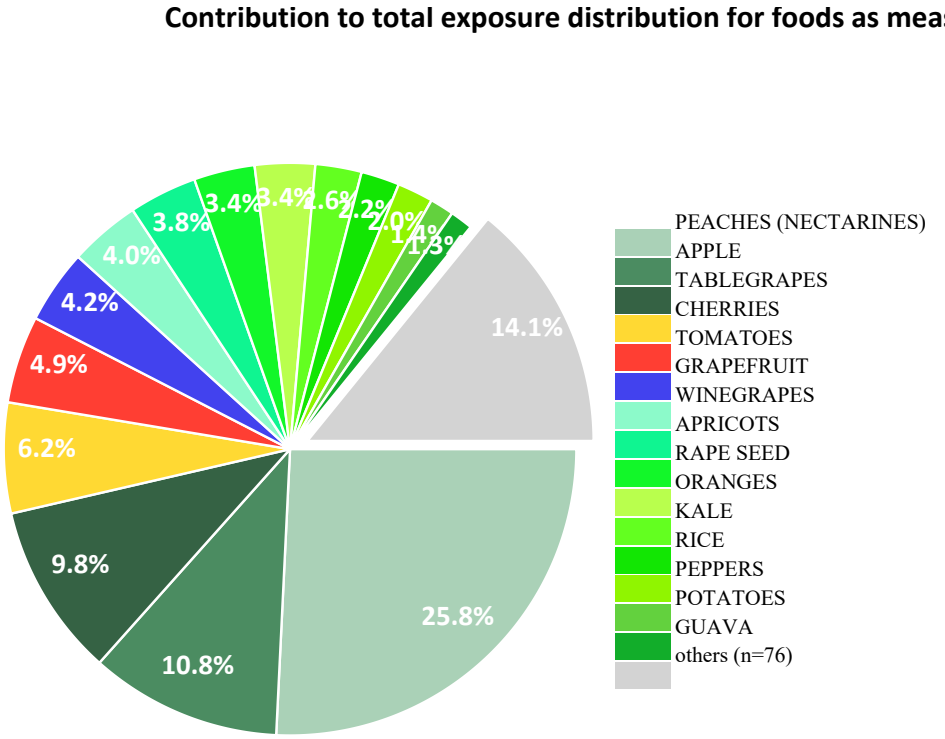


Figure 3.27: Example MCRA dietary exposure contributions modelled foods.

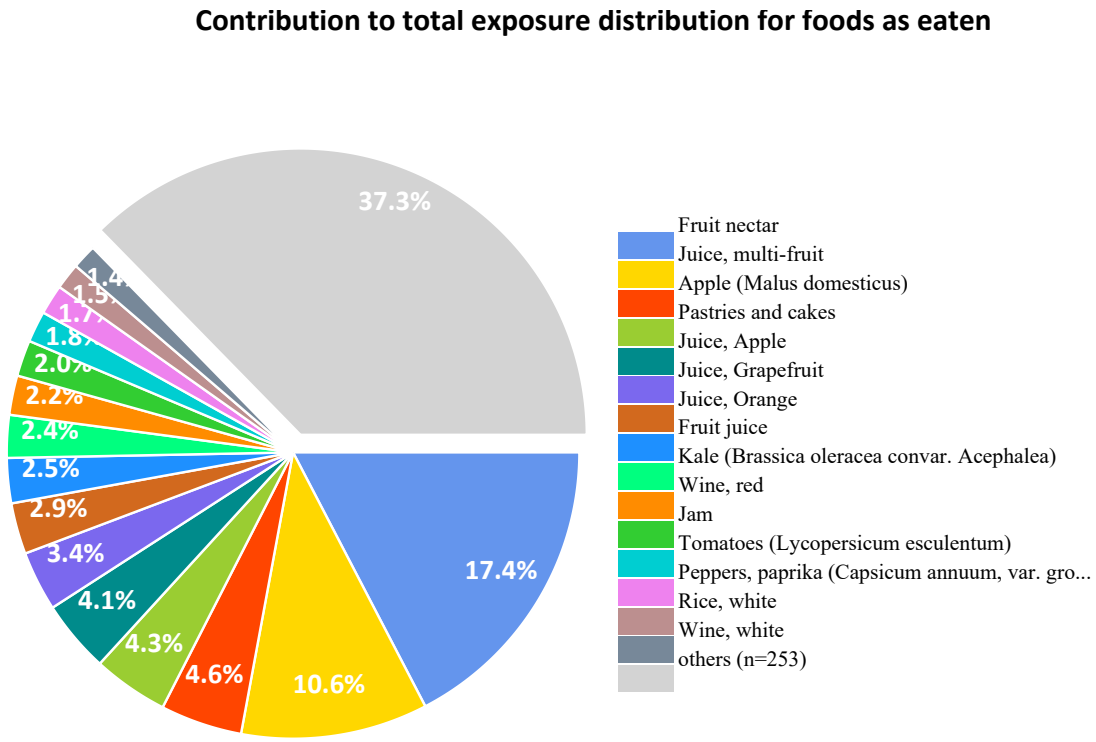


Figure 3.28: Example MCRA dietary exposure contributions foods as eaten

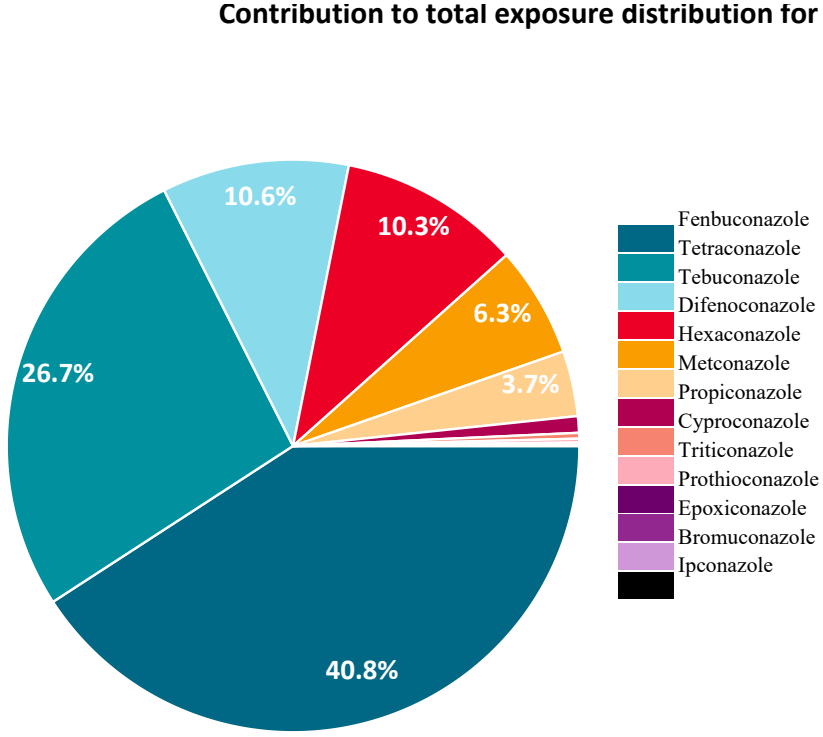


Figure 3.29: Example MCRA dietary exposure contributions substances

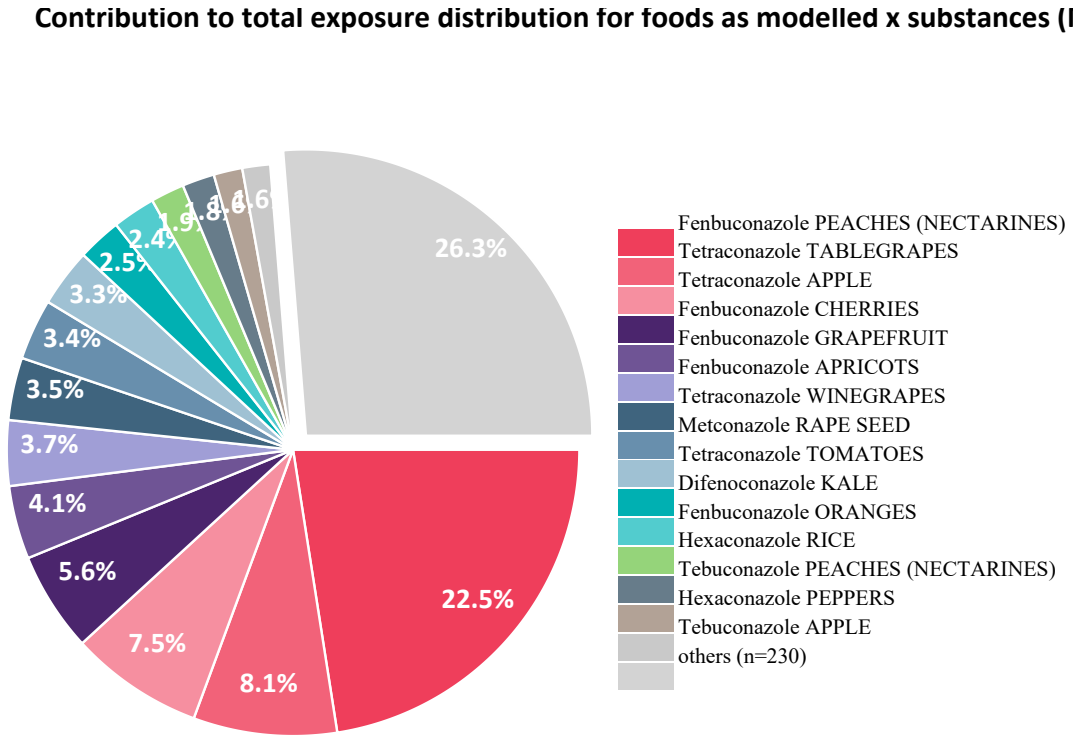


Figure 3.30: Example MCRA dietary exposure contributions modelled foods x substances

Modelling unit-to-unit variation

The basic model for an acute exposure assessment assumes 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. 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 composite sample.

MCRA offers three distributions to sample from:

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

The beta distribution simulates values for a unit in the composite sample. It 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. It 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 either 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 times 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 c_V (standard deviation divided by mean). Following FAO/WHO recommendations, the default variability factor $v = 1$ for small crops (unit weight < 25 g). For large crops (unit weight ≥ 25 g) $v = 5$. For foods which

are processed in large batches, e.g. *juicing, marmalade/jam, sauce/puree, bulking/blending* the 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 (whole food/RAC) 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.

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)). 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

$$c_v^2 = ba^{-1}(a + b + 1)^{-1}$$

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 $(0, cm_{max})$.

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.

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.

Chronic exposure assessment

In a chronic exposure assessment, usual exposure is defined as the long-run average of daily exposure to a substance or group of substances by an individual. The interest is in the distribution of individual exposures and derived statistics like the fraction of individuals that exceed an intake limit or point of departure *PoD*). The PoD is calculated as the average daily intake (ADI) * safety factor (SF). 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 exposure):

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

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 exposures MCRA, provides a number of *exposure models* to calculate the distribution of usual exposure at the person level.

Chronic exposure models

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

1. The observed individual means *observed individual means* (OIM) model;
2. The *logisticnormal-normal (LNN) model*, in a full version that includes the estimation of correlation between exposure frequency and amount, and in a simpler version without this estimation;
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 exposure, 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 exposure amount on consumption days.

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

Observed individual means (OIM)

The usual exposure distribution for a population is estimated with the empirical distribution of individual means. Each mean is the average of all single-day exposures 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 exposure distribution and estimates using the OIM-method are biased, leading to a too high estimate of the fraction of the population with a usual exposure 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) basic assessments.

Model based and model assisted

Following Kipnis et al. (2009), some of the models available in MCRA are extended to predict individual usual exposures. This model assisted approach has been added to BBN and LNN when used without correlation) and may be a useful extension in evaluating the relationship between health outcomes and individual usual exposures of foods. In contrast, the estimation of the usual exposure distribution in the general population is called the model based approach. Summarizing, we get Table 3.92:

Table 3.92: 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 (LNN) without correlation	logisticnormal-normal (LNN) without correlation
logisticnormal-normal (LNN) with correlation	
Iowa State University Foods (ISUF)	

The model assisted approach builds on the proposal of Kipnis et al. (2009), 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 exposure frequencies. The model-assisted usual exposure distribution applies to the population for which the consumption data are representative, and automatically integrates over any covariates present in the model. Model-assisted exposures 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 exposure is presented in graphs and tables as a *function of the covariates* (conditional usual exposure distributions).

Betabinomial-Normal model (BBN)

The *Betabinomial-Normal (BBN)* model for chronic risk assessment is described in de Boer et al. (2009), including its near-identity to the STEM-II model presented in Slob (2006). The BBN model combines a betabinomial model for the exposure frequencies with a normal model for transformed positive exposures.

Logisticnormal-Normal model (LNN with and without correlation)

In the logisticnormal-normal (LNN) model, exposure frequencies are modelled 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 $\sigma^2_{between}$. Similarly as in the BBN model, the positive exposure amounts are modelled, after transformation (logarithmic or Box-Cox), with a normal distribution. This model is referred to as the *LogisticNormal-Normal (LNN)* model. The full *LNN model* model includes the estimation of a correlation between exposure frequency and exposure amount. This is similar to the NCI model described in Tooze et al. (2006). A simple and computationally less demanding version of the LNN method does not estimate the correlation between frequency and amount. 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. (1996) described how to assess chronic risks for data sets with positive exposures (a small fraction of zero exposures was allowed, but then replaced by a small positive value). The modelling 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 exposure from data sets with a substantial amount of zero exposures became feasible by modelling separately zero exposure on part or all of the days via the estimation of exposure probabilities as detailed in Nusser et al. (1997) and Dodd (1996). In MCRA, a discrete/semi-parametric model is implemented allowing for zero exposure and heterogeneity of variance following the basic ideas of Nusser et al. (1996), Nusser et al. (1997) and Dodd (1996). This implementation of the ISUF model for chronic risk assessment is fully described in de Boer et al. (2009).

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 (LNN)). That this principle can work in practice was shown in previous work (de Boer et al. (2009), Slob et al. (2010), Goedhart et al. (2012)) 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), de Boer and van der Voet (2011)). For specific cases involving separate modelling of dietary supplements and the rest of the diet, proposals have been made (Verkaik-Kloosterman et al. (2011)). 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 3.31) 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 3.31).
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 3.31).

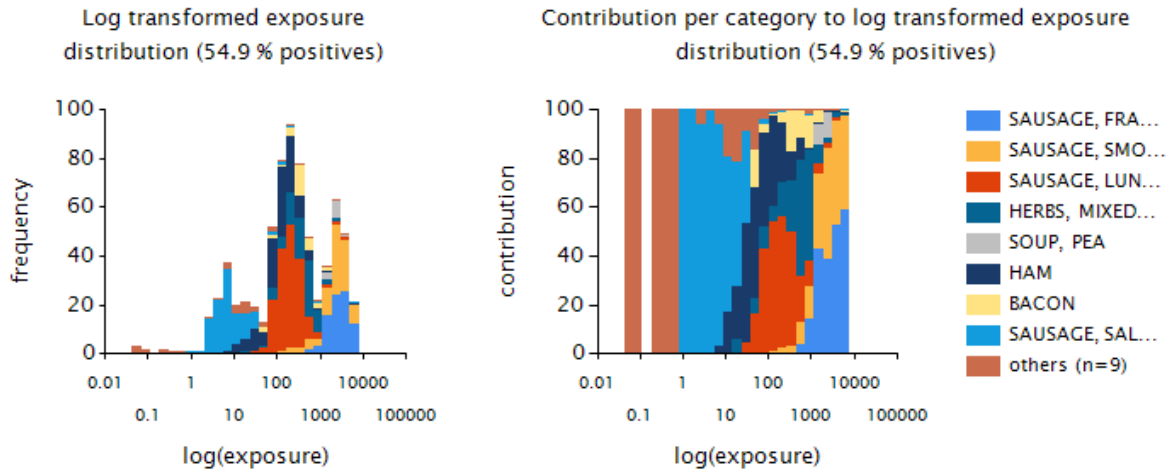


Figure 3.31: Left panel: OIM usual exposure distribution to smoke flavours via the different foods (excluding the zero exposures) in young children; right panel: Contribution of foods to exposures within each bar of the OIM distribution histogram.

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 3.32 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.

After a split-off selection has been made, the OIM distribution is summarised in terms of the defined grouping (Figure 3.33), and the usual exposure distribution per split-off food or food group is fitted according to the chosen modelling settings.

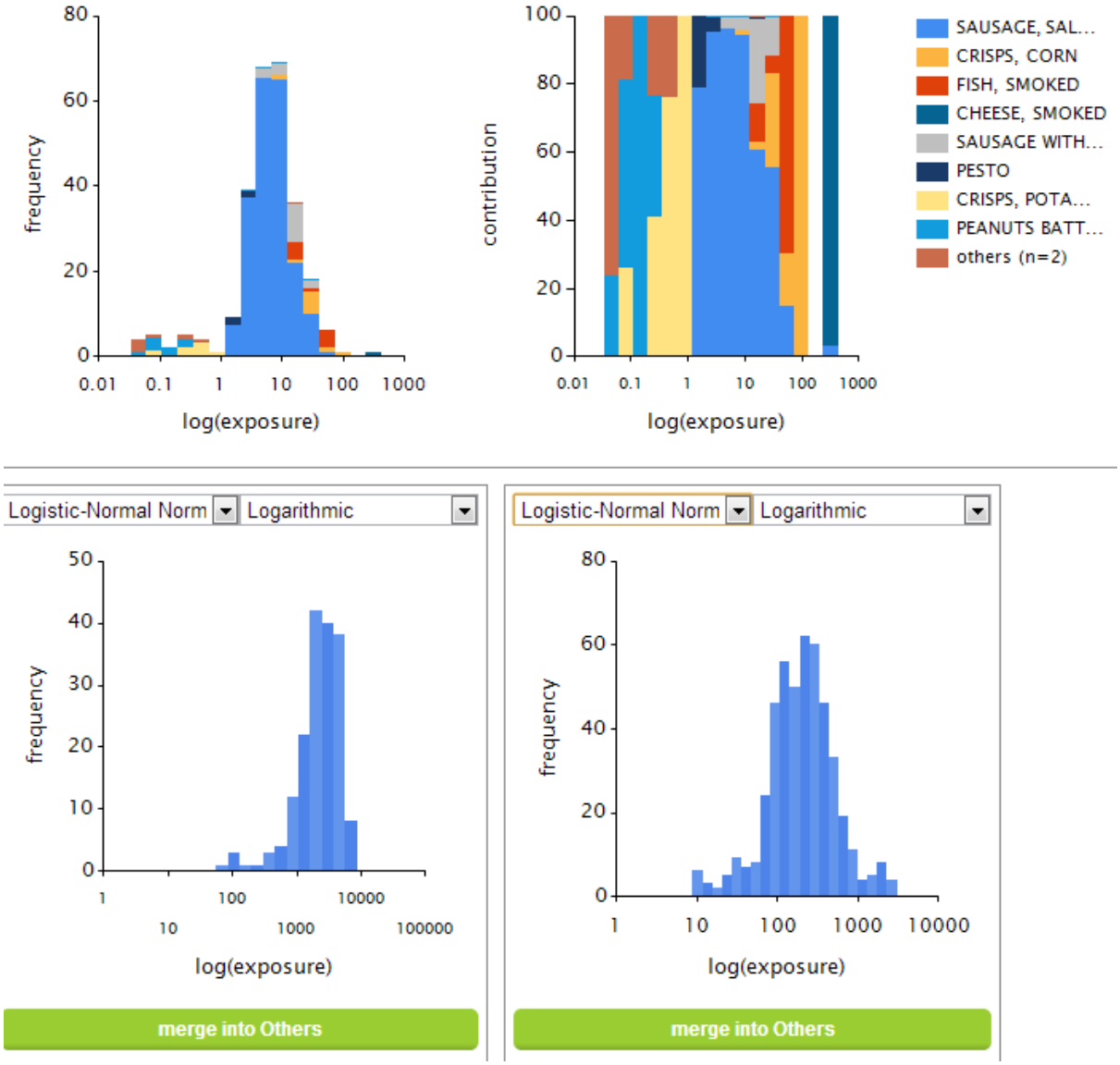


Figure 3.32: Result of a selection into two split-off groups and a rest group. The graph bottom left represents the exposure via a food group containing *'Sausage, frankfurter'* and *'Sausage, smoked cooked'*. The graph bottom right represents the exposure via a food group containing *'Sausage, luncheon meat', Herbs, mixed, main brands, not prepared', 'Soup, pea', 'Ham', and 'Bacon'*. The top graph represents the exposure via the rest group.

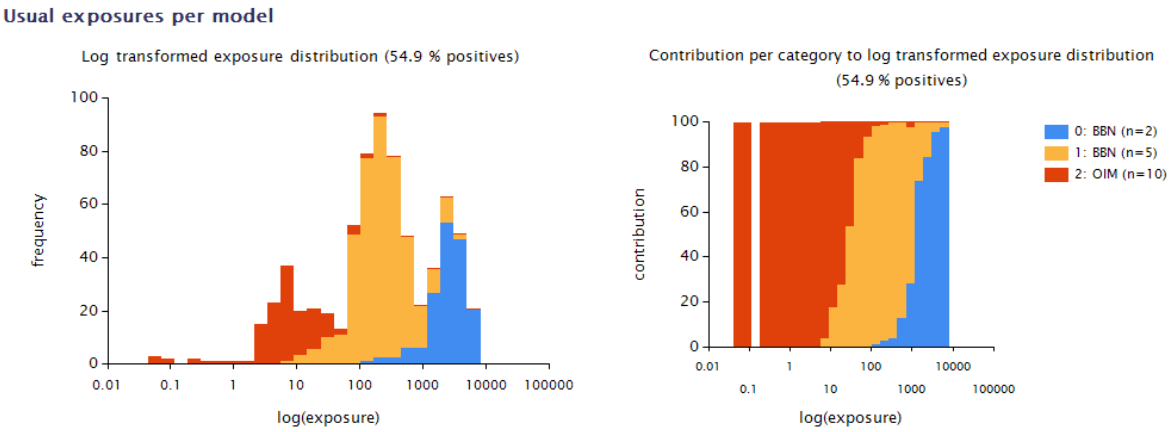


Figure 3.33: OIM usual exposure distribution showing the contributions from the three food groups as constructed in Figure 3.32.

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:

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.

See also, *episodically consumed foods, model-based, model-assisted*.

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) 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.

Chronic exposure as a function of covariates

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

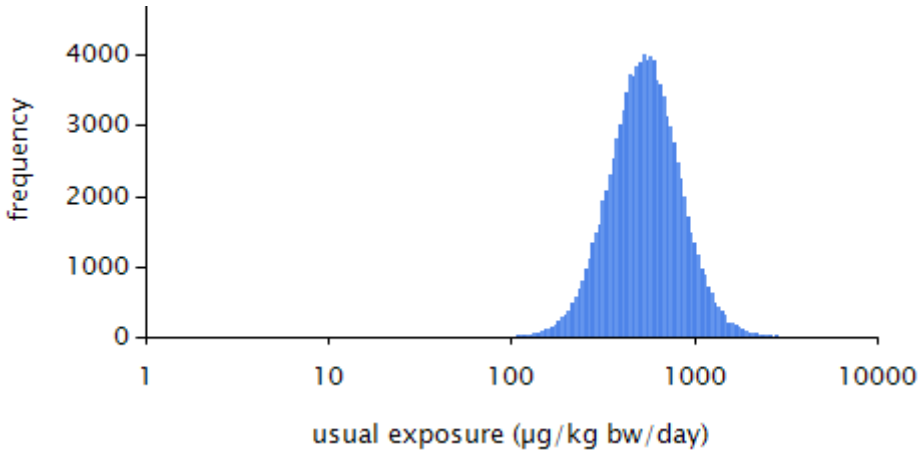


Figure 3.34: Model-assisted estimated usual exposure distributions (excluding the zero exposures).

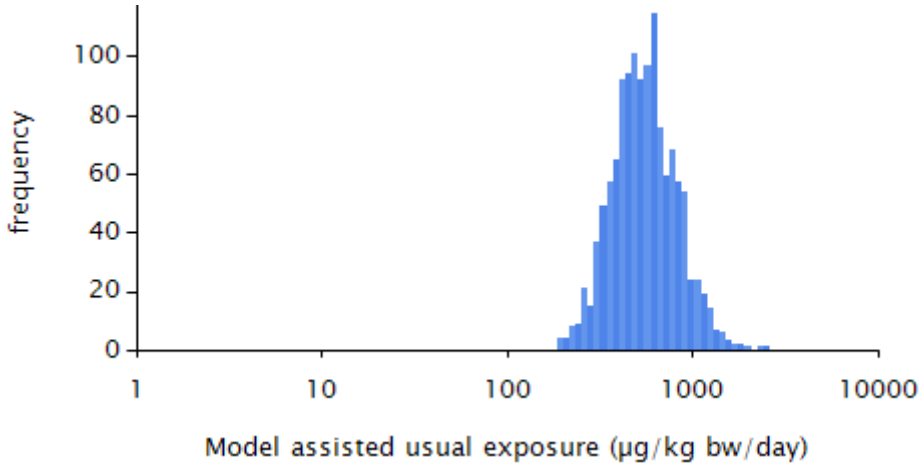


Figure 3.35: Model-based estimated usual exposure distributions (excluding the zero exposures).

Table 3.93: 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 modelled foods 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 modelled food *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.

TDS reduction to limit 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 concentration distributions of the raw agricultural commodities (RACs) that make up the TDS food sample. The decision to intervene or not can be based on comparison between the p95 percentile point of the concentration distribution and a concentration limit value that associated with a high risk.

- 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 by the *concentration distributions module*. In the MCRA interface, a scenario analysis can be checked and the scroll down list allows to select the foods that should be included in the scenario analysis. For the selected foods, concentration reduction factors are computed based on the p95 percentile and the limit value:

$$f_{\texttt{reduction}} = limit/p95$$

These reduction factors (computed for the RACs) are applied to the simulated concentrations on the level of the TDS composite foods of which the conversion paths contain the RACs included in the scenario analysis.

$$f_{\texttt{reduction}} \cdot c_{TDS}$$

Here, c_{TDS} is the concentration value of the TDS food.

Substance concentrations generation

Both *chronic* and *acute* dietary exposure assessments rely on assigning substance concentrations to consumed modelled foods. 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 modelled food of a consumption, substance concentrations are generated by drawing a random sample from the set of all samples available for that modelled food. Assuming that for the drawn sample, substance concentration values are known for all substances of interest (i.e., all missing values and censored values 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 censored 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 censored values, two approaches exist:

1. **Replace by zero:** Censored values are imputated by a zero concentration value. This is an optimistic approach.
2. **Replace by factor times LOR:** Each censored value is replaced by a factor f (e.g., 1 or 1/2) times its LOR.
3. **Replace by factor times LOD LOQ system:** Non-detects are replaced by f * LOD; non-quantifications are replaced by LOD + f * (LOQ - LOD) and factor f is e.g., 1 or 1/2.

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 correction

Concentrations in the consumed food (food as eaten) may be different from concentrations in the modelled food 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 modelled food 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 modelled foods.

Chronic exposure assessment, daily consumed foods

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 $b_i \sim N(0, \sigma^2_b)$ and $w_{ij} \sim N(0, \sigma^2_w)$

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 $Z_i = (Z_{i1}, ..., Z_{ip})$:

$$\mu_i = \beta_0 + \beta^t_1 Z_i$$

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:

$$T_i = E_w[g^{-1}(\mu_i + b_i + w_{ij})|b_i] = F(b_i)$$

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

$$T_i = E_w[g^{-1}(\mu_i + b_i + w_{ij}|b_i)] = \int_{-\infty}^{\infty} g^{-1}(\mu_i + b_i + w_{ij}) \frac{1}{\sqrt{2\pi\sigma_w^2}} \exp\left(-\frac{w^2}{2\sigma_w^2}\right) \mathrm{d}w$$

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

$$\mu_{iT} = E[T_i] = \exp(\mu_i + \sigma_w^2/2 + \sigma_b^2/2)$$

$$\sigma_{iT}^2 = Var[T_i] = [\exp(\sigma_b^2) - 1] \exp(2\mu_i + \sigma_w^2 + \sigma_b^2)$$

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 et al. (2009). 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.

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

$$\mathit{BLUP}_i = \mu_i + (\bar{Y}_i^* - \mu_i)\left(\frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2/n_i}\right)$$

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

$$\mathit{Var}(\bar{Y}_i^*) = \sigma_b^2 + \sigma_w^2/n_i$$

and thus

$$\mathit{Var}(\mathit{BLUP}_i) = \left(\frac{\sigma_b^4}{\sigma_b^2 + \sigma_w^2/n_i}\right)$$

which is smaller than the between individual variance σ_b^2. As an alternative a modified BLUP can be defined by means of

$$\mathit{modifiedBLUP}_i = \mu_i + (\bar{Y}_i^* - \mu_i)\sqrt{\left(\frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2/n_i}\right)}$$

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.

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

$$\mathit{modelassistedBLUP}_i = \mu_i + \frac{\sigma_w^2}{2} + (\bar{Y}_i^* - \mu_i)\sqrt{\left(\frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2/n_i}\right)}$$

So the model assisted individual intake $\exp(\mathit{modelassistedBLUP}_i)$ has the same distribution as the usual intake and is thus the best predictor for usual intake.

Kipnis et al. (2009) 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:

$$(b_i, \bar{Y}_i^*) \sim \mathit{BivariateNormal}(0, \mu_i, \sigma_b^2, \sigma_b^2 + \sigma_w^2/n_i, \sigma_b^2)$$

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

$$\begin{aligned} E[E[F(b_i)|\bar{Y}_i^*]] &= \exp\left(\mu_i + \frac{\sigma_b^2 \sigma_w^2 / n_i}{2(\sigma_b^2 + \sigma_w^2/n_i)} + \frac{\sigma_w^2}{2} + \frac{\sigma_b^4}{2(\sigma_b^2 + \sigma_w^2/n_i)}\right) \\ &= \exp\left(\mu_i + \frac{\sigma_b^2}{2} + \frac{\sigma_w^2}{2}\right) \end{aligned}$$

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) will therefor result in too much shrinkage of the model assisted usual intake.

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) 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

$$P_i^* = \mu_{iT} + \frac{\sigma_{iT}}{\sigma_{iP}}(P_i - \mu_{iP})$$

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.

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

$$X_i = \textit{Binomial}(n_i, p_i)$$

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:

$$P(X_i = x) = \left(\begin{array}{c} n_i \\ r \end{array} \right) \frac{B(\alpha + x, n_i + \beta - x)}{B(\alpha, \beta)}$$

The mean and variance of the betabinomial distribution are given by

$$E[X_i] = n_i \frac{\alpha}{\alpha + \beta}$$

and

$$\textit{Var}[X_i] = n_i \frac{\alpha\beta(\alpha + \beta + n_i)}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

Using the reparameterization $\pi = \alpha/(\alpha + \beta)$ and $\phi = 1/(\alpha + \beta + 1)$, it follows that

$$E[X_i] = n_i \pi$$

and

$$\textit{Var}[X_i] = n_i \pi(1-\pi)[1 + (n_i - 1)\phi]$$

This reparameterization enables to model the probability π_i of consumption for individual i directly as a logistic regression:

$$\textit{logit}(\pi_i) = \gamma_0 + \gamma_1^t Z_i$$

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.

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.

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 every 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:

$$
\begin{aligned}
E(p_i|X_i = x) &= \int_p pf(p|X_i = x)\mathrm{d}p \\
&= \int_p p\frac{f(X_i = x|p)f(p)}{\int f(X_i = x|p)f(p)dp}\mathrm{d}p \\
= \frac{1}{P(x_i = x)}\int_p p\left(\begin{array}{c} n_i \\ r \end{array}\right)p^x(1-p)^{n_i-x}B^{-1}(\alpha_i,\beta_i)p^{\alpha_i-1}(1-p)^{\beta_i-1}\mathrm{d}p \\
&= \frac{B^{-1}(\alpha_i,\beta_i)}{P(x_i = x)}\left(\begin{array}{c} n_i \\ r \end{array}\right)\int_p p^{\alpha_i+x}(1-p)^{n_i+\beta_i-x-1}\mathrm{d}p \\
&= \frac{B(\alpha_i + x + 1, n_i + \beta_i - x)}{B(\alpha_i + x, n_i + \beta_i - x)} \\
&= \frac{\alpha_i + x}{\alpha_i + \beta_i - x}
\end{aligned}
$$

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.

Logistic-Normal model for frequencies (LNN0)

In this model the distribution of X_i is again binomial:

$$X_i = \textit{Binomial}(n_i, p_i)$$

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

$\textit{logit}(p_i) = \lambda_i + v_i$ 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 $\textit{logit}^{-1}(\lambda_i)$ 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 $Se(\lambda_i^*)$. 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:

$$Se(\pi_i^*) \approx \frac{Se(\lambda_i^*)}{\sqrt{\pi}} \sum_{j=1}^{N} w_j \frac{d}{d\lambda_i^*} H(\lambda_i^* + \sqrt{2}\sigma_v x_j)$$

$$= \frac{Se(\lambda_i^*)}{\sqrt{\pi}} \sum_{j=1}^{N} w_j H(\lambda_i^* + \sqrt{2}\sigma_v x_j)[1 - H(\lambda_i^* + \sqrt{2}\sigma_v x_j)]$$

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.

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.

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) with further mathematical details in Kipnis et al. (2009). 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*.

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.

Model assisted usual intake

This requires simultaneous prediction of the random effect for frequency and for amount as outlined in Kipnis et al. (2009). 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 a *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. $L_0 = \mathit{logit}(P_0)$. 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.

Dietary exposures settings

Calculation settings

Table 3.94: Calculation settings for module Dietary exposures.

Name	Type	Description
Dietary exposure calculation tier	*DietaryIntakeCalculationTier*	A tier is a pre-specified set of model configurations. By selectin model tier, MCRA automatically sets all model settings in this module according to this tier. Note that currently tier setting m need to be performed separately in sub-modules. Use the Custo tier when you want to manually set each model setting.
Risk type	*ExposureType*	The type of exposure considered in the assessment; acute (short term) or chronic (long-term).
Total diet study concentration data	Boolean	Specifies whether exposure is based on sampling data from tota diet studies.
Multiple substances analysis	Boolean	Specifies whether the assessment involves multiple substances.
Compute cumulative exposures	Boolean	Specifies whether the assessment involves multiple substances a results should be cumulated over all substances.

continues on next pa

Table 3.94 – continued from previous page

Name	Type	Description
Sample based	Boolean	Include co-occurrence of substances in samples in simulations. checked, substance residue concentrations are sampled using th correlations between values on the same sample. If unchecked, any correlation between substances is ignored, substance residu concentrations are sampled ignoring the correlations between values on the same sample.
Consumptions on the same day come from the same sample	Boolean	if checked, in procedure of EFSA Guidance 2012, section 4.1. 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 individ on the same day are assumed to come from different samples.
Maximise co-occurrence of high values in simulated samples	Boolean	Within each pattern of substance presence. If checked, substan residue concentrations are sorted within co-occurrence patterns substances on the same samples. After sorting, high residue val occur more frequently on the same sample. This choice is conservative. If unchecked, substance residue concentrations ar sampled at random, ignoring any co-occurrence patterns of substances on the same samples. This choice is less conservativ
Apply processing factors	Boolean	Specified in table ProcessingFactor. If checked, processing fact are applied. Concentrations in the consumed food may be different from concentrations in the modelled food in monitorin programs (typically raw food) due to processing, such as peelin washing, cooking etc. If unchecked, no processing information used. This is in most (though not all) cases a worst-case assumption
Use distribution	Boolean	Probabilistic specifications of processing factors will be used
Ignore processing factors less than 1	Boolean	This setting will suppress the use of processing factors lower th 1 (it is used in the EFSA 2012 Pessimistic tier).
Use unit variability	Boolean	Controls whether to use unit variability.
Unit variability model	*UnitVariabilityModelType*	Describes variation between single units when concentration da are from composite samples.
Estimates nature	*EstimatesNature*	Simulated unit concentrations can be higher or lower than composite value (realistic) or only equal or higher (conservative
Unit variability parameter	*UnitVariabilityType*	Use Coefficient of variation or Variability factor, specified in VariabilityFactor table.
Mean of LogNormal simulated values (biasing)	*MeanValueCorrectionType*	Unbiased: correct unit simulations for difference between medi and mean.
Default variability factor for unit weight <= 25g	Numeric	Default variability factor 1 (unit weight <= 25 g, small crops). S requires specification of unit weight (FoodProperties table) and case of beta model, also the Number of units in a composite sample (UnitVariability table).
Default variability factor for unit weight > 25g	Numeric	Default variability factor 5 (unit weight > 25 g, medium/large crops). Still requires specification of unit weight (FoodPropertie table) and, in case of beta model, also the Number of units in a composite sample (UnitVariability table).
Model type	*IntakeModelType*	The parametric model for between-and within-individual variation, and possibly covariates.
Model-then-add	Boolean	Specifies whether to create separate exposure models for specif groups of modelled foods (model-then-add).
Covariate modelling	Boolean	Specifies whether to model exposures as a function of covariate at individual level.
Amount model covariate model	*CovariateModelType*	Specifies whether, and how to model exposures amounts as function of covariates.

continues on next pa

Table 3.94 – continued from previous page

Name	Type	Description
Frequency model covariates model	*CovariateModelType*	Specifies whether, and how to model exposure frequency as function of covariates.
Use occurrence patterns for generating simulated samples	Boolean	When selected, this simulated samples will be based on occurrence patterns.
Details level dietary exposures	*DietaryExposuresDetailsLevel*	Level of detail for summarizing dietary exposure/intakes.
Iterate survey	Boolean	Instead of (re-)sampling the individual days, loop over the entir survey (= 1 iteration). The number of iterations for a survey is calculated as round (number of Monte Carlo iterations /(numbe of individuals * surveys days)).
Monte Carlo iterations	Numeric	The number of iterations for Monte Carlo simulations, e.g. 100.000 (maximum is 100.000).
Impute exposure distributions	Boolean	Impute exposure distributions for substances with missing concentrations.
Include diagnostics analysis for variability	Boolean	For each percentile the variability (standard deviation) of the estimated percentiles versus sample size are plotted.
Allow conversion using food extrapolations	Boolean	Step 3c: try to find read across 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 b used for pineapple (as-eaten or as ingredient). If successful, restart at step 1.
Censored values replacement	*NonDetectsHandlingMethod*	How to replace censored values (when not co-modelled, as in censored models).
Function	*FunctionType*	Functional relation between exposure and covariable.
Maximum degrees of freedom	Numeric	Order of function. Determines the maximum degree of complexity of the function.
Minimum degrees of freedom	Numeric	Order of function. Determines the minimum degree of complexity of the function.
Testing level	Numeric	Significance level for testing the degree of the function. e.g., 0.0
Testing method	*TestingMethodType*	Starting from a full model (backward) or empty model (forward
Default concentration model	*ConcentrationModelType*	The concentration model type that will be used as default for al food/substance combinations. If this model type cannot be fitte e.g., due to a lack of data, a simpler model will be chosen automatically as a fall-back.
Perform MCR analysis	Boolean	Perform a Maximum Cumulative Ratio (MCR) analysis to determine co-exposure between substances.
Substance weighting for MCR	*ExposureApproachType*	Risk based: exposures in equivalents of the reference substance standardised: standardised exposures per substance have varian 1; or unweighted exposures: RPFs are equal to 1.
Display ratio total exposure/ maximum (in MCR plot)	Numeric	For MCR plot: specify ratio total exposure/ maximum for individual(day) exposures .
Show tail percentiles (MCR plot) for:	AlphaNumeric	Give specific percentiles of exposure distribution (%), e.g. 97.5 99 (space separated).
Set minimum percentage contribution per substance to the tail exposure (MCR plot)	Numeric	Set minimum percentage contribution per substance to the tail exposure.

Output settings

Table 3.95: Output settings for module Dietary exposures.

Name	Type	Description
Include drill-down on 9 individuals around specified percentile.	Boolean	Specifies whether drilldown on 9 individuals is to be included in the output.
Summarize simulated data	Boolean	Specifies whether a summary of the simulated consumptions an concentrations should be included in the output.
Store simulated individual day exposures	Boolean	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	AlphaNumeric	Give specific percentiles of exposure distribution (%), e.g. 50 9 95 97.5 99 (space separated).
Percentage for drilldown	Numeric	Gives detailed output for nine individuals near this percentile of the exposure distribution.
Percentage for upper tail	Numeric	Gives detailed output for this upper percentage of the exposure distribution.
Show % of population below level(s)	*ExposureMethod*	This setting is used for reporting the percentages of individuals (chronic) or individual days (acute) exceeding certain exposure levels. These exposure levels can be generated automatically based on the observed exposures (Automatic, default) or specifi explicitly (Manual).
Exposure levels	AlphaNumeric	Specify exposure levels for which to give the percentage of exposure below these levels, e.g. 1 10 50 100 200 500.
Number of levels of covariable to predict exposure	Numeric	Specify the number of levels, e.g. 20. The range of the covariat is divided by the number of levels: range = (max - min)/levels. For these covariable levels exposures are predicted.
Predict exposure at extra covariable levels	Numeric	Specify specific prediction levels in addition to the automaticall generated prediction levels (space separated).
Lower percentage for variability (%)	Numeric	The default value of 25% may be overruled.
Upper percentage for variability (%)	Numeric	The default value of 75% may be overruled.
Report consumptions and exposures per individual instead of per kg body weight	Boolean	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 3.96: Uncertainty settings for module Dietary exposures.

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

Dietary exposures tiers

Overview

Table 3.97: Tier overview for module Dietary exposures.

Name	EFSA 2012 Optimistic	EFSA 2012 Pessimistic - Acute	EFSA 2012 Pessimistic - Chronic	EC 2018 Tier 1	EC 2018 Tier 2
Total diet study concentration data	false		false	false	false
Sample based	true	true	true	true	true
Consumptions on the same day come from the same sample	false	true	true	false	false
Apply processing factors	true	true	true	true	true
Use distribution	false	false	false	false	false
Ignore processing factors less than 1	false	true	true	false	false
Use unit variability	false	true		true	true
Unit variability model	NoUnitVariability	BetaDistribution		BetaDistribution	BetaDistribution
Model type	OIM		OIM	OIM	OIM
Model-then-add	false		false	false	false
Covariate modelling	false	false	false	false	false
Iterate survey	false	false	false	false	false
Report consumptions and exposures per individual instead of per kg body weight	false	false	false	false	false
Default concentration model	Empirical	NonDetectSpikeLogNormal	NonDetectSpikeLogNormal	Empirical	Empirical
Include MRL fallback model	false	true	true	false	false
Censored values replacement	ReplaceByZero	ReplaceByLOR	ReplaceByLOR	ReplaceByLOR	ReplaceByLOR
Impute missing values from available values (if unchecked, missing values are imputed with 0)	false	true	true	true	true
Correlate imputed values with sample potency	false	true	true	true	false
Use occurrence frequencies for imputation	false			true	true
Parametric uncertainty	false	true	false	false	false
Risk type		Acute	Chronic		
Estimates nature		Realistic		Realistic	Realistic

continues on next page

Table 3.97 – continued from previous page

Name	EFSA 2012 Optimistic	EFSA 2012 Pessimistic - Acute	EFSA 2012 Pessimistic - Chronic	EC 2018 Tier 1	EC 2018 Tier 2
Unit variability parameter		VariabilityFactor		VariabilityFactor	VariabilityFactor
Restrict LOR imputation to authorised uses		false	false	false	false
Factor f (f x LOR or f x LOD or LOD + f x (LOQ - LOD)		1	1	0.5	0.5
MRL Factor (f x MRL)		1	1		
Apply occurrence pattern percentages				false	true
Substance conversion method				UseMostToxic	DrawRandom
Retain all allocated substances after active substance allocation				true	true
Account for substance authorisations in substance conversions				false	true
Fix duplicate substance allocation inconsistencies				false	false
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 the five most toxic substances				true	true
Restrict water imputation to authorised uses				false	false
Restrict water imputation to approved substances				false	false
Scale up use frequency to 100%					true
Restrict use percentage up-scaling to authorised uses					true

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 variability model should be applied.

Table 3.98: Tier definition for EFSA 2012 Optimistic.

Name	Setting	From input tier	In module
Total diet study concentration data	false		
Sample based	true		
Consumptions on the same day come from the same sample	false		
Apply processing factors	true		
Use distribution	false		
Ignore processing factors less than 1	false		
Use unit variability	false		
Unit variability model	NoUnitVariability		
Model type	OIM		
Model-then-add	false		
Covariate modelling	false		
Iterate survey	false		
Report consumptions and exposures per individual instead of per kg body weight	false		
Default concentration model	Empirical	*EFSA 2012 Optimistic*	*Concentration models*
Include MRL fallback model	false	*EFSA 2012 Optimistic*	*Concentration models*
Censored values replacement	ReplaceByZero	*EFSA 2012 Optimistic*	*Concentration models*
Sample based	true	*EFSA 2012 Optimistic*	*Concentration models*
Impute missing values from available values (if unchecked, missing values are imputed with 0)	false	*EFSA 2012 Optimistic*	*Concentration models*
Correlate imputed values with sample potency	false	*EFSA 2012 Optimistic*	*Concentration models*
Use occurrence frequencies for imputation	false	*EFSA 2012 Optimistic*	*Concentration models*
Parametric uncertainty	false	*EFSA 2012 Optimistic*	*Concentration models*

EFSA 2012 Pessimistic - Acute

Acute probabilistic exposure assessment using the pessimistic model settings according to the EFSA Guidance 2012. Only processing factors > 1 are applied. For unit variability, the Beta distribution is applied.

Table 3.99: Tier definition for EFSA 2012 Pessimistic - Acute.

Name	Setting	From input tier	In module
Risk type	Acute		
Sample based	true		
Consumptions on the same day come from the same sample	true		
Apply processing factors	true		
Use distribution	false		
Ignore processing factors less than 1	true		
Use unit variability	true		
Unit variability model	BetaDistribution		
Estimates nature	Realistic		
Unit variability parameter	VariabilityFactor		
Covariate modelling	false		
Iterate survey	false		
Report consumptions and exposures per individual instead of per kg body weight	false		
Default concentration model	NonDetectSpikeLogNormal	*EFSA 2012 Pessimistic - Acute*	*Concentration models*
Include MRL fallback model	true	*EFSA 2012 Pessimistic - Acute*	*Concentration models*
Restrict LOR imputation to authorised uses	false	*EFSA 2012 Pessimistic - Acute*	*Concentration models*
Censored values replacement	ReplaceByLOR	*EFSA 2012 Pessimistic - Acute*	*Concentration models*
Factor f (f x LOR or f x LOD or LOD + f x (LOQ - LOD)	1	*EFSA 2012 Pessimistic - Acute*	*Concentration models*
MRL Factor (f x MRL)	1	*EFSA 2012 Pessimistic - Acute*	*Concentration models*
Sample based	true	*EFSA 2012 Pessimistic - Acute*	*Concentration models*
Impute missing values from available values (if unchecked, missing values are imputed with 0)	true	*EFSA 2012 Pessimistic - Acute*	*Concentration models*
Correlate imputed values with sample potency	true	*EFSA 2012 Pessimistic - Acute*	*Concentration models*
Parametric uncertainty	true	*EFSA 2012 Pessimistic - Acute*	*Concentration models*

EFSA 2012 Pessimistic - Chronic

Chronic probabilistic exposure assessment using the pessimistic model settings according to the EFSA Guidance 2012. Only processing factors > 1 are applied.

Table 3.100: Tier definition for EFSA 2012 Pessimistic - Chronic.

Name	Setting	From input tier	In module
Risk type	Chronic		
Total diet study concentration data	false		
Sample based	true		
Consumptions on the same day come from the same sample	true		
Apply processing factors	true		
Use distribution	false		
Ignore processing factors less than 1	true		
Model type	OIM		
Model-then-add	false		
Covariate modelling	false		
Iterate survey	false		
Report consumptions and exposures per individual instead of per kg body weight	false		
Default concentration model	NonDetectSpikeLogNormal	*EFSA 2012 Pessimistic - Chronic*	*Concentration models*
Include MRL fallback model	true	*EFSA 2012 Pessimistic - Chronic*	*Concentration models*
Restrict LOR imputation to authorised uses	false	*EFSA 2012 Pessimistic - Chronic*	*Concentration models*
Censored values replacement	ReplaceByLOR	*EFSA 2012 Pessimistic - Chronic*	*Concentration models*
Factor f (f x LOR or f x LOD or LOD + f x (LOQ - LOD)	1	*EFSA 2012 Pessimistic - Chronic*	*Concentration models*
MRL Factor (f x MRL)	1	*EFSA 2012 Pessimistic - Chronic*	*Concentration models*
Sample based	true	*EFSA 2012 Pessimistic - Chronic*	*Concentration models*
Impute missing values from available values (if unchecked, missing values are imputed with 0)	true	*EFSA 2012 Pessimistic - Chronic*	*Concentration models*
Correlate imputed values with sample potency	true	*EFSA 2012 Pessimistic - Chronic*	*Concentration models*
Parametric uncertainty	false	*EFSA 2012 Pessimistic - Chronic*	*Concentration models*

EC 2018 Tier 1

Table 3.101: Tier definition for EC 2018 Tier 1.

Name	Setting	From input tier	In module
Total diet study concentration data	false		
Sample based	true		
Consumptions on the same day come from the same sample	false		
Apply processing factors	true		
Use distribution	false		
Ignore processing factors less than 1	false		
Use unit variability	true		
Unit variability model	BetaDistribution		
Estimates nature	Realistic		
Unit variability parameter	VariabilityFactor		
Model type	OIM		
Model-then-add	false		
Covariate modelling	false		
Iterate survey	false		
Report consumptions and exposures per individual instead of per kg body weight	false		
Default concentration model	Empirical	*EC 2018 Tier 1*	*Concentration models*
Include MRL fallback model	false	*EC 2018 Tier 1*	*Concentration models*
Restrict LOR imputation to authorised uses	false	*EC 2018 Tier 1*	*Concentration models*
Censored values replacement	ReplaceByLOR	*EC 2018 Tier 1*	*Concentration models*
Factor f (f x LOR or f x LOD or LOD + f x (LOQ - LOD)	0.5	*EC 2018 Tier 1*	*Concentration models*
Sample based	true	*EC 2018 Tier 1*	*Concentration models*
Impute missing values from available values (if unchecked, missing values are imputed with 0)	true	*EC 2018 Tier 1*	*Concentration models*
Correlate imputed values with sample potency	true	*EC 2018 Tier 1*	*Concentration models*
Use occurrence frequencies for imputation	true	*EC 2018 Tier 1*	*Concentration models*
Parametric uncertainty	false	*EC 2018 Tier 1*	*Concentration models*

continues on next page

Table 3.101 – continued from previous page

Name	Setting	From input tier	In module
Apply occurrence pattern percentages	false	*EC 2018 Tier 1*	*Occurrence patterns*
Substance conversion method	UseMostToxic	*EC 2018 Tier 1*	*Concentrations*
Retain all allocated substances after active substance allocation	true	*EC 2018 Tier 1*	*Concentrations*
Account for substance authorisations in substance conversions	false	*EC 2018 Tier 1*	*Concentrations*
Fix duplicate substance allocation inconsistencies	false	*EC 2018 Tier 1*	*Concentrations*
Use extrapolation rules	true	*EC 2018 Tier 1*	*Concentrations*
Threshold for extrapolation	10	*EC 2018 Tier 1*	*Concentrations*
Restrict extrapolations to equal MRLs	true	*EC 2018 Tier 1*	*Concentrations*
Restrict extrapolations to authorised uses	true	*EC 2018 Tier 1*	*Concentrations*
Impute water concentrations	true	*EC 2018 Tier 1*	*Concentrations*
Water concentration value (µg/kg)	0.1	*EC 2018 Tier 1*	*Concentrations*
Restrict water imputation to the five most toxic substances	true	*EC 2018 Tier 1*	*Concentrations*
Restrict water imputation to authorised uses	false	*EC 2018 Tier 1*	*Concentrations*
Restrict water imputation to approved substances	false	*EC 2018 Tier 1*	*Concentrations*

EC 2018 Tier 2

Table 3.102: Tier definition for EC 2018 Tier 2.

Name	Setting	From input tier	In module
Total diet study concentration data	false		
Sample based	true		
Consumptions on the same day come from the same sample	false		
Apply processing factors	true		
Use distribution	false		
Ignore processing factors less than 1	false		
Use unit variability	true		
Unit variability model	BetaDistribution		
Estimates nature	Realistic		
Unit variability parameter	VariabilityFactor		
Model type	OIM		
Model-then-add	false		
Covariate modelling	false		

continues on next page

Table 3.102 – continued from previous page

Name	Setting	From input tier	In module
Iterate survey	false		
Report consumptions and exposures per individual instead of per kg body weight	false		
Default concentration model	Empirical	*EC 2018 Tier 2*	*Concentration models*
Include MRL fallback model	false	*EC 2018 Tier 2*	*Concentration models*
Restrict LOR imputation to authorised uses	false	*EC 2018 Tier 2*	*Concentration models*
Censored values replacement	ReplaceByLOR	*EC 2018 Tier 2*	*Concentration models*
Factor f (f x LOR or f x LOD or LOD + f x (LOQ - LOD)	0.5	*EC 2018 Tier 2*	*Concentration models*
Sample based	true	*EC 2018 Tier 2*	*Concentration models*
Impute missing values from available values (if unchecked, missing values are imputed with 0)	true	*EC 2018 Tier 2*	*Concentration models*
Correlate imputed values with sample potency	false	*EC 2018 Tier 2*	*Concentration models*
Use occurrence frequencies for imputation	true	*EC 2018 Tier 2*	*Concentration models*
Parametric uncertainty	false	*EC 2018 Tier 2*	*Concentration models*
Apply occurrence pattern percentages	true	*EC 2018 Tier 2*	*Occurrence patterns*
Scale up use frequency to 100%	true	*EC 2018 Tier 2*	*Occurrence patterns*
Restrict use percentage up-scaling to authorised uses	true	*EC 2018 Tier 2*	*Occurrence patterns*
Substance conversion method	DrawRandom	*EC 2018 Tier 2*	*Concentrations*
Retain all allocated substances after active substance allocation	true	*EC 2018 Tier 2*	*Concentrations*
Account for substance authorisations in substance conversions	true	*EC 2018 Tier 2*	*Concentrations*
Fix duplicate substance allocation inconsistencies	false	*EC 2018 Tier 2*	*Concentrations*
Use extrapolation rules	true	*EC 2018 Tier 2*	*Concentrations*

continues on next page

Table 3.102 – continued from previous page

Name	Setting	From input tier	In module
Threshold for extrapolation	10	*EC 2018 Tier 2*	*Concentrations*
Restrict extrapolations to equal MRLs	true	*EC 2018 Tier 2*	*Concentrations*
Restrict extrapolations to authorised uses	true	*EC 2018 Tier 2*	*Concentrations*
Impute water concentrations	true	*EC 2018 Tier 2*	*Concentrations*
Water concentration value (µg/kg)	0.05	*EC 2018 Tier 2*	*Concentrations*
Restrict water imputation to the five most toxic substances	true	*EC 2018 Tier 2*	*Concentrations*
Restrict water imputation to authorised uses	false	*EC 2018 Tier 2*	*Concentrations*
Restrict water imputation to approved substances	false	*EC 2018 Tier 2*	*Concentrations*

EFSA 2012 Pessimistic

Note: This tier is deprecated and has been replaced by separate acute/chronic tiers.

Probabilistic exposure assessment using the pessimistic model settings according to the EFSA Guidance 2012. Only processing factors > 1 are applied. For unit variability, the Beta distribution is applied.

Table 3.103: Tier definition for EFSA 2012 Pessimistic.

Name	Setting	From input tier	In module
Total diet study concentration data	false		
Sample based	true		
Consumptions on the same day come from the same sample	true		
Apply processing factors	true		
Use distribution	false		
Ignore processing factors less than 1	true		
Use unit variability	true		
Unit variability model	BetaDistribution		
Estimates nature	Realistic		
Unit variability parameter	VariabilityFactor		
Model type	OIM		
Model-then-add	false		
Covariate modelling	false		
Iterate survey	false		
Report consumptions and exposures per individual instead of per kg body weight	false		
Default concentration model	NonDetectSpikeLogNormal	*EFSA 2012 Pessimistic*	*Concentration models*
Include MRL fallback model	true	*EFSA 2012 Pessimistic*	*Concentration models*
Restrict LOR imputation to authorised uses	false	*EFSA 2012 Pessimistic*	*Concentration models*
Censored values replacement	ReplaceByLOR	*EFSA 2012 Pessimistic*	*Concentration models*
Factor f (f x LOR or f x LOD or LOD + f x (LOQ - LOD)	1	*EFSA 2012 Pessimistic*	*Concentration models*
MRL Factor (f x MRL)	1	*EFSA 2012 Pessimistic*	*Concentration models*
Sample based	true	*EFSA 2012 Pessimistic*	*Concentration models*
Impute missing values from available values (if unchecked, missing values are imputed with 0)	true	*EFSA 2012 Pessimistic*	*Concentration models*
Correlate imputed values with sample potency	true	*EFSA 2012 Pessimistic*	*Concentration models*
Parametric uncertainty	true	*EFSA 2012 Pessimistic*	*Concentration models*

Calculation of dietary exposures

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

- *Dietary exposures calculation*

Inputs used: *Consumptions by modelled food Concentration models Processing factors Unit variability factors High exposure food-substance combinations Active substances Occurrence patterns Relative potency factors Food conversions Concentration distributions*

Settings used

- *Calculation Settings*

3.4.3 High exposure food substance combinations

Identification of food-as-eaten/modelled food/substance combinations that have the highest expected contribution to exposure based on a simple screening model.

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

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

High exposure food substance combinations 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 foods as eaten (e.g. white bread) and modelled foods (e.g. wheat), and
- a drill-down can be made into the exact foods and substances contributing for simulated individuals or individual-days in the upper tail.

The number of combinations of simulation, substance, modelled food 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/substance combinations in the exposure calculation, and will therefore produce exactly the same cumulative exposure distribution, and allow to see contributions of all substances and all modelled foods. Details with respect to foods as eaten are however restricted to the risk drivers selected in the screening step. For more details see *screening calculation for large Cumulative Assessment Groups*.

The two-step approach consists of:

- **Step 1: Data screening and selection of risk drivers** Run a simple analysis for each potential source/substance combination (SCC). Here source means the combination of food as eaten and modelled food, for example apple in apple pie. The screening is based on this combination, and not just modelled foods, 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 censored and positive concentrations, respectively. SCCs (risk driver components) can be combined to measured source/substance 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 at the screening stage. Note, originally SCC stands for Source Compound Combination, MSCC for Measured Source Compound Combination.

- **Step 2: Full MC analysis** Perform the standard MC to all combinations of substances 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/substance 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 censored (non-detect or non-quantified) 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 censored 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 calculation *acute exposure* and *chronic exposure* .

Screening calculation for large Cumulative Assessment Groups

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 modelled food in source S and the concentrations of substance C in source S. A 'non-detect' consumption is assumed to be a zero consumption. A censored 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 3.104. 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 3.104: Parameters for screening models (per source/substance)

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 censored values (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 censored 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 censored 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 satisfies

$$(1 - \pi_c) \cdot \Phi\left(\frac{e_p - \mu_1}{\sigma_1}\right) + \pi_c \cdot \Phi\left(\frac{e_p - \mu_{12}}{\sigma_2}\right) = p'$$

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-substance 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).

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 exposure* 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 censored values. 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 (*OIM*), 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

$$P_i = Pr(e > e_{p,max}) = \pi_x \cdot \Phi\left(\frac{e_{p,max} - \mu_2}{\sigma_2}\right)$$

High exposure food substance combinations settings

Calculation settings

Table 3.105: Calculation settings for module High exposure food-substance combinations.

Name	Type	Description
Percentage defining the exposure percentile of interest per food-as-eaten/food-as-measured/substance combination	Numeric	Percentage defining the exposure percentile of interest per food-as-eaten/food-as-measured/substance combination.
Include risk drivers to include minimally a percentage	Numeric	The selection criterion for the risk drivers. The cumulative contribution percentage of the selected risk drivers will be this percentage.
Censored value replacement: factor x LOR	Numeric	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 censored values.

Calculation of high exposure food-substance combinations

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

- *High exposure food-substance combinations calculation*

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

Settings used

- *Calculation Settings*

3.4.4 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 Biological matrix concentration comparisons Risks*

Exposures calculation

The main goal of the exposures module is to calculate the exposure at the internal level. External exposures from dietary and non-dietary routes are aggregated at a specified target compartment (organ) in two steps:

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

Absorption factors, just simple multiplication factors or more advanced *PBK models* aggregate the exposures from multiple routes into exposure at the target compartment. Currently, only dietary exposures or dietary exposures combined with non-dietary exposures are aggregated at the target compartment. Internal exposure calculation of non-dietary sources only, is yet not available but will be implemented in the future.

In cumulative internal 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 substances are individuals exposed? Just the co-occurence of substances is observed.
2. quantitative approach: *maximum cumulative ratio (MCR)*. To what degree are mixtures more important than single substances? The relative exposure levels of the substances present in a measurement, e.g. an individual (chronic) or individual day (acute) are taken into account.

In the *exposures mixtures module*, two more advanced approaches are available to analyse the co-occurence of substances, the *SNMU approach* and a *network analysis*.

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 surveys, 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.

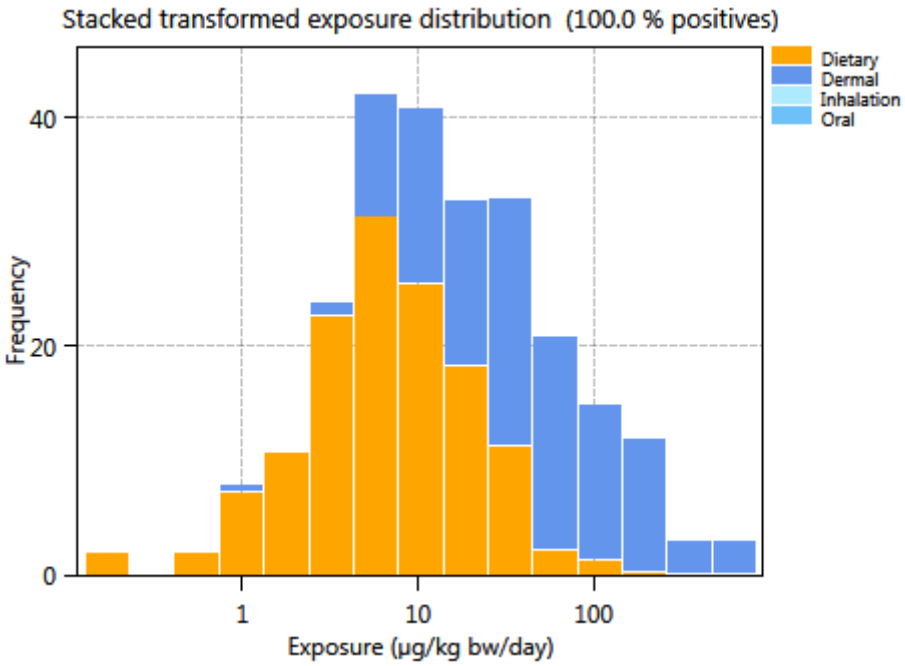


Figure 3.36: Aggregate exposure distributions.

Example 1

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

Table 3.106: 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 3.107: NonDietarySurveys

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

Table 3.108: 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 3.106 and the user has provided Table 3.108 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 3.108 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 **IndividualProperties** 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 3.109: 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 3.110: NonDietarySurveys

idNonDietary-Survey	Description	Loca-tion	Date	NonDietaryIntakeU-nit
1	BROWSE, acute, cumulative, opera-tors	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 3.109 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 3.109) and the individual matches the specified demographic criteria for the survey, as specified in Table 3.108. (In this example, a default exposure would apply to all individuals not listed in Table 3.109 because Table 3.108 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 3.111: 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 3.112: NonDietarySurveys

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

Table 3.113: 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 3.111 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 3.113. (In fact for the same result rather than changing the values in the *idIndividual* column in Table 3.109 from the previous example may be used with matching switched off). Exposures in Table 3.111 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 3.114: 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.

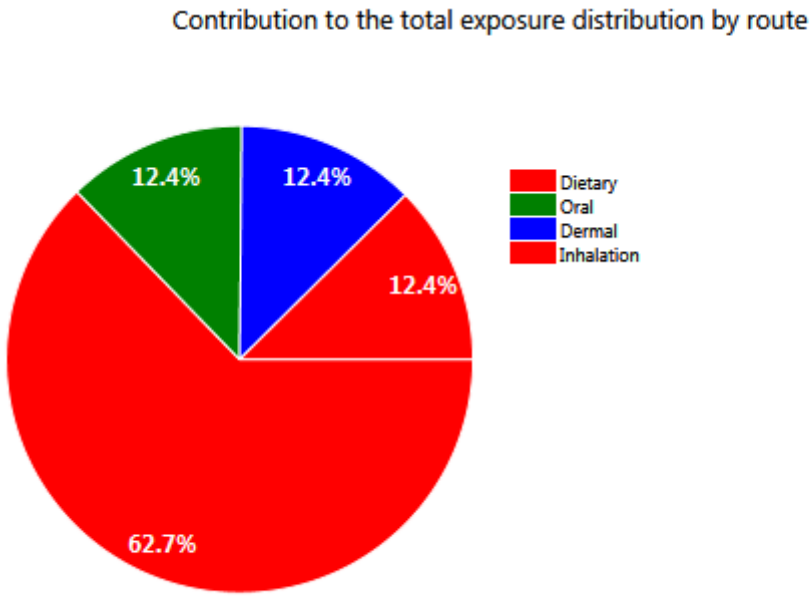


Figure 3.37: Contributions by route to aggregate exposure distributions.

See *non-dietary exposure settings*.

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 3.38 for acute exposure assessments, and in Figure 3.39 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.

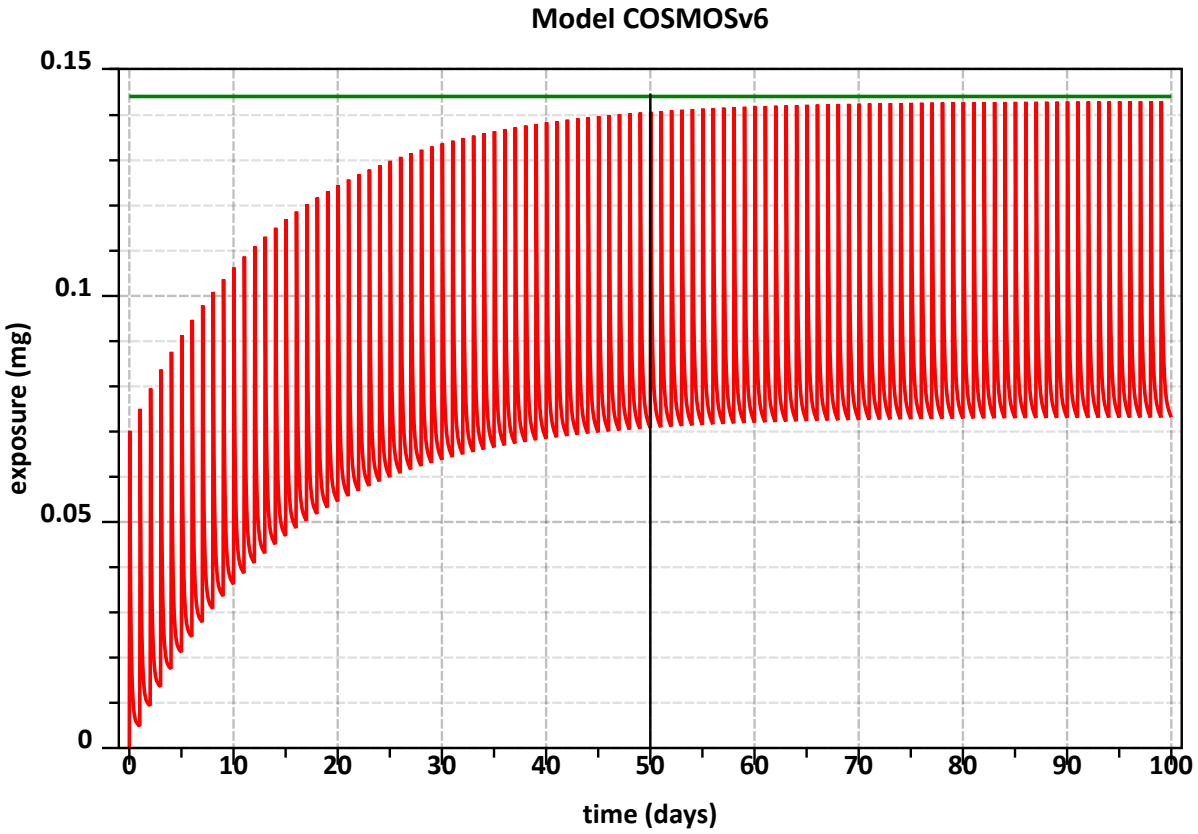


Figure 3.38: Time course of the internal substance amount when applying the same single dose on each day. The acute internal concentration is derived as the peak substance amount (the green line in the figure) divided by the compartment weight. The vertical line at 50 indicates the selected end of an 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_{peak}) for deriving acute internal exposures is as follows:

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

Here, $d(t)$ denotes the substance amount at time t, t_{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_{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

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

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

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

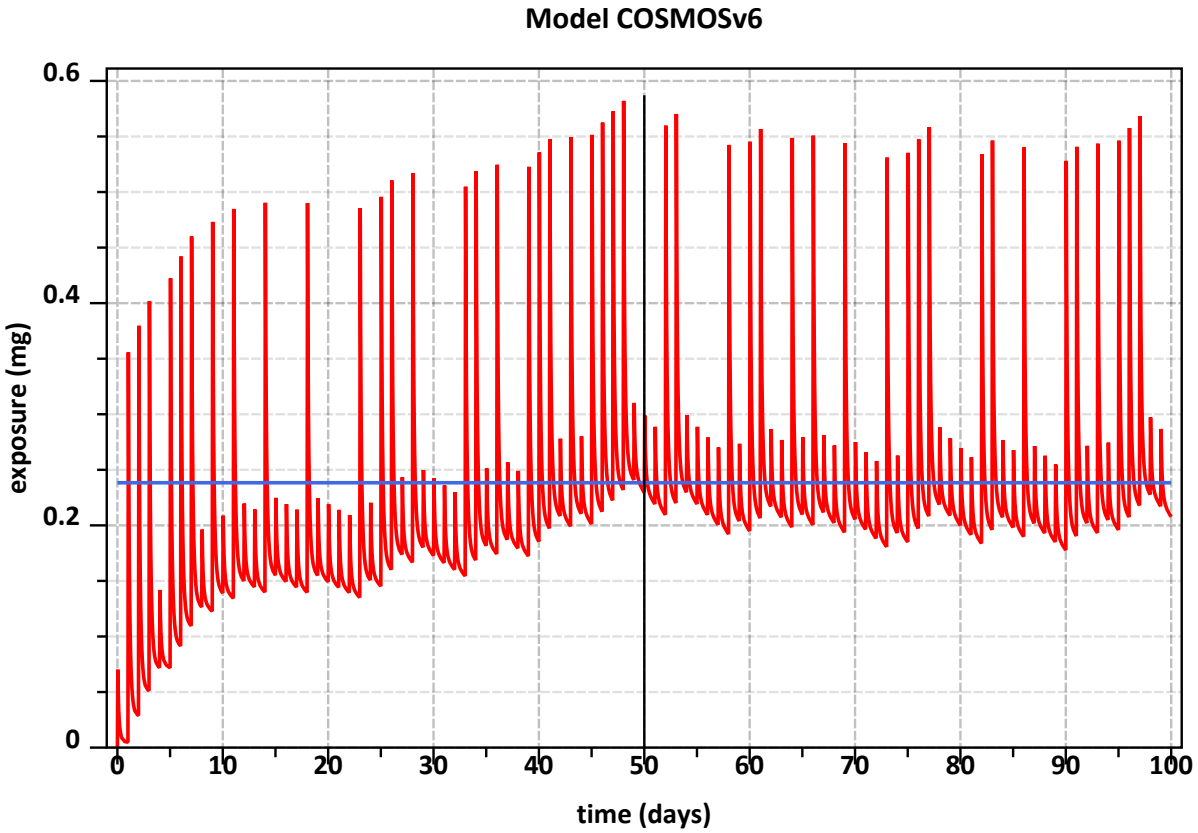


Figure 3.39: Time course of the internal substance amount when randomly applying one of the individual-day doses for a number days. The chronic internal concentration is derived as the average substance amount (the blue line in the figure), divided by the compartment weight. The vertical line at 50 indicates 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 3.115. There is no cut-off level, the only criterion is the presence of a substance in the simulated daily diet (eventually aggregated withe the exposure from non-dietary sources) 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 3.115: 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 3.116, the frequency and percentage for the number of substances occurring together are shown.

Table 3.116: 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 3.117, the mixtures containing the substance(s) including all other combinations with the specified combination of substance(s), (a maximum number of 15 records is shown).

Table 3.117: 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 (2011) 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:

$$\text{MCR} = \frac{\text{Cumulative exposure}}{\text{Maximum exposure}}$$

This MCR statistic is also picked up as a practical device in a recent JRC report, Bopp et al. (2015), 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. For the chronic case long-term individual exposures (estimated by aggregating over the available survey days of each individual) are used.

Risk based, standardised or unweighted exposures

Before calculating the MCR statistics, three optional choices are available, *see settings, MCR exposure approach type*.

- Risk based exposures: exposures are multiplied by the *relative potency factor* (RPF) of each substance and thus exposures for different substances are on the same and comparable scale.
- Standardised exposures: all exposures are standardised to equal variance (unit variance).
- Unweighted exposures: exposures are taken as such, this is equivalent to RPF s equal to 1 for each substance.

Table 3.118 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 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, with the substance with the highest exposure in parenthesis. 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 3.118: 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 this artificial example, all days have equal values for total exposure (= 1). 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 3.40, French steatosis data (39 substances, 4079 persons) are used to calculate the chronic exposure matrix. For each individual the MCR is calculated and plotted against the total exposure. The different colours 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 p_5, p_{50} and p_{95} percentiles are indicated with the black line segments. The red line indicates the ratio with value 5. The dashed green lines indicate the p_{95} percentiles for the MCR value for different ranges of the total exposure.

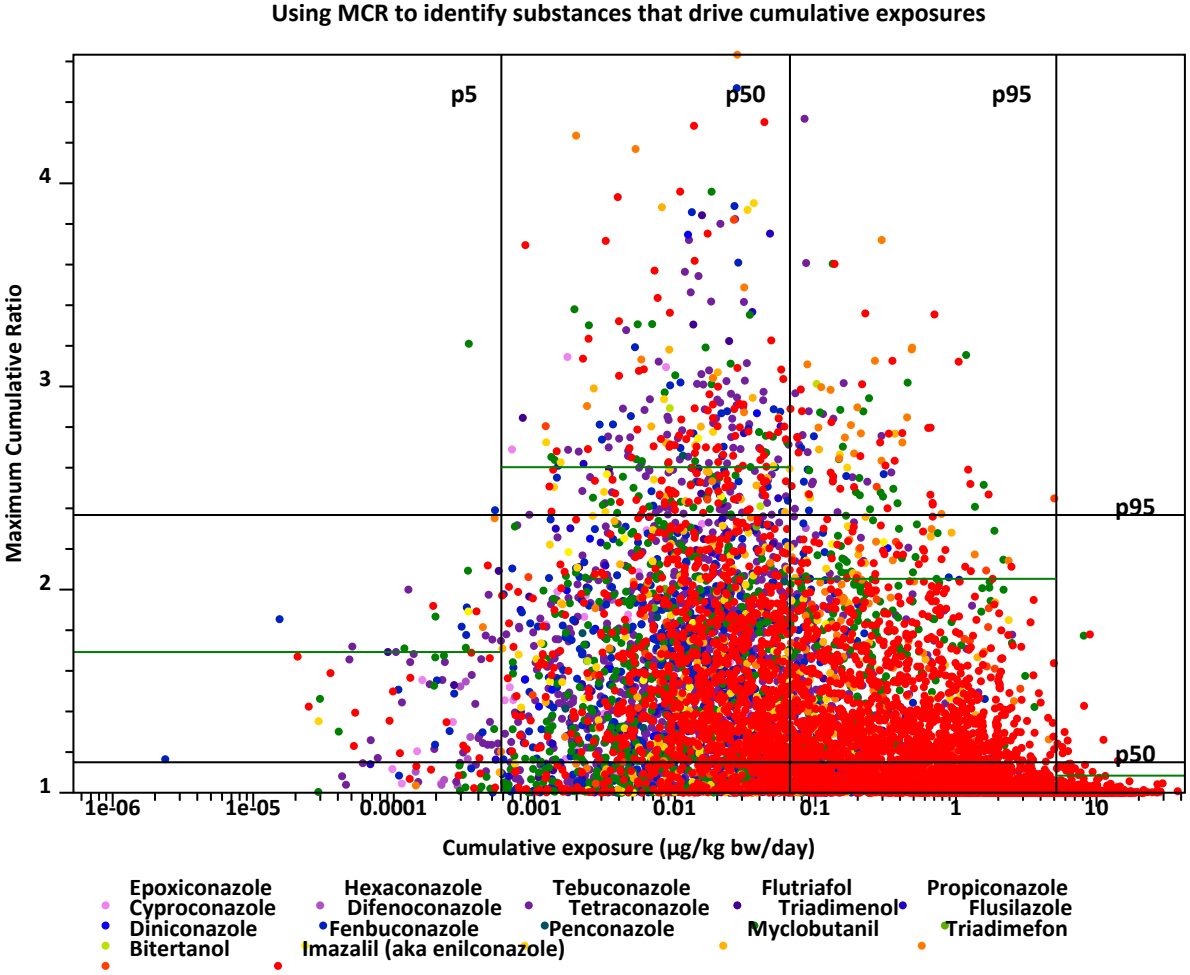


Figure 3.40: Maximum Cumulative Ratios vs total exposure

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.

Because Figure 3.40 can be very dense, in Figure 3.41, 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).

In Figure 3.42 and Figure 3.43 scattered MCR distributions for the total and upper tail (here 37%) that drive the cumulative exposure are shown. The red line indicates the MCR threshold, 1.5. The black lines represent the regression lines MCR vs ln(Cumulative exposure) for each tail. Substances with an exposure contribution less than 15% are not displayed.

In Table 3.119 contributions to tail exposures at various percentile are shown. Column MCR = 1 shows the percent-

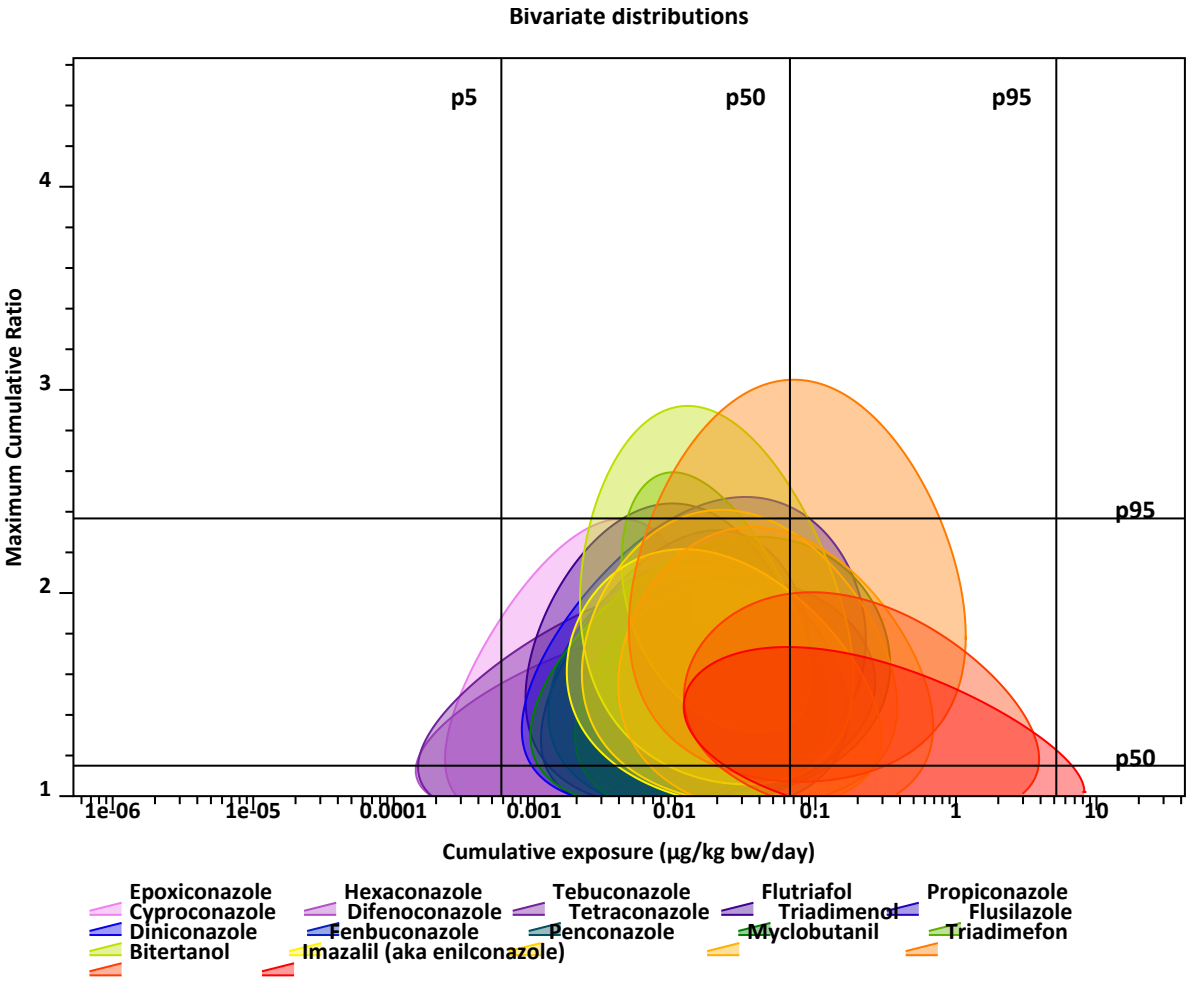


Figure 3.41: Bivariate distributions MCR vs total exposure

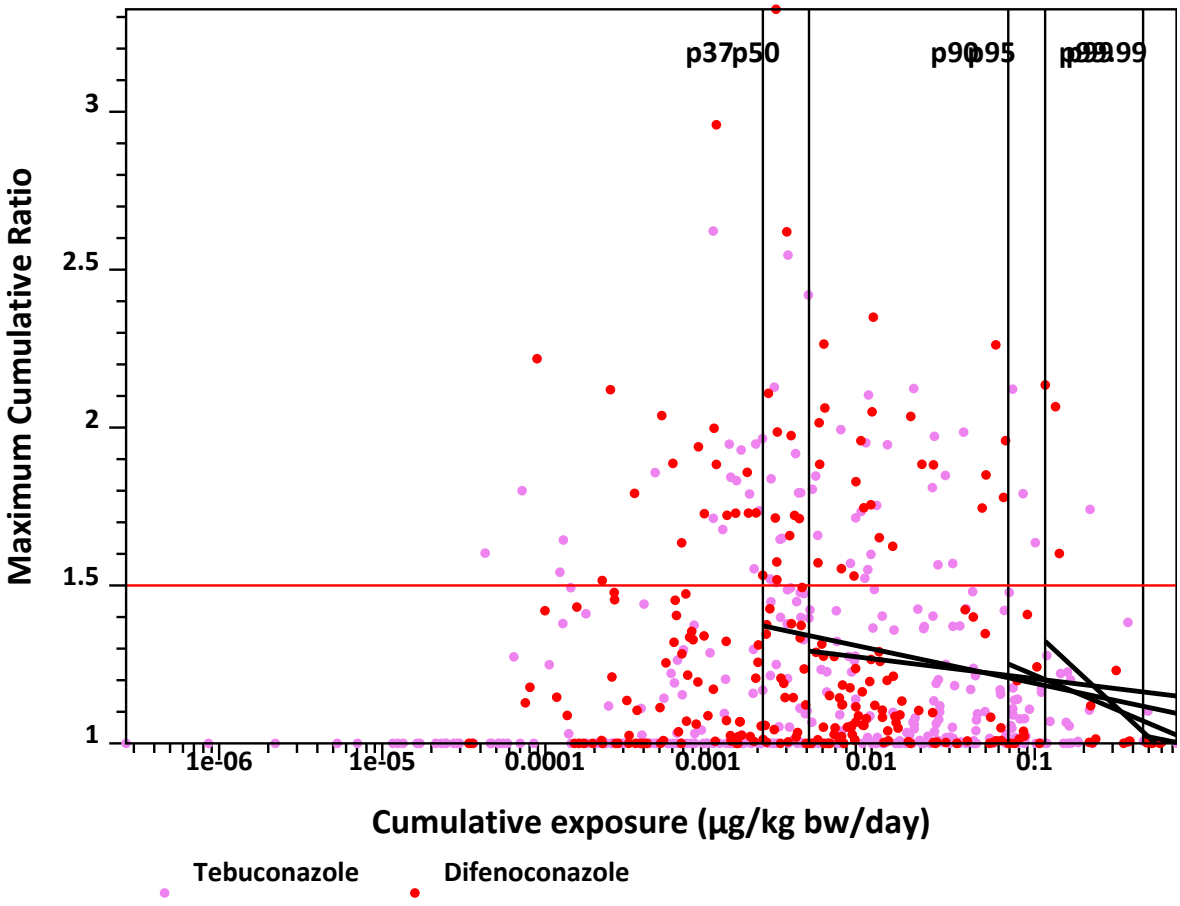


Figure 3.42: Using MCR to identify substances that drive cumulative exposures, scatter distributions (total).

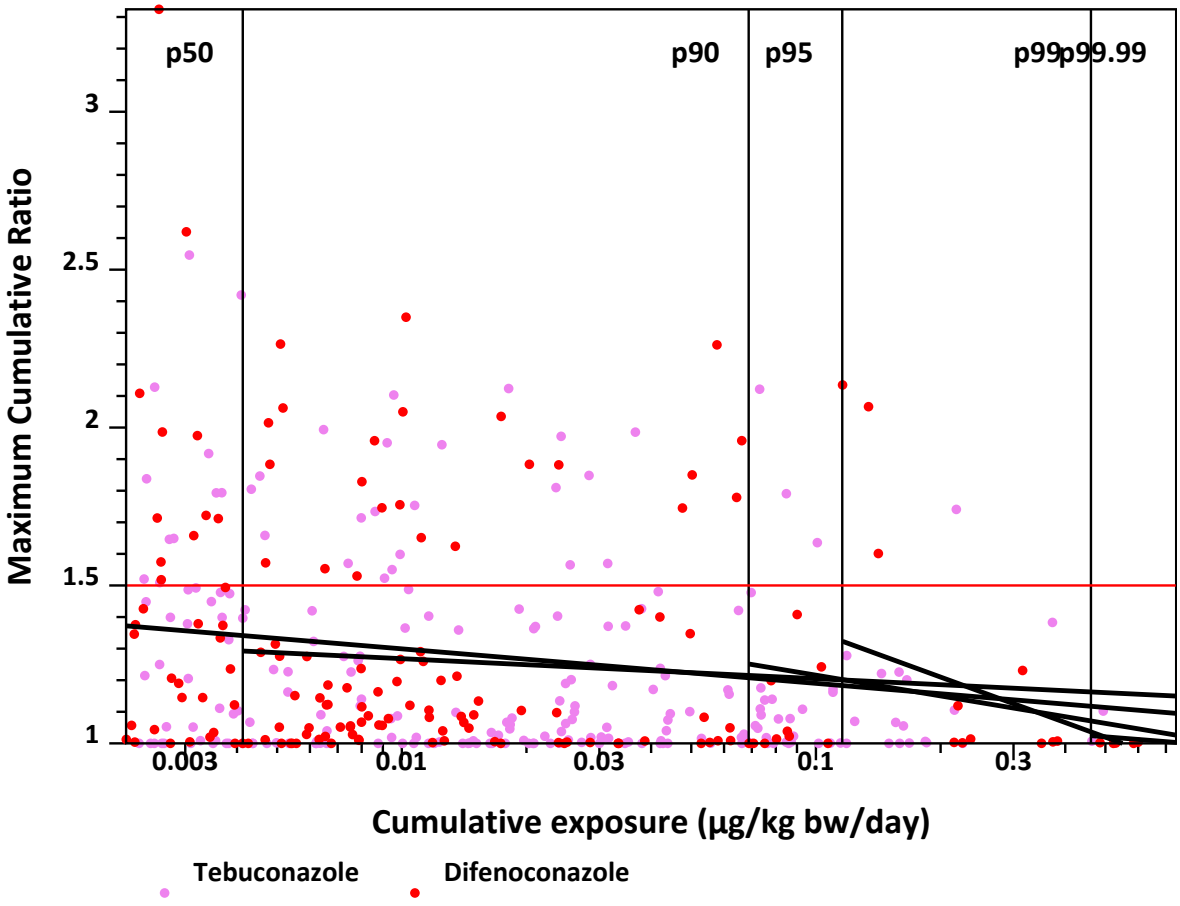


Figure 3.43: Using MCR to identify substances that drive cumulative exposures, scatter distributions (upper tail 37%).

age of tail exposure due to individual(day)s with a single substance. Column $1 < \text{MCR} \leq 2$ shows the percentage of tail exposure due to individual(day)s with multiple substances, but the MCR ≤ 2. Column MCR > 2 shows the percentage of tail exposure due to individual(day)s with multiple substances with MCR > 2.

Table 3.119: Maximum Cumulative Ratio summary

Tail %	% with MCR = 1	Sub-stances	% with 1 < MCR<=2	Substances	% with MCR > 2	Sub-stances
37	20.6	Difeno, Tebu	73.7	Difeno, Tebu	5.7	Difeno, Tebu
50	19.2	Difeno, Tebu	75.6	Difeno, Tebu	5.2	Difeno, Tebu
90	16.3	Difeno, Tebu	78.8	Difeno, Tebu	5.0	Difeno, Tebu
95	15.0	Difeno, Tebu	82.5	Difeno, Tebu	2.5	Difeno, Tebu
99	25.0	Difeno	75.0	Difeno, Tebu Propi	0.0	

To configure the MCR plot, see *dietary exposures settings*, *human monitoring analysis settings* and *exposures settings* with options to display the ratio total exposure/ maximum for individual(day) exposures (MCR plot), to specify tail percentiles of the exposure distribution, e.g. 95, 97.5 and 99% (MCR plot) or to set the minimum percentage contribution per substance to the tail exposure (MCR plot).

Exposures settings

Calculation settings

Table 3.120: Calculation settings for module Exposures.

Name	Type	Description
Risk type	*ExposureType*	The type of exposure considered in the assessment; acute (short term) or chronic (long-term).
Multiple substances analysis	Boolean	Specifies whether the assessment involves multiple substances.
Compute cumulative exposures	Boolean	Specifies whether the assessment involves multiple substances a results should be cumulated over all substances.
Include dietary and non-dietary routes of exposure	Boolean	Specifies whether the assessment involves both dietary and non-dietary (oral, inhalatory or dermal) routes of exposure.
Target level	*TargetLevelType*	Select to express hazard characterisations at external or internal exposure level. For an aggregate assessment, that is dietary and nondietary exposure data are combined, the target dose level is always internal. When only dietary exposures are available, the target dose level is optional, i.c. external or internal.
Match non-dietary to dietary survey individuals	Boolean	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	Boolean	If checked, exposures from identical individuals in non-dietary surveys are aggregated to obtain the overall nondietary exposure If unchecked, exposures from random individuals in all non-dietary surveys are aggregated.
Model-then-add	Boolean	Specifies whether to create separate exposure models for specif groups of modelled foods (model-then-add).
Kinetic model	AlphaNumeric	Code Kinetic Model.
Specify the type of kinetic model	*InternalModelType*	Specify the type of model to convert external exposure to the internal level.
Substance weighting for MCR	*ExposureApproachType*	Risk based: exposures in equivalents of the reference substance standardised: standardised exposures per substance have varian 1; or unweighted exposures: RPFs are equal to 1.
Perform MCR analysis	Boolean	Perform a Maximum Cumulative Ratio (MCR) analysis to determine co-exposure between substances.
Display ratio total exposure/ maximum (in MCR plot)	Numeric	For MCR plot: specify ratio total exposure/ maximum for individual(day) exposures .
Show tail percentiles (MCR plot) for:	AlphaNumeric	Give specific percentiles of exposure distribution (%), e.g. 97.5 99 (space separated).
Set minimum percentage contribution per substance to the tail exposure (MCR plot)	Numeric	Set minimum percentage contribution per substance to the tail exposure.

Output settings

Table 3.121: Output settings for module Exposures.

Name	Type	Description
Include drill-down on 9 individuals around specified percentile.	Boolean	Specifies whether drilldown on 9 individuals is to be included in the output.
Summarize simulated data	Boolean	Specifies whether a summary of the simulated consumptions an concentrations should be included in the output.
Store simulated individual day exposures	Boolean	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	AlphaNumeric	Give specific percentiles of exposure distribution (%), e.g. 50 9 95 97.5 99 (space separated).
Percentage for drilldown	Numeric	Gives detailed output for nine individuals near this percentile of the exposure distribution.
Percentage for upper tail	Numeric	Gives detailed output for this upper percentage of the exposure distribution.
Show % of population below level(s)	*ExposureMethod*	This setting is used for reporting the percentages of individuals (chronic) or individual days (acute) exceeding certain exposure levels. These exposure levels can be generated automatically based on the observed exposures (Automatic, default) or specifi explicitly (Manual).
Exposure levels	AlphaNumeric	Specify exposure levels for which to give the percentage of exposure below these levels, e.g. 1 10 50 100 200 500.
Number of levels of covariable to predict exposure	Numeric	Specify the number of levels, e.g. 20. The range of the covariat is divided by the number of levels: range = (max - min)/levels. For these covariable levels exposures are predicted.
Predict exposure at extra covariable levels	Numeric	Specify specific prediction levels in addition to the automaticall generated prediction levels (space separated).
Lower percentage for variability (%)	Numeric	The default value of 25% may be overruled.
Upper percentage for variability (%)	Numeric	The default value of 75% may be overruled.
Report consumptions and exposures per individual instead of per kg body weight	Boolean	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 3.122: Uncertainty settings for module Exposures.

Name	Type	Description
Resample kinetic model parameter values	Boolean	Specifies whether kinetic model parameter values are resample

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*

3.4.5 Exposure mixtures

Exposure mixtures will select sets of co-occurring substances (one or more) that contribute most to the overall exposure patterns.

This module has as primary entities: *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 substances 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 substances to which an individual in a population is exposed can be numerous. The exposures mixtures module is used to identify the most relevant components to which a population is exposed.

Risk based, standardised or unweighted exposures

Before performing the mixture exposure assessment, the data are preprocessed. Three optional choices are available, *see settings, exposure approach type*.

- Risk based exposures: exposures are multiplied by the *relative potency factor* (RPF) of each substance and thus exposures for different substances are on the same and comparable scale.
- Standardised exposures: all exposures are standardised to equal variance (unit variance).
- Unweighted exposures: exposures are taken as such, this is equivalent to RPF s equal to 1 for each substance.

Exposure mixtures are identified using a quantitative approach: *sparse non-negative matrix underapproximation (SNMU)* (Gillis and Plemmons (2013)) and answers the question what combination of substances predominantly determine the underlying patterns in the exposure matrix (substance x person (day)).

After identifying the components, a cluster analysis is applied to group individuals with similar profiles of exposure to the obtained component (Crépet et al. (2022)).

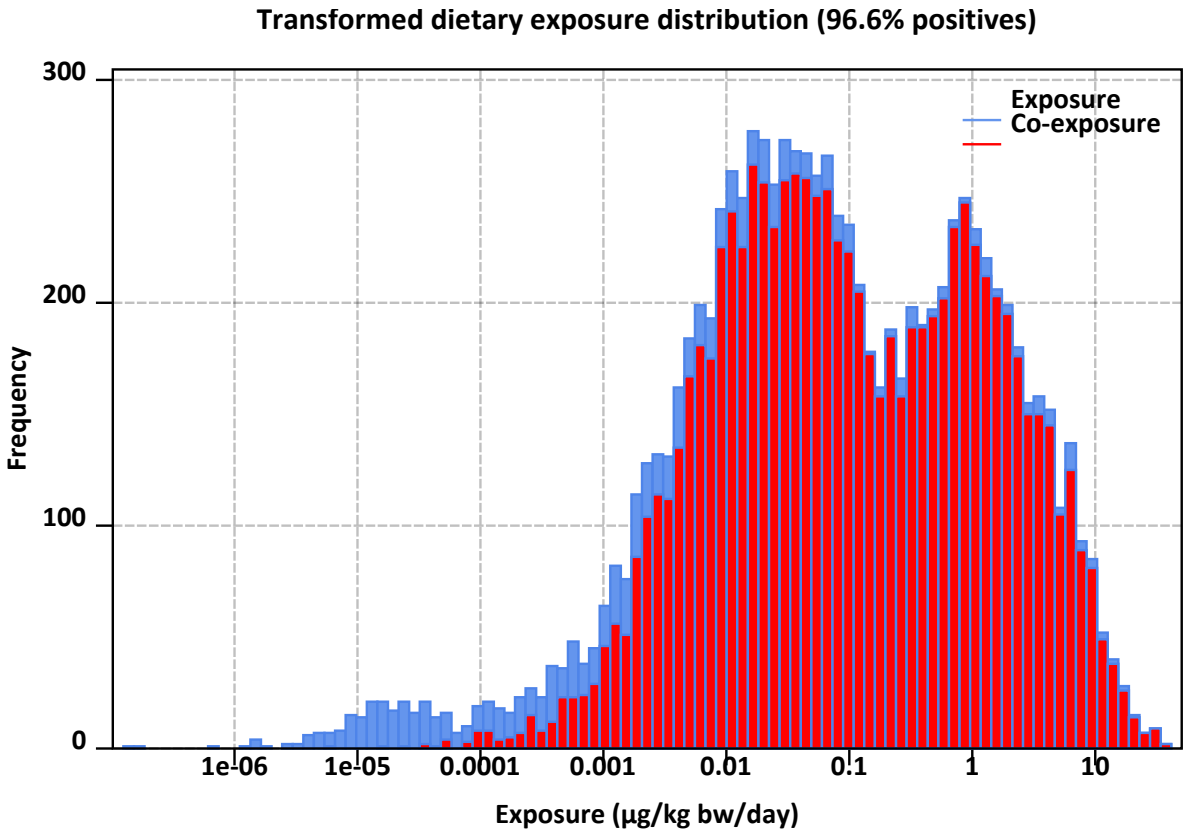


Figure 3.44: Example of co-exposure distribution (from >1 substance per individual-day, red) super-imposed on the total exposure distribution (blue).

Sparse nonnegative matrix underapproximation

Starting point to identify major components of combinations of substances using exposure data was to use Non-negative Matrix Factorization (NMF). This algorithm was proposed by Lee and Seung (1999) and Saul and Lee (2002) and deals specifically with non-negative data that have excess zeros such as exposure data. Zetlaoui et al. (2011), introduced this method in food risk assessment to define diet clusters.

The NMF method was then implemented by Béchaux et al. (2013) in order to identify food consumption patterns and main components of combinations 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 substances per component experiments were made using Sparse Nonnegative Matrix Factorization (SNMF) (Hoyer (2004)). This method was found not to give stable solutions. Better results were obtained with Sparse Nonnegative Matrix Underapproximation (SNMU) (Gillis and Plemmons (2013)). This model also fits better to the problem situation because also the residual matrix after extracting some components 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 3.45, the exposure matrix M with dimensions m (number of substances) 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 components. Matrix U contains weights of the substances per component, matrix V contains the coefficients of presence of components 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 criterion is: $||M–UV||2$ such that $U \geq 0$ and $V \geq 0$.

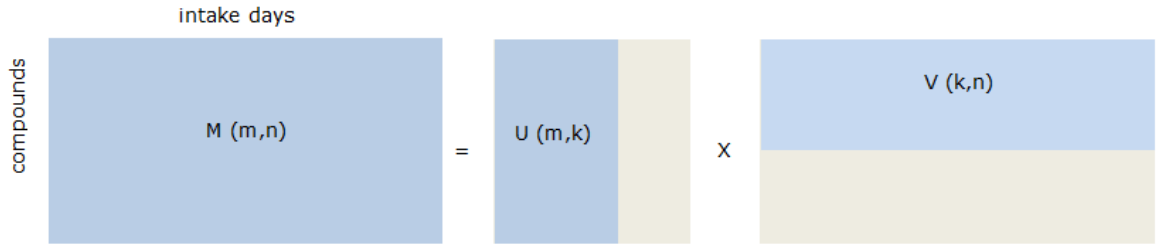


Figure 3.45: NMF approximation of exposure data

The solution found by NMF contains zeros, but components still contain many substances which complicates interpretability. Therefore, the Sparse Nonnegative Matrix Underapproximation (SNMU) (Gillis and Plemmons (2013)) 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 component, 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 component:

- u is $m \times 1$ vector, contains weights of the component.
- v is $1 \times n$ vector, contains presence of component 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 component $(k = 1)$ is:

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

For the second component:

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 component $(k = 2)$ 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 component. Continue with the third component etc….

The given factorization by SNMU is not unique. A matrix D and its inverse can be used to transform the two lower rank matrices U and V by, e.g.

$$M = UV = UDD^{-1}V$$

where matrix D is non-negative and corresponds to a scaling of matrix U and V to matrix $\tilde{U} = UD$ and $\tilde{V} = D^{-1}V$.

In Figure 3.46, SNMU is applied on the exposure matrix. The SNMU solution after scaling results in a matrix that represents the mixture composition and a matrix representing the individual scores. The first matrix is interpreted as the set of contributions of the substances to a component., the second matrix, as the set of exposures of the individuals to the mixtures.

Exposure matrix

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

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

where y_{ijc} is the exposure to substance 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 substance 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

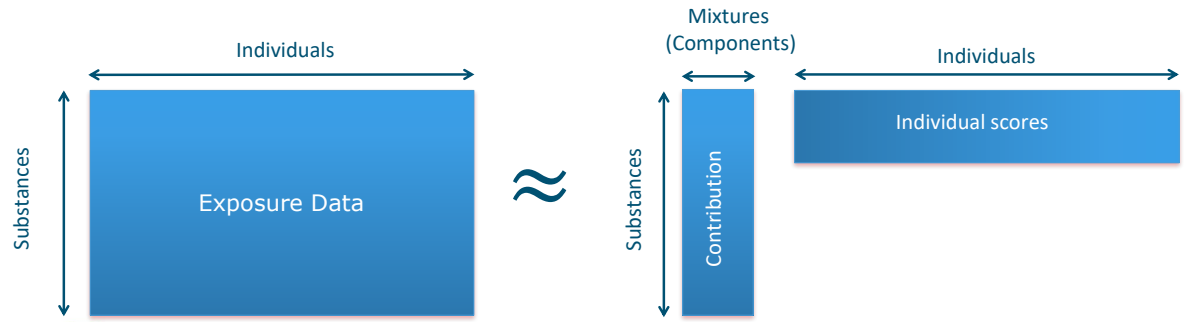


Figure 3.46: SNMU approximation of exposure data after scaling.

matrix summarises the y_{ijc} with columns denoting the individual-days (ij) and rows denoting the substances (c). Each cell represents the sum of the exposures for a substance 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 substance averaged over the total number of intake days of an individual (substances x persons). So each row represents a substance and a column an individual. Each cell represents the observed individual mean exposure for a substance for that individual. Note that in the exposure calculation, the concentration is the average of all residue values of a substance.

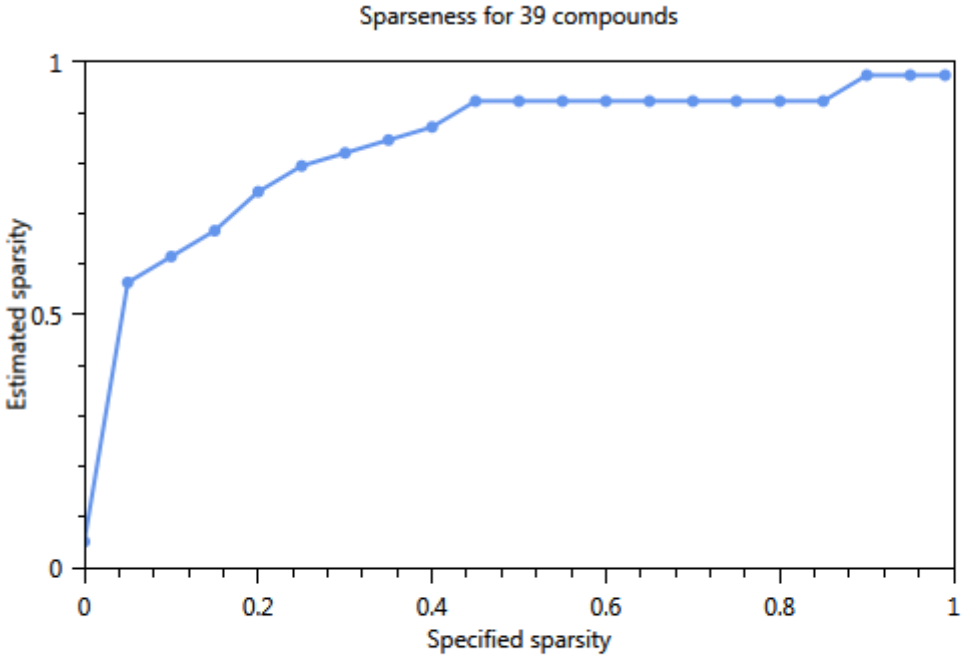
When *relative potency factors* (RPF) are available then exposures are multiplied by the RPF and thus exposures to the different substances are on the same and comparable scale (toxicological scale). In this case, the selection of the components is risk-based. In some cases, RPFs may not be available. In this case exposure to different substances, even in the same unit, may lead to very different effects. To give all substances an equal weight a priori and avoid scaling effects, a standardisation of the data can be applied as done in Béchaux et al. (2013). In this case, all substances are assigned equal mean and variance, and the selection of the components will work on patterns of correlation only. Then component selection is not risk-based anymore but, what could be called, co-exposure-based.

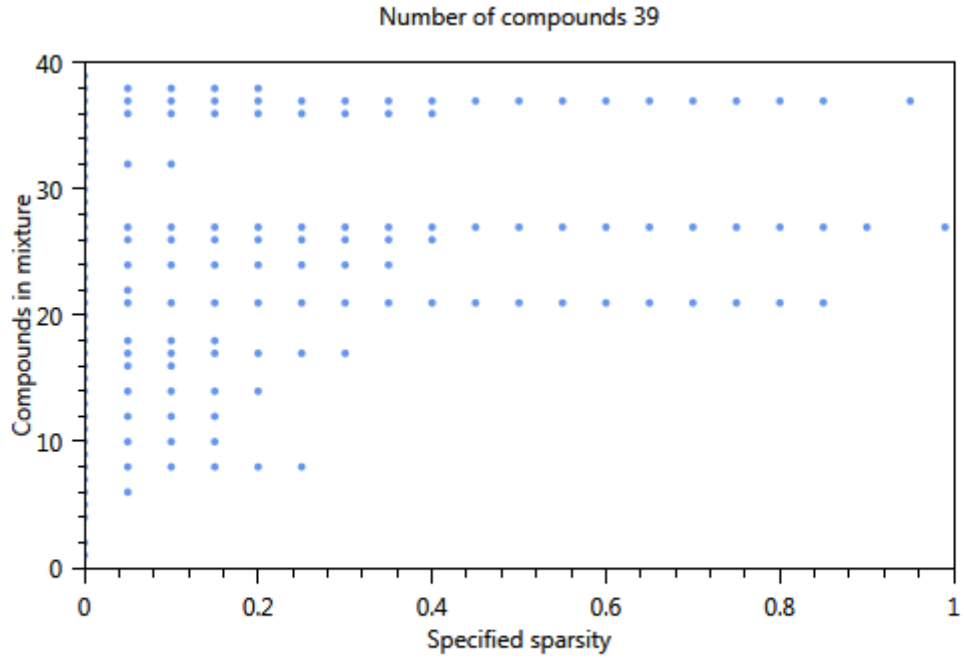
Finally, in the mixture selection module of MCRA, the SNMU expects RPF data for a risk-based selection. If not available, the user should provide alternative RPF data, e.g. values 1 for a purely exposure-based selection, or lower-tier estimates (e.g. a maximum value on RPF thought possible).

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 components will be more sparse (sparsity close to 0: not sparse, many substances; sparsity close to 1: sparse, few substances). 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 substance (ratio close to 1). To illustrate the combined use of MCR and the sparsity parameter, the French steatosis data (39 substances, 4079 persons) are used and person days with a ratio > 5 (see Figure 3.40) are selected yielding a subset of 139 records.

In Figure 3.47, the effect of using a cut-off level is immediately clear. The number of substances of the first component is 17 whereas in the unselected case (not shown) only 4 substances were found The three plots show the influence of increasing the sparsity parameter from 0 to 1 on the number of substances in the component. The first plot shows the estimated sparsity versus the specified sparsity. The second and third plot the the numer of substances in a component. For values close to 0, the component contains 17 substances. For values > 0.4 the number of substances in the component drops to 3 and for a sparsity close to 1, only one substance is found in a component.


Sparseness for 39 compounds
Estimated sparsity
Specified sparsity
0
0.5
1
0
0.2
0.4
0.6
0.8
1


Number of compounds 39
Compounds in mixture
Specified sparsity
0
10
20
30
40
0
0.2
0.4
0.6
0.8
1

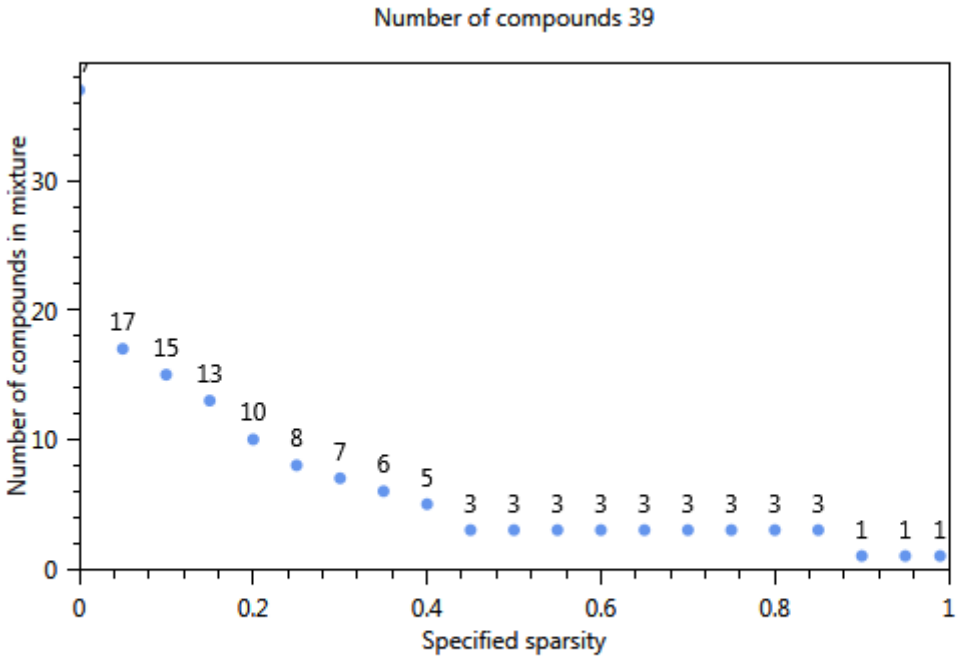


Figure 3.47: Influence of the specified sparsity parameter on the realized sparsity, n = 139.

Substance contributions to components

The SNMU solution of matrix U can be displayed in a heatmap. The heatmap shows the relative contribution of each substance to a component.

In Figure 3.49 and Figure 3.50 the sparsity parameter is set to 0.1 (not sparse) and 0.8 (sparse), respectively. This leads to components containing different number of substances.

In Figure 3.51, the relative contributions of the substances to the first component are displayed in a piechart.

As mentioned before, one of the nice features of the SNMU algorithm is its recursive character which results in identical components. In Figure 3.52, the U matrix is visualized using three components. Compare this solution with Figure 3.50, the first three components are identical. Because of the ordering the plots look slightly different, but a closer inspection of the first 3 components of each solution shows that they are the same. In both figures, component 1 contains mbzp, A, B, C and D; component 2, mibp, E, F and ohmehp; and component 3 mnbp and G.

In paragraph *network analysis*, an alternative to the SNMU approach is proposed.

For selection of individual(day) exposures with a *maximum cumulative ratio* above a cutoff and/or above a cutoff percentage in the set of individual(day)s ranked on total exposure, see *'Cutoff MCR' and 'Cutoff percentage' settings*.

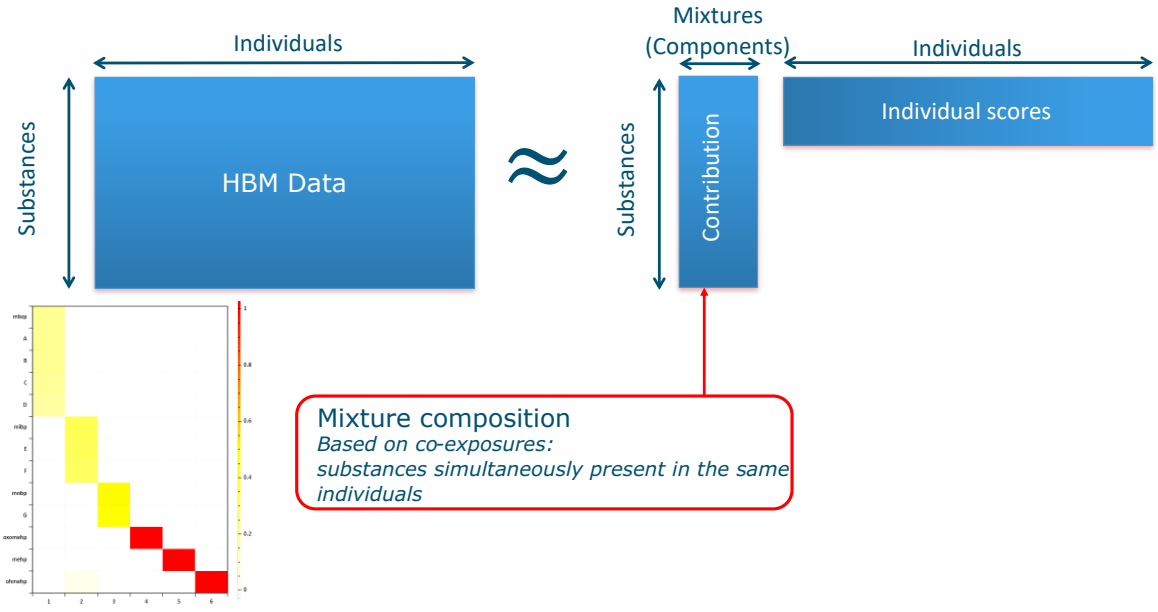


Figure 3.48: SNMU: matrix U, substance contributions to components.

Component exposures in population and subgroups

The SNMU solution of matrix V is used to group individuals with similar mixture exposure profiles. In Figure 3.53, the idea of clustering is shown.

Crépet et al. (2022) propose to identify components by coupling statistical criteria with the relevance of combined exposure profiles and component composition. First, the optimal choice for k, the number of components, is determined using a trade off between the decrease of the residual sum of squares and number of components. Then, hierarchical clustering was applied to the matrix of individual scores V to group individual(day)s with similar exposure profiles to the k components. This identification of components is repeated for different values of k where inspection of components not relevant to characterize a cluster, or concerned with only a small part of the population leads to rejection of the mixture.

In MCRA, two clustering methods are availabe. The first, hierarchical clustering, is implemented as described in Crépet et al. (2022). Ward's clustering criterion is implemented using Euclidean distances (Ward.D2, Murtagh and Legendre (2014)) . Specification of the optimal number of clusters is not needed. Results of the clustering are displayed in a dendrogram, Figure 3.54. The second one, based on K-means, requires specification of the number of clusters. The results of the clustering are represented in a scatter plot using principal components and convex envelopes to identify the clusters, Figure 3.56.

In Figure 3.55, the relative exposure to componaents in the total populations are shown. These plots are also available for the subgroups resulting from the clustering.

Advantages of K-means clustering is that it is simple and fast and large datasets can be handeled. Visualisation for large data sets is straightforward but for hierarchical clustering dendrograms maybe very dense. Disadvantage of K-means is that it requires the number of clusters set before. For large datasets, hierarchical clustering maybe slow, O(n^2), but for small datasets, the dendrogram helps in interpreting the results and in selecting the optimal number of clusters.

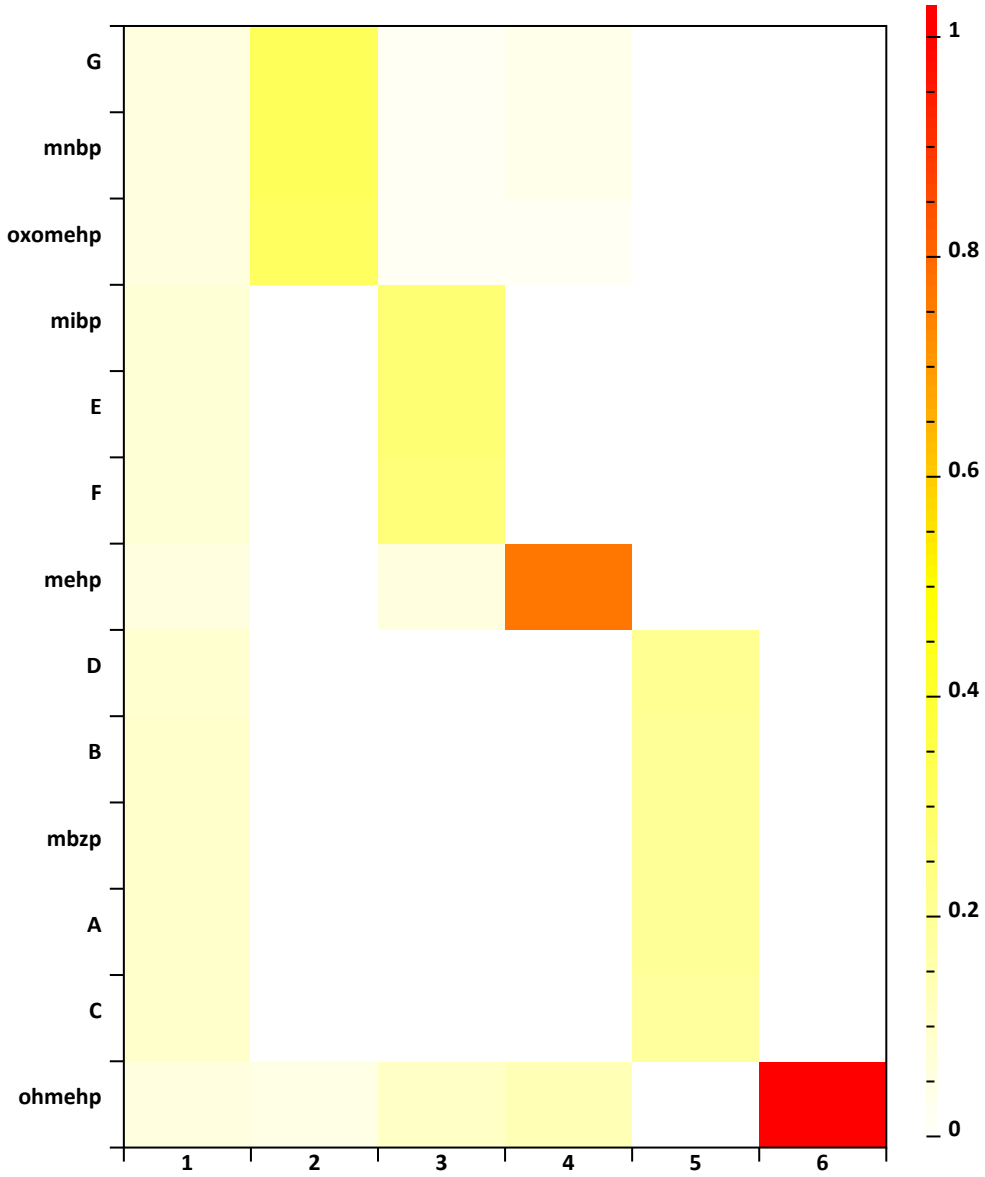


Figure 3.49: Co-exposure of substances. Heatmap for a solution with 6 components. The sparsity = 0.1. Each component, especially the first, contains many substances (see also Figure 3.50).

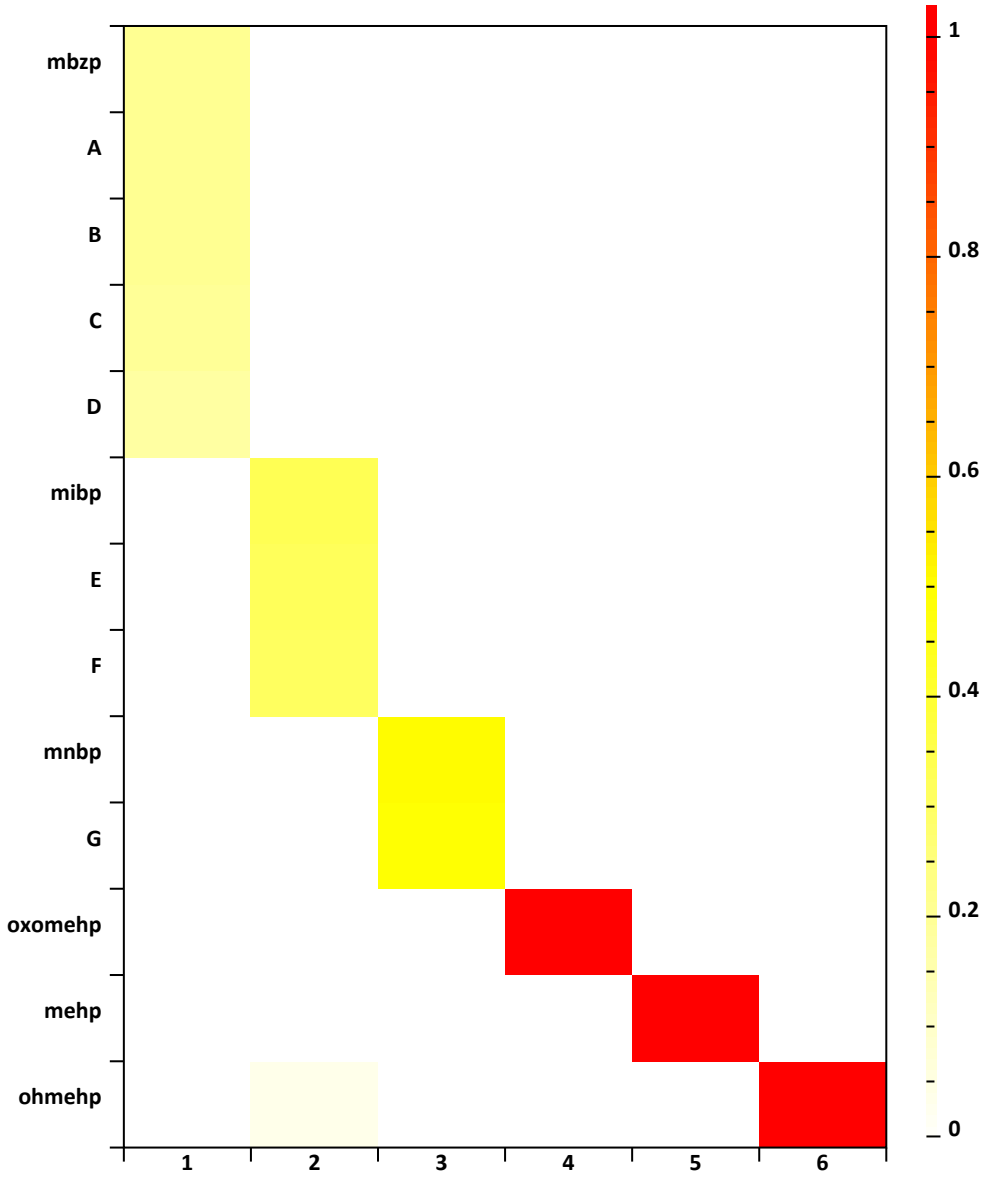


Figure 3.50: Co-exposure of substances. Heatmap for a solution with 6 components. The sparsity = 0.8. Components contain less substances compared to Figure 3.49.

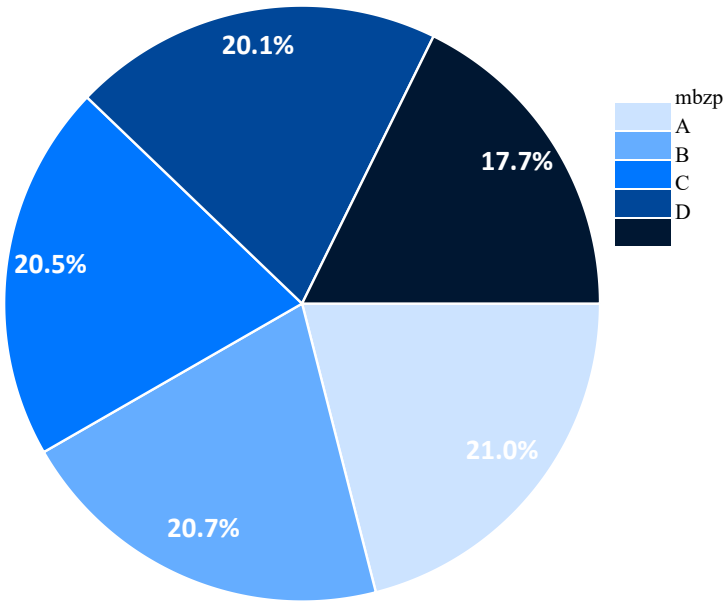


Figure 3.51: Relative contributions substances to component 1. The sparsity is set to 0.8 (sparse), estimated sparsity = 0.62.

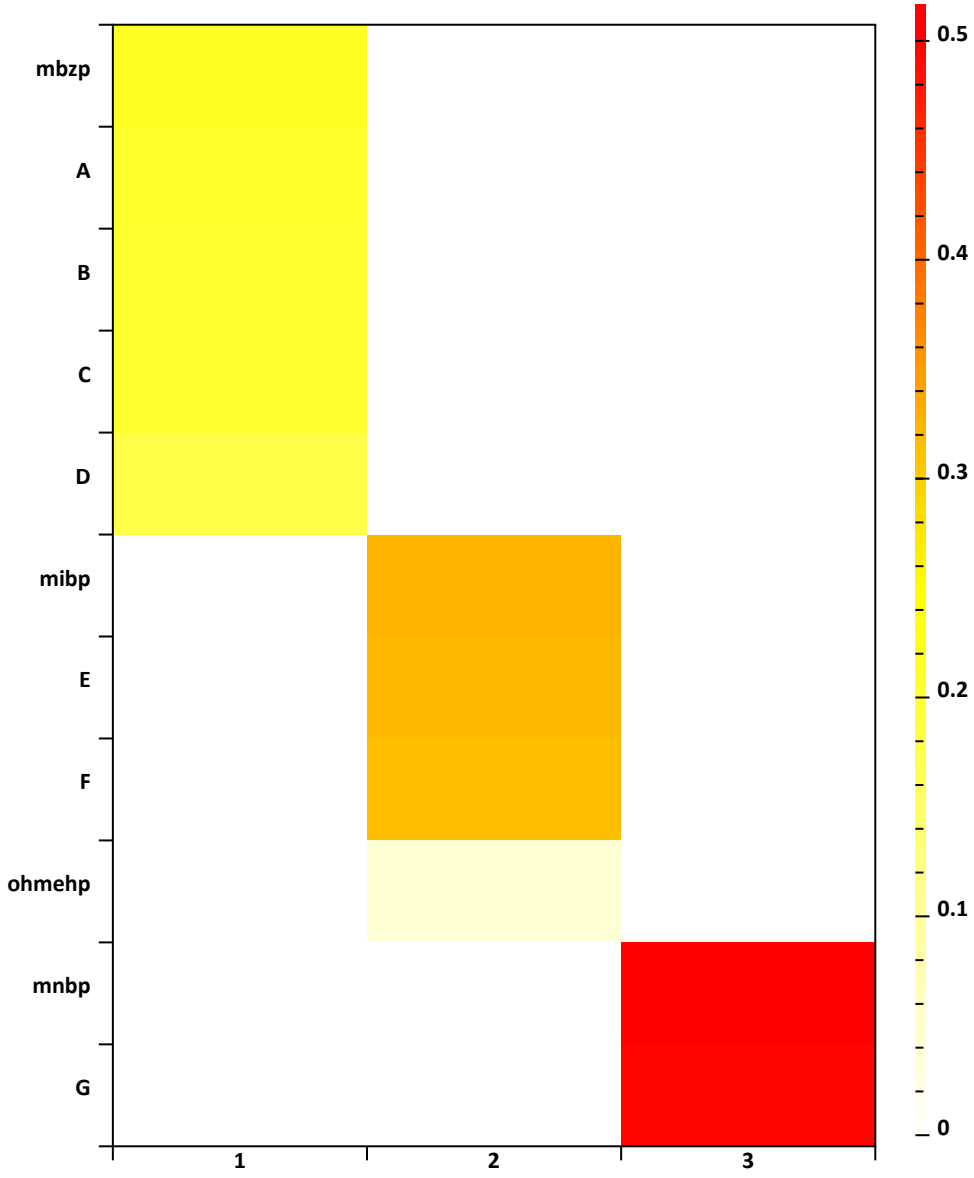


Figure 3.52: Heatmap for solution with 3 components. The first 3 components in Figure 3.52 and Figure 3.50 contain the same substances.

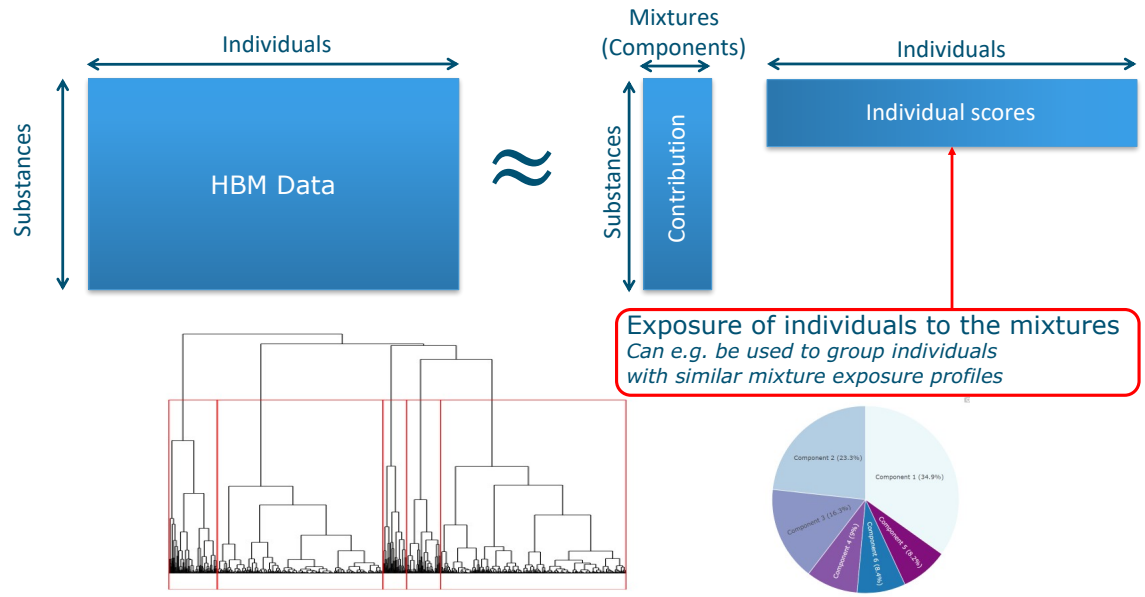


Figure 3.53: SNMU: matrix V, individual scores to components.

Network analysis

As an alternative to *SNMU*, a network analysis is proposed. The outcome of a network analysis is graphically displayed by a network. This is a collection of nodes, that may have pairwise relationships. Each node represents a substance, and an edge represents pairwise dependence between substances (e.g. correlation or partial correlation). In MCRA, the network is estimated using a Gaussian graphical model (GLASSO) based on partial correlation and a sparseness penalty to control the number of nonzero edges (Friedman et al. (2008)). Parameters are automatically tuned. The communities are detected using a Walkman algorithm. In Figure 3.57, using HBM data, a network is displayed with 6 communities. The largest community contains 5 substances: mbzp, A, B, C and D. Compared to the results of the *SNMU mixture analysis*, the communities are almost identical to the components found in the SNMU approach.

The exposure data may be log transformed before the network analysis. Zeros are replaced by the logarithm of the minimum of the non-zero exposure values per substance multiplied by a factor 0.01.

Exposure mixtures settings

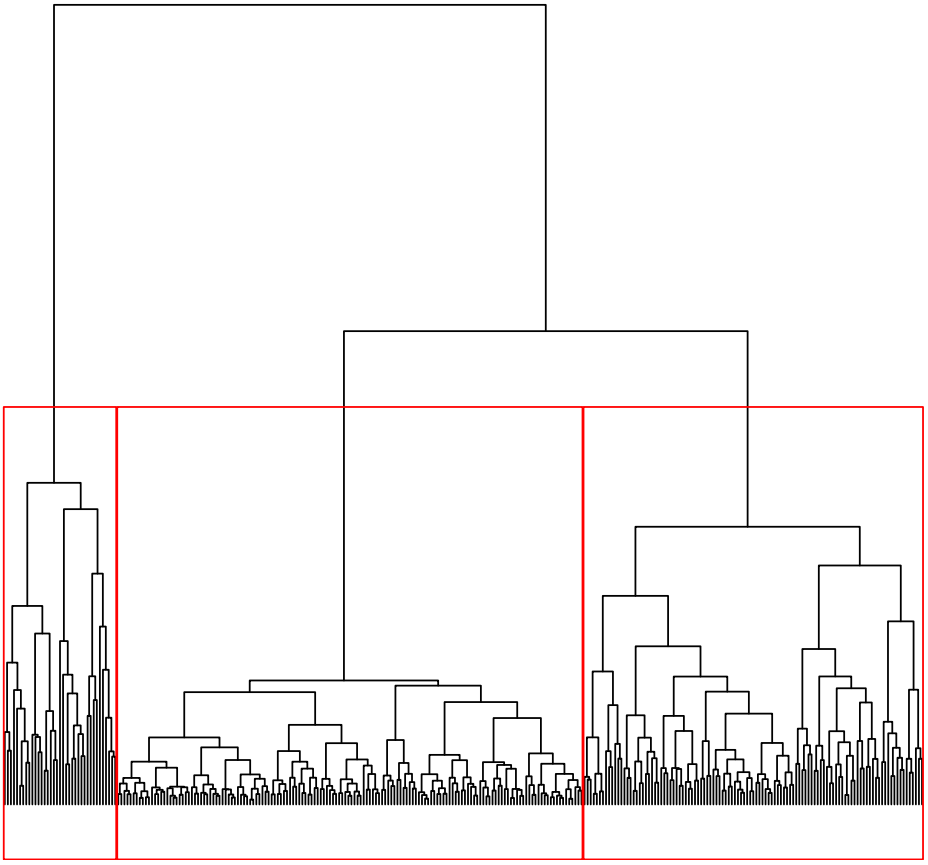


Figure 3.54: Hierarchical clustering of human monitoring data, 3 clusters, largest and smallest cluster contain 152 and 37 individuals, respectively

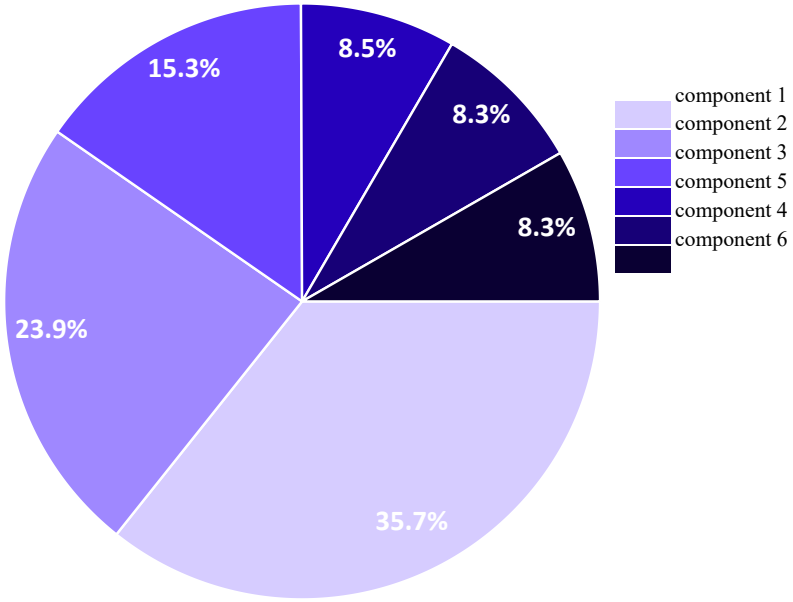


Figure 3.55: Relative exposure to components in the population

Calculation settings

Table 3.123: Calculation settings for module Exposure mixtures.

Name	Type	Description
Risk type	*ExposureType*	The type of exposure considered in the assessment; acute (short term) or chronic (long-term).
Target level	*TargetLevelType*	Select to express hazard characterisations at external or internal exposure level. For an aggregate assessment, that is dietary and nondietary exposure data are combined, the target dose level is always internal. When only dietary exposures are available, the target dose level is optional, i.c. external or internal.
Internal concentration	*InternalConcentrationType*	Internal concentrations are derived form dietary and/or non-dietary concentrations and aggregated using a kinetic or absorption factor model or are human monitoring concentration
Substance weighting in mixtures	*ExposureApproachType*	Risk based: exposures in equivalents of the reference substance standardised: standardised exposures per substance have varian 1; or unweighted exposures: RPFs are equal to 1.
Sparseness parameter	Numeric	Sparseness parameter. Value between 0 (not sparse, many substances) and 1 (sparse, few substances).
SNMU: number of components	Numeric	The number of components to select in SNMU.
Iterations	Numeric	Maximum number of iterations, e.g. 1000.
Seed for pseudo-random number generator.	Numeric	Random seed for initialising matrix W and H.
Convergence criterion	Numeric	Convergence criterion for factorisation algorithm.
Cutoff MCR	Numeric	For selection of individual(day) exposures with maximum cumulative ratio (MCR = total exposure/maximum) above the cutoff.
Cutoff percentage in population ranked on total exposure	Numeric	For selection of individual(day) exposures above the cutoff percentage in the set of individual(day)s ranked on total exposu
Number of clusters	Numeric	Number of clusters for hierarchical cluster analysis or clustering minimizing within-cluster variance (k-means).

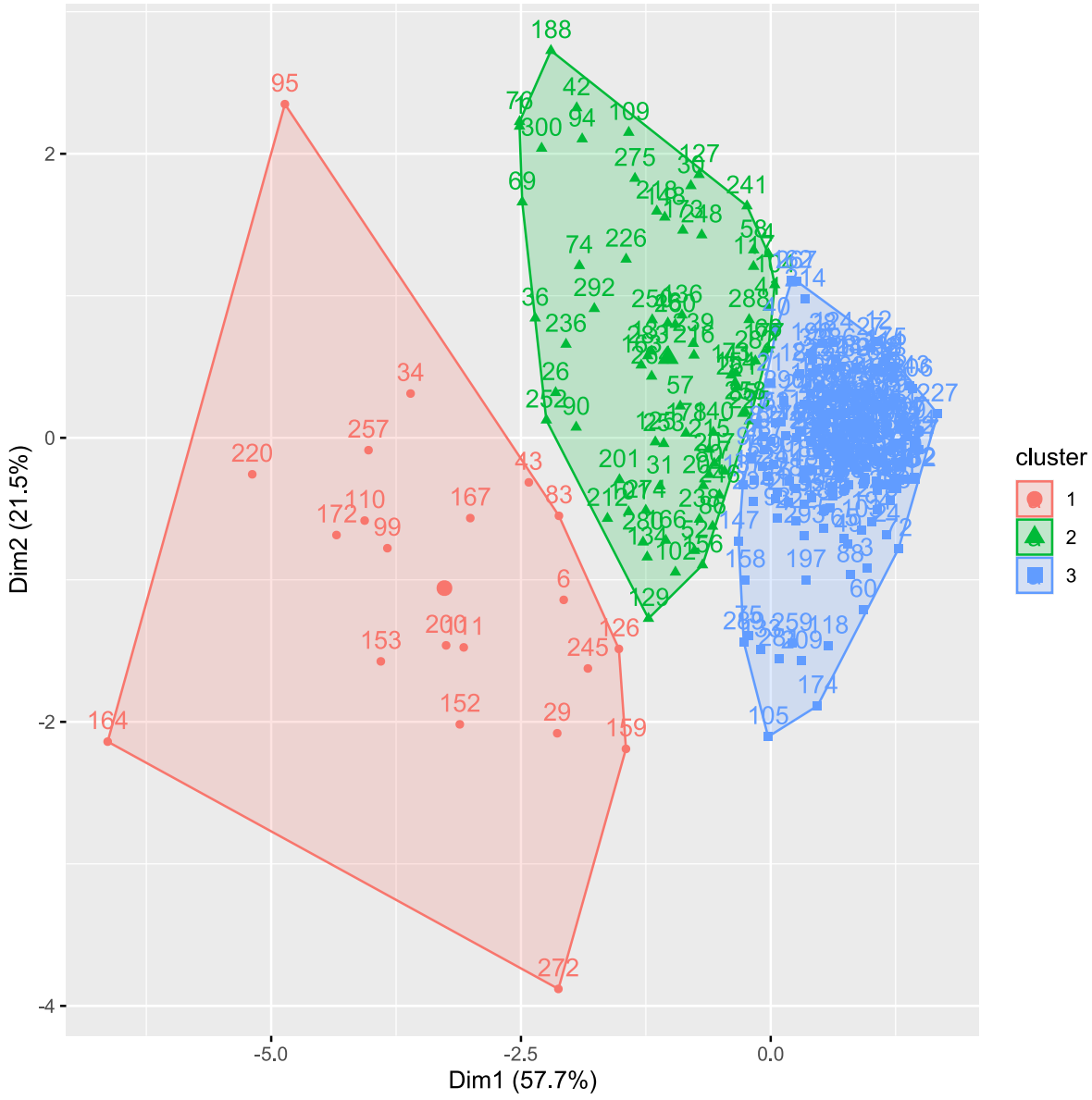


Figure 3.56: K-means clustering of human monitoring data, 3 clusters, largest and smallest cluster contain 204 and 21 individuals, respectively

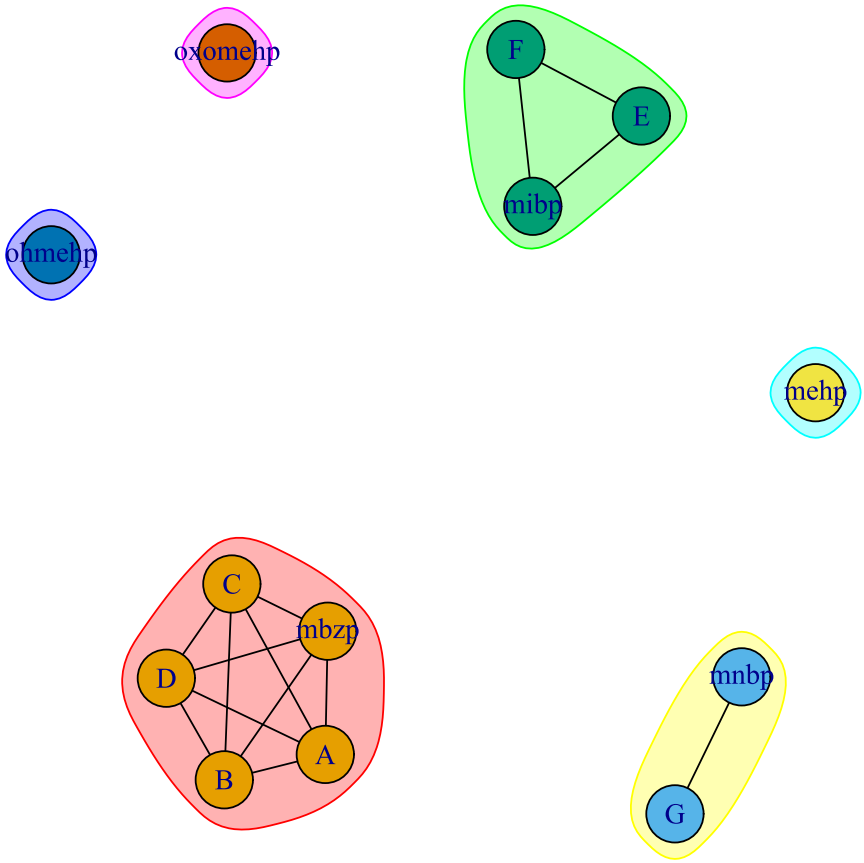


Figure 3.57: Network analysis.

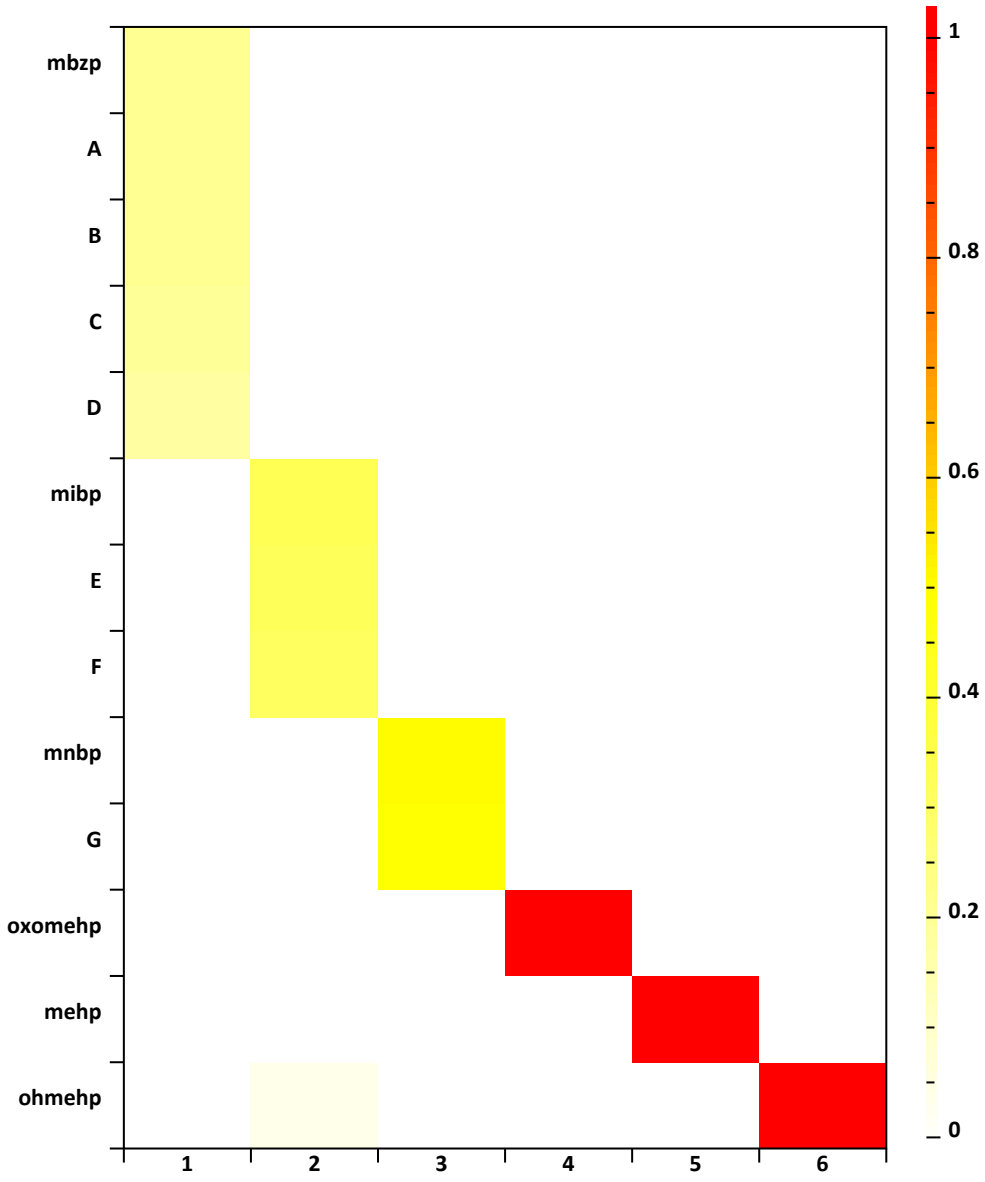


Figure 3.58: SNMU solution with 6 components. Sparsity = 0.8.

Calculation of exposure mixtures

Exposure mixtures or components can be computed from (external) dietary exposures, or from (internal) exposures (possibly from combined dietary- and non-dietary sources) or human monitoring concentrations. A multivariate decomposition method, sparse non-negative matrix underapproximation (SNMU), is applied to the matrix of exposures per substance and per individual (chronic) or individual-day (acute) to find component containing substances that contribute most to the cumulative exposure. Exposures per substance are preprocessed either by multiplication with relative potency factors (RPFs) to make the analysis risk-based, or by standardisation to variance 1 to make the analysis correlation-based. An alternative to SNMU is network analysis. This method estimates communities of substances that have pairwise relationships.

- *Exposure mixtures calculation*

Inputs used: *Dietary exposures Exposures Relative potency factors Human monitoring analysis*

Settings used

- *Calculation Settings*

3.4.6 Food conversions

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

This module has as primary entities: *Foods Substances*

Output of this module is used by: *Consumptions by modelled food Dietary exposures*

Food conversions calculation

Food conversions are computed using a recursive search algorithm to link foods-as-eaten to modelled foods, possibly through intermediate conversion steps. For instance, if (unpeeled) apple and grapes are the modelled foods, 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 apple pie, there are two conversions, one to apple (with processing type 'peeled') and one to grapes (with processing type 'dried'), each with its own conversion path of intermediate conversion steps.

Substance independent conversion

The current implementation of the food conversion algorithm can be run substance independent. The *Find processing link (deprecated)* is skipped from the algorithm (default = false) and is only retained for backwards compatibility reasons only (see Advanced, set to true). Processed foods are easily recognized in the food translation step and retrieving the processing factor that belongs to a certain processing type is done outside the algorithm. In fact, finding processing types with corresponding processing factors is not a task of the conversion algorithm: conversion is about converting food codes to other food codes.

When the processing step is skipped, there is no need to run the conversion algorithm on a substance basis. The only information that is needed is whether a food code is a modelled food or not (i.c. is there a concentration available or not). This information can be computed beforehand: for each substance all modelled foods are collected and supplied to the conversion algorithm in a common dictionary containing all modelled food codes. As soon as a food code is found in the dictionary, the conversion ends and the next food code is converted.

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

1. *Substance independent conversion:* the conversion algorithm is substance independent. Check whether the current food is a modelled food. If successful, the food has been found, and the current search stops.
2. Check whether the current food translates to one or more foods through composition or read-across. Identify any processing types of facets.

a. *Food recipe link:* 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. Simultaneously check, whether the current food is a processed food or not. If so, determine the processing type or facets.

b. *TDS food sample composition link:* 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:* try to find a food extrapolation rule for the current food, a default translation proportion of 100% for 'idToFood' is assumed.

Note that in the *food recipe link* processed foods are recognized and that the translation proportion to correct for a weight reduction or increase is stored.

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

3. *Marketshares link:* try to find subtype codes, e.g. 'xxx$*' in the MarketShares table. In general, marketshares should sum to 100%. Foods with marketshares not summing to 100% are ignored in the analysis unless the checkbox *Allow marketshares not summing to 100%* is checked. This step is optional, see advanced settings. If successful, restart at step 1 with each of the new codes of the subtype foods.
4. *Supertype link:* try to find supertypes, e.g. 'xxx$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.
5. *Default processing factor:* remove processing part (-xxx) of the code. If successful, restart at step 1 with the new code without processing part.
6. Check whether the food is a FoodEx 2 code. If so, strip all food facets of the code. The code that remains i.e the food base code of FoodEx 2 classification is returned and the algorithm restarts at step 1 with the base code (without facets).

Substance dependent conversion

The original conversion algorithm contains two steps which are substance dependent. For each substance all food codes are supplied to the conversion algorithm and for each food code it is checked whether there is:

- a concentration,
- a processing type.

When a concentration is available for the food, this food is a modelled food (formerly known as food as measured). The food may be a food as eaten as such, like apple, or an ingredient of a food as eaten like tomato sauce on pizza which is converted to tomato. If concentrations are available, the food code is found and the conversion algorithm starts with converting the next food code. Otherwise, the conversion proceeds to the *processing link (deprecated)*. Here, basically, processed foods are converted to an unprocessed food and processing type with corresponding processing factor. This processing step may be substance specific and, occasionally, this results in different conversion paths for different substances. This is undesirable behaviour and normally not the case (dependent on the supplied data in the food processing factor table). However, on rare occasions this might happen.

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

Match processing factor: try to find the code in the processing factors table.

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

Warning: the *find processing link (deprecated)* step is not recommended and is currently maintained for backwards compatibility reasons only. Finding different conversions paths depending on the substance is undesirable behaviour.

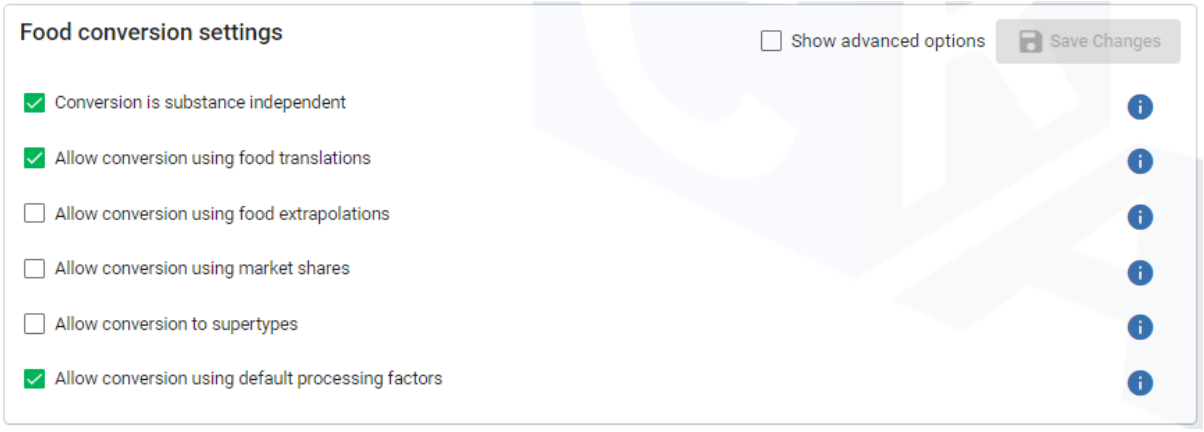


Figure 3.59: Default settings conversion.

Food conversion settings

Calculation settings

Table 3.124: Calculation settings for module Food conversions.

Name	Type	Description
Allow conversion using processing info	Boolean	Warning, the processing step is deprecated and is currently only maintained for backwards compatibility reasons. See documentation for more details how processed foods are converted in the upgraded conversion algorithm. Step 2a: try to find the code in the processing table. Try to find the code in the FoodTranslation table (step 3a) to account for weight reduction/increase (translation proportion). If unchecked (default), processing table is ignored. If successful, restart at ste 1.
Allow conversion using food translations	Boolean	Step 3a: try to find food translations for the current food (i.e., t ingredients of a composite food). This may result in one or mo food codes for ingredients, and the iterative algorithm will proceed with each of the ingredient food codes in turn.
Allow conversion using TDS food sample compositions	Boolean	Step 3b: try to find the code in the TDS food sample compositi table (idFood), a default translation proportion of 100% is assumed. The iterative algorithm will proceed with a TDS food (column idTDSFood) sample.
Allow conversion using food extrapolations	Boolean	Step 3c: try to find read across codes. If unchecked, read acros 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 o 100%), the TDS sample concentration value of FruitMix will b used for pineapple (as-eaten or as ingredient). If successful, restart at step 1.
Allow conversion using market shares	Boolean	Step 4: try to find subtype codes, e.g. 'xxx$*' in the market sha table.
Allow marketshares not summing to 100%	Boolean	Specify whether to rescale market share percentages that do no sum to 100%. If checked, then foods with marketshares not summing to 100% are allowed. If not, then these foods are ignored in the analysis.
Allow conversion to supertypes	Boolean	Step 5: try to find supertypes, e.g. 'xxx$yyy' is converted to 'xx (optional, check box if you want to use this). If checked, allows for linkage of consumed foods coded at a lower hierarchical lev 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 allow to use the concentration data of a supertype for all underlying food codes. If successful, restart at step 1.
Allow conversion using default processing factors	Boolean	Step 6: remove processing part. If unchecked, no default processing factors are assumed, default is 'Use default processin factors'. If successful, restart at step 1.
Conversion is substance independent	Boolean	Conversion of foods is independent of the substance.

Calculation of food conversions

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

- *Food conversions calculation*

Inputs used: *Consumptions Modelled foods Processing factors Food recipes Market shares Food extrapolations Total diet study sample compositions Active substances*

Settings used

- *Calculation Settings*

3.4.7 Biological matrix concentration comparisons

Substances in the human body are absorbed, excreted without transformation, excreted after metabolization or stored in various tissues, bones or body fluids. The term biological matrix refers to all human specimens where concentrations of a chemical can be measured like bodily fluids, such as blood, urine, saliva, breast milk, sweat, and other specimens, such as faeces, hair, teeth, and nails. Biological matrix concentration comparisons 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*

Biological matrix concentration comparisons calculation

In this module, concentration estimates from measurements in biological (human) matrices are compared with modelled or predicted concentrations obtained from *dietary* and/or *non-dietary* surveys via *exposure* assessments. The comparison provides insight in the coherence between modelled exposures and the measured reality. It is required that monitoring data and dietary/non-dietary exposure data are available for the same individual or individual-day. In Figure 3.60, summary statistics are visualised for the monitoring and modelled concentrations for bisphenols BPA, BPS and BPF.

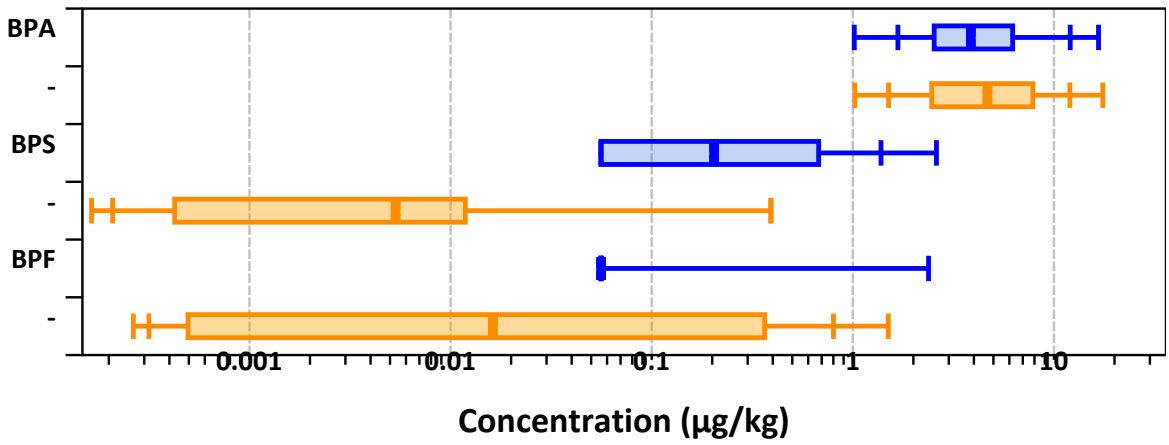


Figure 3.60: Boxplots for monitoring and modelled concentrations for bisphenols BPA, BPS and BPF. Lower whiskers indicate the p5 and p10 percentiles, upper whiskers the p90 and p95. The edges of the box indicate the p25 and p75 percentiles with the median in the centre of the box.

When both human monitoring data and exposure data are available for the same individuals in a population, a direct comparison can be made between the monitoring and modelled concentrations. In Figure 3.61, an example is shown.

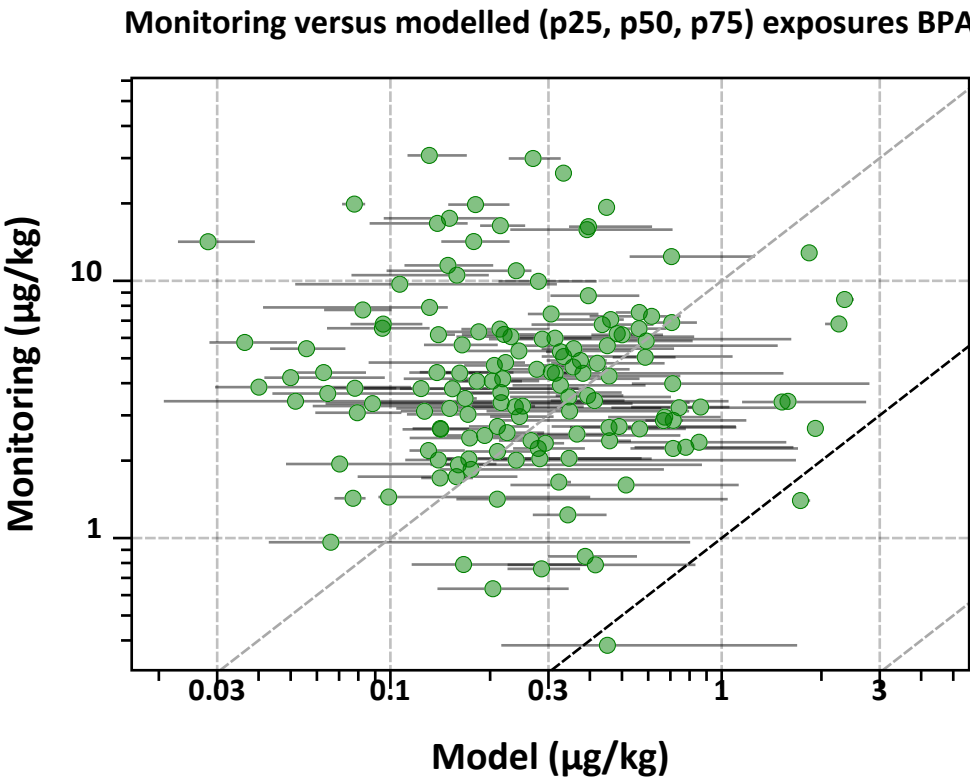


Figure 3.61: Monitoring versus modelled concentrations for bisphenol BPA

Biological matrix concentration comparisons settings

Calculation settings

Table 3.125: Calculation settings for module Biological matrix concentration comparisons.

Name	Type	Description
Risk type	*ExposureType*	The type of exposure considered in the assessment; acute (short term) or chronic (long-term).
Correlate monitoring with modelled concentrations	Boolean	Correlate monitoring with modelled concentrations. This is opt prevents that monitoring and modelled individuals that are code unintentionally with identical id's or codes are correlated.

Output settings

Table 3.126: Output settings for module Biological matrix concentration comparisons.

Name	Type	Description
Store simulated individual day exposures	Boolean	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.

Calculation of biological matrix concentration comparisons

Biological matrix concentration comparisons 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 is to relate the human monitoring concentrations to modelled concentrations from exposure assessments.

- *Biological matrix concentration comparisons calculation*

Inputs used: *Human monitoring analysis Exposures*

Settings used

- *Calculation Settings*

3.4.8 Human monitoring analysis

Human monitoring concentrations are substance concentration estimates for a biological matrix (e.g., urine or blood) derived from data obtained from human monitoring studies.

This module has as primary entities: *Populations Substances*

Output of this module is used by: *Exposure mixtures Biological matrix concentration comparisons Risks*

Human monitoring analysis calculation

Human monitoring analysis computes substance concentration estimates for a biological matrix (e.g., urine or blood) based on *human monitoring data*. These estimates are specified at the level of long term average concentrations for individuals in case of *chronic assessments*, or concentrations for individual-days in case of *acute assessments*. The concentrations are computed independently for each substance and biological matrix.

The main steps, and also performed in this order, for computing human monitoring concentration estimates are:

1. Imputation of censored values.
2. Imputation of missing values.
3. Standardise blood for lipid content.
4. Standardise/normalise urine for creatinine or specific gravity.
5. Apply conversion of substance concentrations from other biological matrices.
6. Calculation of individual concentrations (chronic) or individual day concentrations (acute).

Imputation of censored values

Similar to *concentrations measurements in food*, human monitoring measurements contain measurements below the limit of reporting and similar to *concentrations modelling in foods*, human monitoring analysis addresses these censored values and replaces them with imputed concentration values. Two approaches are available:

1. Impute using a *non-detects handling method*.
2. Impute using a draw from the left tail of the *censored lognormal distribution*. See also *concentration models* and *concentration model types*.

The available non-detects handling methods for deterministic imputation are:

- Replace censored values by zero.
- Replace censored values by a factor * LOR, the factor is set between zero and one.

- Replace non-detects by a factor * LOD and non-quantifications by LOD + factor * (LOQ - LOD), the factor is set between zero and one.

Note that for the option based on the censored lognormal distribution also a non-detects handling method needs to be specified. Occasionally, fitting the censored lognormal model fails and the deterministic imputation method is used as fallback.

In Figure 3.62, the *human monitoring concentrations data* for three bisphenols are imputed with a factor * LOR and the summary statistics are visualized.

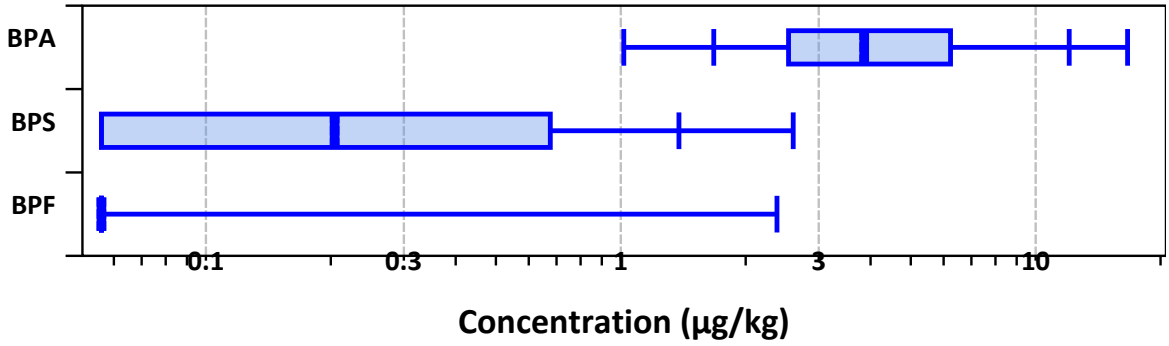


Figure 3.62: Boxplots for imputed concentration data for bisphenols BPA, BPS and BPF. Lower whiskers indicate the p5 and p10 percentiles, upper whiskers the p90 and p95. The edges of the box indicate the p25 and p75 percentiles with the median in the centre of the box.

Imputation of missing values

For missing concentration measurements, *three imputation methods* are available. The third opion ignores imputation and all missing values remain in the data set.

1. Replace missing values by zero.
2. For each substance and combination of biological matrix and sampling type, replace missing values by a random draw from the non-missing concentration values (samples). This is conditional on the specified minimum percentage of non-missing values.
3. Do not impute missing values

Imputation of missing values by a random draw from the data is conditional on the specified percentage of non-missing values. When the percentage of non-missing values for a specific substance and combination of biological matrix and sampling type in the data is smaller than the specified percentage, imputation is ignored. This is to prevent that imputation takes place using a set of imputation values that is not representative or unrealistically small (e.g. one or a few values). Note that for the second imputation method, more refined methods could be used. E.g., when for a given day multiple samples are available and one is missing, then this sample might be neglected in the computation of an average exposure. Also, when samples have been taken at different time-slots, impute the missing records using samples from the same time-slot.

Standardise blood for lipid content

Lipid-soluble substances measured in blood data are typically standardised by total lipid content. Three methods are in available, namely:

1. Standardisation based on *gravimatic analysis*.
2. Standardisation based on *enzymatic summation analysis*.
3. Standardisation based on Bernert et al. (2007), *total lipids (mg/dL) = 2.27 * total cholesterol + triglycerides + 62.3 mg/dL.*

The standardisation is only applied to lipid-soluble substances, see *Substances data formats*. After standardisation, the amount of substance is expressed per g lipid e.g. $\mu g/g\,\texttt{lipid}$. Note that substance concentrations in blood samples with unmeasured lipid concentrations are set to missing after specifying option blood standardisation.

Standardise/normalise urine for creatinine or specific gravity

Two methods are available for correcting spot urine measurements for creatinine of specific gravity.

1. Normalisation based on *specific gravity*.
2. Standardisation based on *creatinine content*.

Urine's specific gravity is determined by the concentration of excreted molecules in the urine. In adult humans, normal specific gravity values range from 1.010 to 1.030. The specific gravity normalisation used here is equal to $(1.024 - 1)/(specific\,gravity - 1)$. The specific gravity value should be available in the HBM data, otherwise urine sample concentrations are set to missing.

After standardisation for creatinine content, the amount of substance is expressed per g creatinine e.g. $\mu g/g\,\texttt{creatinine}$. Note that substance concentrations in urine samples where the creatinine content is not measured are set to missing values after specifying option urine standardisation.

Apply conversion of substance concentrations from other biological matrices

Conversion of substances measured in biological matrices other than the target matrix is used when the number of substances in the target biological matrix is limited. For instance, in target matrix urine (spot) five substances are measured. For the same individuals also the blood (serum) concentrations are analysed resulting in a additional set of concentrations for five different substances. These five substances from blood (source matrix) are converted to the target matrix urine by checking the option *Apply conversion of substance concentrations from other biological matrices* and specifying the 'between matrix conversion' multiplication factor. Then, the analysis continues with ten substances. Conversion of substance concentrations is performed after imputation of censored and missing values. Selecting specific substances for conversion is possible through the selection options in the primary entity *substances* module.

Note that after choosing *Do not impute missing values* or *Impute from data*, missing value imputation, the data still contains missing values. These values may be imputed from other substance concentrations after applying matrix conversion.

Restricting the monitoring concentrations

After imputation of non-detects and missing values, standardisation/normalisation, eventually followed by conversion of substance concentrations from other biological matrices, some individual day monitoring concentration records may contain missing values (e.g. after '*Do not impute missing values* or *Impute from data*, with the minimum percentage of non-missing values set to high). All individual day samples with missing values for one or more substances are removed from the data set and the analysis continues.

Occasionally, removing all records with missing values results in empty datasets. Then a warning will be thrown 'All HBM individual day records were removed for having non-imputed missing substance concentrations'. To circumvent this warning, inspect your data and remove substances with too many missing values, lower the minimum percentage of non-missing vales (impute from data) or impute with zero.

Calculation of acute human monitoring concentrations

For acute assessments, the monitoring concentrations are computed for each substance and biological matrix as average individual-day concentrations. The computation is done after imputation of censored and missing values, eventually followed by a conversion of substance concentrations from other biological matrices. For a given substance and biological matrix, the acute individual-day concentration c_{ij} for individual i on day j is:

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

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

After urine normalisation for specific gravity:

$$c'_{ij} = c_{ij} \cdot sg$$

where *sg* denotes the specific gravity correction factor for that individual day.

After standardisation for blood lipid content:

$$c'_{ij} = \frac{c_{ij}}{c_{ij}\,\texttt{lipid}}$$

where c'_{ij} denotes the lipid concentration per g `lipid`. The standardisation is only performed for lipid soluble substances. After standardisation the concentration of the substance is is expressed as substance amount, with an user specified unit, per g `lipid`.

The standardisation for creatinine is similar to the above equation replacing `lipid` by `creatinine`.

Calculation of chronic human monitoring concentrations

For chronic assessments, the monitoring concentrations are computed as the average monitoring concentrations of multiple individual-days for each substance and biological matrix. The computation is done after imputation of censored and missing values, eventually followed by a conversion of substance concentrations from other biological matrices. The chronic concentration c_i for individual i is computed as:

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

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.

Standardisation and normalisation of blood and urine samples respectively, are similar to the expressions for the calculation of individual day concentrations (acute).

For co-exposure of substances, see *maximum cumulative ratio (MCR)* and the *exposure mixtures module*.

Human monitoring analysis settings

Calculation settings

Table 3.127: Calculation settings for module Human monitoring analysis.

Name	Type	Description
Risk type	*ExposureType*	The type of exposure considered in the assessment; acute (short term) or chronic (long-term).
Multiple substances analysis	Boolean	Specifies whether the assessment involves multiple substances.
Compute cumulative exposures	Boolean	Specifies whether the assessment involves multiple substances a results should be cumulated over all substances.
Censored values handling method	*NonDetectsHandlingMethod*	Method for dealing with censored value samples in human monitoring data. Note that this method is also used as a fallbac when fitting a censored lognormal model to the concentration d fails.
Fraction for censored value replacement	Numeric	Factor used for replacing the censored value.
Imputation method for non detect values	*NonDetectImputationMethod*	Imputation method for non detect values: replace nondetects based on by f*LOD/LOQ) or from left tail censored lognormal distribution.
Missing value imputation method	*MissingValueImputationMethod*	Imputation method for missing values.
Biological matrix	AlphaNumeric	Biological matrix.
Apply conversion of substance concentrations from other biological matrices	Boolean	Apply conversion of substance human monitoring concentration from other biological matrices to the target biological matrix. T conversion is applied when the number of substances measured the target biological matrix is limited. Substances measured on other matrices are converted and muliplied by a 'between matri conversion' factor.
Between matrix concentration conversion factor	Numeric	Conversion factor to use when extrapolating concentrations of other biological matrices to concentrations of the selected targe biological matrix.
Specify the minimum percentage of non-missing values (%)	Numeric	Specify the minimum percentage of non-missing values require for imputation. No imputation is done when the percentage of non-missing values in the data is smaller than the specified percentage.
Standardise blood concentrations for lipid soluble substances	bool	Standardise blood concentrations for lipid soluble substances.
Specify the standardisation method of blood concentrations for lipid soluble substances	*StandardiseBloodMethod*	Specify the standardisation method of blood concentrations for lipid soluble substances.
Standardise urine concentrations for creatinine or specific gravity	bool	Standardise urine concentrations for creatinine or specific gravi
Specify the standardisation method of urine concentrations for creatinine or specific gravity	*StandardiseUrineMethod*	Specify the standardisation method of urine concentrations for creatinine or specific gravity.
Substance weighting for MCR	*ExposureApproachType*	Risk based: exposures in equivalents of the reference substance standardised: standardised exposures per substance have varian 1; or unweighted exposures: RPFs are equal to 1.
Perform MCR analysis	Boolean	Perform a Maximum Cumulative Ratio (MCR) analysis to determine co-exposure between substances.
Display ratio total exposure/ maximum (in MCR plot)	Numeric	For MCR plot: specify ratio total exposure/ maximum for individual(day) exposures .
Show tail percentiles (MCR plot) for:	AlphaNumeric	Give specific percentiles of exposure distribution (%), e.g. 97.5 99 (space separated).
Set minimum percentage contributions per substance to the tail exposure (MCR plot)	Numeric	Set minimum percentage contribution per substance to the tail exposure.

Output settings

Table 3.128: Output settings for module Human monitoring analysis.

Name	Type	Description
Store simulated individual day exposures	Boolean	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.

Calculation of human monitoring analysis

Human monitoring concentration estimates are computed from data obtained from human monitoring studies. These concentration estimates are computed per substance for a selected (human) biological matrix. Modelling includes imputation of missing values and non-detects, but also correction of concentrations for, e.g., specific gravity. Occasionally, the number of substances measured on the target biological matrix is limited or to low to perform a risk assessment or mixture analysis. If so, matrix conversion is applied: substances measured on other biological matrices are translated to the target matrix after multiplication with a 'between matrix conversion' factor. Cumulative concentration estimates can be computed using RPFs.

- *Human monitoring analysis calculation*

Inputs used: *Human monitoring data Active substances Relative potency factors*

Settings used

- *Calculation Settings*

3.4.9 Human monitoring data

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

This module has as primary entities: *Populations Substances*

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

Human monitoring data data formats

Human (bio)monitoring data are analytical measurements of chemical substances or markers of health effects in body fluids or tissues. This data can be used for assessing (combined) exposure to chemicals and the risks associated with this exposure. MCRA supports upload of this data in two formats: a relational table structure that matches the internal representation of MCRA and the PARC harmonized HBM data format.

Relational human monitoring concentration data format

The relational human monitoring concentrations data format is the format that is used internally in MCRA. The data format consists of a number of tables for specification of the study, the individuals of the study (and their specific characteristics), and the human (bio)monitoring samples and sample analyses.

Download empty dataset template: Zipped CSV Excel

Human monitoring surveys

Contains the definitions of the human (bio)monitoring surveys/studies.

Table 3.129: Table definition for Human monitoring surveys.

Name	Type	Description	Aliases	Required
idSurvey	AlphaNumeric (50)	Unique identification code of the survey.	idSurvey, idStudy	Yes
Name	AlphaNumeric (100)	Name of the study/survey.	Name	No
Description	AlphaNumeric (200)	Description of the study/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
Lipid-Concentration-Unit	*ConcentrationUnit*	The unit of the lipid concentration (defaults mg/dL).	Lipid-Concentration-Unit, LipidUnit	No
Triglyc-Concentration-Unit	*ConcentrationUnit*	The unit of the triglycerides concentration (defaults mg/dL).	Triglyc-Concentration-Unit, Triglycerides-Concentration-Unit, Triglycerides-Unit	No
Cholest-Concentration-Unit	*ConcentrationUnit*	The unit of the cholesterol concentration (defaults mg/dL).	Cholest-Concentration-Unit, Cholesterol-Concentration-Unit, CholesterolUnit	No
Creat-Concentration-Unit	*ConcentrationUnit*	The unit of the creatinine concentration (defaults mg/dL).	Creat-Concentration-Unit, Creatinine-Concentration-Unit, CreatinineUnit	No

Accepted table names: HumanMonitoringSurveys, HumanMonitoringSurvey.

Individuals

The individuals of a survey are recorded in the Individuals table. Add additional properties like Region, Breastfeeding to further describe an individual. In table IndividualProperties, each property in the Individuals table is described (recommended way). Note that only those properties that are available in the Individuals table are used in module Populations, table Populations or PopulationIndividualPropertyValues to subset the individuals. This is only relevant when the UseData option in the population module is used.

Table 3.130: 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
Name	AlphaNumeric (100)	Name or label of the individual.	Name	No
Description	AlphaNumeric (200)	Additional description of the individual.	Description	No
Individual properties		Other individual properties can be added like the fields Age, Gender, Region etc. These properties are automatically parsed as co-factors or co-variables.		No

Accepted table names: Individuals, SurveyIndividuals, ConsumptionSurveyIndividuals, FoodConsumptionSurveyIndividuals.

Individual properties

This table is used to describe the properties used in the Populations or PopulationIndividualPropertyValues table characterising the population (table Populations) and/or the properties used in the Individuals table characterising an individual. Properties like Age, Gender, Region are describing an individual (PropertyLevel = Individual). Properties like Period (for populations) or Month (sampling date for an individual day) are describing an individual day (PropertyLevel = IndividualDay).

Table 3.131: Table definition for Individual properties.

Name	Type	Description	Aliases	Required
idIndividual-Property	AlphaNumeric (50)	The code of the property.	idIndividual-Property, Individual-PropertyId, Individual-Property	Yes
Name	AlphaNumeric (100)	The name of the property.	Name	No
PropertyLevel	*PropertyLevelType*	The level of the property. This type follows a controlled terminology, with possible values: Individual or IndividualDay.	PropertyLevel, LevelProperty	No
Description	AlphaNumeric (200)	Description of the property.	Description	No
Type	*IndividualProperty-Type*	This field specifies the type of the values of this individual property. This type follows a controlled terminology, with possible values: Boolean, Categorical (default), Numeric, Nonnegative, Integer, NonnegativeInteger, Month, Datetime, Gender.	Type	No

Accepted table names: IndividualProperties, IndividualProperty.

Individual property values

Not recommended. This table describes individual property values. Property values are describing an individual for properties like e.g. Region, Breastfeeding. The recommended way is to add these columns as additional columns in the Individuals table. In table IndividualProperties, each property in the IndividualPropertyValues table is described.

Table 3.132: Table definition for Individual property values.

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

Accepted table names: IndividualPropertyValues, IndividualPropertyValue.

Analytical methods

The analytical methods used for analysing 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 3.133: Table definition for Analytical methods.

Name	Type	Description	Aliases	Required
idAnalytical-Method	AlphaNumeric (50)	The code for the method of analysis.	idAnalytical-Method, Analytical-MethodId, Analytical-MethodName, Id	Yes
Name	AlphaNumeric (100)	Name of the analytical method.	Name	No
Description	AlphaNumeric (255)	Additional description of method of analysis.	Description	No

Accepted table names: AnalyticalMethod, AnalyticalMethods.

Analytical method properties for substances

This table describes the substances analysed by an analytical method. For each substance analysed by an analytical method a record should be included that describes the unit of measurement and the reporting limits (LOQ/LOD).

Table 3.134: Table definition for Analytical method properties for substances.

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	Yes
LOD	Numeric	The limit of detection (LOD) is the lowest concentration of an substance in a sample that can be consistently detected.	LOD	No
LOQ	Numeric	The limit of quantification (LOQ) is the lowest concentration of a substance that can be quantified. The LOQ should be larger than the LOD.	LOQ, LOR	No
Concentration-Unit	*ConcentrationUnit*	The unit used for reporting the LOD, LOQ, and the substance concentrations. When not specified, then a default unit of mg/kg is assumed.	Concentration-Unit, Units, Unit	No

Accepted table names: AnalyticalMethodSubstances, AnalyticalMethodSubstance, AnalyticalMethodCompounds, AnalyticalMethodCompound.

Human monitoring samples

Describes the samples taken during the study. Each sample has a unique identifier/code.

Table 3.135: Table definition for Human monitoring samples.

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	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	Correction factor for the concentration to account for the specific gravity of the measured person for this particular sample.	SpecificGravity-Correction-Factor	No
Name	AlphaNumeric (100)	Name of the human monitoring sample.	Name	No
Description	AlphaNumeric (200)	Additional description of the human monitoring sample.	Description	No
LipidGrav	Numeric	Lipid content based on gravimatic analysis of the measured person for this particular sample.	LipidGrav, LipidGravimatic	No
LipidEnz	Numeric	Lipid content based on enzymatric summation of the measured person for this particular sample.	LipidEnz, LipidEnzymatic	No
Triglycerides	Numeric	Triglycerides total of the measured person for this particular sample.	Triglyc, Triglycerides	No
Cholesterol	Numeric	Cholesterol total of the measured person for this particular sample.	Cholest, Cholesterol	No
Creatinine	Numeric	Creatinine content of the measured person for this particular sample.	Creat, Creatinine	No
Osmotic-	Numeric	Osmotic concentration of the	Osm, Osmotic-	No

Accepted table names: HumanMonitoringSamples, HumanMonitoringSample.

Human monitoring sample analyses

Contains the measurements of the samples of human monitoring studies.

Table 3.136: Table definition for Human monitoring sample analyses.

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	DateTime	Date of analysis.	AnalysisDate, DateAnalysis	No
Substance concentration(s)	AlphaNumeric (100)	Substance concentrations can be uploaded via the sample concentrations table or via additional columns of the sample analyses table. For the latter, one or more columns with the measured concentrations of the substances in the unit as specified by the analytical method should be included in the data table. 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 censored with measurement values below LOQ or LOD.		No
Name	AlphaNumeric (100)	Name of the human monitoring sample analysis.	Name	No
Description	AlphaNumeric (200)	Additional description of the human monitoring sample analysis.	Description	No

Accepted table names: HumanMonitoringSampleAnalyses, HumanMonitoringSampleAnalysis.

Sample concentrations

The positive concentration values for substances from analysis in the unit specified in table human monitoring sample analyses. Censored values (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 3.137: Table definition for Sample concentrations.

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	Yes
Concentration	Numeric	The measured concentration.	Concentration	No
ResType	*ResType*	The type of residue. Should be VAL (= default), LOQ, LOD or MV.	ResType	No

Accepted table names: HumanMonitoringSampleConcentrations, HumanMonitoringSampleConcentration, HumanMonitoringConcentrations.

PARC HBM data format

The PARC HBM data format is the data format proposed in the EU Partnership for the Assessment of Risks from Chemicals (PARC) project. MCRA supports upload of files in this format, of which the data is mapped to the MCRA internal data format during the upload of the files.

PARC harmonised HBM data format

Note: The PARC harmonised data format as well as the MCRA data format are under developement and may be changed in the future.

In PARC, a harmonised data format for human biomonitoring data is being developed by VITO. Data harmonization improves the comparability of data from different HBM studies and interoperability for use with different analysis tools such as MCRA and the tool for the calculation of summary statistics of the HBM data, which can be made available via the IPCHEM portal and/or integrated into the European HBM dashboard. More information about this data format and instructions on preparing data files compliant with this format can be found at the PARC HBM data harmonization web page. This page also contains a tool for validating data files prepared in this format and an example data file that can be uploaded to and used in MCRA for testing.

Excel data files provided to MCRA in this format are mapped during the file upload process to the internal data structure/format of MCRA. For this, MCRA uses a custom mapping/conversion procedure. For a large part, this mapping is fairly straightforward. However, for some fields and entities explicit choices are made that users need to be aware of.

- **Survey/study**
  - Each data file corresponds to one HBM survey/study in MCRA.
  - Survey *StartDate* refers to the first sampling date of all repeated samples.

  - Survey *EndDate* refers to the last sampling date of all repeated samples.
  - When all reported *country* values of the subjects are the same, then this is used as location of the survey/study.
- **Subjects/individuals**
  - Each subject maps to MCRA individuals of the study/survey.
  - A selection of the subject/individual properties is mapped to the MCRA data format. Since MCRA does not support repeated properties, a choice has to be made when mapping individual properties that are repeated individual/subject properties in the harmonised HBM data format.
  - Repeated recordings for each subject's *weight* are averaged to one value: *BodyWeight*.
  - Repeated recordings for each subject's *height* are averaged to one value: *Height*.
  - From the repeated recordings for *smoking* only the last recording is taken: *Smoking_status*.
- **Samples**
  - Each sample maps to an MCRA sample.
  - The *matrix* code is translated to a *biological matrix* and *sampling type*, e.g. *US* translates to *Urine* and *Spot*, *BP* to *Blood* and *Plasma*.
  - *id_timepoint* refers to the sampling time and is translated to *DayOfSurvey*.
- **Analytical methods**
  - MCRA analytical methods are derived/reconstructed from the concentration data sheets of the harmonised HBM data format based on the reported detection limits (LOQs/LODs) and substances and the matrix of the samples.
- **Sample concentrations/measurements**
  - Concentrations without value (blanks) are recorded as missing value (MV).
  - Concentrations recorded as -1 are interpreted as a value below limit of detection (LOD).
  - Concentrations recorded as -2 or -3 are interpreted as values below the limit of quantification (LOQ).

Human monitoring data calculation

Human monitoring data characterize human exposure to substances, their metabolites in body fluids or tissues. In Figure 3.63, the positive (> zero) human monitoring concentrations for bisphenols in urine are visualized through summary statistics.

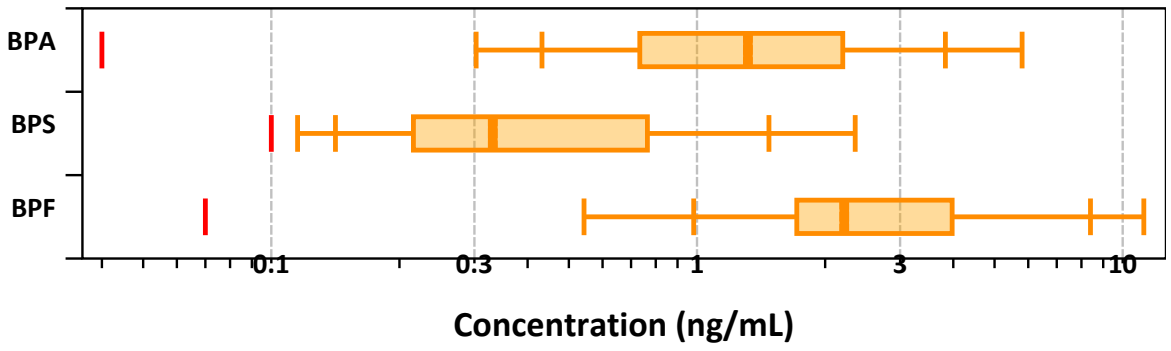


Figure 3.63: Boxplots for positives concentrations of bisphenols BPA, BPS and BPF. Lower whiskers indicate the p5 and p10 percentiles, upper whiskers the p90 and p95. The edges of the box indicate the p25 and p75 percentiles with the median in the centre of the box. LORs are displayed by a vertical red bar.

Find in Figure 3.64, boxplots for all human monitoring concentrations for bisphenols in urine.

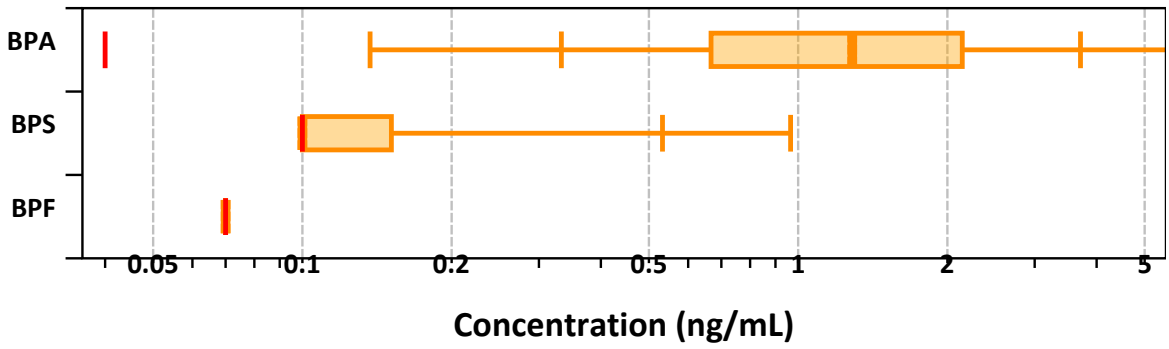


Figure 3.64: Boxplots for all concentrations of bisphenols BPA, BPS and BPF. Lower whiskers indicate the p5 and p10 percentiles, upper whiskers the p90 and p95. The edges of the box indicate the p25 and p75 percentiles with the median in the centre of the box. LORs are displayed by a vertical red bar.

Human monitoring data settings

Selection settings

Table 3.138: Selection settings for module Human monitoring data.

Name	Type	Description
Survey	AlphaNumeric	The survey that should be included in the action.
Sampling method	AlphaNumeric	The sampling method that should be included in the action.
Match HBM individuals selection to population definition options	*IndividualSubsetType*	Match HBM individuals selection to population definition. Use population definitions (default), ignore all population definitions use a selection of properties.
Select one or more individual(day) properties to filter the population	AlphaNumeric	Select one or more individual(day) properties to filter the individuals(days) in the population.
Use sampling weights	Boolean	If checked, individual sampling weights are used. If unchecked the individual sampling weights are not in the calculations.

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.

- *Human monitoring data data formats*
- *Human monitoring data calculation*

3.4.10 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. (Kennedy et al. (2012), Kennedy et al. (2015a), Kennedy et al. (2015b), Kennedy and Butler Ellis (2017)) 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, 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, see Kennedy and Butler Ellis (2017). Volatilisation is also included through the PEARL-OPS model (van den Berg et al. (2016)) to account for inhalation of vapours. Bream2 is an updated version of the original Bream model (Kennedy et al. (2012)) 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 (Butler Ellis et al. (2018)). Currently, the Browse software is outdated and replaced by Bream2.

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. Uncertainty 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 substances. 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 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.

Download empty dataset template: Zipped CSV Excel

Non-dietary surveys

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

Table 3.139: Table definition for Non-dietary surveys.

Name	Type	Description	Aliases	Required
idNonDietary-Survey	AlphaNumeric (50)	The survey identification number.	idNonDietary-Survey	Yes
Name	AlphaNumeric (100)	Name of the non-dietary survey.	Name	No
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	*ExposureUnit*	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

Accepted table names: NonDietarySurveys, NonDietarySurvey.

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 3.140: Table definition for Non-dietary survey properties.

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

Accepted table names: NonDietarySurveyProperties, NonDietarySurveyProperty.

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 3.141: Table definition for Non-dietary exposures.

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	Yes
Dermal	Numeric	The dermal (non-dietary) exposure value.	Dermal	Yes
Oral	Numeric	The oral (non-dietary) exposure value.	Oral	Yes
Inhalation	Numeric	The inhalation (non-dietary) exposure value.	Inhalation	Yes

Accepted table names: NonDietaryExposures, NonDietaryExposure.

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 3.142: Table definition for Non-dietary exposure uncertainty records.

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).	idSubstance, SubstanceId, SubstanceCode, Substance	Yes
id	AlphaNumeric (50)	Uncertainty set identification number.	id	Yes
Dermal	Numeric	Dermal non-dietary exposure value.	Dermal	Yes
Oral	Numeric	Oral non-dietary exposure value.	Oral	Yes
Inhalation	Numeric	Inhalation (non-dietary) exposure value.	Inhalation	Yes

Accepted table names: NonDietaryExposuresUncertain, NonDietaryExposureUncertain.

Non-dietary exposures settings

Uncertainty settings

Table 3.143: Uncertainty settings for module Non-dietary exposures.

Name	Type	Description
Resample non-dietary exposures	Boolean	Specifies whether non-dietary exposures are resampled. Note th 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*

See also *Combining dietary and non dietary exposures*.

3.4.11 Single value dietary exposures

Single value dietary exposures are based on the single value concentrations of substances, expressed per standard (kg) bodyweight and/or single value amounts of consumed modelled food. Depending on the exposure type, dietary exposures can be short-term/acute exposures.

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

Output of this module is used by: *Single value risks*

Single value dietary exposures data formats

Single value dietary exposures are IESTI etc.

Dietary exposure data is specified through dietary exposure models. To each dietary exposure model, exposure distributions are linked.

Download empty dataset template: Zipped CSV Excel

Dietary exposure models

High level description of the dietary exposure models, specifying the id, name, description and the (reference) substance and exposure unit used for reporting the exposures. To this models, exposure percentiles and bootstrap values of the percentile may be linked.

Table 3.144: Table definition for Dietary exposure models.

Name	Type	Description	Aliases	Required
idDietary-ExposureModel	AlphaNumeric (50)	Identifier of the dietary exposure model.	id, idDietary-Exposure, idExposure-Model	Yes
Name	AlphaNumeric (100)	The name of the dietary exposure model.	Name	No
Description	AlphaNumeric (200)	Description of dietary exposure model.	Description	No
idSubstance	AlphaNumeric (50)	The code of the substance.	idSubstance, SubstanceId, SubstanceCode, Substance, idCompound, CompoundId, Compound-Code, Compound	Yes
ExposureUnit	AlphaNumeric (50)	The intake/exposure unit of the dietary exposures reported by this model. If not specified, then a default exposure unit of mg/kg BW/day is assumed.	Unit, ExposureUnit, IntakeUnit	Yes

Accepted table names: DietaryExposureModels, DietaryExposures.

Dietary exposure percentiles

Exposure percentiles linked to a dietary exposure model. The percentiles are reported in the unit specified by the exposure model to which they belong.

Table 3.145: Table definition for Dietary exposure percentiles.

Name	Type	Description	Aliases	Required
idDietary-ExposureModel	AlphaNumeric (50)	The code of the dietary exposure model to which this record belongs.	idDietary-ExposureModel	Yes
Percentage	Numeric	The percentage to which the percentile value belongs.	Individual-PropertyDouble-ValueMin	Yes
Exposure	Numeric	The percentile value. I.e., the exposure value belonging to the specified percentage.	Exposure	Yes

Accepted table names: DietaryExposurePercentiles.

Dietary exposure percentile bootstrap values

Uncertainty values, obtained from bootstrap runs, of the dietary exposure percentiles.

Table 3.146: Table definition for Dietary exposure percentile bootstrap values.

Name	Type	Description	Aliases	Required
idDietary-ExposureModel	AlphaNumeric (50)	The code of the dietary exposure model to which this record belongs.	idDietary-ExposureModel	Yes
idUncertainty-Set	AlphaNumeric (50)	The uncertainty set identifier.	idUncertainty-Set, UncertaintyId	Yes
Percentage	Numeric	The percentage to which the percentile value belongs.	Individual-PropertyDouble-ValueMin	Yes
Exposure	Numeric	The percentile value. I.e., the exposure value belonging to the specified percentage.	Exposure	Yes

Accepted table names: DietaryExposurePercentilesUncertain, DietaryExposurePercentileUncertains.

Single value dietary exposures calculation

Either the short-(acute) or long-term (chronic) dietary exposure to a substances via food can be estimated as a single value calculated from single value inputs. This is often referred to as deterministic estimation. MCRA implements the IESTI, TMDI, IEDI and NEDI (Rees-Day) calculation methods that are also available in the EFSA PRIMo (Pesticide Residue Intake Model) tool revision 3 (EFSA (2018)).

The implementation in MCRA allows more choices than EFSA PRIMo by choosing other inputs or input combinations for the calculation formula. Moreover, the calculations can in all cases be adapted for processing factors or occurrence frequencies. For the chronic estimates, also the contributions per food or processed food are reported.

Acute single value dietary exposure assessment

The short term (acute) exposure assessment is usually the exposure related to a consumption of food over a single day. MCRA applies in principle the IESTI equations as shown in EFSA PRIMo revision 3 (EFSA (2018)), but the equations are extended with a factor $\mathtt{OF}$ to allow adaptation for an occurrence frequency lower than 1. So the inputs to the equations are not necessarily the same as used in PRIMo. For example, the large portion ($\mathtt{LP}$) and body weight ($\mathtt{BW}$) can be computed instead of just being standard values.

IESTI (International Estimated Short-Term Intake)

The IESTI (International Estimated Short-Term Intake) is calculated according to different equations depending on the unit weight of the raw agricultural commodity (RAC) and the unit weight of the edible portion (EP). The following cases are distinguished.

Case 1 refers to commodities with unit weight of the raw agricultural commodity $\mathtt{U_{RAC}} \leq 25$ g (e.g. walnuts, strawberries and peas. It is also used for meat, liver, kidney, edible offal, eggs and for post-harvest uses in cereal grains, oilseeds and pulses).

Case 2a for food product with a $\mathtt{U_{RAC}} > 25$ g, where the meal portion is $> \mathtt{U_{ep}}$ (unit weight edible portion).

Case 2b for food products with a $\mathtt{U_{RAC}} > 25$ g, where the meal portion is $\leq \mathtt{U_{ep}}$.

Case 3 for food products that are usually bulked or blended before they are consumed (e.g. cereals, pulses, oilseeds and milk).

The calculations are as follows.

Case 1

$$\frac{\mathtt{LP} \cdot \mathtt{HR} \cdot \mathtt{PF} \cdot \mathtt{CF} \cdot \mathtt{OF}}{\mathtt{BW}}$$

Case 2a

$$\frac{\mathtt{U_{ep}} \cdot \mathtt{HR} \cdot \mathtt{PF} \cdot \mathtt{CF} \cdot \mathtt{VF} \cdot \mathtt{OF} + (\mathtt{LP} - \mathtt{U_{ep}}) \cdot \mathtt{HR} \cdot \mathtt{PF} \cdot \mathtt{CF} \cdot \mathtt{OF}}{\mathtt{BW}}$$

Case 2b

$$\frac{\mathtt{LP} \cdot \mathtt{HR} \cdot \mathtt{PF} \cdot \mathtt{CF} \cdot \mathtt{VF} \cdot \mathtt{OF}}{\mathtt{BW}}$$

Case 3

$$\frac{\mathtt{LP} \cdot \mathtt{STMR} \cdot \mathtt{PF} \cdot \mathtt{CF} \cdot \mathtt{OF}}{\mathtt{BW}}$$

New Case 1 and 3:

$$\frac{\mathtt{LP} \cdot \mathtt{MRL} \cdot \mathtt{PF} \cdot \mathtt{CF} \cdot \mathtt{OF}}{\mathtt{BW}}$$

New Case 2a and 2b

$$\frac{\mathtt{LP} \cdot \mathtt{MRL} \cdot \mathtt{PF} \cdot \mathtt{CF} \cdot \mathtt{VF} \cdot \mathtt{OF}}{\mathtt{BW}}$$

Parameters used in the equations

MRL: Maximum residue level for the RAC concerned (default in mg/kg);

STMR: Supervised Trials Median Residue for raw agricultural commodity (RAC) concerned (default in mg/kg);

CF: Conversion factor residue definition enforcement to residue definition risk assessment (calculated as the ratio of residues according to the residue definition for risk assessment divided by the residue concentration according to the residue definition for enforcement);

OF: Use Frequency of the raw agricultural commodity (RAC),

BW: body weight of the population related to the LP (default in kg);

LP: Large portion reported (in kg/day) (97.5th percentile of eaters (or alternative percentile, depending on the number of reported eating occasions);

HR: Highest residue according to residue definition for enforcement in composite sample (default in mg/kg);

$\mathtt{U}_{\mathtt{ep}}$: Unit weight of edible portion (in kg), provided by the country from which the LP was reported (or mean unit weight calculated from all available unit weight data, if no unit weight is available from the country matching the highest LP;

PF: Processing factor or peeling factor (calculated as the ratio of residues in processed/peeled product, divided by residue concentration in unprocessed/unpeeled product);

VF: variability factor, depending on the unit weight of the whole product $(\mathtt{U}_{\mathtt{RAC}})$, different default VF are used in the calculations.

$(\mathtt{U}_{\mathtt{RAC}}) < 25$ g, the calculations are performed according to case 1 (VF = 1).

$(\mathtt{U}_{\mathtt{RAC}})$ between 25 and 250 g: VF = 7.

$(\mathtt{U}_{\mathtt{RAC}})$ greater than 250: VF = 5.

In $\mathtt{IESTI}_{new}$, a default VF of 3 is used.

In case the empirically derived variability factors are available, the default VF is to be replaced.

Alternative IESTI-styled assessments

If consumption survey data for a specific population are available, the LP values in the IESTI equations may be replaced by statistics calculated from these data (at the consumed modelled food level).

If concentration monitoring data (retrospective) or concentration field trial data (prospective) are available, the MRL, HR, STMR values in the IESTI equations may be replaced by statistics calculated from these data (at the consumed modelled food level).

In the current use of IESTI, the occurrence frequency (use frequency) OF is assumed to be 1. In alternative assessments, a more realistic estimate may be used. Such an estimate could be derived for example as the highest occurrence frequency observed in a retrospective assessment for either the same substance or the same food.

IESTI special cases

For some foods, substances are applied after harvest, i.c. post-harvest use. For those combinations of food and substance, Case 1 should be used in the calculation. However, commodities with post-harvest use like cereal grains, oilseeds and pulses are typically bulked or blended (Case 3). To overrule Case 3, specify in table *IESTISpecialCases* the food and substance combination with 'PostHarvest' as application type. For those food and substance combinations with a unit weight of the raw agricultural commodity $\mathrm{U}_{\mathrm{RAC}} \leq 25$ g, Case 1 is applied. When substances are applied before harvest, i.c. pre-harvest use, Case 1 should be overruled by Case 3. Specify in table *IESTISpecialCases* the food and substance combination with 'PreHarvest' as application type. See also *IESTISpecialCases table format*.

Chronic single value dietary exposure assessment

The long term (chronic) exposure assessment is usually the exposure related to a consumption over a longer period of time. MCRA applies in principle the TMDI, IEDI or NEDI (Rees-Day) equations as shown in EFSA PRIMo revision 3 (EFSA (2018)). However, the equations are extended with factors PF_i and OF_i to allow adaptation for processing factors and occurrence frequencies lower than 1. Also, the inputs to the equations are not necessarily the same as used in PRIMo. For example, the consumption statistics (MC, $p_{97.5}$) and body weight (BW) can be computed instead of just being standard values. Note that TMDI, IEDI and NEDI (Rees-Day) estimates are summations over foods (raw agricultural products). In addition to the summations, MCRA will also report the individual terms (single value dietary exposures per food).

TMDI (Theoretical Maximum Dietary Intake)

$$\sum_{X=i}^{n} \frac{MRL_i \cdot CF_i \cdot PF_i \cdot OF_i \cdot MC_i}{BW}$$

$i, j, k, ...n$: individual raw agricultural products

IEDI (International Estimated Dietary Intake)

$$\sum_{X=i}^{n} \frac{STMR_i \cdot CF_i \cdot PF_i \cdot OF_i \cdot MC_i}{BW}$$

$i, j, k, ...n$: individual raw agricultural products

NEDI (National Estimated Dietary Intake): Rees-Day model (I)

$$\sum_{X=i}^{j} \frac{MRL_i \cdot CF_i \cdot PF_i \cdot OF_i \cdot p_{97.5}consumption_i}{BW} + \sum_{X=k}^{n} \frac{MRL_k \cdot CF_k \cdot PF_i \cdot OF_i \cdot MC_k}{BW}$$

i, j: two raw agricultural products leading to the highest intake;

$k, l, m, ...n$: remaining raw agricultural commodities consumed

NEDI (National Estimated Dietary Intake): Rees-Day model (II)

$$\sum_{X=i}^{j} \frac{STMR_i \cdot CF_i \cdot PF_i \cdot OF_i \cdot p_{97.5}consumption_i}{BW} + \sum_{X=k}^{n} \frac{STMR_k \cdot CF_i \cdot PF_i \cdot OF_i \cdot MC_k}{BW}$$

i, j: two raw agricultural products leading to the highest intake;

$k, l, m, ...n$: remaining raw agricultural commodities consumed

Parameters used in the equations

MRL$_i$: Maximum residue level for the RAC concerned (default in mg/kg);

STMR$_i$: Supervised Trials Median Residue for raw agricultural commodity (RAC) concerned (default in mg/kg);

CF$_i$: Conversion factor residue definition enforcement to residue definition risk assessment (calculated as the ratio of residues according to the residue definition for risk assessment divided by the residue concentration according to the residue definition for enforcement);

MC$_i$: mean consumption for a given raw agricultural product (RAC) calculated for the whole survey/subgroup of the survey, including processed products (recalculated to the unprocessed RAC) (default in kg/day);

$p_{97.5}$ *consumption*$_i$ for a given raw agricultural product (RAC), calculated from the individual consumption reported by the participants of the whole survey/subgroup of the survey, including processed products (recalculated to the unprocessed RAC) (default in kg/day);

BW: mean body weight of the population related to the LP or mean consumption (default in kg).

OF$_i$: Occurrence Frequency of the substance on the food (typically, a raw agricultural commodity, RAC),

PF$_i$: Processing factor or peeling factor (calculated as the ratio of residues in processed/peeled product, divided by residue concentration in unprocessed/unpeeled product);

Alternative TMDI-, IEDI- or NEDI-styled assessments

If consumption survey data for a specific population are available, the $MC, p_{97.5}$*consumption* values in the IESTI equations may be replaced by statistics calculated from these data (at the consumed modelled food level).

If concentration monitoring data (retrospective) or concentration field trial data (prospective) are available, the $MRL, STMR$ values in the IESTI equations may be replaced by statistics calculated from these data (at the consumed modelled food level).

In the current use of IESTI, the occurrence frequency (use frequency) OF is assumed to be 1. In alternative assessments, a more realistic estimate may be used. Such an estimate could be derived for example as the highest occurrence frequency observed in a retrospective assessment for either the same substance or the same food.

Single value dietary exposures settings

Selection settings

Table 3.147: Selection settings for module Single value dietary exposures.

Name	Type	Description
Risk type	*ExposureType*	The type of exposure considered in the assessment; acute (short term) or chronic (long-term).
Dietary exposure calculation tier	*DietaryIntakeCalculationTier*	A tier is a pre-specified set of model configurations. By selectin model tier, MCRA automatically sets all model settings in this module according to this tier. Note that currently tier setting m need to be performed separately in sub-modules. Use the Custo tier when you want to manually set each model setting.

Calculation settings

Table 3.148: Calculation settings for module Single value dietary exposures.

Name	Type	Description
Single value dietary exposure calculation method	*SingleValueDietaryExposuresCalculationMethod*	Method for computing single value dietary exposures.
Apply processing factors	Boolean	Specified in table ProcessingFactor. If checked, processing fact are applied. Concentrations in the consumed food may be different from concentrations in the modelled food in monitorin programs (typically raw food) due to processing, such as peelin washing, cooking etc. If unchecked, no processing information used. This is in most (though not all) cases a worst-case assumption
Use occurrence frequencies	Boolean	Account for occurrence frequencies for combinations of food a substance in the exposure calculations.

Calculation of single value dietary exposures

Single value dietary exposures are calculated from single value consumptions per modelled food and single value concentrations. Optionally, also processing factors, unit variability models and use frequencies are applied.

- *Single value dietary exposures calculation*

Inputs used: *Single value consumptions Single value concentrations Processing factors Unit variability factors Occurrence frequencies*

Settings used

- *Calculation Settings*

3.5 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 modelled 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*.

3.5.1 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 Single value concentrations Occurrence patterns Occurrence frequencies Substance conversions Non-dietary exposures Kinetic models Relative potency factors Hazard characterisations Inter-species conversions Intra species factors Concentration models Food conversions High exposure food-substance combinations Dietary exposures Exposures Human monitoring analysis*

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 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.

Download empty dataset template: Zipped CSV Excel

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 3.149: Table definition for Assessment group membership models.

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
idIndex-Substance	AlphaNumeric (50)	The id/code of the index substance.	idIndex-Substance, idReference-Substance, IndexSubstance-Id, Reference-SubstanceId	No
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

Accepted table names: AssessmentGroupMembershipModels, AssessmentGroupMembershipModel.

Assessment group memberships

Substances belong to an assessment group with certainty (probability 1), or the membership are 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 3.150: Table definition for Assessment group memberships.

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	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

Accepted table names: AssessmentGroupMemberships, AssessmentGroupMembership.

Active substances calculation

Depending on the *model settings*, the set of active substances for a specified effect is 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 docking models are first translated to crisp 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. (2020) 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 from the assessment group. Set option *Include substances without membership information* to include them in the assessment group.

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 3.151: Calculation settings for module Active substances.

Name	Type	Description
Filter by certain assessment group membership	Boolean	Filter substances by certain assessment group membership.
Filter by possible assessment group membership	Boolean	Filter substances by possible assessment group membership.
Restrict active substances to substances with available PODs	Boolean	Restrict assessment group membership based on presence/abse of points of departure.
Restrict active substances to substances with available hazard characterisations	Boolean	Restrict assessment group membership based on presence/abse of hazard characterisations.
Derive memberships from QSAR membership data	Boolean	Specifies whether QSAR membership data is used for computi the assessment group memberships.
Derive memberships from molecular docking data	Boolean	Specifies whether molecular docking data is used for computin the assessment group memberhips.
Include substances without membership information	Boolean	For non-probabilistic methods: specifies whether substances fo 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	*CombinationMethodMembershipInfoAndPodPresence*	Specifies whether to take the intersection or the union of the se substances with available PoDs and the set of substances with positive/probable (in-silico) membership score.
Membership calculation method	*AssessmentGroupMembershipCalculationMethod*	Calculation method for computing assessment group memberships: majority/any (crisp methods), ratio/Bayesian (probabilistic methods)
Default/prior membership probability	Numeric	Default substance membership probability for which no membership information is available in the specified inputs. Pri probability for Bayesian method.
Use probabilistic assessment group memberships	Boolean	Specifies whether substance memberships should be expressed terms of probabilities (probabilistic). Otherwise, substance memberships are expressed as in or out (crisp).

Uncertainty settings

Table 3.152: Uncertainty settings for module Active substances.

Name	Type	Description
Resample assessment group memberships	Boolean	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 Hazard characterisations*

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*

3.5.2 AOP networks

Effects are 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 initiating 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 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.

Download empty dataset template: Zipped CSV Excel

AOP networks

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

Table 3.153: Table definition for AOP networks.

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	*ExposureType*	The risk type of the adverse outcome. Acute or chronic.	RiskType	No

Accepted table names: AOPNetworks, AOPNetwork.

Effect relations

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

Table 3.154: Table definition for Effect relations.

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

Accepted table names: EffectRelations, EffectRelation, EffectRelationships, EffectRelationship, KeyEventRelationships, KeyEventRelationship.

AOP Networks calculation

Find below a simplified AOP network based on a MCRA analysis for focal effect Steatosis-liver and related effects Fatty liver cells (tissue level), Cytoplasm displacement liver, Endoplasmatic reticulum stress liver, Mitochondrial disruption liver, Nucleus distortion liver and Triglyceride accumulation liver (cellular level).

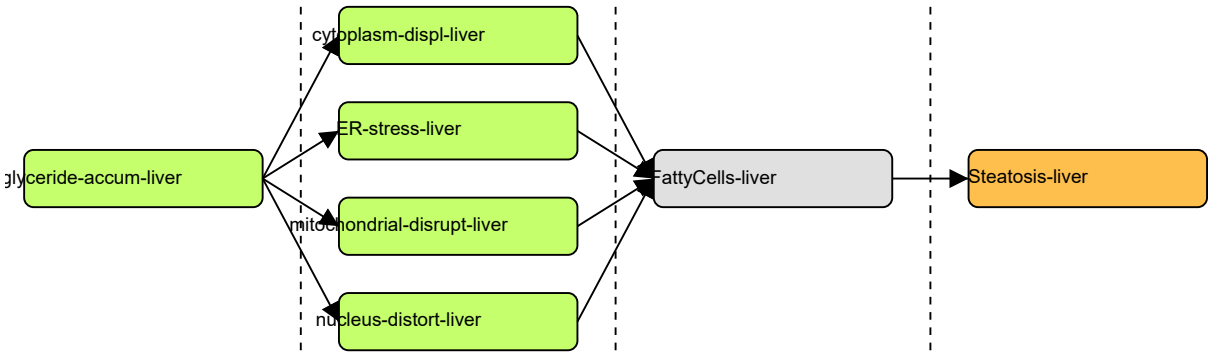


Figure 3.65: AOP network liver steatosis

AOP networks settings

Selection settings

Table 3.155: Selection settings for module AOP networks.

Name	Type	Description
AOP Network	AlphaNumeric	The AOP networks of interest.
Restrict AOP network by focal upstream event	Boolean	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	AlphaNumeric	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*
- *AOP networks calculation*

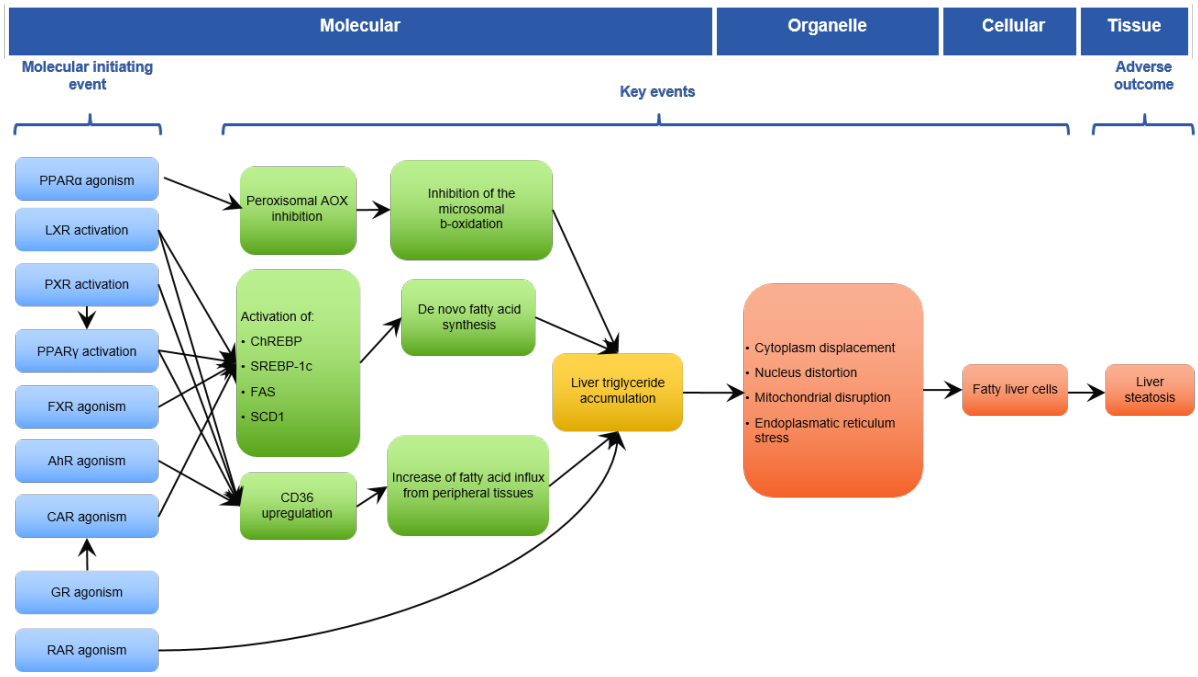


Figure 3.66: AOP network

3.5.3 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 are 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 tabular format 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.

Tabular dose response data format

In the tabular dose response data format, the substance doses and response measurements are provided in the same data table as columns and automatically parsed based on the specification of the substances and responses of the experiments in the dose response experiments table.

Download empty dataset template: Zipped CSV Excel

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 are used to map the column header names of the columns of the single data table to these their respective types.

Table 3.156: Table definition for Dose response experiments.

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 (50)	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	Yes
DoseUnit	*DoseUnit*	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 of the experiment. E.g. 'Gender, Inhibitor,	Covariates, Covariate	No

Accepted table names: DoseResponseExperiments, DoseResponseExperiment.

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 are specified as [Datatype:Response]. E.g., 'SD:Y', 'CV:Y', 'N:Y'. Uncertainty upper 95%limits are 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 3.157: Table definition for Dose response data.

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.	Experimental-Unit, Experimental-Units, 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

Accepted table names: TwoWayDoseResponseData, DoseResponseDataTwoWay, DoseResponseData.

Relational dose response data format

The relational dose response data format is the internal format of MCRA. In this format, the dose response experiments, doses, responses, and experimental units are provided in separate data tables.

Download empty dataset template: Zipped CSV Excel

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 are used to map the column header names of the columns of the single data table to these their respective types.

Table 3.158: Table definition for Dose response experiments.

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 (50)	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	Yes
DoseUnit	*DoseUnit*	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 of the experiment. E.g. 'Gender, Inhibitor,	Covariates, Covariate	No

Accepted table names: DoseResponseExperiments, DoseResponseExperiment.

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 3.159: Table definition for Dose response experiment doses.

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	Yes
Dose	Numeric	The dose that was administered.	Dose	Yes

Accepted table names: DoseResponseExperimentDoses, DoseResponseExperimentDose.

Experimental unit properties

The table ExperimentalUnitProperties are 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 3.160: Table definition for Experimental unit properties.

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

Accepted table names: ExperimentalUnitProperties, ExperimentalUnitProperty.

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 3.161: Table definition for Dose response experiment measurements.

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

Accepted table names: DoseResponseExperimentMeasurements, DoseResponseExperimentMeasurement, DoseResponseMeasurements, DoseResponseMeasurement.

Dose response data settings

Selection settings

Table 3.162: Selection settings for module Dose response data.

Name	Type	Description
Experiments	AlphaNumeric	The dose response experiments of interest.
Merge dose response data of multiple experiments	Boolean	Specifies whether the dose response data of multiple experimen should be merged into one large dose response data set.

Dose response data as data

Dose response data are 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*

3.5.4 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 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.

Download empty dataset template: Zipped CSV Excel

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 3.163: Table definition for Dose response models.

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 (50)	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-ResponseType*	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

Accepted table names: DoseResponseModels, DoseResponseModel.

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 3.164: Table definition for Dose response model benchmark doses.

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	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

Accepted table names: DoseResponseModelBenchmarkDoses.

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 3.165: Table definition for Dose response model benchmark dose bootstraps.

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
idUncertainty-Set	AlphaNumeric (50)	The uncertainty set identifier.	idUncertainty-Set, UncertaintyId	Yes
idSubstance	AlphaNumeric (50)	The code of the substance.	idSubstance, SubstanceId, SubstanceCode, Substance	Yes
Covariates	AlphaNumeric (500)	Comma separated list of the covariate values for which this benchmark dose applies.	Covariates	No
Benchmark-Dose	Numeric	Benchmark dose (BMD).	Benchmark-Dose, BMD, CED	Yes

Accepted table names: DoseResponseModelBenchmarkDosesBootstraps,
DoseResponseModelBenchmarkDosesUncertain.

Dose response models calculation

Dose response models are uploaded as data or retrieved from the PROASTweb through *linked remote repositories*. A second possibility is to compute dose response models using an integrated version of PROAST: for each response in a dose response experiment a dose response model is fitted. 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.

PROAST (copyright RIVM National Institute for Public Health and the Environment) is a software package for the statistical analysis of dose-response data. Its main purpose is dose-response modelling of toxicological data, and the derivation of a Benchmark dose (BMDL) in human risk assessment (or an ECx in ecotoxicological risk assessment). More generally, it can be used for (nonlinear) regression (with covariates) (see Slob (2002), Slob and Setzer (2013))

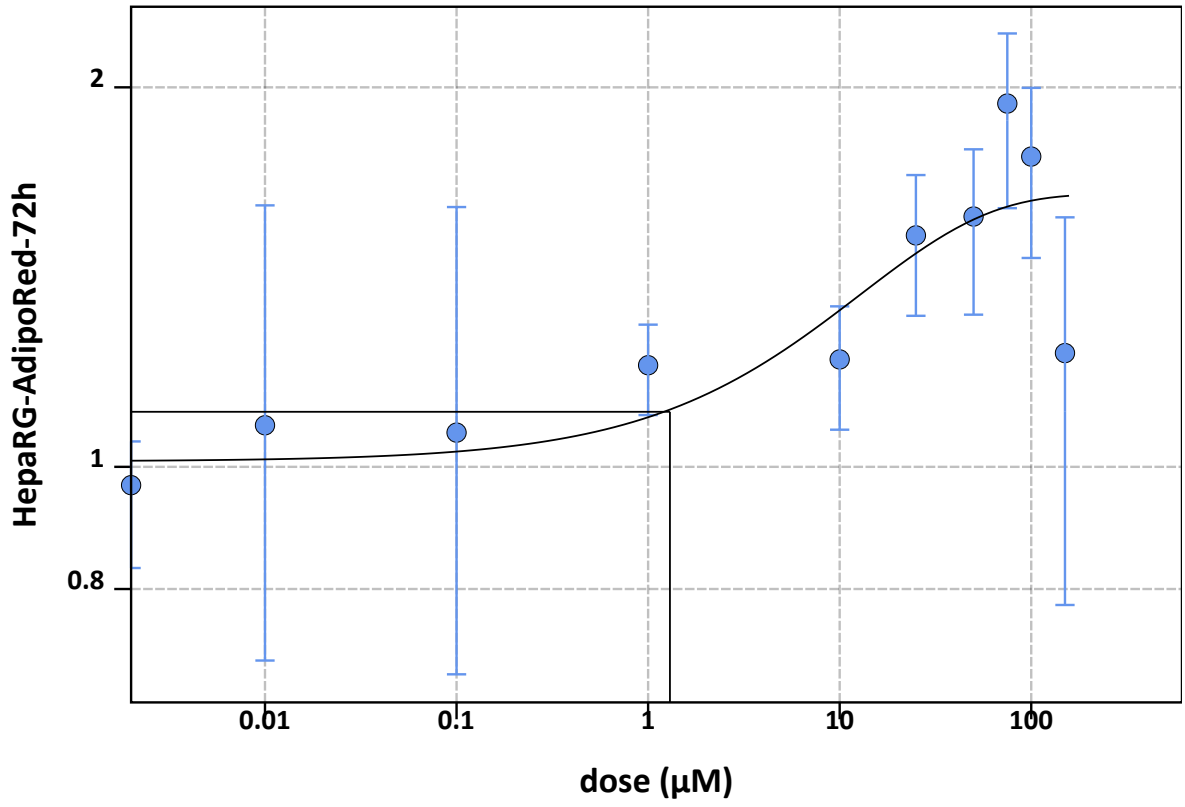


Figure 3.67: Dose response model: HepaRG Adipo72 72h.

Proast models

In Proast, a family of (nested) dose-response models are available that can be used for describing the change in any continuous endpoint as a function of dose. The likelihood ratio test is used to select one of the available models (model selection to prevent overparameterization).

- Model 1: $y = a$ with $a > 0$
- Model 2: $y = a \cdot exp(x/b)$ with $a > 0$
- Model 3: $y = a \cdot exp(\pm(x/b)^d)$ with $a > 0, b > 0, d \geq 1$
- Model 4: $y = a[c - (c-1)exp(-x/b)]$ with $a > 0, b > 0, c > 0$
- Model 5: $y - (c-1)exp(-(x/b)^d)]$ with $a > 0, b > 0, c > 0, d \geq 1$

where y is any continuous endpoint and x denotes the dose. In all models parameter a represents the level of the endpoint at dose 0, and b is considered as the parameter reflecting the efficacy of the substance or the sensitivity of the subject. At high doses model 4 and 5 level of to the value $a \cdot c$, so the parameter c can be interpreted as the maximum relative change. Model 3 and 5 have the flexibility to mimic threshold-like responses. All these model are nested to each other, except models 3 and 4, which both have three parameters.

In all models the parameter a is constrained to being positive for obvious reasons (it denotes the value of the endpoint at dose 0). The parameter d is constrained to values larger than (or equal to) 1, to prevent the slope of the function at dose 0 being infinite, which seems biologically implausible. The parameter b is constrained to be positive in all models. Parameter c in models 4 and 5 determines whether the function increases or decreases, by being larger or smaller than unity, respectively. To make model 3 a decreasing function a minus sign has to be inserted in the exponent (Slob (2002), Slob and Setzer (2013)).

Dose response models settings

Calculation settings

Table 3.166: Calculation settings for module Dose response models.

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

Uncertainty settings

Table 3.167: Uncertainty settings for module Dose response models.

Name	Type	Description
Resample hazard characterisations or RPFs	Boolean	Specifies whether to resample the hazard characterisations or relative potency factors. Requires hazard characterisation or RF uncertainty to be quantified in DoseResponseModelsUncertain RelativePotencyFactorsUncertain tables.

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*

Settings used

- *Calculation Settings*

3.5.5 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 specify responses that may represent the effect.

Download empty dataset template: Zipped CSV Excel

Effect representations

One response can be set as the canonical response (golden standard). For a quantitative or stochastically qualitative canonical response a benchmark response should be defined.

Table 3.168: Table definition for Effect representations.

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-ResponseType*	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

Accepted table names: EffectRepresentations, EffectRepresentation.

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*

3.5.6 Hazard characterisations

Hazard characterisations are reference exposure values for active substances at the chosen biological target level (external or internal). Hazard characterisations may be specified for specific effects or for the critical effect as defined in hazard characterisation. Hazard characterisations are specified as external values (e.g. human based guidance values, such as ADI or ARfD) or are based on points of departure, such as BMDs from dose-response models or externally specified points of departure (NOAEL, LOAEL, MDS). The computation may involve assessment factors, e.g. for inter-species conversion, intra-species variation or additional sources of uncertainty. The calculation may also use kinetic models or absorption factors to convert external doses to internal doses or vice versa.

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

Output of this module is used by: *Active substances Relative potency factors Risks Single value risks*

Hazard characterisations data formats

Hazard characterisations provide reference threshold values associated with the hazard of interest. Examples are health-based guidance values such as ADI or ARfD, and points of departure such as BMD or NOAEL.

Download empty dataset template: Zipped CSV Excel

Hazard characterisations

Hazard characterisations are specified for combinations of hazard characterisation type, effect, substance, population type, target level, and exposure route (for external) or target organ (for internal). Effects can be specific, but can also be labelled as being the critical effect and used as such if this has been specified in the hazard characterisation settings.

Table 3.169: Table definition for Hazard characterisations.

Name	Type	Description	Aliases	Required
idHazard-Characterisation	AlphaNumeric (50)	Id of the hazard characterisation.	id, idHazard-Characterisation	Yes
idEffect	AlphaNumeric (50)	Code of the (critical) effect linked to this hazard characterisation.	idEffect, EffectId, Effect, EffectCode	No
idSubstance	AlphaNumeric (50)	The code of the substance.	idSubstance, SubstanceCode, SubstanceId, Substance	Yes
idPopulation-Type	AlphaNumeric (50)	The code of the population type for which this reference value is defined. If not specified, PS06A, Consumers is assumed.	idPopulation-Type, PopulationType, Population-Group, Population-Subgroup	No
TargetLevel	*TargetLevelType*	The target level. I.e., internal or external. If omitted, external is assumed	TargetLevel	No
ExposureRoute	*ExposureRouteType*	The exposure route (only applicable if target level is external). If not specified, Dietary is assumed.	ExposureRoute	No
Matrix	AlphaNumeric (50)	The biological matrix or organ (should be specified when target level is internal).	Matrix, BiologicalMatrix	No
IsCriticalEffect	Boolean	Specifies whether this value is the value associated with the critical effect. If omitted, No is assumed	IsCriticalEffect	No
ExposureType	*ExposureType*	The exposure type associated with the hazard characterisation (i.e., chronic or acute).	ExposureType	Yes
Hazard-Characterisation-Type	*Hazard-Characterisation-Type*	The type of the hazard characterisation (e.g., ARfD, ADI, NOAEL, BMD).	Hazard-Characterisation-Type	Yes
Qualifier	*ValueQualifier*	Qualifier of the hazard characterisation value, e.g. equal-to (=) or smaller-than (<). If omitted, = is assumed.	QualifierType	No
Value	Numeric	Reference value that characterises the hazard.	Value, Hazard-Characterisation-Value	Yes
DoseUnit	*DoseUnit*	Unit of the hazard characterisation value.	DoseUnit, Unit	Yes
idPointOf-Departure	AlphaNumeric (50)	The code of the point of departure from which this hazard characterisation was derived.	idHazardDose, idPod	No
Combined-Assessment-Factor	Numeric	Combined assessment factor (includes, e.g., safety factor, but also other extrapolation factors that may be used to derive the hazard characterisation from the underlying PoD).	Combined-Assessment-Factor, Assessment-Factor, Uncertainty-Factor	No
PublicationTitle	AlphaNumeric (250)	Title of the publication of the study in which this hazard characterisation was established.	PublicationTitle, Title	No

Accepted table names: HazardCharacterisations.

Hazard characterisations calculation

Hazard characterisations are defined as deterministic threshold values (e.g. ADI, ARfD) or as distributions (using probabilistic models). They are linked to an effect of interest or alternatively are defined for the critical effect. Hazard characterisations depend on the *risk type* (acute or chronic) and the biological *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 points of departure (PoD, e.g. BMD, NOAEL, LOAEL).
2. Calculation of hazard characterisations from PoDs (in this case BMD) calculated from dose response data.
3. 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)* (cumulative assessments only) .

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:

$$HC = f_{\texttt{expression-type}} \cdot f_{\texttt{kinetic}} \cdot \frac{1}{f_{\texttt{inter-species}}} \cdot \frac{1}{f_{\texttt{intra-species}}} \cdot \frac{1}{f_{\texttt{additional}}} \cdot PoD$$

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.
- $f_{\texttt{kinetic}}$ denotes the kinetic conversion factor for *conversion from internal to external or external to internal hazard characterisations*.
- $f_{\texttt{inter-species}}$ denotes the inter-species factor for *extrapolation from animal to human (inter-species)*.
- $f_{\texttt{intra-species}}$ denotes the intra-species factor for *extrapolation from the average to the sensitive human or probabilistic calculation of the distribution of human individuals (intra-species)*.
- $f_{\texttt{additional}}$ denotes the additional assessment factor for *extrapolation from the POD to the hazard characterisation in humans for sources where appropriate data or information is scarce or missing (additional)*.
- *PoD* denotes the point of departure.

Note that inter- and intra-species extrapolation and the use of an additional assessment factor are optional. However, expression type correction and the kinetic conversion are always applied (when relevant) whatever option is chosen.

Occasionally, for some substances multiple hazard characterisations are available (e.g., obtained from multiple experiments) and for others substance 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*. Set the *selection method in case of multiple candidate hazard characterisations* from *MostToxic* to *Aggregate*.
- *Imputation of missing hazard characterisations*. Check the option *Imput missing hazard characterisations*, to select the Imputation method.

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 3.170 are used, roughly based on the WHO guidance document on evaluating and expressing uncertainty in hazard characterization, see WHO (2018).

Table 3.170: 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 MCRA:

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 conversion factor of 1.
2. **Default distribution:** Use a conversion factor drawn from a default, substance and species independent lognormal uncertainty distribution specified (as *model settings*) by a geometric mean (GM) and geometric 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 lognormal uncertainty distributions specified (as *data*) by a geometric mean (GM) and geometric 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 $f_{\texttt{inter-species}}$ 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.

Additional assessment factors

In cases where where appropriate data or information is scarce or missing, an additional assessment factor is used for extrapolation from the POD to the hazard characterisation in humans. This factor can be used as a worst case value (preferably AF = 100) for inter- and intra-species extrapolation, but it may equally well serve as an additional extrapolation factor next to inter- and/or intra-species extrapolation. In the latter case the factor merely serves to account for differences in for example route to route, metabolic rates in interspecies, completeness and consistency of available data, reliability of alternative data (e.g. read-across) or quality of data in general.

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 3.68 shows the conceptual model that forms the basis of the IVIVE methodology of MCRA.

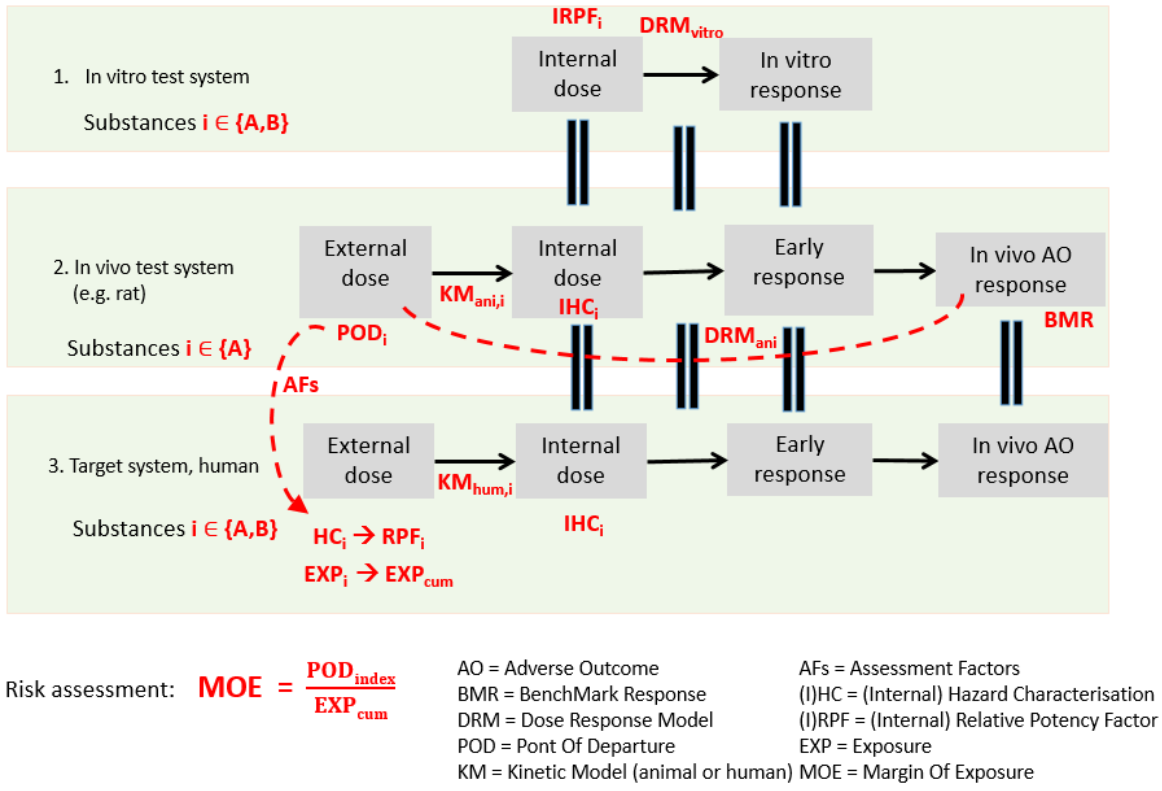


Figure 3.68: Conceptual model IVIVE.

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.,

$$RPF_{\texttt{mass-based},i} = RPF_{\texttt{mol-based},i} \cdot \frac{MW_{\texttt{ref}}}{MW_i}$$

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.,

$$HC_i = HC_{\texttt{ref}} \cdot RPF_{\texttt{mass-based},i}$$

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 MCRA, 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 MCRA, 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 $HC_{\texttt{int}}$ can be derived from an external hazard characterisation $HC_{\texttt{ext}}$ as

$$HC_{\texttt{int}} = f_{\texttt{abs},r} \cdot HC_{\texttt{ext},r}$$

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 preferred 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 coming 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 corresponding 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 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 substances that is already in common use (see Munro et al. (1996)). 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 Cramer et al. (1976)), 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 (HC) of the available hazard characterisations. I.e.,

$$\mathrm{HC} = \min_{i=1,\ldots,n} \mathrm{HC}_i$$

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.

$$\mathrm{HC} = \left(\sum_{i}^{n} \frac{1}{\mathrm{HC}_i} \right)^{-1}$$

Hazard characterisations settings

Selection settings

Table 3.171: Selection settings for module Hazard characterisations.

Name	Type	Description
Risk type	*ExposureType*	The type of exposure considered in the assessment; acute (short term) or chronic (long-term).
Target level	*TargetLevelType*	Select to express hazard characterisations at external or internal exposure level. For an aggregate assessment, that is dietary and nondietary exposure data are combined, the target dose level is always internal. When only dietary exposures are available, the target dose level is optional, i.c. external or internal.
Consider critical effect	Boolean	Specifies whether the analysis should look at critical effects such as specified in the Hazard characterisation data source.

Calculation settings

Table 3.172: Calculation settings for module Hazard characterisations.

Name	Type	Description
Index substance	AlphaNumeric	The substance of interest or index substance.
Method	*TargetDosesCalculationMethod*	Choose method for computing the hazard characterisations: fro in-vivo or in-vitro points of departure or both.
Expression type	*PointOfDeparture*	Specifies how hazard characterisations are expressed: as BMD, NOAEL, or the expression type is ignored.
Selection method in case of multiple candidate hazard characterisations	*TargetDoseSelectionMethod*	Choose either the most toxic (default) or an aggregated hazard characterisation when in nominal runs there are multiple availal candidates. In uncertainty runs, multiple candidates are resampled.
Impute missing hazard characterisations	Boolean	If checked, missing hazard characterisations are imputed based Munro NOELs or on other available points of departure.
Imputation method	*HazardDoseImputationMethod-Type*	Imputation of Hazard characterisations: use low percentile (P5) unbiased central estimate from either the Munro set or the available POD collection.
Use BMDs from dose response models	Boolean	If checked, preferably BMDs from dose response models will b used. If these data are not available, other POD data are used.
Use inter-species conversions	Boolean	Use inter-species conversion factors (default value, e.g. 10, or data).
Use intra-species factors	Boolean	Use intra-species conversion factors (default value, e.g. 10, or data).
Use additional assessment factor	Boolean	Use additional assessment factor for extrapolation of PODs to (human) hazard characterisations.
Include dietary and non-dietary routes of exposure	Boolean	Specifies whether the assessment involves both dietary and non-dietary (oral, inhalatory or dermal) routes of exposure.

Uncertainty settings

Table 3.173: Uncertainty settings for module Hazard characterisations.

Name	Type	Description
Resample intra-species factor	Boolean	Specifies whether intra-species factors are resampled from a parametric uncertainty distribution.
Resample hazard characterisations or RPFs	Boolean	Specifies whether to resample the hazard characterisations or relative potency factors. Requires hazard characterisation or RP uncertainty to be quantified in DoseResponseModelsUncertain RelativePotencyFactorsUncertain tables.

Hazard characterisations as data

Hazard characterisations can be provided as data e.g., in the form of ADI or ARfD.

- *Hazard characterisations data formats*

Inputs used: *AOP networks Active substances Points of departure*

Calculation of hazard characterisations

Hazard characterisations can be computed from points of departure. The computation may involve assessment factors, e.g. for inter-species conversion, intra-species variation or additional sources of uncertainty. The additional assessment factor can be used to bypass inter- and intra species conversion, or as an additional assessment factor to account for extrapolation for sources where appropriate data or information is scarce or missing (e.g. to implement a mixture assessment factor). The hazard characterisation calculation may also use kinetic models or absorption factors to convert external doses to internal doses or vice versa.

- *Hazard characterisations calculation*

Inputs used: *Dose response models Effect representations Inter-species conversions Intra species factors Kinetic models*

Settings used

- *Calculation Settings*

3.5.7 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 conversion models specify how to convert a hazard dose for a given species to a hazard dose for humans.

Download empty dataset template: Zipped CSV Excel

Inter-species model parameters

Inter-species extrapolation factors are described using a lognormal distribution specified by a geometric mean (GM) and geometric standard deviation (GSD). Inter-species factors are defined for an effect and a species and may optionally be specified specifically for a substance.

Table 3.174: Table definition for Inter-species model parameters.

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	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	*BodyWeightUnit*	The unit of the human body weight specification (kg is assumed if not defined).	HumanBody-WeightUnit	No
Standard-AnimalBody-Weight	Numeric	The standard animal body weight.	Standard-AnimalBody-Weight	Yes
AnimalBody-WeightUnit	*BodyWeightUnit*	The unit of the animal body weight specification (kg is assumed if not defined).	AnimalBody-WeightUnit	No

Accepted table names: InterSpeciesModelParameters, InterSpeciesModelParameter, InterSpeciesFactors, InterSpeciesFactor.

Inter-species conversions settings

Selection settings

Table 3.175: Selection settings for module Inter-species conversions.

Name	Type	Description
Interspecies factor geometric mean	Numeric	Interspecies factor geometric mean.
Interspecies factor geometric standard deviation	Numeric	Interspecies factor geometric standard deviation.

Uncertainty settings

Table 3.176: Uncertainty settings for module Inter-species conversions.

Name	Type	Description
Resample inter-species factor	Boolean	Specifies whether inter-species factors are resampled from a parametric uncertainty distribution.

Inter-species conversions as data

Data are provided in the form of a geometric mean (GM) and geometric standard deviation (GSD)

- *Inter-species conversions data formats*

Inputs used: *Active substances*

3.5.8 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.

Download empty dataset template: Zipped CSV Excel

Intra-species model parameters

Intra species factors.

Table 3.177: Table definition for Intra-species model parameters.

Name	Type	Description	Aliases	Required
idEffect	AlphaNumeric (50)	The effect code.	idEffect, EffectId, Effect	Yes
idSubstance	AlphaNumeric (50)	The code of the substance.	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

Accepted table names: IntraSpeciesModelParameters, IntraSpeciesModelParameter, IntraSpeciesFactors, IntraSpeciesFactor.

Intra species factors settings

Selection settings

Table 3.178: Selection settings for module Intra species factors.

Name	Type	Description
Intra-species factor	Numeric	Intra-species 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*

Inputs used: *Active substances*

3.5.9 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, 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.

Download empty dataset template: Zipped CSV Excel

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 3.179: Table definition for Points of departure.

Name	Type	Description	Aliases	Required
idModel	AlphaNumeric (50)	The dose response model code.	idDose-ResponseModel, idModel, idPod, idPointOf-Departure	No
idEffect	AlphaNumeric (50)	The effect code.	idEffect, EffectId, Effect	Yes
idSubstance	AlphaNumeric (50)	The code of the substance.	idSubstance, SubstanceId, SubstanceCode, Substance	Yes
Species	AlphaNumeric (50)	The species used to obtain this point of departure.	Species, System	No
Value	Numeric	Point of departure, can be of various types, e.g. NOAEL, LOAEL, BMD, CED	PointOf-Departure, LimitDose, HazardDose, Value, CED	Yes
Type	*PointOfDeparture-Type*	The type of the point of departure: e.g. NOAEL, LOAEL, BMD (default).	Type, PODType, HazardDose-Type, LimitDoseType	No
DoseUnit	*DoseUnit*	The dose unit (if not specified, then mg/kg is assumed).	DoseUnit, Unit, UnitDose	No
Benchmark-Response	AlphaNumeric (100)	The benchmark response or effect size.	Benchmark-Response, CriticalEffect-Size, HazardEffect-Size	No
ExposureRoute	*ExposureRouteType*	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
IsCriticalEffect	Boolean	Specifies whether this value is the value associated with the critical effect. If omitted, No is assumed	IsCriticalEffect	No

Accepted table names: PointsOfDeparture, PointOfDeparture, HazardDoses, HazardDose.

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 3.180: Table definition for Points of departure uncertainty.

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	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

Accepted table names: PointsOfDepartureUncertain, PointOfDepartureUncertain, HazardDosesUncertain, HazardDoseUncertain.

Points of departure settings

Uncertainty settings

Table 3.181: Uncertainty settings for module Points of departure.

Name	Type	Description
Resample hazard characterisations or RPFs	Boolean	Specifies whether to resample the hazard characterisations or relative potency factors. Requires hazard characterisation or RF uncertainty to be quantified in DoseResponseModelsUncertain RelativePotencyFactorsUncertain tables.

Points of departure as data

Points of departure are provided as data for combinations of substance and effect and each is minimally described by a reference value and a type (e.g., NOAEL or LOAEL). In addition, the exposure route, specifies, and references may be specified.

- *Points of departure data formats*

Inputs used: *AOP networks*

3.5.10 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 High exposure food-substance combinations Dietary exposures Exposures Exposure mixtures Human monitoring analysis Risks*

Relative potency factors data formats

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.

Download empty dataset template: Zipped CSV Excel

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 3.182: Table definition for Relative potency factors.

Name	Type	Description	Aliases	Required
idSubstance	AlphaNumeric (50)	The code of the substance.	idSubstance, SubstanceId, SubstanceCode, Substance	Yes
idEffect	AlphaNumeric (50)	The effect code.	idEffect, EffectId, Effect	Yes
RPF	Numeric	The relative potency factor.	RPF, Relative-PotencyFactor	Yes

Accepted table names: RelativePotencyFactors, RelativePotencyFactor.

Relative potency factor uncertainty

This table contains sets of values representing the uncertainty for relative potency factors.

Table 3.183: Table definition for Relative potency factor uncertainty.

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	Yes
RPF	Numeric	The relative potency factor.	RPF, Relative-PotencyFactor	Yes

Accepted table names: RelativePotencyFactorsUncertain, RelativePotencyFactorUncertain.

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

Calculation settings

Table 3.184: Calculation settings for module Relative potency factors.

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

Uncertainty settings

Table 3.185: Uncertainty settings for module Relative potency factors.

Name	Type	Description
Resample hazard characterisations or RPFs	Boolean	Specifies whether to resample the hazard characterisations or relative potency factors. Requires hazard characterisation or RF uncertainty to be quantified in DoseResponseModelsUncertain RelativePotencyFactorsUncertain tables.

Relative potency factors as data

Data are provided in the form of a RPF for a specific substance and effect.

- *Relative potency factors data formats*

Inputs used: *Active substances AOP networks*

Calculation of relative potency factors

RPFs are computed from hazard characterisations.

- *Relative potency factors calculation*

Inputs used: *Hazard characterisations*

Settings used

- *Calculation Settings*

3.6 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.

3.6.1 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

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.

Download empty dataset template: Zipped CSV Excel

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 3.186: Table definition for Molecular docking models.

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).		Yes
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

Accepted table names: MolecularDockingModels, MolecularDockingModel, BindingEnergyModels, BindingEnergyModel.

Molecular docking binding energies

Molecular docking model binding energies per substance

Table 3.187: Table definition for Molecular docking binding energies.

Name	Type	Description	Aliases	Required
idMolecular-DockingModel	AlphaNumeric (50)	The id of the molecular docking model or source.	idMolecular-Docking, Molecular-DockingModel	Yes
idSubstance	AlphaNumeric (50)	The code of the substance.	idSubstance, SubstanceId, SubstanceCode, Substance	Yes
BindingEnergy	Numeric	Molecular Docking binding energy.	Molecular-Docking-BindingEnergy	Yes

Accepted table names: MolecularBindingEnergies, MolecularBindingEnergy, BindingEnergies, BindingEnergy, MolecularDockingBindingEnergies, MolecularDockingBindingEnergy.

Molecular docking models as data

Binding energies for substances in specific molecular docking models related to a specific health effect (adverse outcome) are provided as data.

- *Molecular docking models data formats*

Inputs used: *AOP networks*

3.6.2 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.

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.

Download empty dataset template: Zipped CSV Excel

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 3.188: Table definition for QSAR membership models.

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

Accepted table names: QSAR, QSARMembershipModels, QSARMembershipModel, QSARModels, QSARModel.

QSAR membership scores

Substance membership score according to the QSAR model.

Table 3.189: Table definition for QSAR membership scores.

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	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

Accepted table names: QSARMembershipScores, QSARMembershipScore, QSARMemberships, QSARMembership.

QSAR membership models as data

QSAR memberships models are provided as data, per QSAR model assessment group memberships for active substances related to a specific health effect are specified.

- *QSAR membership models data formats*

Inputs used: *AOP networks*

3.7 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.

3.7.1 Kinetic models

External exposure can be from one or 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 (EFSA (2014), EFSA (2017a)): 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 calculation 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 are specified as kinetic model instances that contain parameter specifications of built in kinetic models or as simple absorption factors.

Download empty dataset template: Zipped CSV Excel

Kinetic model instances

Kinetic model instances.

Table 3.190: Table definition for Kinetic model instances.

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 (50)	The species on which the experiment was performed.	System, TestSystem	Yes
Substances	AlphaNumeric (150)	Code or comma separated list of the codes of the substances. Unique identification code of substance, Default: valid for all substances. Should be omitted for parameters in the class Physiological	idSubstance, idSubstances, SubstanceId, SubstanceCode, SubstanceCodes, Substance, Substances	No
Reference	AlphaNumeric (100)	Reference or author.	References	No
Name	AlphaNumeric (100)	Name of the kinetic model instance.	Name	No
Description	AlphaNumeric (200)	Additional description of the kinetic model instance.	Description	No

Accepted table names: KineticModelInstances, KineticModelInstance.

Kinetic model instance parameters

Kinetic model parameters

Table 3.191: Table definition for Kinetic model instance parameters .

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 (50)	Name of the parameter in the kinetic model.		Yes
Description	AlphaNumeric (200)	Description of or reference for the parameter values in the kinetic model.		No
Value	Numeric	Mean.	MEAN, mean	Yes
Distribution-Type	*Probability-Distribution*	Distribution.	Distribution-Type, Distribution	No
CvVariability	Numeric	Variability.		No
CvUncertainty	Numeric	Uncertainty.		No

Accepted table names: KineticModelInstanceParameters, KineticModelInstanceParameter.

Kinetic model absorption factors

Kinetic absorption factors

Table 3.192: Table definition for Kinetic model absorption factors .

Name	Type	Description	Aliases	Required
idCompound	AlphaNumeric (50)	code of substance (must correspond to values in id column of Substances table)	idSubstance, SubstanceId, SubstanceCode, Substance	No
Route	*ExposureRouteType*	Non-dietary route or pathway, use 'Oral', 'Dermal', or 'Inhalation' to specify the route.	Route, Pathway	Yes
Absorption-Factor	Numeric	absorption factor value	Absorption-Factor, Factor	Yes

Accepted table names: KineticAbsorptionFactors, KineticAbsorptionFactor, AbsorptionFactors, AbsorptionFactor.

Kinetic models settings

Calculation settings

Table 3.193: Calculation settings for module Kinetic models.

Name	Type	Description
Default oral absorption factor for non-dietary exposure	Numeric	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) do
Default oral absorption factor for dietary exposure	Numeric	When there is no kinetic model and absorption factors are not specified in file, dietary exposures (external doses) are multiplie by this factor to determine the absorbed (internal) dose .
Default dermal absorption factor for non-dietary exposure	Numeric	When there is no kinetic model and absorption factors are not specified in file, dermal exposures (external doses) are multiplie by this factor to determine the absorbed (internal) dose.
Default inhalation absorption factor for non-dietary exposure	Numeric	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) do
Number of days	Numeric	The number of days.
Number of events per day for the ORAL dietary dose	Numeric	The daily dose is administered in equal portions (dose / number events) per event.
Number of initial days skipped	Numeric	This period is skipped in the calculation of the mean internal exposure.
Kinetic model	AlphaNumeric	Code Kinetic Model.
Use parameter variability	Boolean	When specified, use parameter variability.
Specify the type of kinetic model	*InternalModelType*	Specify the type of model to convert external exposure to the internal level.
Biological matrix	AlphaNumeric	Biological matrix.
Number of events per day for the DERMAL nondietary dose	Numeric	The daily dose is administered in equal portions (dose / number events) per event.
Number of events per day for the INHALATION nondietary dose	Numeric	The daily dose is administered in equal portions (dose / number events) per event.
Number of events per day for the ORAL nondietary dose	Numeric	The daily dose is administered in equal portions (dose / number events) per event.
Specify the hours (events).	AlphaNumeric	Specify the hours (events) on which a dose is applied. Allowed numbers are 1, 2, 3, 4,…, 24 separated by spaces.
Specify events	Boolean	if checked, a sequence of hours can be specified, otherswise the events are derived based on the number of doses per day.

Uncertainty settings

Table 3.194: Uncertainty settings for module Kinetic models.

Name	Type	Description
Resample kinetic model parameter values	Boolean	Specifies whether kinetic model parameter values are resample

Kinetic models as data

Specify nondietary absorption factors as data.

- *Kinetic models data formats*

Inputs used: *Active substances*

Available kinetic models

Physiologically based kinetic (PBK) 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. PBK models can simulate both internal doses from exposure scenarios (forward dosimetry) and external dose from biomonitoring data (reverse dosimetry).

The following generic PBK models are currently implemented in MCRA:

- *EuroMix generic PBK model* (Tebby et al. (2020)).
- *bisphenol PBK model ETHZ* (Karrer et al. (2019)).

The MCRA interface allows to run PBK models for any number of days and supplies an option to skip an initial number of days (build-up phase PBK model) in the calculation of the internal exposure.

In the original bisphenol PBK model of Karrer et al. (2019), doses were applied on fixed time points and only for the first four days. For oral exposure, three dosings per day with t = 0, 6 and 12 h, for dermal exposure to PCPs and TP, two dosings per day with t = 0 and 12 h. The steady state was reached after dosing on four consecutive days. Therefore, the bisphenol PBK model described in *Karrer et al.* should be run for four days to reached steady state and the number of initial days to be skipped should be set to 0. Note, no dosing is applied from day 5th onwards.

These restrictions were relaxed in a new implementation of the bisphenol PBK model (General Model BPA Re-implementation). In this new version the number of days is unlimited. Dosings are specified through the user interface including the definition on the non-stationary period, see *settings kinetic models*.

EuroMix Bisphenols PBPK model (v1)

EuroMix Bisphenols PBPK model by Karrer et al. (23 July 2018).

Table 3.195: Exposure routes (forcings)

Id	Description	Unit
Dietary	Dietary exposure	nmoles
Oral	Oral exposure	nmoles
Dermal	Dermal exposure	nmoles
Inhalation	Inhalation exposure	nmoles

Table 3.196: Model outputs

Id	Description	ScalingFactor	Multiplication-Factor	Unit
CPlasmaOut	Concentration in plasma			nmol/L
CGonadOut	Concentration in gonads			nmol/L
AurinebpaOut	Cumulative excretion of BPA in urine			nmoles
AurinegOut	Cumulative excretion of BPA-g in urine			nmoles
AurineTotalOut	Cumulative excretion of BPA and metabolites in urine			nmoles

Table 3.197: Model parameters

Id	Description	Default	Unit	Type
BW	Bodyweight		kg	Physiological
QCC	Cardiac output		L/min	Physiological
QgonadC	Fractional blood flow to gonads			Physiological
QliverC	Fractional blood flow to liver			Physiological
QfatC	Fractional blood flow to fat tissue			Physiological
QbrainC	Fractional blood flow to brain			Physiological
QskinC	Fractional blood flow to skin			Physiological
QmuscleC	Fractional blood flow to gonads			Physiological
VplasmaC	Fractional volume of plasma			Physiological
VfatC	Fractional volume of fat tissue			Physiological
VliverC	Fractional volume of liver tissue			Physiological
VbrainC	Fractional volume of brain tissue			Physiological
VskinC	Fractional volume of skin tissue			Physiological
VgonadC	Fractional volume of gonads			Physiological
VmuscleC	Fractional volume of muscle tissue			Physiological
VrichC	Fractional volume of richly perfused tissue			Physiological
VbodygC	Distribution volume of BPA-g			Physiological

continues on next page

Table 3.197 – continued from previous page

Id	Description	Default	Unit	Type
VbodysC	Distribution volume of BPA-s			Other
MW	Molecular weight		g/mol	ChemicalProperty
pliver	Partition coefficient liver to blood			PartitionCoefficient
pfat	Partition coefficient fat to blood			PartitionCoefficient
pslow	Partition coefficient slowly perfused tissue to blood			PartitionCoefficient
prich	Partition coefficient richly perfused tissue to blood			PartitionCoefficient
pgonad	Partition coefficient gonads to blood			PartitionCoefficient
pbrain	Partition coefficient brain to blood			PartitionCoefficient
pskin	Partition coefficient skin to blood			PartitionCoefficient
geC	Gastric emptying		1/h/kg bw^-0.25	Metabolic
k0C	Oral uptake from the stomach into the liver		1/h/kg bw^-0.25	Metabolic
k1C	Oral uptake from the small intestine into the liver		1/h/kg bw^-0.25	Metabolic
k4C	Fecal elimination from small intestine after oral administration		1/h/kg bw^-0.25	Metabolic
kGIingC	Transport of glucuronide from enterocytes into serum		1/h/kg bw^-0.25	Metabolic
kGIinsC	Transport of sulfate from enterocytes into serum		1/h/kg bw^-0.25	Metabolic
kmgutg	Km of Glucuronidation in the gut		nM	Metabolic
vmaxgutgC	Vmax of Glucuronidation in the gut		nmol/h/kg bw	Metabolic
fgutg	Correction factor of glucuronidation in the gut			Metabolic
kmguts	Km of Sulfation in the gut		nM	Metabolic
vmaxgutsC	Vmax of Sulfation in the gut		nmol/h/kg bw	Metabolic
fguts	Correction factor of sulfation in the gut			Metabolic

continues on next page

Table 3.197 – continued from previous page

Id	Description	Default	Unit	Type
met1g	Fraction of glucuronide in the liver taken up directly into serum (the rest undergoes EHR)			Metabolic
met1s	Fraction of sulfate in the liver taken up directly into serum			Metabolic
enterocytes	Sum of enterocytes weights in duodenum, jejunum and ileum		L	Metabolic
kmliver	Km of Glucuronidation in the liver		nM	Metabolic
vmaxliverC	Vmax of Glucuronidation in the liver		nmol/h/g liver	Metabolic
fliverg	Correction factor of glucuronidation in the liver			Metabolic
kmlivers	Km of Sulfation in the liver		nM	Metabolic
vmaxliversC	Vmax of Sulfation in the liver		nmol/h/g liver	Metabolic
flivers	Correction factor of sulfation in the liver			Metabolic
EHRtime	Time until EHR occurs		h	Metabolic
EHRrateC	EHR of glucuronide		1/h/kg bw^-0.25	Metabolic
k4C_IV	Fecal elimination of glucuronide from the EHR compartment		1/h/kg bw^-0.25	Metabolic
kurinebpaC	Clearance, urine excretion of parent compound		L/h/kg bw^0.75	Metabolic
kurinebpagC	Clearance, urine excretion of glucuronide		L/h/kg bw^0.75	Metabolic
kurinebpasC	Clearance, urine excretion of sulfate		L/h/kg bw^0.75	Metabolic
vreabsorptiong-C	Vmax for renal reabsorption of glucuronide		nmol/h/kg bw^0.75	Metabolic
vreabsorptionsC	Vmax for renal reabsorption of sulfate		nmol/h/kg bw^0.75	Metabolic
kreabsorptiong	Km for renal reabsorption of glucuronide		nM	Metabolic

continues on next page

Table 3.197 – continued from previous page

Id	Description	Default	Unit	Type
kreabsorptions	Km for renal reabsorption of sulfate		nM	Metabolic
kenterobpagC	EHR of parent compound due to biliary excretion of glucuronide		1/h/kg bw^-0.25	Metabolic
kenterobpasC	EHR of parent compound due to biliary excretion of sulfate		1/h/kg bw^-0.25	Metabolic
D_o	oral dose		ng/kg bw/dosing	Other
dose_O	oral dose		nmol/kg bw/dosing	Other
EoA_O	Extent of oral absorption			Physiological
uptake_O	amount of oral uptake		nmol/dosing	Other
period_O	uptake period		h	Other
koa	uptake rate		nmol/h	Other
t0_O	time point at which dosing starts		h	Other
t1_O	time point at which dosing ends		h	Other
D_d	dermal dose from thermal paper (TP)		ng/kg bw/dosing	Other
EoA_D	Extent of dermal absorption from TP			Physiological
dose_D	dermal dose from TP		nmol/kg bw/dosing	Other
aHL_D	Half-life for dermal penetration		h	Other
uptake_D	amount of dermal uptake from TP		nmol/dosing	Other
period_D	Uptake period dermal exposure from TP		h	Other
kda	Uptake rate of dermal exposure from TP		nmol/h	Other
t0_D	Time points at which dermal dosing from TP starts		h	Other
t1_D	Time points at which dermal dosing from TP ends		h	Other
D_d2	Dermal dose from PCPs		ng/kg bw/dosing	Other
EoA_D2	Extent of dermal absorption from PCPs			Physiological
dose_D2	Dermal dose from PCPs		nmol/kg bw/dosing	Other

continues on next page

Table 3.197 – continued from previous page

Id	Description	Default	Unit	Type
aHL_D2	Half-life for dermal penetration from PCPs		h	Other
uptake_D2	amount of dermal uptake from PCPs		nmol/dosing	Other
period_D2	Uptake period dermal exposure from PCPs		h	Other
kda2	uptake rate of dermal exposure from PCPs		nmol/h	Other
t0_D2	Time points at which dermal dosing from PCPs starts		h	Other
t1_D2	time points at which dermal dosing from TP ends		h	Other
QC	Cardiac output		L/h	Other
Qfat	Blood flow to the fat tissue		L/h	Other
Qliver	Blood flow to the liver tissue		L/h	Other
Qgonad	Blood flow to the gonads		L/h	Other
Qbrain	Blood flow to the brain		L/h	Other
Qskin	Blood flow to the skin tissue		L/h	Other
Qslow	Blood flow to the slowly perfused tissue		L/h	Other
Qrich	Blood flow to the richly perfused tissue		L/h	Other
Vliver	Volume of the liver		L	Other
Vfat	Volume of the fat tissue		L	Other
Vgonad	Volume of the gonads		L	Other
Vplasma	Volume of the plasma		L	Other
Vbrain	Volume of the brain		L	Other
Vskin	Volume of the skin		L	Other
Vslow	Volume of the slowly perfused tissue		L	Other
Vrich	richly perfused tissue		L	Other
Vbodyg	Volume of the distribution for BPAG		L	Other
Vbodys	Volume of the distribution for BPAS		L	Other
BW075	BW^0.75		kg^0.75	Other

continues on next page

Table 3.197 – continued from previous page

Id	Description	Default	Unit	Type
BW025	BW^0.25		kg^0.25	Other
vmaxlivers-Cnew	scaled Vmax of Sulfation of BPA in the liver		nmol/h/kg bw^0.75	Other
vmaxliverCnew	scaled Vmax of Glucuronidation of BPA in the liver		nmol/h/kg bw^0.75	Other
vmaxgutgCnew	scaled Vmax of Glucuronidation of BPA in the gut		nmol/h/kg bw^0.75	Other
vreabsorptiong	scaled vmax of renal resorption of BPAG		nmol/h	Other
vreabsorptions	scaled vmax of renal resorption of BPAS		nmol/h	Other
EHRrate	scaled EHR of BPAG		1/h	Other
k0	scaled Uptake of BPA from the stomach into the liver		1/h	Other
ge	scaled Gastric emptying of BPA		1/h	Other
k1	scaled Uptake of BPA from small intestine into the liver		1/h	Other
k4	scaled Fecal excretion of BPA after oral administration from small intestine		1/h	Other
k4_IV	scaled Fecal excretion of BPAG from the EHR compartment		1/h	Other
vmaxliver	rescaled and corrected vmax of BPA glucuronidation in the liver		nmol/h	Other
kGIing	scaled Uptake of BPAG from small intestine into serum		1/h	Other
met2g	Fraction of BPAG formed subject to EHR			Other
met2s	Fraction of BPAS formed subject to EHR			Other
kurinebpa	scaled Clearance of BPA via urine		L/h	Other
kurinebpag	scaled Clearance of BPAg via urine		L/h	Other
kurinebpas	scaled Clearance of BPAs via urine		L/h	Other

continues on next page

Table 3.197 – continued from previous page

Id	Description	Default	Unit	Type
vmaxlivers	rescaled and corrected vmax of BPA sulfation in the liver		nmol/h	Other
kGIins	scaled Uptake of BPAS from small intestine into serum		1/h	Other
vmaxgutg	rescaled and corrected vmax of BPA glucuronidation in the gut		nmol/h	Other
vmaxguts	rescaled and corrected vmax of BPA sulfation in the gut		nmol/h	Other
kenterobpag	scaled EHR of BPA due to biliary excretion of BPAG		1/h	Other
kenterobpas	scaled EHR of BPA due to biliary excretion of BPAS		1/h	Other
t0_D1_day1	time points at which dermal dosing from Thermal paper starts		h	Other
t0_D2_day1	time points at which dermal dosing from Thermal paper starts		h	Other
t0_D1_day2	time points at which dermal dosing from Thermal paper starts		h	Other
t0_D2_day2	time points at which dermal dosing from Thermal paper starts		h	Other
t0_D1_day3	time points at which dermal dosing from Thermal paper starts		h	Other
t0_D2_day3	time points at which dermal dosing from Thermal paper starts		h	Other
t0_D1_day4	time points at which dermal dosing from Thermal paper starts		h	Other
t0_D2_day4	time points at which dermal dosing from Thermal paper starts		h	Other
t0_D21_day1	time points at which dermal dosing from PCPs starts		h	Other
t0_D22_day1	time points at which dermal dosing from PCPs starts		h	Other
t0_D21_day2	time points at which dermal dosing from PCPs starts		h	Other

continues on next page

Table 3.197 – continued from previous page

Id	Description	Default	Unit	Type
t0_D22_day2	time points at which dermal dosing from PCPs starts		h	Other
t0_D21_day3	time points at which dermal dosing from PCPs starts		h	Other
t0_D22_day3	time points at which dermal dosing from PCPs starts		h	Other
t0_D21_day4	time points at which dermal dosing from PCPs starts		h	Other
t0_D22_day4	time points at which dermal dosing from PCPs starts		h	Other
t0_O1_day1	time points at which oral dosing starts		h	Other
t0_O2_day1	time points at which oral dosing starts		h	Other
t0_O3_day1	time points at which oral dosing starts		h	Other
t0_O1_day2	time points at which oral dosing starts		h	Other
t0_O2_day2	time points at which oral dosing starts		h	Other
t0_O3_day2	time points at which oral dosing starts		h	Other
t0_O1_day3	time points at which oral dosing starts		h	Other
t0_O2_day3	time points at which oral dosing starts		h	Other
t0_O3_day3	time points at which oral dosing starts		h	Other
t0_O1_day4	time points at which oral dosing starts		h	Other
t0_O2_day4	time points at which oral dosing starts		h	Other
t0_O3_day4	time points at which oral dosing starts		h	Other
ksiLiver	Ksi of glucuronidation in liver		nM	Other
ksiGut	Ksi of glucuronidation in gut		nM	Other
age	age			Other
gender	gender			Other
QCC_adult_f	QCC_adult_f			Other
Qgonad-C_adult_f	QgonadC_adult_f			Other
QliverC_adult_f	QliverC_adult_f			Other
QfatC_adult_f	QfatC_adult_f			Other
Qbrain-C_adult_f	QbrainC_adult_f			Other

continues on next page

Table 3.197 – continued from previous page

Id	Description	Default	Unit	Type
QskinC_adult_f	QskinC_adult_f			Other
Qmuscle-C_adult_f	QmuscleC_adult_f			Other
Vplasma-C_adult_f	VplasmaC_adult_f			Other
VfatC_adult_f	VfatC_adult_f			Other
VliverC_adult_f	VliverC_adult_f			Other
Vbrain-C_adult_f	VbrainC_adult_f			Other
VskinC_adult_f	VskinC_adult_f			Other
Vgonad-C_adult_f	VgonadC_adult_f			Other
Vmuscle-C_adult_f	VmuscleC_adult_f			Other
VrichC_adult_f	VrichC_adult_f			Other
Vbodyg-C_adult_f	VbodygC_adult_f			Other
Vbodys-C_adult_f	VbodysC_adult_f			Other
QCC_adult_m	QCC_adult_m			Other
Qgonad-C_adult_m	QgonadC_adult_m			Other
Qliver-C_adult_m	QliverC_adult_m			Other
QfatC_adult_m	QfatC_adult_m			Other
Qbrain-C_adult_m	QbrainC_adult_m			Other
Qskin-C_adult_m	QskinC_adult_m			Other
Qmuscle-C_adult_m	QmuscleC_adult_m			Other
Vplasma-C_adult_m	VplasmaC_adult_m			Other
VfatC_adult_m	VfatC_adult_m			Other
Vliver-C_adult_m	VliverC_adult_m			Other
Vbrain-C_adult_m	VbrainC_adult_m			Other
Vskin-C_adult_m	VskinC_adult_m			Other
Vgonad-C_adult_m	VgonadC_adult_m			Other
Vmuscle-C_adult_m	VmuscleC_adult_m			Other
Vrich-C_adult_m	VrichC_adult_m			Other
Vbodyg-C_adult_m	VbodygC_adult_m			Other
Vbodys-C_adult_m	VbodysC_adult_m			Other
QC-C_adolescent_f	QCC_adolescent_f			Other
Qgonad-C_adolescent_f	Qgonad-C_adolescent_f			Other

continues on next page

Table 3.197 – continued from previous page

Id	Description	Default	Unit	Type
Qliver-C_adolescent_f	Qliver-C_adolescent_f			Other
Qfat-C_adolescent_f	QfatC_adolescent_f			Other
Qbrain-C_adolescent_f	Qbrain-C_adolescent_f			Other
Qskin-C_adolescent_f	Qskin-C_adolescent_f			Other
Qmuscle-C_adolescent_f	Qmuscle-C_adolescent_f			Other
Vplasma-C_adolescent_f	Vplasma-C_adolescent_f			Other
Vfat-C_adolescent_f	VfatC_adolescent_f			Other
Vliver-C_adolescent_f	Vliver-C_adolescent_f			Other
Vbrain-C_adolescent_f	Vbrain-C_adolescent_f			Other
Vskin-C_adolescent_f	Vskin-C_adolescent_f			Other
Vgonad-C_adolescent_f	Vgonad-C_adolescent_f			Other
Vmuscle-C_adolescent_f	Vmuscle-C_adolescent_f			Other
Vrich-C_adolescent_f	Vrich-C_adolescent_f			Other
Vbodyg-C_adolescent_f	Vbodyg-C_adolescent_f			Other
Vbodys-C_adolescent_f	Vbodys-C_adolescent_f			Other
QC-C_adolescent_m	QCC_adolescent_m			Other
Qgonad-C_adolescent_m	Qgonad-C_adolescent_m			Other
Qliver-C_adolescent_m	Qliver-C_adolescent_m			Other
Qfat-C_adolescent_m	Qfat-C_adolescent_m			Other
Qbrain-C_adolescent_m	Qbrain-C_adolescent_m			Other
Qskin-C_adolescent_m	Qskin-C_adolescent_m			Other
Qmuscle-C_adolescent_m	Qmuscle-C_adolescent_m			Other
Vplasma-C_adolescent_m	Vplasma-C_adolescent_m			Other
Vfat-C_adolescent_m	VfatC_adolescent_m			Other
Vliver-C_adolescent_m	Vliver-C_adolescent_m			Other
Vbrain-C_adolescent_m	Vbrain-C_adolescent_m			Other
Vskin-C_adolescent_m	Vskin-C_adolescent_m			Other

continues on next page

Table 3.197 – continued from previous page

Id	Description	Default	Unit	Type
Vgonad-C_adolescent_m	Vgonad-C_adolescent_m			Other
Vmuscle-C_adolescent_m	Vmuscle-C_adolescent_m			Other
Vrich-C_adolescent_m	Vrich-C_adolescent_m			Other
Vbodyg-C_adolescent_m	Vbodyg-C_adolescent_m			Other
Vbodys-C_adolescent_m	Vbodys-C_adolescent_m			Other
QCC_child_f	QCC_child_f			Other
Qgonad-C_child_f	QgonadC_child_f			Other
QliverC_child_f	QliverC_child_f			Other
QfatC_child_f	QfatC_child_f			Other
Qbrain-C_child_f	QbrainC_child_f			Other
QskinC_child_f	QskinC_child_f			Other
Qmuscle-C_child_f	QmuscleC_child_f			Other
Vplasma-C_child_f	VplasmaC_child_f			Other
VfatC_child_f	VfatC_child_f			Other
VliverC_child_f	VliverC_child_f			Other
Vbrain-C_child_f	VbrainC_child_f			Other
VskinC_child_f	VskinC_child_f			Other
Vgonad-C_child_f	VgonadC_child_f			Other
Vmuscle-C_child_f	VmuscleC_child_f			Other
VrichC_child_f	VrichC_child_f			Other
Vbodyg-C_child_f	VbodygC_child_f			Other
Vbodys-C_child_f	VbodysC_child_f			Other
QCC_child_m	QCC_child_m			Other
Qgonad-C_child_m	QgonadC_child_m			Other
Qliver-C_child_m	QliverC_child_m			Other
QfatC_child_m	QfatC_child_m			Other
Qbrain-C_child_m	QbrainC_child_m			Other
Qskin-C_child_m	QskinC_child_m			Other
Qmuscle-C_child_m	QmuscleC_child_m			Other
Vplasma-C_child_m	VplasmaC_child_m			Other
VfatC_child_m	VfatC_child_m			Other
Vliver-C_child_m	VliverC_child_m			Other

continues on next page

Table 3.197 – continued from previous page

Id	Description	Default	Unit	Type
Vbrain-C_child_m	VbrainC_child_m			Other
Vskin-C_child_m	VskinC_child_m			Other
Vgonad-C_child_m	VgonadC_child_m			Other
Vmuscle-C_child_m	VmuscleC_child_m			Other
Vrich-C_child_m	VrichC_child_m			Other
Vbodyg-C_child_m	VbodygC_child_m			Other
Vbodys-C_child_m	VbodysC_child_m			Other

Model aliases: EuroMix_Bisphenols_PBPK_model_V1, PBPKModel_BPA.

EuroMix Bisphenols PBPK model (v2)

EuroMix Bisphenols PBPK model by Karrer et al. (2019).

Table 3.198: Exposure routes (forcings)

Id	Description	Unit
Dietary	Dietary exposure	nmoles
Oral	Oral exposure	nmoles
Dermal	Dermal exposure	nmoles
Inhalation	Inhalation exposure	nmoles

Table 3.199: Model outputs

Id	Description	ScalingFactor	Multiplication-Factor	Unit
CPlasmaOut	Concentration in plasma			nmol/L
CGonadOut	Concentration in gonads			nmol/L
AurinebpaOut	Cumulative excretion of BPA in urine			nmoles
AurinegOut	Cumulative excretion of BPA-g in urine			nmoles
AurineTotalOut	Cumulative excretion of BPA and metabolites in urine			nmoles

Table 3.200: Model parameters

Id	Description	Default	Unit	Type
BW	Bodyweight		kg	Physiological
QCC	Cardiac output		L/min	Physiological

continues on next page

Table 3.200 – continued from previous page

Id	Description	Default	Unit	Type
QgonadC	Fractional blood flow to gonads			Physiological
QliverC	Fractional blood flow to liver			Physiological
QfatC	Fractional blood flow to fat tissue			Physiological
QbrainC	Fractional blood flow to brain			Physiological
QskinC	Fractional blood flow to skin			Physiological
QmuscleC	Fractional blood flow to gonads			Physiological
VplasmaC	Fractional volume of plasma			Physiological
VfatC	Fractional volume of fat tissue			Physiological
VliverC	Fractional volume of liver tissue			Physiological
VbrainC	Fractional volume of brain tissue			Physiological
VskinC	Fractional volume of skin tissue			Physiological
VgonadC	Fractional volume of gonads			Physiological
VmuscleC	Fractional volume of muscle tissue			Physiological
VrichC	Fractional volume of richly perfused tissue			Physiological
VbodygC	Distribution volume of BPA-g			Physiological
VbodysC	Distribution volume of BPA-s			Other
pliver	Partition coefficient liver to blood			PartitionCoefficient
pfat	Partition coefficient fat to blood			PartitionCoefficient
pslow	Partition coefficient slowly perfused tissue to blood			PartitionCoefficient
prich	Partition coefficient richly perfused tissue to blood			PartitionCoefficient
pgonad	Partition coefficient gonads to blood			PartitionCoefficient
pbrain	Partition coefficient brain to blood			PartitionCoefficient
pskin	Partition coefficient skin to blood			PartitionCoefficient
geC	Gastric emptying		1/h/kg bw^-0.25	Metabolic
k0C	Oral uptake from the stomach into the liver		1/h/kg bw^-0.25	Metabolic

continues on next page

Table 3.200 – continued from previous page

Id	Description	Default	Unit	Type
k1C	Oral uptake from the small intestine into the liver		1/h/kg bw^-0.25	Metabolic
k4C	Fecal elimination from small intestine after oral administration		1/h/kg bw^-0.25	Metabolic
kGIingC	Transport of glucuronide from enterocytes into serum		1/h/kg bw^-0.25	Metabolic
kGIinsC	Transport of sulfate from enterocytes into serum		1/h/kg bw^-0.25	Metabolic
kmgutg	Km of Glucuronidation in the gut		nM	Metabolic
vmaxgutgC	Vmax of Glucuronidation in the gut		nmol/h/kg bw	Metabolic
fgutg	Correction factor of glucuronidation in the gut			Metabolic
kmguts	Km of Sulfation in the gut		nM	Metabolic
vmaxgutsC	Vmax of Sulfation in the gut		nmol/h/kg bw	Metabolic
fguts	Correction factor of sulfation in the gut			Metabolic
met1g	Fraction of glucuronide in the liver taken up directly into serum (the rest undergoes EHR)			Metabolic
met1s	Fraction of sulfate in the liver taken up directly into serum			Metabolic
enterocytes	Sum of enterocytes weights in duodenum, jejunum and ileum		L	Metabolic
kmliver	Km of Glucuronidation in the liver		nM	Metabolic
vmaxliverC	Vmax of Glucuronidation in the liver		nmol/h/g liver	Metabolic
fliverg	Correction factor of glucuronidation in the liver			Metabolic
kmlivers	Km of Sulfation in the liver		nM	Metabolic

continues on next page

Table 3.200 – continued from previous page

Id	Description	Default	Unit	Type
vmaxliversC	Vmax of Sulfation in the liver		nmol/h/g liver	Metabolic
flivers	Correction factor of sulfation in the liver			Metabolic
EHRtime	Time until EHR occurs		h	Metabolic
EHRrateC	EHR of glucuronide		1/h/kg bw^-0.25	Metabolic
k4C_IV	Fecal elimination of glucuronide from the EHR compartment		1/h/kg bw^-0.25	Metabolic
kurinebpaC	Clearance, urine excretion of parent compound		L/h/kg bw^0.75	Metabolic
kurinebpagC	Clearance, urine excretion of glucuronide		L/h/kg bw^0.75	Metabolic
kurinebpasC	Clearance, urine excretion of sulfate		L/h/kg bw^0.75	Metabolic
vreabsorptiong-C	Vmax for renal reabsorption of glucuronide		nmol/h/kg bw^0.75	Metabolic
vreabsorptionsC	Vmax for renal reabsorption of sulfate		nmol/h/kg bw^0.75	Metabolic
kreabsorptiong	Km for renal reabsorption of glucuronide		nM	Metabolic
kreabsorptions	Km for renal reabsorption of sulfate		nM	Metabolic
kenterobpagC	EHR of parent compound due to biliary excretion of glucuronide		1/h/kg bw^-0.25	Metabolic
kenterobpasC	EHR of parent compound due to biliary excretion of sulfate		1/h/kg bw^-0.25	Metabolic
koa	uptake rate		nmol/h	Other
EoA_D	Extent of dermal absorption from TP			Physiological
aHL_D	Half-life for dermal penetration		h	Other
kda	Uptake rate of dermal exposure from TP		nmol/h	Other
EoA_D2	Extent of dermal absorption from PCPs			Physiological
aHL_D2	Half-life for dermal penetration from PCPs		h	Other

continues on next page

Table 3.200 – continued from previous page

Id	Description	Default	Unit	Type
kda2	Uptake rate of dermal exposure from PCPs		nmol/h	Other
QC	Cardiac output		L/h	Other
Qfat	Blood flow to the fat tissue		L/h	Other
Qliver	Blood flow to the liver tissue		L/h	Other
Qgonad	Blood flow to the gonads		L/h	Other
Qbrain	Blood flow to the brain		L/h	Other
Qskin	Blood flow to the skin tissue		L/h	Other
Qslow	Blood flow to the slowly perfused tissue		L/h	Other
Qrich	Blood flow to the richly perfused tissue		L/h	Other
Vliver	Volume of the liver		L	Other
Vfat	Volume of the fat tissue		L	Other
Vgonad	Volume of the gonads		L	Other
Vplasma	Volume of the plasma		L	Other
Vbrain	Volume of the brain		L	Other
Vskin	Volume of the skin		L	Other
Vslow	Volume of the slowly perfused tissue		L	Other
Vrich	richly perfused tissue		L	Other
Vbodyg	Volume of the distribution for BPAG		L	Other
Vbodys	Volume of the distribution for BPAS		L	Other
vreabsorptiong	scaled vmax of renal resorption of BPAG		nmol/h	Other
vreabsorptions	scaled vmax of renal resorption of BPAS		nmol/h	Other
EHRrate	scaled EHR of BPAG		1/h	Other
k0	scaled Uptake of BPA from the stomach into the liver		1/h	Other
ge	scaled Gastric emptying of BPA		1/h	Other

continues on next page

Table 3.200 – continued from previous page

Id	Description	Default	Unit	Type
k1	scaled Uptake of BPA from small intestine into the liver		1/h	Other
k4	scaled Fecal excretion of BPA after oral administration from small intestine		1/h	Other
k4_IV	scaled Fecal excretion of BPAG from the EHR compartment		1/h	Other
vmaxliver	rescaled and corrected vmax of BPA glucuronidation in the liver		nmol/h	Other
kGIing	scaled Uptake of BPAG from small intestine into serum		1/h	Other
met2g	Fraction of BPAG formed subject to EHR			Other
met2s	Fraction of BPAS formed subject to EHR			Other
kurinebpa	scaled Clearance of BPA via urine		L/h	Other
kurinebpag	scaled Clearance of BPAg via urine		L/h	Other
kurinebpas	scaled Clearance of BPAs via urine		L/h	Other
vmaxlivers	rescaled and corrected vmax of BPA sulfation in the liver		nmol/h	Other
kGIins	scaled Uptake of BPAS from small intestine into serum		1/h	Other
vmaxgutg	rescaled and corrected vmax of BPA glucuronidation in the gut		nmol/h	Other
vmaxguts	rescaled and corrected vmax of BPA sulfation in the gut		nmol/h	Other
kenterobpag	scaled EHR of BPA due to biliary excretion of BPAG		1/h	Other
kenterobpas	scaled EHR of BPA due to biliary excretion of BPAS		1/h	Other

continues on next page

Table 3.200 – continued from previous page

Id	Description	Default	Unit	Type
ksiLiver	Ksi of glucuronidation in liver		nM	Other
ksiGut	Ksi of glucuronidation in gut		nM	Other
age	age			Other
gender	gender			Other
QCC_adult_f	QCC_adult_f			Other
Qgonad-C_adult_f	QgonadC_adult_f			Other
QliverC_adult_f	QliverC_adult_f			Other
QfatC_adult_f	QfatC_adult_f			Other
Qbrain-C_adult_f	QbrainC_adult_f			Other
QskinC_adult_f	QskinC_adult_f			Other
Qmuscle-C_adult_f	QmuscleC_adult_f			Other
Vplasma-C_adult_f	VplasmaC_adult_f			Other
VfatC_adult_f	VfatC_adult_f			Other
VliverC_adult_f	VliverC_adult_f			Other
Vbrain-C_adult_f	VbrainC_adult_f			Other
VskinC_adult_f	VskinC_adult_f			Other
Vgonad-C_adult_f	VgonadC_adult_f			Other
Vmuscle-C_adult_f	VmuscleC_adult_f			Other
VrichC_adult_f	VrichC_adult_f			Other
Vbodyg-C_adult_f	VbodygC_adult_f			Other
Vbodys-C_adult_f	VbodysC_adult_f			Other
QCC_adult_m	QCC_adult_m			Other
Qgonad-C_adult_m	QgonadC_adult_m			Other
Qliver-C_adult_m	QliverC_adult_m			Other
QfatC_adult_m	QfatC_adult_m			Other
Qbrain-C_adult_m	QbrainC_adult_m			Other
Qskin-C_adult_m	QskinC_adult_m			Other
Qmuscle-C_adult_m	QmuscleC_adult_m			Other
Vplasma-C_adult_m	VplasmaC_adult_m			Other
VfatC_adult_m	VfatC_adult_m			Other
Vliver-C_adult_m	VliverC_adult_m			Other
Vbrain-C_adult_m	VbrainC_adult_m			Other

continues on next page

Table 3.200 – continued from previous page

Id	Description	Default	Unit	Type
Vskin-C_adult_m	VskinC_adult_m			Other
Vgonad-C_adult_m	VgonadC_adult_m			Other
Vmuscle-C_adult_m	VmuscleC_adult_m			Other
Vrich-C_adult_m	VrichC_adult_m			Other
Vbodyg-C_adult_m	VbodygC_adult_m			Other
Vbodys-C_adult_m	VbodysC_adult_m			Other
QC-C_adolescent_f	QCC_adolescent_f			Other
Qgonad-C_adolescent_f	Qgonad-C_adolescent_f			Other
Qliver-C_adolescent_f	Qliver-C_adolescent_f			Other
Qfat-C_adolescent_f	QfatC_adolescent_f			Other
Qbrain-C_adolescent_f	Qbrain-C_adolescent_f			Other
Qskin-C_adolescent_f	Qskin-C_adolescent_f			Other
Qmuscle-C_adolescent_f	Qmuscle-C_adolescent_f			Other
Vplasma-C_adolescent_f	Vplasma-C_adolescent_f			Other
Vfat-C_adolescent_f	VfatC_adolescent_f			Other
Vliver-C_adolescent_f	Vliver-C_adolescent_f			Other
Vbrain-C_adolescent_f	Vbrain-C_adolescent_f			Other
Vskin-C_adolescent_f	Vskin-C_adolescent_f			Other
Vgonad-C_adolescent_f	Vgonad-C_adolescent_f			Other
Vmuscle-C_adolescent_f	Vmuscle-C_adolescent_f			Other
Vrich-C_adolescent_f	Vrich-C_adolescent_f			Other
Vbodyg-C_adolescent_f	Vbodyg-C_adolescent_f			Other
Vbodys-C_adolescent_f	Vbodys-C_adolescent_f			Other
QC-C_adolescent_m	QCC_adolescent_m			Other
Qgonad-C_adolescent_m	Qgonad-C_adolescent_m			Other
Qliver-C_adolescent_m	Qliver-C_adolescent_m			Other
Qfat-C_adolescent_m	Qfat-C_adolescent_m			Other

continues on next page

Table 3.200 – continued from previous page

Id	Description	Default	Unit	Type
Qbrain-C_adolescent_m	Qbrain-C_adolescent_m			Other
Qskin-C_adolescent_m	Qskin-C_adolescent_m			Other
Qmuscle-C_adolescent_m	Qmuscle-C_adolescent_m			Other
Vplasma-C_adolescent_m	Vplasma-C_adolescent_m			Other
Vfat-C_adolescent_m	VfatC_adolescent_m			Other
Vliver-C_adolescent_m	Vliver-C_adolescent_m			Other
Vbrain-C_adolescent_m	Vbrain-C_adolescent_m			Other
Vskin-C_adolescent_m	Vskin-C_adolescent_m			Other
Vgonad-C_adolescent_m	Vgonad-C_adolescent_m			Other
Vmuscle-C_adolescent_m	Vmuscle-C_adolescent_m			Other
Vrich-C_adolescent_m	Vrich-C_adolescent_m			Other
Vbodyg-C_adolescent_m	Vbodyg-C_adolescent_m			Other
Vbodys-C_adolescent_m	Vbodys-C_adolescent_m			Other
QCC_child_f	QCC_child_f			Other
Qgonad-C_child_f	QgonadC_child_f			Other
QliverC_child_f	QliverC_child_f			Other
QfatC_child_f	QfatC_child_f			Other
Qbrain-C_child_f	QbrainC_child_f			Other
QskinC_child_f	QskinC_child_f			Other
Qmuscle-C_child_f	QmuscleC_child_f			Other
Vplasma-C_child_f	VplasmaC_child_f			Other
VfatC_child_f	VfatC_child_f			Other
VliverC_child_f	VliverC_child_f			Other
Vbrain-C_child_f	VbrainC_child_f			Other
VskinC_child_f	VskinC_child_f			Other
Vgonad-C_child_f	VgonadC_child_f			Other
Vmuscle-C_child_f	VmuscleC_child_f			Other
VrichC_child_f	VrichC_child_f			Other
Vbodyg-C_child_f	VbodygC_child_f			Other
Vbodys-C_child_f	VbodysC_child_f			Other
QCC_child_m	QCC_child_m			Other

continues on next page

Table 3.200 – continued from previous page

Id	Description	Default	Unit	Type
Qgonad-C_child_m	QgonadC_child_m			Other
Qliver-C_child_m	QliverC_child_m			Other
QfatC_child_m	QfatC_child_m			Other
Qbrain-C_child_m	QbrainC_child_m			Other
Qskin-C_child_m	QskinC_child_m			Other
Qmuscle-C_child_m	QmuscleC_child_m			Other
Vplasma-C_child_m	VplasmaC_child_m			Other
VfatC_child_m	VfatC_child_m			Other
Vliver-C_child_m	VliverC_child_m			Other
Vbrain-C_child_m	VbrainC_child_m			Other
Vskin-C_child_m	VskinC_child_m			Other
Vgonad-C_child_m	VgonadC_child_m			Other
Vmuscle-C_child_m	VmuscleC_child_m			Other
Vrich-C_child_m	VrichC_child_m			Other
Vbodyg-C_child_m	VbodygC_child_m			Other
Vbodys-C_child_m	VbodysC_child_m			Other

Model aliases: EuroMix_Bisphenols_PBPK_model_V2, PBPKModel_BPA_Reimplementation.

EuroMix Generic PBTK model (v5)

Cosmos version 5 (adapted 9/11/2018)

Table 3.201: Exposure routes (forcings)

Id	Description	Unit
Dietary	Dietary exposure	mmoles
Dermal	Dermal exposure	mmoles
Inhalation	Inhalatory exposure	mmoles

Table 3.202: Model outputs

Id	Description	ScalingFactor	Multiplication-Factor	Unit
CVen	Venous blood	scVBlood	0.66667	mM
CArt	Arterial blood	scVBlood	0.33333	mM
CFat	Fat tissues	scVFat		mM
CPoor	Muscle tissues			mM
CRich	Viscera	scVRich		mM
CLiver	Liver	scVLiver		mM
CSkin_u	Viable skin, unexposed			mM
CSkin_e	Viable skin, exposed	BSA, Height_vs, fsA_exposed		mM
CSkin_sc_u	Skin stratum corneum, unexposed			mM
CSkin_sc_e	Skin stratum corneum, exposed	BSA, Height_vs, fsA_exposed		mM

Table 3.203: Model parameters

Id	Description	Default	Unit	Type
BM	Body mass		kg	Physiological
BSA	Body skin surface area		dm2	Physiological
scVFat	Fat as fraction of total body volume			Physiological
scVRich	Richly perfused tissues (viscera) as fraction of total body volume			Physiological
scVLiver	Liver as fraction of total body volume			Physiological
scVBlood	Blood as fraction of total body volume			Physiological
Height_sc	Skin thickness		decimeter	Physiological
Height_vs	Viable skin			Physiological
scFBlood	Total blood flow per unit mass		L/h/kg	Physiological
scFFat	Fat fraction of total blood flow going to compartments			Physiological
scFPoor	Poorly perfused tissues (muscles) fraction of total blood flow going to compartments			Physiological
scFLiver	Liver fraction of total blood flow going to compartments			Physiological
scFSkin	Skin fraction of total blood flow going to compartments			Physiological
Falv	Alveolar ventilation rate		L/h	Physiological

continues on next page

Table 3.203 – continued from previous page

Id	Description	Default	Unit	Type
mic	Microsomal proteins content		mg/gr liver	Physiological
PCAir	Partition coefficient: blood over air			PartitionCoefficient
log_PCFat	Scaled parameter, partition coefficient: fat over blood			PartitionCoefficient
log_aPoor	Scaled parameter, partition coefficient: muscle over blood (poorly perfused tissue)			PartitionCoefficient
log_aRich	Scaled parameter, partition coefficient: viscera over blood (richly perfused tissue)			PartitionCoefficient
log_aLiver	Scaled parameter, partition coefficient: liver over blood			PartitionCoefficient
log_aSkin	Scaled parameter, partition coefficient: viable skin / blood			PartitionCoefficient
log_aSkin_sc	Scaled parameter, partition coefficient: viable skin / stratum corneum			PartitionCoefficient
Kp_sc_vs	Diffusion rate from stratum corneum to viable skin		decimeter/h	Metabolic
Ke	Renal excretion rate		L/h	Metabolic
Michaelis	Flag for Michaelis-Menten vs linear metabolism (0 = linear)			Metabolic
Vmax	Maximum rate of metabolism		mmoles/h/L liver	Metabolic
Km	Michaelis-Menten constant		mM	Metabolic
CLH	Hepatic clearance			Metabolic
fup	Unbound fraction in blood			Metabolic
Frac	Fraction absorbed by the gut			Metabolic
kGut	Oral 1st order absorption rate constant		1/h	Metabolic
Cinh	Inhalation			Other
Tinh	Inhalation duration			Other
OralDose			mmol	Other
DermalDose			mmol	Other
fSA_exposed	Fraction of skin surface area actually exposed			Metabolic

continues on next page

Table 3.203 – continued from previous page

Id	Description	Default	Unit	Type
FBlood	Blood flow			Other
FFat	Scaled parameters			Other
FPoor	Scaled parameters			Other
FRich	Scaled parameters			Other
FLiver	Scaled parameters			Other
FSkin	Scaled parameters			Other
VFat	Scaled parameters			Other
VRich	Scaled parameters			Other
VLiver	Scaled parameters			Other
VSkin_e	Scaled parameters			Other
VSkin_u	Scaled parameters			Other
VSkin_sc_e	Scaled parameters			Other
VSkin_sc_u	Scaled parameters			Other
VBlood	Scaled parameters			Other
VPoor	Scaled parameters			Other
VArt	Scaled parameters			Other
VVen	Scaled parameters			Other
FSkin_e	Scaled parameters			Other
FSkin_u	Scaled parameters			Other
PCFat	Partition coefficient: fat over blood			PartitionCoefficient
PCPoor	Partition coefficient: muscle over blood (poorly perfused tissue)			PartitionCoefficient
PCRich	Partition coefficient: viscera over blood (richly perfused tissue)			PartitionCoefficient
PCLiver	Partition coefficient: liver over blood			PartitionCoefficient
PCSkin	Partition coefficient: viable skin / blood			PartitionCoefficient
PCSkin_sc	Partition coefficient: viable skin / stratum corneum			PartitionCoefficient
ResampledPC-Fat	Resampled value PCFat			PartitionCoefficient

Model aliases: EuroMix_Generic_PBTK_model_V5, CosmosV4, CosmosV5.

EuroMix Generic PBTK model (v6)

Cosmos version 6 (received 3/27/2019)

Table 3.204: Exposure routes (forcings)

Id	Description	Unit
Dietary	Dietary exposure	mmoles
Dermal	Dermal exposure	mmoles
Inhalation	Inhalatory exposure	mmoles

Table 3.205: Model outputs

Id	Description	ScalingFactor	Multiplication-Factor	Unit
CTotal	Total concentration			mM
CVen	Venous blood concentration	scVBlood	0.66667	mM
CArt	Arterial blood concentration	scVBlood	0.33333	mM
CFat	Fat (adipose) tissue concentration	scVFat		mM
CPoor	Poorly perfused tissue (muscle) concentration			mM
CRich	Richly perfused tissue (viscera) concentration	scVRich		mM
CLiver	Liver concentration	scVLiver		mM
CSkin_u	Viable unexposed skin concentration			mM
CSkin_e	Viable exposed skin concentration	BSA, Height_vs, fsA_exposed		mM
CSkin_sc_u	Skin unexposed stratum corneum concentration			mM
CSkin_sc_e	Skin exposed stratum corneum concentration	BSA, Height_vs, fsA_exposed		mM

Table 3.206: Model parameters

Id	Description	Default	Unit	Type
BM	Body mass		kg	Physiological
BSA	Body surface area (internally scaled by an allometric scaling factor s = 70/BM^0.3)		dm2	Physiological
scVFat	Fat as fraction of total body volume			Physiological
scVRich	Richly perfused tissues (viscera) as fraction of total body volume			Physiological
scVLiver	Liver as fraction of total body volume			Physiological
scVBlood	Blood as fraction of total body volume			Physiological
Height_sc	Skin thickness		decimeter	Physiological
Height_vs	Viable skin			Physiological
scFBlood	Total blood flow per unit mass		L/h/kg	Physiological
scFFat	Fat fraction of total blood flow going to compartments			Physiological

continues on next page

Table 3.206 – continued from previous page

Id	Description	Default	Unit	Type
scFPoor	Poorly perfused tissues (muscles) fraction of total blood flow going to compartments			Physiological
scFLiver	Liver fraction of total blood flow going to compartments			Physiological
scFSkin	Skin fraction of total blood flow going to compartments			Physiological
Falv	Alveolar ventilation rate		L/h	Physiological
mic	Microsomal proteins content		mg/gr liver	Physiological
PCAir	Partition coefficient: blood over air			PartitionCoefficient
log_PCFat	Scaled parameter, partition coefficient: fat over blood			PartitionCoefficient
log_aPoor	Scaled parameter, partition coefficient: muscle over blood (poorly perfused tissue)			PartitionCoefficient
log_aRich	Scaled parameter, partition coefficient: viscera over blood (richly perfused tissue)			PartitionCoefficient
log_aLiver	Scaled parameter, partition coefficient: liver over blood			PartitionCoefficient
log_aSkin	Scaled parameter, partition coefficient: viable skin over blood			PartitionCoefficient
log_aSkin_sc	Scaled parameter, partition coefficient: viable skin stratum corneum over blood			PartitionCoefficient
Kp_sc_vs	Diffusion rate from stratum corneum to viable skin		decimeter/h	Metabolic
Ke	Renal excretion rate		L/h	Metabolic
Michaelis	Flag for Michaelis-Menten vs linear metabolism (0 = linear)			Metabolic
Vmax	Maximum rate of metabolism		mmoles/h/L liver	Metabolic
Km	Michaelis-Menten constant for metabolism		mM	Metabolic

continues on next page

Table 3.206 – continued from previous page

Id	Description	Default	Unit	Type
CLH	Hepatic metabolic clearance			Metabolic
fub	Unbound fraction in blood			Metabolic
Frac	Fraction absorbed by the gut			Metabolic
kGut	Oral 1st order absorption rate constant		1/h	Metabolic
Cinh	Inhalation			Other
Tinh	Inhalation duration			Other
OralDose			mmol	Other
DermalDose			mmol	Other
fSA_exposed	Fraction of skin surface area actually exposed			Metabolic
FBlood	Blood flow			Other
FFat	Scaled parameters, blood flow to the fat			Other
FPoor	Scaled parameters, blood flow to poorly perfused tissues			Other
FRich	Scaled parameters, blood flow to richly perfused tissues			Other
FLiver	Scaled parameters, blood flow to the liver			Other
FSkin	Scaled parameters, blood flow to the skin			Other
VFat	Scaled parameters			Other
VRich	Scaled parameters, richly perfused tissue volume			Other
VLiver	Scaled parameters, liver volume			Other
VSkin_e	Scaled parameters, exposed skin volume			Other
VSkin_u	Scaled parameters, unexposed skin volume			Other
VSkin_sc_e	Scaled parameters, stratum corneum exposed skin volume			Other
VSkin_sc_u	Scaled parameters, stratum corneum unexposed skin volume			Other
VBlood	Scaled parameters, blood volume			Other
VPoor	Scaled parameters, poorly perfused tissue volume			Other

continues on next page

Table 3.206 – continued from previous page

Id	Description	Default	Unit	Type
VArt	Scaled parameters, arterial blood volume			Other
VVen	Scaled parameters, venous blood volume			Other
FSkin_e	Scaled parameters, blood flow to exposed skin			Other
FSkin_u	Scaled parameters, blood flow to unexposed skin			Other
PCFat	Partition coefficient: fat over blood			PartitionCoefficient
PCPoor	Partition coefficient: muscle over blood (poorly perfused tissue)			PartitionCoefficient
PCRich	Partition coefficient: viscera over blood (richly perfused tissue)			PartitionCoefficient
PCLiver	Partition coefficient: liver over blood			PartitionCoefficient
PCSkin	Partition coefficient: viable skin over blood			PartitionCoefficient
PCSkin_sc	Partition coefficient: viable skin / stratum corneum			PartitionCoefficient
ResampledPC-Fat	Resampled value PCFat			PartitionCoefficient

Model aliases: EuroMix_Generic_PBTK_model_V6, CosmosV6.

PBK model chlorpyrifos (v1)

PBK model chlorpyrifos (v1)

Table 3.207: Exposure routes (forcings)

Id	Description	Unit
Dietary	Dietary description	umoles

Table 3.208: Model outputs

Id	Description	ScalingFactor	Multiplication-Factor	Unit
O_CV	Venous blood	VVc		uM
O_CP	Plasma	VVc	0.6	uM
O_CU	Uterus tissue	VUc		uM
O_ACL	Cleared renally		0.03	umoles
O_CS	Slowly perfused tissue			umoles
O_CR	Richly perfused tissue			umoles
O_CF	Fat	VFc		umoles
O_CL	Liver	VLc		umoles
O_CK	Kidney	VKc		umoles
O_CM	Muscle	VMc		umoles
O_CH	Heart	VHc		umoles
O_CLu	Lung	VLuc		umoles
O_CBrb	Brain blood	VBrc	0.05	umoles
O_CBrt	Brain tissue	VBrc	0.95	umoles
O_CBr	Brain total	VBrc		umoles
O_CA	Arterial blood	VAc		umoles

Table 3.209: Model parameters

Id	Description	Default	Unit	Type
VLc	Fraction liver tissue of BW	0.0257		Physiological
VFc	Fraction fat tissue of BW	0.2142		Physiological
VLuc	Fraction lung tissue of BW	0.0076		Physiological
VAc	Fraction arterial blood of BW (0.074*1/4)	0.0198		Physiological
VVc	Fraction venous blood of BW (0.074*3/4)	0.0593		Physiological
VKc	Fraction kidney tissue of BW	0.004		Physiological
VMc	Fraction muscle tissue of BW	0.04		Physiological
VUc	Fraction uterus tissue of BW	0.0018		Physiological
VBrc	Fraction brain tissue of BW	0.02		Physiological
VHc	Fraction heart tissue of BW	0.0047		Physiological
QLc	Fraction of blood flow to liver	0.227		Physiological
QFc	Fraction of blood flow to fat	0.052		Physiological
QKc	Fraction of blood flow to kidneys	0.175		Physiological
QMc	Fraction of blood flow to muscle	0.12		Physiological

continues on next page

Table 3.209 – continued from previous page

Id	Description	Default	Unit	Type
QUc	Fraction of blood flow to uterus	0.2		Physiological
QBrc	Fraction of blood flow to brain	0.114		Physiological
QHc	Fraction of blood flow to heart	0.04		Physiological
MWP	Molecular weight	350.59	g/mol	PhysicoChemical
MWM1	Molecular weight	334.52	g/mol	PhysicoChemical
MWM2	Molecular weight	198.43	g/mol	PhysicoChemical
LogPP	Octanol/water partition coefficient	4.784		PartitionCoefficient
LogPM1	Octanol/water partition coefficient	3.894		PartitionCoefficient
LogPM2	Octanol/water partition coefficient	1.856		PartitionCoefficient
Fa	Fraction absorbed. Obtained from Nolan 1984	0.7		Metabolic
KaS	Absorption constant stomach. Obtained from Timchalk et al. 2002 (stomach; /h)	0.00000733	/h	Metabolic
KaI	Absorption constant intestine. Obtained from Timchalk et al. 2002 (intestine; /h)	1.00033	/h	Metabolic
KsI	Absorption constant (transfer stomach to intestine). Obtained from Timchalk et al. 2002 (transfer stomach to intestine; /h)	0.967749	/h	Metabolic
fuP	Unbound fraction in plasma. Obtained from SimCyp	0.021		Metabolic
fuM1	Unbound fraction in plasma. Obtained from SimCyp	0.15		Metabolic
fuM2	Unbound fraction in plasma. Obtained from SimCyp	0.082		Metabolic
BPP	B/P ratio obtained from Hsu, 2013. If no data available, set to 1	1.3		Metabolic
BPM1	B/P ratio obtained from Hsu, 2013. If no data available, set to 1	2.7		Metabolic
BPM2	B/P ratio obtained from Hsu, 2013. If no data available, set to 1	1		Metabolic

continues on next page

Table 3.209 – continued from previous page

Id	Description	Default	Unit	Type
KurineP	Urinary excretion rate constant (/h)	0	/h	Metabolic
KurineM1	Urinary excretion rate constant (/h)	0	/h	Metabolic
KurineM2	Urinary excretion rate constant (/h)	0.026	/h	Metabolic
CYPabundance-CYP1A2	CYP1A2 abundance (pmolCYP/mg protein) ;(calculated based on database in Simcyp; sum of EM, PM and UM)	52		Metabolic
CYPabundance-CYP2B6	CYP2B6 abundance (pmolCYP/mg protein) ;(calculated based on database in Simcyp; sum of EM, PM and UM)	15.8		Metabolic
CYPabundance-CYP2C19	CYP2C19 abundance (pmolCYP/mg protein) ;(calculated based on database in Simcyp; sum of EM, PM and UM)	5.4		Metabolic
CYPabundance-CYP3A4	CYP3A4 abundance (pmolCYP/mg protein) ;(calculated based on database in Simcyp; sum of EM, PM and UM)	137		Metabolic
ISEFCYP1A2	Scaling factor CYP1A2 ISEF (Vmax) (non compound-specific) (calculated based on probe incubation, lab specific)	0.072		Metabolic
ISEFCYP2B6	Scaling factor CYP2B6 ISEF (Vmax) (non compound-specific) (calculated based on probe incubation, lab specific)	0.476		Metabolic
ISEFCYP2C19	Scaling factor CYP2C19 ISEF (Vmax) (non compound-specific) (calculated based on probe incubation, lab specific)	0.209		Metabolic

continues on next page

Table 3.209 – continued from previous page

Id	Description	Default	Unit	Type
ISEFCYP3A4	Scaling factor CYP3A4 ISEF (Vmax) (non compound-specific) (calculated based on probe incubation, lab specific)	0.107		Metabolic
MPL	Scaling factor of human liver microsome in mg to gram liver (mg microsomal protein /g liver) (Al-Malahmeh, A. J et al., 2017) (Barter et al., 2007)	32	mg/g	Metabolic
VMaxCYP1-A2mP1	Vmax of CYP1A2 at supersome level (pmol/min/pmol CYP) (Experimentally determined using supersomes, Chen et al., 2022)	3.963		Metabolic
VMaxCYP2-B6mP1	Vmax of CYP2B6 at supersome level (pmol/min/pmol CYP) (Experimentally determined using supersomes, Chen et al., 2022)	7.755		Metabolic
VMaxCYP2-C19mP1	Vmax of CYP2C19 at supersome level (pmol/min/pmol CYP) (Experimentally determined using supersomes, Chen et al., 2022)	2.744		Metabolic
VMaxCYP3-A4mP1	Vmax of CYP3A4 at supersome level (pmol/min/pmol CYP) (Experimentally determined using supersomes, Chen et al., 2022)	17.78		Metabolic
KmCYP1A2P1	Affinity constant of CYP1A2 (umoles/L) (Experimentally determined using supersomes, Chen et al., 2022)	0.61	umoles/L	Metabolic

continues on next page

Table 3.209 – continued from previous page

Id	Description	Default	Unit	Type
KmCYP2B6P1	Affinity constant of CYP2B6 (umoles/L) (Experimentally determined using supersomes, Chen et al., 2022)	0.14	umoles/L	Metabolic
KmCYP2C19-P1	Affinity constant of CYP2C19 (umoles/L) (Experimentally determined using supersomes, Chen et al., 2022)	1.89	umoles/L	Metabolic
KmCYP3A4P1	Affinity constant of CYP3A4 (umoles/L) (Experimentally determined using supersomes, Chen et al., 2022)	29.77	umoles/L	Metabolic
VMaxCYP1-A2mP2	Vmax of CYP1A2 at supersome level (pmol/min/pmol CYP) (Experimentally determined using supersomes, Chen et al., 2022)	2.957		Metabolic
VMaxCYP2-B6mP2	Vmax of CYP2B6 at supersome level (pmol/min/pmol CYP) (Experimentally determined using supersomes, Chen et al., 2022)	5.492		Metabolic
VMaxCYP2-C19mP2	Vmax of CYP2C19 at supersome level (pmol/min/pmol CYP) (Experimentally determined using supersomes, Chen et al., 2022)	17.51		Metabolic
VMaxCYP3-A4mP2	Vmax of CYP3A4 at supersome level (pmol/min/pmol CYP) (Experimentally determined using supersomes, Chen et al., 2022)	23.86		Metabolic

continues on next page

Table 3.209 – continued from previous page

Id	Description	Default	Unit	Type
KmCYP1A2P2	Affinity constant of CYP1A2 (umoles/L) (Experimentally determined using supersomes, Chen et al., 2022)	1.25	umoles/L	Metabolic
KmCYP2B6P2	Affinity constant of CYP2B6 (umoles/L) (Experimentally determined using supersomes, Chen et al., 2022)	1.28	umoles/L	Metabolic
KmCYP2C19-P2	Affinity constant of CYP2C19 (umoles/L) (Experimentally determined using supersomes, Chen et al., 2022)	1.37	umoles/L	Metabolic
KmCYP3A4P2	Affinity constant of CYP3A4 (umoles/L) (Experimentally determined using supersomes, Chen et al., 2022)	18.13	umoles/L	Metabolic
VMax3c	Maximum velocity constant (nmol/min/ml plasma) (Experimentally determined using microsomes, Chen et al., 2022)	37.98	nmoles/min/ml	Metabolic
Km3	Affinity constant (umoles/L) (Experimentally determined)	627.9		Metabolic
VMax4c	Maximum velocity constant (nmol/min/ml plasma) (Experimentally determined using microsomes, Chen et al., 2022)	1844	nmoles/min/ml	Metabolic
Km4	Affinity constant (umoles/L) (Experimentally determined)	289.8	umoles/L	Metabolic
BW		70	kg	Physiological

Model aliases: PBK_Chlorpyrifos_V1.

Note: Additional kinetic models can be implemented, please contact the MCRA administrator.

EuroMix generic PBK model

Reference: Tebby et al. (2020)

In MCRA updated versions (version 4b, 6) of the PBK 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.

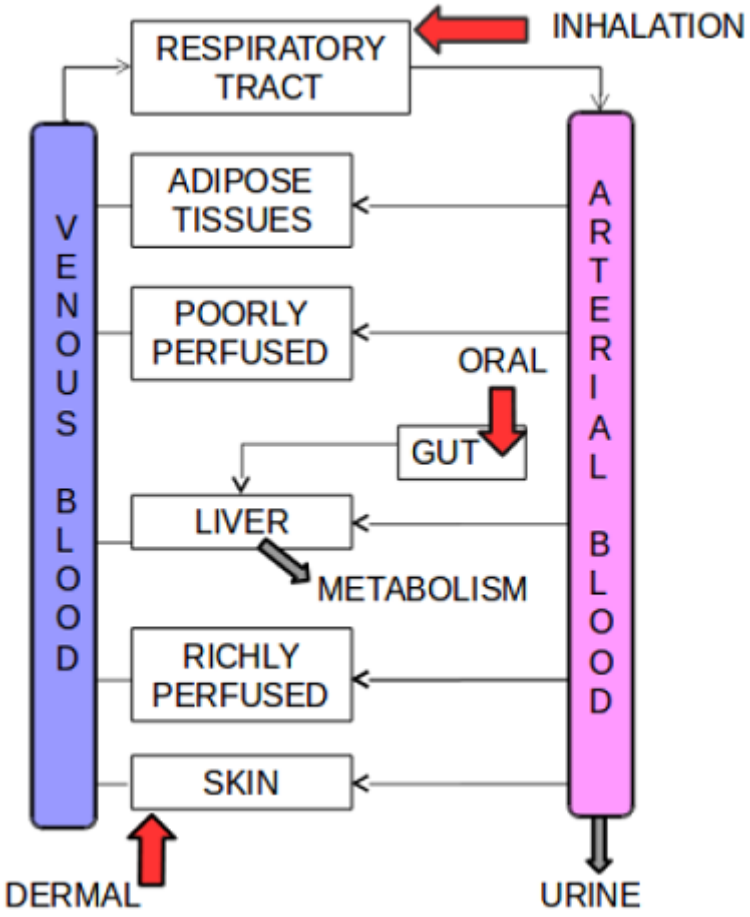


Figure 3.69: Schematic representation of the EuroMix Generic PBK model.

The EuroMix generic PBK 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 Tebby et al. (2020).

In Figure 3.70 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).

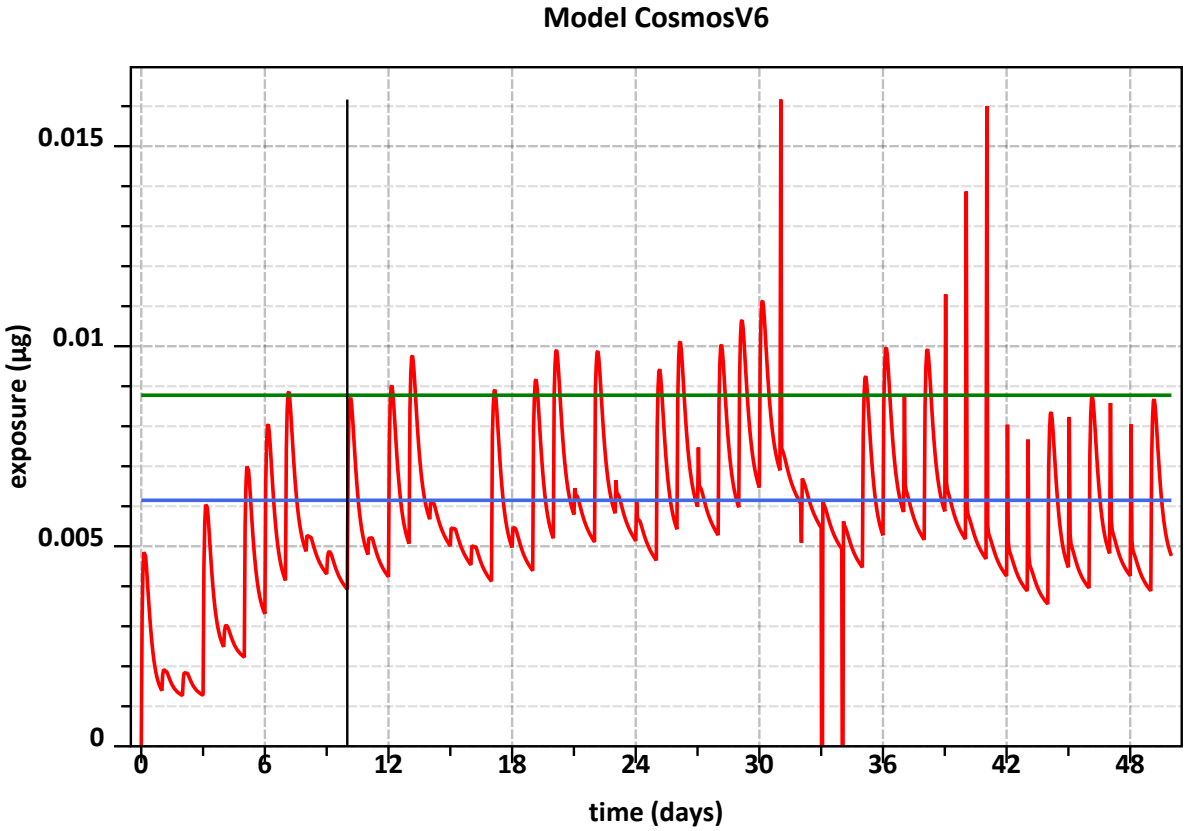


Figure 3.70: Time course of exposure (μg) for Clothianidin in the liver (EuroMix generic PBK model version 6).

In Figure 3.71, 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: Karrer et al. (2019)

'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 modelled 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 modelling 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

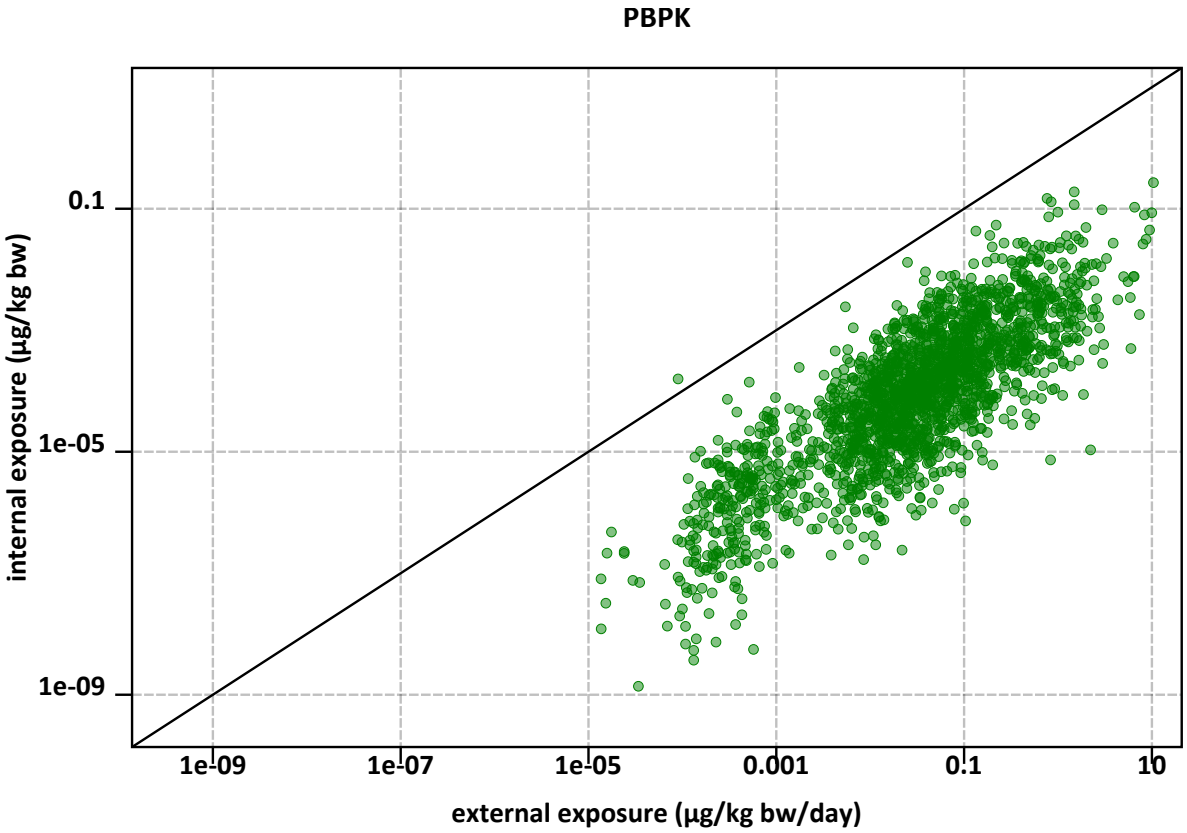


Figure 3.71: Internal versus external exposure for Clothianidin in the liver (EuroMix Generic PBK model version 6).

3.8 Risk modules

Exposures and *hazard characterisations* are compared in risk metrics. If both exposure and hazard characterisation are characterised by a single value, the risk metric (e.g. a traditional margin of exposure, hazard quotient or hazard index) can be calculated using module Single value risks. Module Risks allows for probabilistic risk calculations. In both cases a threshold can be specified to assist in interpretation. The threshold value should be chosen in relation to the assessment factors used in the hazard characterisation, e.g. a threshold MOE=100 (of HI=0.01) is often used if no assessment factors have been used, but a threshold 1 would be appropriate if assessment factors have already been used to address relevant uncertainties.

3.8.1 Risks

Risks (health impacts) are defined as a function of exposure and hazard characterisation at a chosen biological level (external or internal). Risk metrics are margins of exposure (MOE) or hazard indices (HI) or more generalised MOE or HI distributions.

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

Output of this module is used by: *Single value risks*

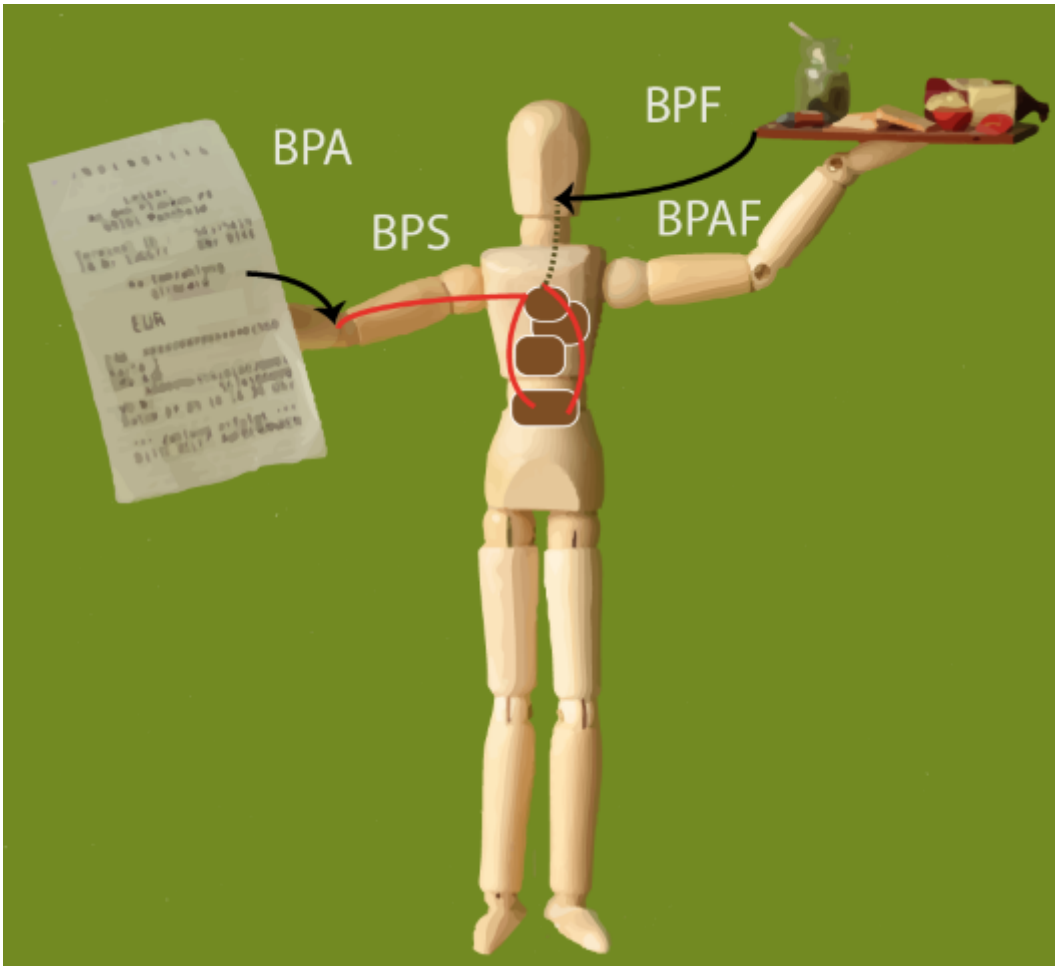


Figure 3.72: Graphical abstract 'Linking probabilistic exposure and pharmacokinetic modelling to assess the cumulative risk from the bisphenols BPA, BPS, BPF, and BPAF for Europeans.'

Risks calculation

A (cumulative) risk assessment aims to characterise the health impact due to exposure to one or multiple substances causing common adverse health effects. The health impact is characterized by a distribution of individual risks, expressed by a *risk metric* (i.e., a margins of exposure (*MOE*) or a hazard index (*HI*)) comparing exposures and hazard characterizations at the chosen level (external or internal). Hazard characterisations are included as single values or 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. The exposure level that results in exactly that critical effect in a particular person is that person's individual critical hazard dose (*ICED*). Individuals in a population typically show variation, both in their individual exposure and in their hazard characterization. Both the variation in exposure and the variation in hazard characterization are quantified in the form of probability distributions. Assuming independence between both distributions, they are combined by Monte Carlo methods. The proportion of the MOE distribution below the (safety/uncertainty) threshold (or the proportion of the HI distribution above 1) 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 hazard characterisation) are quantified using Monte Carlo and bootstrap methods. This results in an uncertainty distribution for any statistic of interest.

In Figure 3.73, margin of exposures for a number of substances are shown. As shown, the distinction between variability (grey bars, 90% probability) and uncertainty (whiskers) is retained. This is discussed in van der Voet and Slob (2007) and van der Voet et al. (2009).

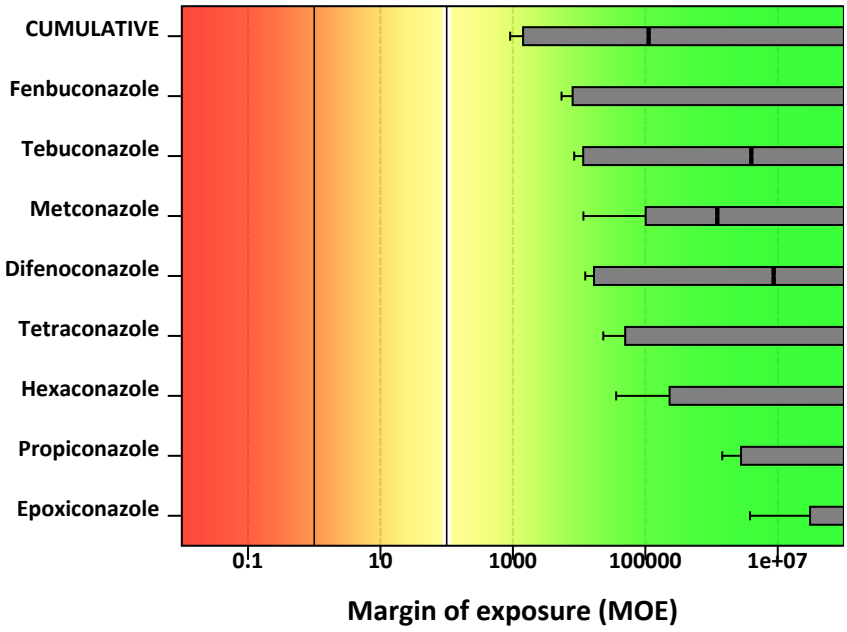


Figure 3.73: Individual margin of exposure (MOE) plot for multiple substances.

In Figure 3.74, hazards versus exposures are plotted for the same substances.

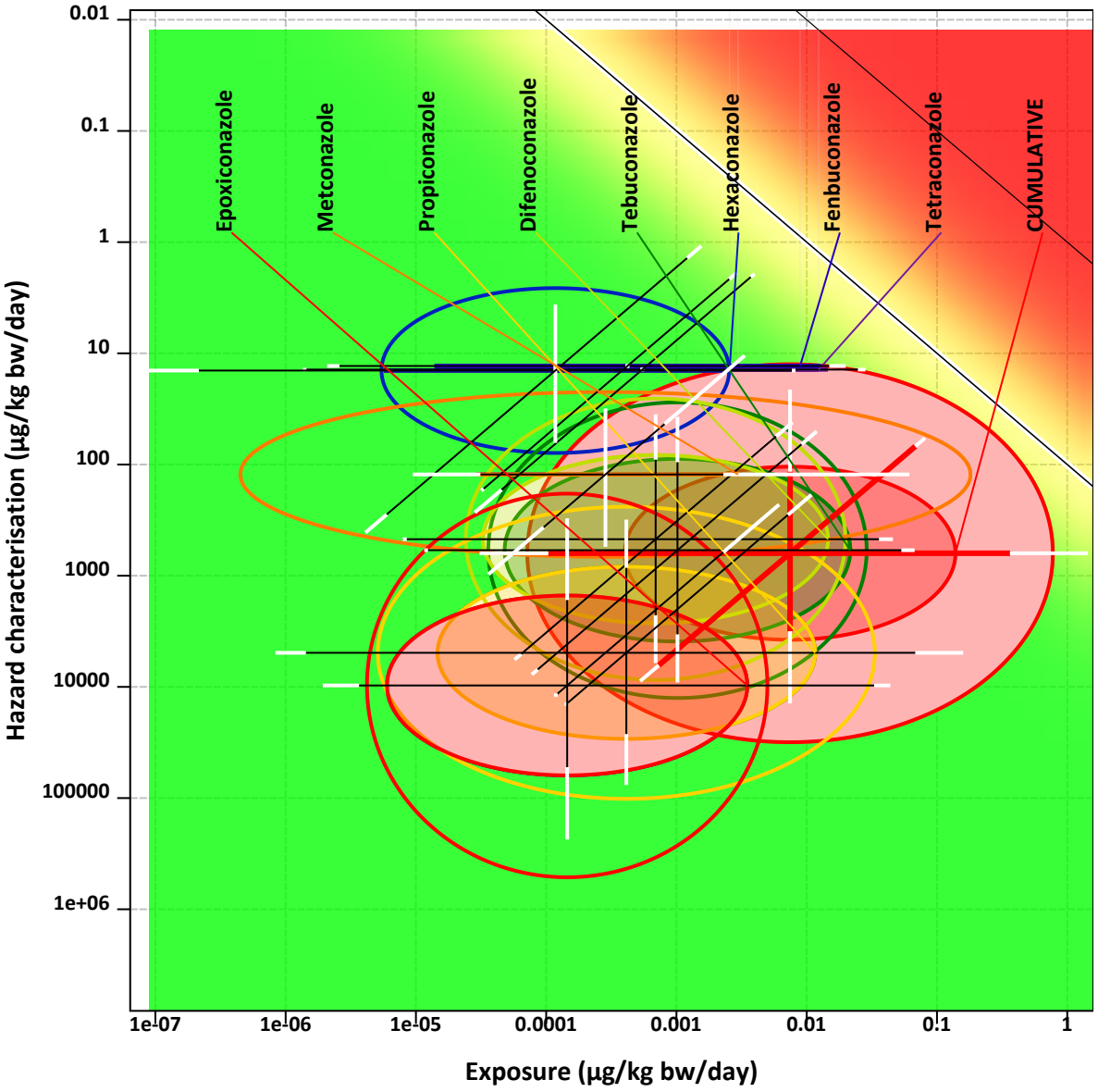


Figure 3.74: Hazard vs. exposure plot for multiple substances. 95% bivariate confidence areas for target hazard dose distribution and exposure distribution. Inner ellipses express variability, outer ellipses uncertainty.

Risk metric calculation type

Currently, two types of risk metric calculation types are available. Both for the margin of exposure and hazard index:

- exposures are cumulated over substancse using RPFs and the risk distribution is estimated based on the hazard characterisation of the reference substance and the cumulative exposure. See also *cumulative RPF weighted risk distribution*.
- risk is calculated per pubstance as a ratio of each hazard characterization and the exposure. Then, the risk is estimated as the cumulated *sum of ratios* over all substances. In formula:

For the hazard index HI:

$$HI = \sum_{s=1}^{S} HQ_s$$

where summation is over the number of substances per individual(day).

For the margin of exposure or more precise $MOET$:

$$MOET = \frac{1}{\sum_{s=1}^{S} \frac{1}{MOE_s}}$$

where summation is over the number of substances per individual(day).

Inverse distribution

Risk can be calculated as a distribution of either margin of exposure (MOE) or hazard index (HI), if at least one of the inputs exposure and hazard characterisation is a distribution. The risk distribution is characterised by percentiles. To accommodate for matching results of MOE and HI in the case of percentiles, there is an option to calculate percentiles via the complementary percentile of the *inverse distribution* in order to handle numerical differences when calculating percentiles for a left or right tail. For example, the p1 percentile of the MOE distribution can optionally be calculated as 1 divided by p99 of the corresponding $1/MOE$ (i.e., HI) distribution.

Risk by food

The option *calculate risks by modelled foods* is available when the target dose level is external. Dietary exposures preserving all the information of exposures of modelled foods are used to calculate risks statistics for modelled foods and to calculate the percentages at risk of modelled foods in the background and foreground based on the specified *threshold* in the safety plot.

Risks settings

Calculation settings

Table 3.210: Calculation settings for module Risks.

Name	Type	Description
Multiple substances analysis	Boolean	Specifies whether the assessment involves multiple substances.
Compute cumulative risks	Boolean	Specifies whether to compute the combined/cumulative risk ove all substances.
Health effect type	*HealthEffectType*	Specifies whether the health effect is a risk (negative) or benefit (positive).
Risk metric type	*RiskMetricType*	Report risks in terms of hazard index (HI = 1/MOE) or margin exposure.
Show equivalent animal dose output	Boolean	Specifies whether equivalent animal doses should be reported in the output.
Threshold safety plot	Numeric	Threshold for interpretation in the margin of exposure or hazard index plot, e.g. MOE = HI = 1 or MOE = 100 and HI = 0.01.
Use inverse distribution to calculate percentile	Boolean	Calculate percentile via the complementary percentage of the inverse distribution (default: no). Description: E.g., P0.1 of MC distribution is calculated via P99.9 of 1/MOE distribution. Not This option is provided because percentile calculation in small data sets is asymmetric in both tails.
Target level	*TargetLevelType*	Select to express hazard characterisations at external or internal exposure level. For an aggregate assessment, that is dietary and nondietary exposure data are combined, the target dose level is always internal. When only dietary exposures are available, the target dose level is optional, i.c. external or internal.
Risk type	*ExposureType*	The type of exposure considered in the assessment; acute (short term) or chronic (long-term).
Calculate risks by modelled foods, substances or a combination of the two	Boolean	When the dose target level is external, dietary exposures are directly used as input to risks. Dietary exposures preserve the information of exposure by modelled foods, substances or the combination. Summarizing this information may time consumi
Internal concentration	*InternalConcentrationType*	Internal concentrations are derived form dietary and/or non-dietary concentrations and aggregated using a kinetic or absorption factor model or are human monitoring concentration
Cumulative risk calculation method	*RiskMetricCalculationType*	Specify method for computing cumulative risks of multiple substances (e.g., via RPF weighted exposures or as a sum of ratios).

Output settings

Table 3.211: Output settings for module Risks.

Name	Type	Description
Number of plot labels	Numeric	Maximum number of labels to plot in hazard vs exposure plot.
Number of substances in hazard vs. exposure plot	Numeric	Maximum number of substances to plot in hazard vs exposure plot.
Left margin safety plot	Numeric	Left margin of the plot for margins of exposure or hazard indic
Right margin safety plot	Numeric	Right margin of the plot for margins of exposure of hazard indices.
Inclusion percentage variability interval	Numeric	The central percentage of the variability distribution to include intervals for exposure, hazard and MOE (e.g. 90 means p5-p95
Include drill-down on 9 individuals around specified percentile.	Boolean	Specifies whether drilldown on 9 individuals is to be included in the output.
Summarize simulated data	Boolean	Specifies whether a summary of the simulated consumptions an concentrations should be included in the output.
Store simulated individual day exposures	Boolean	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	AlphaNumeric	Give specific percentiles of exposure distribution (%), e.g. 50 9 95 97.5 99 (space separated).
Percentage for drilldown	Numeric	Gives detailed output for nine individuals near this percentile of the exposure distribution.
Percentage for upper tail	Numeric	Gives detailed output for this upper percentage of the exposure distribution.
Number of levels of covariable to predict exposure	Numeric	Specify the number of levels, e.g. 20. The range of the covariat is divided by the number of levels: range = (max - min)/levels. For these covariable levels exposures are predicted.
Predict exposure at extra covariable levels	Numeric	Specify specific prediction levels in addition to the automaticall generated prediction levels (space separated).
Lower percentage for variability (%)	Numeric	The default value of 25% may be overruled.
Upper percentage for variability (%)	Numeric	The default value of 75% may be overruled.
Report consumptions and exposures per individual instead of per kg body weight	Boolean	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.

Calculation of risks

Risk (health impact) is quantified as exposure relative to hazard characterisation, which in MCRA is called a hazard index (HI) for any type of inputs, or as hazard characterisation relative to exposure, which in MCRA is called a margin of exposure (MOE) for any type of inputs. Exposures or hazards can be single values or distributions, the risk metric is a distribution if at least one of the inputs is a distribution (if both are single values, see the module Single value risks). Risk metrics are valid for a specific biological level (external or internal at a specific organ).

- *Risks calculation*

Inputs used: *Dietary exposures Exposures Hazard characterisations Human monitoring analysis Relative potency factors*

Settings used

- *Calculation Settings*

Risks are expressed as distribution of margin of exposure or hazard index. The distribution is summarised by percentiles, and by the probability to exceed the specified threshold value (e.g. 1 or 100). The hazard vs. exposure plot compares 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.

By using probabilistic tiers for both exposure and hazard characterisation, the calculated MOE distribution is equal to the Integrated Margin Of Exposure (IMOE) distribution, as described for the Integrated Probabilistic Risk Assessment (IPRA) approach in van der Voet and Slob (2007) and van der Voet et al. (2009).

3.8.2 Single value risks

Single value risks are risk estimates obtained from combining single value exposures with single value hazard characterisations or as a percentile from a risk distribution.

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

Single value risks calculation

Single value risks can be calculated in two ways.

- If the option 'Single value risk calculation method' is set to 'From single value risks' then *single value exposures* are combined with (single value) *hazard characterisations*.
- If the option 'Single value risk calculation method' is set to 'As percentile from risks distribution' then a percentile can be selected from a *risks* distribution.

See below for a more detailed explanation.

Combining single value exposures and hazard characterisations

Single value risks are computed by combining *single value exposures* by route/source and substance with (single value) *hazard characterisations* by substance. They are computed as margin of exposure (hazard characterisation / exposure), hazard quotient or hazard index (exposure / hazard characterisation), or as a percentage of the reference dose (100 * exposure / hazard characterisation).

Single value risks from individual risks

In this option, a percentage point can be specified for the chosen risk metric (margin of exposure (MOE) or hazard index (HI)). The corresponding percentile is calculated from the distribution of individual *risks*. The default percentiles are a margin of exposure at 0.1% or a hazard index at 99.9%, but another value can be chosen. It can also be indicated whether the risk metric is calculated using the inverse distribution or not. This option is provided because percentile calculation in small data sets is asymmetric in both tails. When this option is set, the percentile is calculated as the inverse of the complementary percentage of the inverse distribution. E.g., the $p_{0.1}$ of the MOE distribution is calculated as 1/($p_{99.9}$ of 1/MOE distribution); the $p_{99.9}$ of the HI distribution is calculated as 1/($p_{0.1}$ of 1/HI distribution).

Adjustment factors and uncertainty specification

Many sources of uncertainty that may affect input data, model assumptions and assessment methodology do not enter the assessment. In EFSA (2020a) and EFSA (2020b), thirty-four sources of uncertainty were identified and the impact of each source on the MOE was quantified. Some uncertainties tend to overestimate the MOE, others tend to underestimate it. Following the guidance of the EFSA Scientific Committee, specific MOE and/or HI percentiles are adjusted using adjustment factors for exposure and hazard, e.g. from expert elicitation. They may be available as fixed values or as parametric uncertainty distributions. In the nominal run, the percentile is adjusted with the median of the uncertainty distribution. In each uncertainty run, adjustment factors are sampled from the uncertainty distribution. In the MCRA interface, for both exposure and hazard distribution separately, a fixed value or a parametric uncertainty distribution is specified. The available parametric uncertainty distributions are the same as available in the SHELF package that was used by EFSA. The SHeffield ELicitation Framework (SHELF) is a package of documents, templates and software to carry out elicitation of probability distributions for uncertain quantities from a group of experts (http://www.tonyohagan.co.uk/shelf/).

Options for specifying uncertainty distributions are:

- Lognormal(μ, s) with offset c. Parameters μ and s specify the mean and standard deviation of the underlying normal.
- Log Student t(μ, s, ν) with offset d. Parameters μ and s specify the mean and standard deviation of the underlying normal, ν the degrees of freedom, $\nu > 0$
- Beta(a, b) scaled to the interval [c, d], with shape parameters a and b > 0.
- Gamma(a, b) with offset c, with shape and rate parameters a and b > 0.

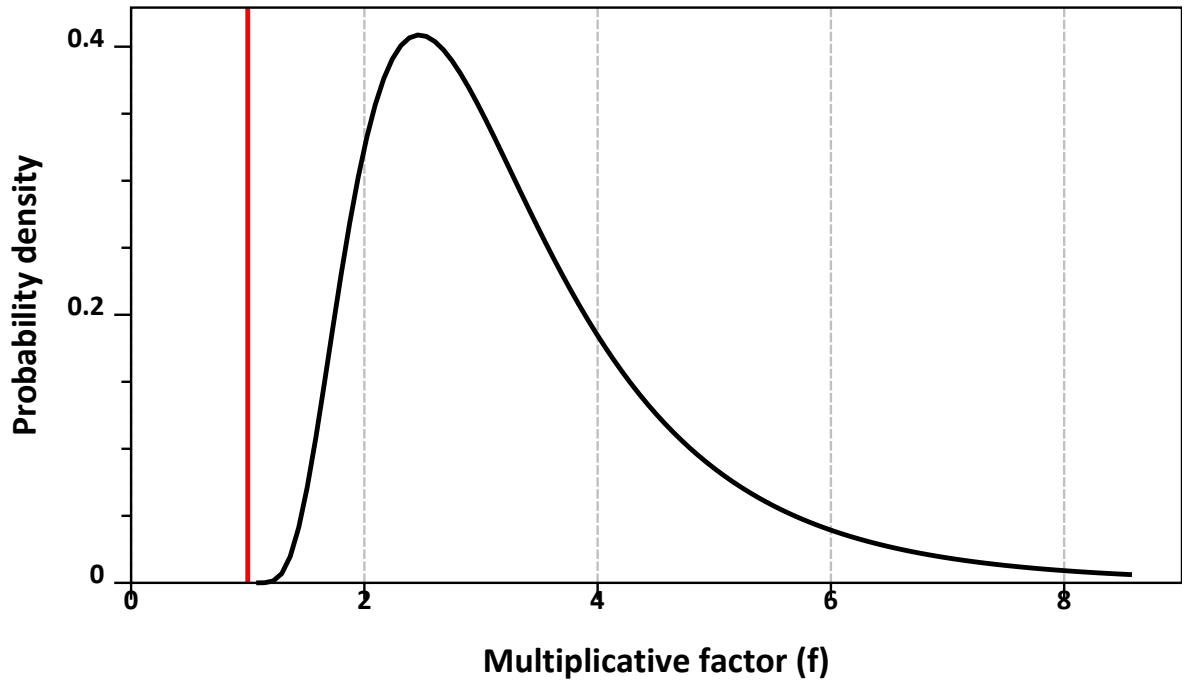


Figure 3.75: Scaled lognormal ($\mu = 0.705$, $s = 0.566$, offset=1), table 8, EFSA (2020b).

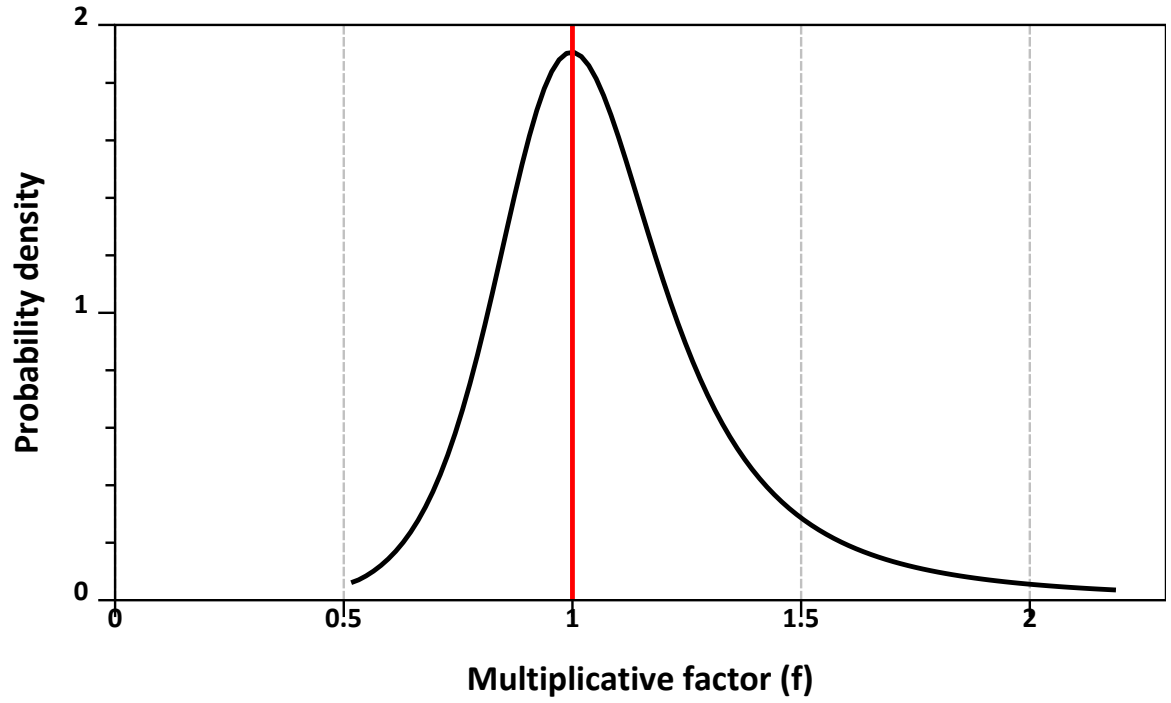


Figure 3.76: Scaled logstudents t ($\mu = -0.593$, $s = 0.367$, $\nu = 3$, offset=0.5), table 9, EFSA (2020b).

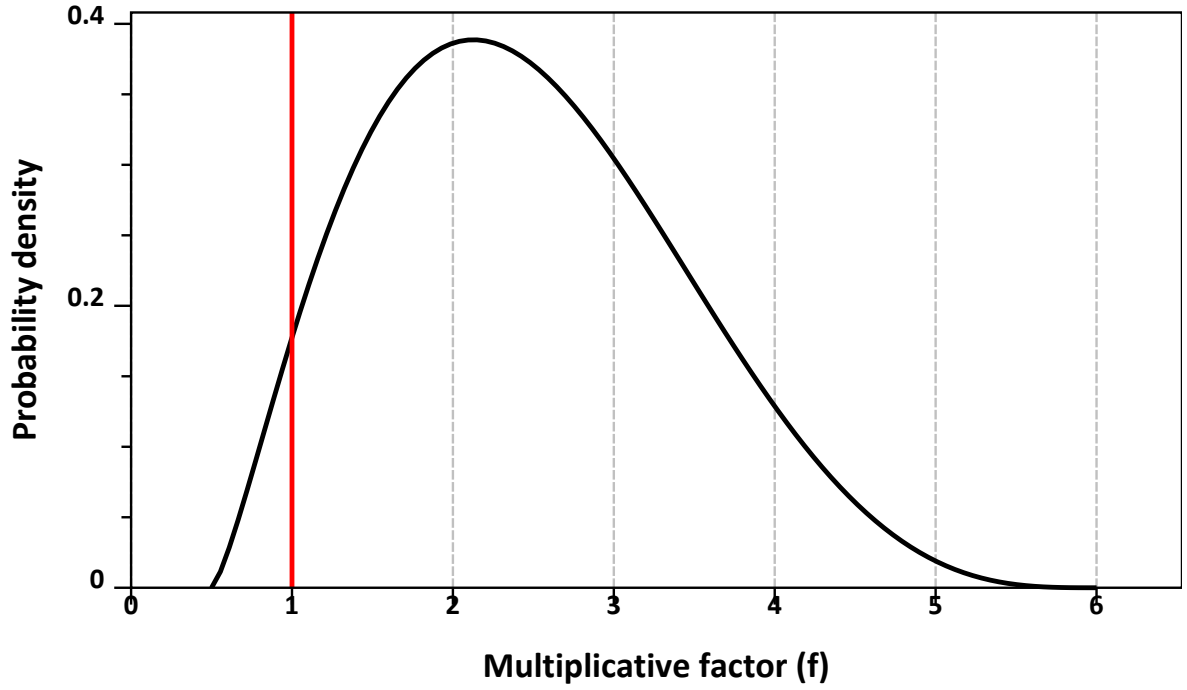


Figure 3.77: Scaled beta (a=2.37, b=4.26, lowerbound=0.5, upperbound=6), table 7, EFSA (2020a).

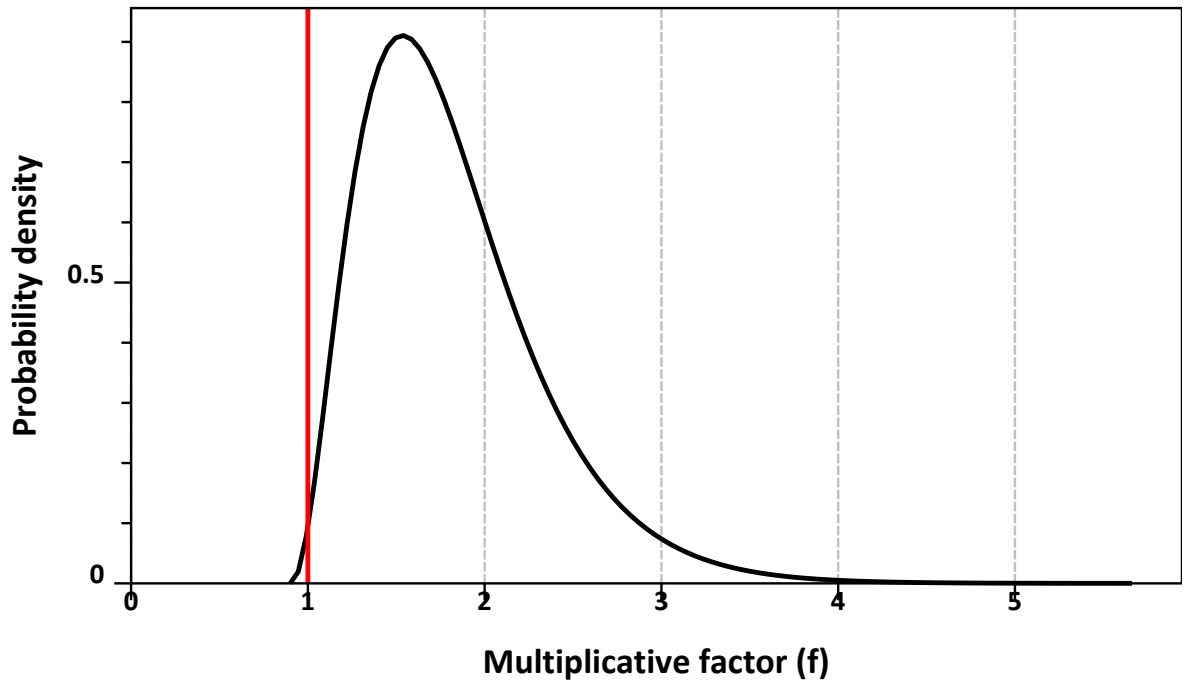


Figure 3.78: Scaled gamma (a=3.26, b=3.56, offset=0.9), table 6, EFSA (2020a).

Background-only adjustment factor

When exposures are calculated by *combining focal food/substance concentrations with background concentrations*, it may be appropriate to have a separate adjustment for the foreground and background. A pragmatic solution agreed with EFSA is to estimate the contribution of the foreground in the tail above the selected percentile. Suppose this contribution is c. Note that c will vary in uncertainty runs. Then, the adjustment factor should be multiplied by $(1-c)$, i.e. no adjustment for the focal part.

The calculation proceeds as follows:

$$
\begin{aligned}
p_{\text{MOE,adjusted}} &= p_{\text{MOE}} \cdot (c + (1-c) \cdot \texttt{AdjustmentFactor}_{\texttt{exposure}} \cdot \texttt{AdjustmentFactor}_{\texttt{hazard}}) \\
p_{\text{HI,adjusted}} &= \frac{p_{\text{HI}}}{c + (1-c) \cdot \texttt{AdjustmentFactor}_{\texttt{exposure}} \cdot \texttt{AdjustmentFactor}_{\texttt{hazard}}}
\end{aligned}
$$

Note that when the focal substance measurements are converted to active substances using *substance conversions* or *deterministic substance conversions*, then c is the sum of the contributions of the focal food in and all active substances to which the substance translates.

In Figure 3.79, an example is shown where the margin of exposure is adjusted for the exposure and hazard distribution based on expert elicitation. The median adjustment factors for exposure and hazard are respectively, 1.77 and 3.01. The overall adjustment factor is 5.33.

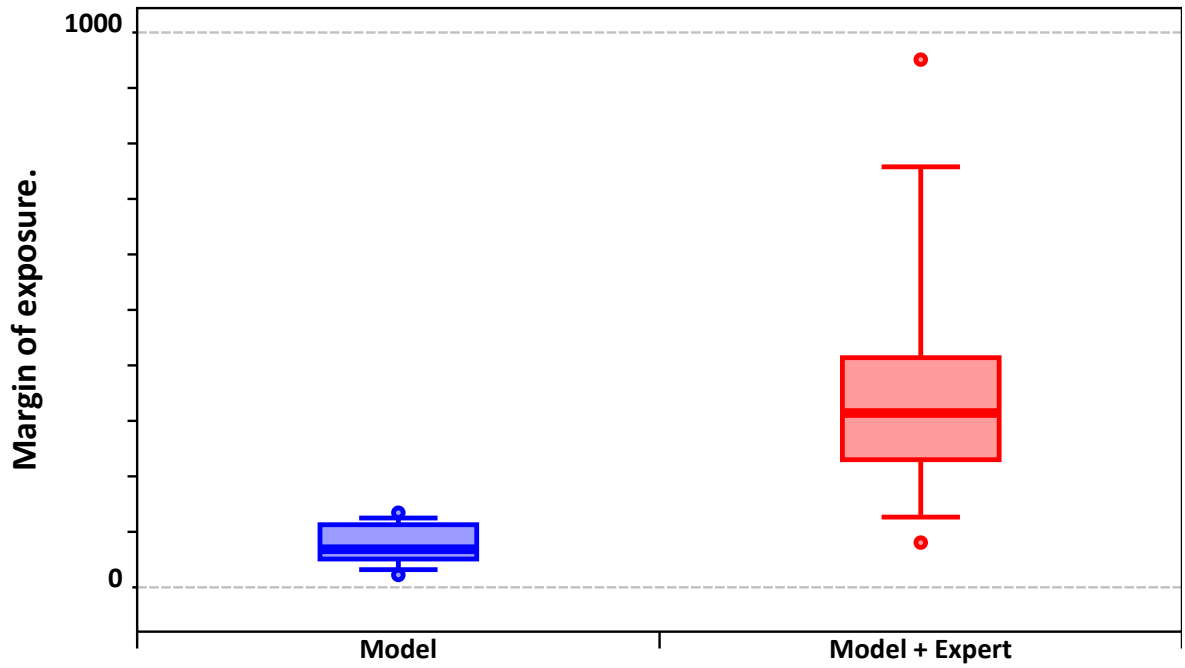


Figure 3.79: Margin of exposure (model) and adjusted margin of exposure (model + expert) with uncertainty bounds.

Single value risks settings

Calculation settings

Table 3.212: Calculation settings for module Single value risks.

Name	Type	Description
Multiple substances analysis	Boolean	Specifies whether the assessment involves multiple substances.
Compute cumulative exposures	Boolean	Specifies whether the assessment involves multiple substances a results should be cumulated over all substances.
Risk type	*ExposureType*	The type of exposure considered in the assessment; acute (short term) or chronic (long-term).
Risk metric type	*RiskMetricType*	Report risks in terms of hazard index (HI = 1/MOE) or margin exposure.
Single value risk calculation method	*SingleValueRiskCalculation-Method*	Calculate single value from exposures and hazard or from an individual risks distribution.
Percentage for percentile	Numeric	Percentage for percentile (default 0.1 for MOE or 99.9 for HI).
Use inverse distribution to calculate percentile	Boolean	Calculate percentile via the complementary percentage of the inverse distribution (default: no). Description: E.g., P0.1 of MO distribution is calculated via P99.9 of 1/MOE distribution. Not This option is provided because percentile calculation in small data sets is asymmetric in both tails.
Apply adjustment factors to the specified risk percentile	Boolean	Specify adjustment factors, e.g. based on expert knowledge elicitation, to a specified MOE percentile (default 0.1%). If the selected risk metric is HI, the adjustment factors should still be specified for the complementary percentile of MOE (e.g. P0.1 MOE if P99.9 of HI is selected).
Adjustment type related to exposure	*AdjustmentFactorDistribution-Method*	Specify the factor and/or distribution of the adjustment factor f the MOE percentile. Default is no adjustment. Alternatives are fixed factor or an uncertainty distribution. If distributions are selected, default values are set based on EFSA cumulative risk reports 2020.
Parameter A (Fixed factor, mean Lognormal or LogStudent-t, or shape parameter Beta or Gamma)	Numeric	This parameter can be: 1) the fixed adjustment factor; 2) for Lognormal or LogStudent-t, the mean of the underlying normal distribution; 3) For Beta or Gamma. the shape parameter.
Parameter B (standard deviation Lognormal or LogStudent-t or second shape parameter Beta or rate parameter Gamma)	Numeric	This parameter can be: 1) for Lognormal or LogStudent-t, the standard deviation of the underlying normal distribution; 2) For Beta, the second shape parameter; 3) for Gamma, the rate parameter.
Parameter C (Lower bound Beta, offset Gamma or Lognormal or degrees of freedom Logstudent-t)	Numeric	This parameter can be: 1) for Beta, the lower bound value; 2) f Gamma or Lognormal, the offset; 3) for LogStudent-t, the degr of freedom.
Parameter D (Upper bound Beta or offset LogStudent-t)	Numeric	This parameter can be: 1) for Beta, the upper bound value; 2) f LogStudent-t, the offset.
Adjustment type related to hazard	*AdjustmentFactorDistribution-Method*	Specify the factor and/or distribution of the adjustment factor f the MOE percentile. Default is no adjustment. Alternatives are fixed factor or an uncertainty distribution. If distributions are selected, default values are set based on EFSA cumulative risk reports 2020.
Parameter A (Fixed factor, mean Lognormal or LogStudent-t, or shape parameter Beta or Gamma)	Numeric	This parameter can be: 1) the fixed adjustment factor; 2) for Lognormal or LogStudent-t, the mean of the underlying normal distribution; 3) For Beta or Gamma. the shape parameter.
Parameter B (standard deviation Lognormal or LogStudent-t or second shape parameter Beta or rate parameter Gamma)	Numeric	This parameter can be: 1) for Lognormal or LogStudent-t, the standard deviation of the underlying normal distribution; 2) For Beta, the second shape parameter; 3) for Gamma, the rate parameter.
Parameter C (Lower bound Beta, offset Gamma or Lognormal or degrees of freedom Logstudent-t)	Numeric	This parameter can be: 1) for Beta, the lower bound value; 2) f Gamma or Lognormal, the offset; 3) for LogStudent-t, the degr of freedom.
Parameter D (Upper bound	Numeric	This parameter can be: 1) for Beta, the upper bound value; 2) f

Calculation of single value risks

Single value risk can be computed by route and substance in the form of hazard quotients or margins of exposure. Single value risks are risk estimates obtained from combining single value exposures with single value hazard characterisations or as a percentile from a risk distribution. Optionally, the exposure and hazard behind the percentile calculation can be adjusted with fixed adjustment factors or factors drawn from parametric uncertainty distributions.

- *Single value risks calculation*

Inputs used: *Single value dietary exposures Hazard characterisations Risks*

Settings used

- *Calculation Settings*

continues on next page

Table 3.213 – continued from previous page

Category	Module	Inputs	Used by	Description

Table 3.213: Overview of MCRA modules.

Category	Module	Inputs	Used by	Description
Primary entity	*Foods*		*Consumptions*, *Single value consumptions*, *Market shares*, *Food recipes*, *Concentrations*, *Concentration distributions*, *Single value concentrations*, *Processing factors*, *Unit variability factors*, *Occurrence patterns*, *Occurrence frequencies*, *Substance authorisations*, *Deterministic substance conversion factors*, *Concentration limits*, *Concentration models*, *Modelled foods*, *Focal food concentrations*, *Total diet study sample compositions*, *Food extrapolations*, *Food conversions*, *Consumptions by modelled food*, *High exposure food-substance combinations*, *Dietary exposures*, *Single value dietary exposures*, *Exposures*.	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) modelled foods, foods as coded in concentration data (e.g. wheat, tomato); 3) any other type of food (e.g. ingredients like flour, tomato sauce).

continues on next page

Table 3.213 – continued from previous page

Category	Module	Inputs	Used by	Description
	Non-dietary exposure sources			Non-dietary exposure sources are the sources containing chemical substances to which individuals in a population are exposed via any of three non-dietary routes: dermal, inhalation or oral, per day.

continues on next page

Table 3.213 – continued from previous page

Category	Module	Inputs	Used by	Description
	Substances		*Concentrations*, *Concentration distributions*, *Single value concentrations*, *Processing factors*, *Unit variability factors*, *Occurrence patterns*, *Occurrence frequencies*, *Substance authorisations*, *Substance approvals*, *Substance conversions*, *Deterministic substance conversion factors*, *Concentration limits*, *Concentration models*, *Modelled foods*, *Focal food concentrations*, *Food conversions*, *Consumptions by modelled food*, *High exposure food-substance combinations*, *Dietary exposures*, *Single value dietary exposures*, *Non-dietary exposures*, *Exposures*, *Exposure mixtures*, *Human monitoring data*, *Human monitoring analysis*,	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 3.213 – continued from previous page

Category	Module	Inputs	Used by	Description
	Effects		*Concentration models*, *High exposure food-substance combinations*, *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*, *Single value 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.

continues on next page

Table 3.213 – continued from previous page

Category	Module	Inputs	Used by	Description
	Populations		*Consumptions*, *Single value consumptions*, *Concentrations*, *Consumptions by modelled food*, *Dietary exposures*, *Single value dietary exposures*, *Non-dietary exposures*, *Exposures*, *Human monitoring data*, *Human monitoring analysis*, *Biological matrix concentration comparisons*, *Hazard characterisations*, *Risks*, *Single value 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 (with a start and end date), age range (with a minimum and maximum age) and gender. Example: the French population in 2005-2007 (= time period) of women (= gender) of child-bearing age 18-45 yr (= age range).
	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.

continues on next page

Table 3.213 – continued from previous page

Category	Module	Inputs	Used by	Description
Consumption	*Consumptions*	*Populations*, *Foods*.	*Food conversions*, *Consumptions by modelled food*.	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 persons, so preferably two or more days per individual are needed.
	Single value consumptions	*Consumptions by modelled food*.	*Single value dietary exposures*.	Single value consumption data are the single value amounts (Large Portion, Mean Consumption, p97.5Consumption) of modelled foods (foods-as-measured) consumed in a population.
	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 modelled foods) by specifying proportions in the form of a percentage.

continues on next page

Table 3.213 – continued from previous page

Category	Module	Inputs	Used by	Description
Occurrence	*Concentrations*	*Foods*, *Substances*, *Populations*, *Focal food concentrations*, *Food extrapolations*, *Substance conversions*, *Deterministic substance conversion factors*, *Relative potency factors*, *Substance authorisations*, *Active substances*, *Concentration limits*, *Substance approvals*.	*Single value concentrations*, *Occurrence patterns*, *Concentration models*, *Modelled foods*.	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.
	Concentration distributions	*Foods*, *Substances*.	*Concentration models*, *Dietary exposures*.	Concentration distributions describe substance concentrations on foods in the form of summary statistics.
	Single value concentrations	*Active substances*, *Concentrations*, *Concentration limits*, *Deterministic substance conversion factors*.	*Modelled foods*, *Single value dietary exposures*.	Single value concentrations data are the single value estimates (High Residue, Maximum Residue Limit, Supervised Trials Median Residue) of residue concentrations on modelled foods.
	Processing factors	*Foods*, *Substances*.	*Food conversions*, *Dietary exposures*, *Single value 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.

continues on next page

Table 3.213 – continued from previous page

Category	Module	Inputs	Used by	Description
	Unit variability factors	*Foods*, *Substances*.	*Dietary exposures*, *Single value 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.
	Occurrence patterns	*Substance authorisations*, *Active substances*, *Concentrations*.	*Occurrence frequencies*, *Dietary exposures*.	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 are 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.
	Occurrence frequencies	*Active substances*, *Occurrence patterns*.	*Concentration models*, *Single value dietary exposures*.	Occurrence frequencies specify how often substances occur on foods. Frequencies are expressed as percentages.

continues on next page

Table 3.213 – continued from previous page

Category	Module	Inputs	Used by	Description
	Substance authorisations	*Foods*, *Substances*.	*Concentrations*, *Occurrence patterns*, *Concentration models*.	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 LOQ or LOD could be assumed true zeros. I.e., if a food/substance combinations is assumed to be unauthorised, then the LOQ, LOD may be assumed to be a zero.
	Substance approvals	*Substances*.	*Concentrations*.	Substance approvals specify which substances are approved within the definition under regulation (EC) No 1107/2009. This information may, for instance, be used to to restrict water imputation to approved substances only.
	Substance conversions	*Substances*, *Active substances*.	*Concentrations*.	Substance conversions specify how measured substances are converted into active substances, which are the substances assumed to cause health effects. In pesticide legislation such measured substances and the substance conversion rules are known as residue definitions.
	Deterministic substance conversion factors	*Substances*, *Foods*.	*Concentrations*, *Single value concentrations*.	Deterministic substance conversion factors.
	Concentration limits	*Foods*, *Substances*.	*Concentrations*, *Single value concentrations*, *Concentration models*, *Modelled foods*.	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.

continues on next page

Table 3.213 – continued from previous page

Category	Module	Inputs	Used by	Description
	Concentration models	*Concentrations*, *Concentration limits*, *Active substances*, *Modelled foods*, *Substance authorisations*, *Occurrence frequencies*, *Relative potency factors*, *Concentration distributions*, *Total diet study sample compositions*.	*High exposure food-substance combinations*, *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.
	Modelled foods	*Concentrations*, *Single value concentrations*, *Concentration limits*.	*Concentration models*, *Food conversions*.	Modelled foods are foods within the foods scope for which concentration data or MRLs of substances are available (or expected).
	Focal food concentrations	*Foods*, *Substances*.	*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*.	*Concentration models*, *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).

continues on next page

Table 3.213 – continued from previous page

Category	Module	Inputs	Used by	Description
Exposure	*Food conversions*	*Consumptions*, *Modelled foods*, *Processing factors*, *Food recipes*, *Market shares*, *Food extrapolations*, *Total diet study sample compositions*, *Active substances*.	*Consumptions by modelled food*, *Dietary exposures*.	Food conversions relate foods-as-eaten, as found in the consumption data, to modelled foods (foods-as-measured), which are the foods for which concentration data are available. A food-as-eaten can be linked to one, or multiple modelled foods using various conversion steps (e.g., using food recipes to translate a composite food into its ingredients). There are several types of conversion steps, and a conversion path may comprise multiple conversion steps between a food-as-eaten and a modelled food.
	Consumptions by modelled food	*Consumptions*, *Food conversions*.	*Single value consumptions*, *High exposure food-substance combinations*, *Dietary exposures*.	Consumptions by modelled food are consumptions of individuals expressed on the level of the foods for which concentration data are available (i.e., the modelled-foods). These are calculated from consumptions of foods-as-eaten and food conversions that link the foods-as-eaten amounts to modelled-foods amounts.
	High exposure food-substance combinations	*Consumptions by modelled food*, *Concentration models*, *Active substances*, *Relative potency factors*.	*Dietary exposures*.	Identification of food-as-eaten/modelled food/substance combinations that have the highest expected contribution to exposure based on a simple screening model.

continues on next page

Table 3.213 – continued from previous page

Category	Module	Inputs	Used by	Description
	Dietary exposures	*Consumptions by modelled food*, *Concentration models*, *Processing factors*, *Unit variability factors*, *High exposure food-substance combinations*, *Active substances*, *Occurrence patterns*, *Relative potency factors*, *Food conversions*, *Concentration distributions*.	*Exposures*, *Exposure mixtures*, *Risks*.	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.
	Single value dietary exposures	*Single value consumptions*, *Single value concentrations*, *Processing factors*, *Unit variability factors*, *Occurrence frequencies*.	*Single value risks*.	Single value dietary exposures are based on the single value concentrations of substances, expressed per standard (kg) bodyweight and/or single value amounts of consumed modelled food. Depending on the exposure type, dietary exposures can be short-term/acute exposures.
	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.

continues on next page

Table 3.213 – continued from previous page

Category	Module	Inputs	Used by	Description
	Exposures	*Dietary exposures*, *Non-dietary exposures*, *Active substances*, *Relative potency factors*, *Kinetic models*.	*Exposure mixtures*, *Biological matrix concentration comparisons*, *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.
	Exposure mixtures	*Dietary exposures*, *Exposures*, *Relative potency factors*, *Human monitoring analysis*.		Exposure mixtures will select sets of co-occurring substances (one or more) that contribute most to the overall exposure patterns.
	Human monitoring data	*Populations*, *Substances*.	*Human monitoring analysis*.	Human monitoring data quantify substance concentrations found in humans collected in human monitoring surveys.
	Human monitoring analysis	*Human monitoring data*, *Active substances*, *Relative potency factors*.	*Exposure mixtures*, *Biological matrix concentration comparisons*, *Risks*.	Human monitoring concentrations are substance concentration estimates for a biological matrix (e.g., urine or blood) derived from data obtained from human monitoring studies.

continues on next page

Table 3.213 – continued from previous page

Category	Module	Inputs	Used by	Description
	Biological matrix concentration comparisons	*Human monitoring analysis*, *Exposures*.		Substances in the human body are absorbed, excreted without transformation, excreted after metabolization or stored in various tissues, bones or body fluids. The term biological matrix refers to all human specimens where concentratrions of a chemical can be measured like bodily fluids, such as blood, urine, saliva, breast milk, sweat, and other specimens, such as faeces, hair, teeth, and nails. Biological matrix concentration comparisons 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).

continues on next page

Table 3.213 – continued from previous page

Category	Module	Inputs	Used by	Description
Kinetic	*Kinetic models*	*Substances*, *Active substances*.	*Exposures*, *Hazard characterisations*.	Kinetic models relate exposures or hazard characterisations from one or more external routes (dietary, non-dietary oral, dermal, inhalation) to an internal (target) compartment. Kinetic models can be simple absorption factors or differential-equation based PBK models. MCRA currently includes the EuroMix generic PBK model and the bisphenol model of ETHZ. Absorption factors and parameters for instances of PBK models can be specified as data. Alternatively, (default) absorption factors can be set in the interface.

continues on next page

Table 3.213 – continued from previous page

Category	Module	Inputs	Used by	Description
Hazard	*Active substances*	*AOP networks*, *Points of departure*, *Hazard characterisations*, *Molecular docking models*, *QSAR membership models*.	*Concentrations*, *Single value concentrations*, *Occurrence patterns*, *Occurrence frequencies*, *Substance conversions*, *Non-dietary exposures*, *Kinetic models*, *Relative potency factors*, *Hazard characterisations*, *Inter-species conversions*, *Intra species factors*, *Concentration models*, *Food conversions*, *High exposure food-substance combinations*, *Dietary exposures*, *Exposures*, *Human monitoring analysis*.	Active substances are substances that may lead (P>0) to a specific health effect (adverse outcome). Active substances are 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).

continues on next page

Table 3.213 – continued from previous page

Category	Module	Inputs	Used by	Description
	Relative potency factors	*Active substances*, *AOP networks*, *Hazard characterisations*.	*Concentrations*, *Concentration models*, *High exposure food-substance combinations*, *Dietary exposures*, *Exposures*, *Exposure mixtures*, *Human monitoring analysis*, *Risks*.	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.

continues on next page

Table 3.213 – continued from previous page

Category	Module	Inputs	Used by	Description
	Hazard characterisations	*AOP networks*, *Active substances*, *Points of departure*, *Dose response models*, *Effect representations*, *Inter-species conversions*, *Intra species factors*, *Kinetic models*.	*Active substances*, *Relative potency factors*, *Risks*, *Single value risks*.	Hazard characterisations are reference exposure values for active substances at the chosen biological target level (external or internal). Hazard characterisations may be specified for specific effects or for the critical effect as defined in hazard characterisation. Hazard characterisations are specified as external values (e.g. human based guidance values, such as ADI or ARfD) or are based on points of departure, such as BMDs from dose-response models or externally specified points of departure (NOAEL, LOAEL, MDS). The computation may involve assessment factors, e.g. for inter-species conversion, intra-species variation or additional sources of uncertainty. The calculation may also use kinetic models or absorption factors to convert external doses to internal doses or vice versa.
	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).

continues on next page

Table 3.213 – continued from previous page

Category	Module	Inputs	Used by	Description
	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.
	Inter-species conversions	*Substances*, *Effects*, *Active substances*.	*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*, *Active substances*.	*Hazard characterisations*.	Intra-species factors specify how to convert a hazard characterisation from the average to a sensitive human individual.

continues on next page

Table 3.213 – continued from previous page

Category	Module	Inputs	Used by	Description
	AOP networks	*Effects*.	*QSAR membership models*, *Molecular docking models*, *Active substances*, *Relative potency factors*, *Hazard characterisations*, *Points of departure*, *Effect representations*.	Effects are 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 initiating 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.
Risks	*Risks*	*Dietary exposures*, *Exposures*, *Hazard characterisations*, *Human monitoring analysis*, *Relative potency factors*.	*Single value risks*.	Risks (health impacts) are defined as a function of exposure and hazard characterisation at a chosen biological level (external or internal). Risk metrics are margins of exposure (MOE) or hazard indices (HI) or more generalised MOE or HI distributions.
	Single value risks	*Single value dietary exposures*, *Hazard characterisations*, *Risks*.		Single value risks are risk estimates obtained from combining single value exposures with single value hazard characterisations or as a percentile from a risk distribution.

CHAPTER

FOUR

STANDARD ACTIONS

A standard action is a user friendly way to perform a complex probabilistic calculation. By using a standard action predefined settings are used and the user can set only a limited number of selections. All settings (pre-defined and set by the user) are visible in the output. As a result a short output is presented. More detailed output is still available.

4.1 Chronic mixture risk assessment of metals relevant for chronic kidney disease (nephrotoxicity)

This standard action is of type: *Risks*

In the context of the European projects HBM4EU and PARC, the National Institute for Public Health and the Environment of the Netherlands (RIVM) performed a case study on the risk assessment of the combined exposure of four metals relevant for chronic kidney disease, i.e. cadmium, lead, in-organic arsenic and inorganic mercury. The exposure assessment used chemical concentration in foods of 14 European countries over the years 2014-2018, obtained from the EFSA data warehouse and individual Dutch food consumption data. Since chronic kidney disease is relevant for the adult population, adults in the age of 18-64 years were selected in the standard action. Exposure estimates were obtained using the observed individual means (OIM) model implemented in MCRA. Using relative potency factors, the exposure to the metals was expressed as equivalents of cadmium and summed per individual in the food consumption database, yielding a distribution of summed cadmium-equivalents. The summed cadmium-equivalents per individual were divided by the reference point of cadmium, which resulted in a distribution of hazard indexes (or better modified reference point indexes) from which the mean, P50 and P95 are obtained. A hazard index or modified reference point index > 1 means either a risk cannot be excluded or refinement of the assessment is needed, depending on the assesses uncertainties. The standard action is meant for training and demonstration purposes, to demonstrate assessment of the impact of setting (new) maximum limits concentrations (MLs).

Table 4.1: Datasources for Chronic mixture risk assessment of metals relevant for chronic kidney disease (nephrotoxicity).

Table Group	Name	Repository	Type
AssessmentGroupMemberships	Catalogues_NEFmetals.xlsx	Standard Actions/Chronic mixture risk assessment of metals relevant for chronic kidney disease (nephrotoxicity)	Fixed
Compounds	Catalogues_NEFmetals.xlsx	Standard Actions/Chronic mixture risk assessment of metals relevant for chronic kidney disease (nephrotoxicity)	Fixed
Effects	Catalogues_NEFmetals.xlsx	Standard Actions/Chronic mixture risk assessment of metals relevant for chronic kidney disease (nephrotoxicity)	Fixed
FoodTranslations	Catalogues_NEFmetals.xlsx	Standard Actions/Chronic mixture risk assessment of metals relevant for chronic kidney disease (nephrotoxicity)	Fixed
Foods	Catalogues_NEFmetals.xlsx	Standard Actions/Chronic mixture risk assessment of metals relevant for chronic kidney disease (nephrotoxicity)	Fixed
HazardCharacterisations	Catalogues_NEFmetals.xlsx	Standard Actions/Chronic mixture risk assessment of metals relevant for chronic kidney disease (nephrotoxicity)	Fixed
Populations	Catalogues_NEFmetals.xlsx	Standard Actions/Chronic mixture risk assessment of metals relevant for chronic kidney disease (nephrotoxicity)	Fixed
RelativePotencyFactors	Catalogues_NEFmetals.xlsx	Standard Actions/Chronic mixture risk assessment of metals relevant for chronic kidney disease (nephrotoxicity)	Fixed
Survey	ConsumptionData1-2yr.mdb	Standard Actions/Chronic mixture risk assessment of metals relevant for chronic kidney disease (nephrotoxicity)	Fixed
Survey	ConsumptionData10-17yr.mdb	Standard Actions/Chronic mixture risk assessment of metals relevant for chronic kidney disease (nephrotoxicity)	Fixed
Survey	ConsumptionData18-64yr.mdb	Standard Actions/Chronic mixture risk assessment of metals relevant for chronic kidney disease (nephrotoxicity)	Fixed
Survey	ConsumptionData3-9yr.mdb	Standard Actions/Chronic mixture risk assessment of metals relevant for chronic kidney disease (nephrotoxicity)	Fixed
Survey	ConsumptionData65-74yr.mdb	Standard Actions/Chronic mixture risk assessment of metals relevant for chronic kidney disease (nephrotoxicity)	Fixed
Survey	ConsumptionData75plusyr.mdb	Standard Actions/Chronic mixture risk assessment of metals relevant for chronic kidney disease (nephrotoxicity)	Fixed
MaximumResidueLimits	NEF-MLsmetal.xlsx	Standard Actions/Chronic mixture risk assessment of metals relevant for chronic kidney disease (nephrotoxicity)	Variable
MaximumResidueLimits	NEF-MLsExtendedmetal.xlsx	Standard Actions/Chronic mixture risk assessment of metals relevant for chronic kidney disease (nephrotoxicity)	Variable
Concentrations	NEF-SSDmetals.xlsx	Standard Actions/Chronic mixture risk assessment of metals relevant for chronic kidney disease (nephrotoxicity)	Fixed

References:

1. Amzal, B., Julin, B., Vahter, M., Wolk, A., Johanson, G., & Akesson, A. (2009). Population toxicokinetic modeling of cadmium for health risk assessment. Environ Health Perspect, 117(8), 1293-1301. https://doi.org/10.1289/ehp.0800317.
2. Carlisle, J. C., & Wade, M. J. (1992). Predicting blood lead concentrations from environmental concentrations. Regul Toxicol Pharmacol, 16(3), 280-289. https://doi.org/10.1016/0273-2300(92)90008-w.
3. EFSA (2009a) Cadmium in food - Scientific opinion of the Panel on Contaminants in the Food Chain. The EFSA Journal (2009) 980, 1-139. Available online: www.efsa.europa.eu.
4. EFSA. (2009b). Scientific Opinion on Arsenic in Food. EFSA J, 7(10). https://doi.org/0.2903/j.efsa.2009.1351.
5. EFSA. (2010b). Scientific Opinion on Lead in Food. EFSA J, 8(8), 1570
6. EFSA. (2012a). Scientific Opinion on the risk for public health related to the presence of mercury and methylmercury in food. EFSA J, 10(12).

7. EFSA. (2021a). Chronic dietary exposure to inorganic arsenic. EFSA J, 19(1), e06380. https://doi.org/10.2903/j.efsa.2021.6380.

8. Lin, Y. J., Hsiao, J. L., & Hsu, H. T. (2020). Integration of biomonitoring data and reverse dosimetry modeling to assess population risks of arsenic-induced chronic kidney disease and urinary cancer. Ecotoxicol Environ Saf, 206, 111212. https://doi.org/10.1016/j.ecoenv.2020.111212 Scoel 2007.

9. Sprong et al. Combined chronic dietary exposure to cadmium, lead, inorganic arsenic and inorganic mercury may pose a risk for nephrotoxicity in the adult population of ten European countries. manuscript in preparation.

4.2 Chronic cumulative exposure assessment PA

This standard action is of type: *Risks*

In 2017 EFSA published a statement *Risks for human health related to the presence of pyrrolizidine alkaloids in honey, tea, herbal infusions and food supplements*, see EFSA (2017b). Occurrence data used in this opinion was published in 2015 in the external scientific report *Occurrence of Pyrrolizidine Alkaloids in food*, see Mulder et al. (2015). The occurrence data in tea and herbal infusions is linked to the Consumption of 6 different population groups of the DNFCS 2012-2016. An lower bound (LB) and upper bound (UB) chronic cumulative exposure assessment can be calculated for different PAs. One scenario is assuming equipotency, another scenario is using provisional RPFs (Merz and Schrenk (2016)).

4.3 Chronic cumulative exposure assessment PFAS

This standard action is of type: *Risks*

This standard action can be used to calculate a chronic cumulative exposure for four PFAS. Assuming equipotency, and using proposed RPFs.

Table 4.2: Datasources for Chronic cumulative exposure assessment PFAS.

Table Group	Name	Repository	Type
Survey	ConsumptionData-1-2yr.mdb	Standard Actions/Chronic cumulative exposure assessment PFAS	Fixed
Survey	ConsumptionData-10-17yr.mdb	Standard Actions/Chronic cumulative exposure assessment PFAS	Fixed
Survey	ConsumptionData-18-64yr.mdb	Standard Actions/Chronic cumulative exposure assessment PFAS	Fixed
Survey	ConsumptionData-3-9yr.mdb	Standard Actions/Chronic cumulative exposure assessment PFAS	Fixed
Survey	ConsumptionData-65-74yr.mdb	Standard Actions/Chronic cumulative exposure assessment PFAS	Fixed
Survey	ConsumptionData-75plusyr.mdb	Standard Actions/Chronic cumulative exposure assessment PFAS	Fixed
Concentrations	PFAS-Occurrencedata-EFSA-LB.mdb	Standard Actions/Chronic cumulative exposure assessment PFAS	Variable
Concentrations	PFAS-Occurrencedata-EFSA-UB.mdb	Standard Actions/Chronic cumulative exposure assessment PFAS	Variable
AssessmentGroupMemberships	PFAS-OtherData.mdb	Standard Actions/Chronic cumulative exposure assessment PFAS	Fixed
Compounds	PFAS-OtherData.mdb	Standard Actions/Chronic cumulative exposure assessment PFAS	Fixed
Effects	PFAS-OtherData.mdb	Standard Actions/Chronic cumulative exposure assessment PFAS	Fixed
Foods	PFAS-OtherData.mdb	Standard Actions/Chronic cumulative exposure assessment PFAS	Fixed
HazardCharacterisations	PFAS-OtherData.mdb	Standard Actions/Chronic cumulative exposure assessment PFAS	Fixed
RelativePotencyFactors	PFAS-RelativePotencyFactorsEquipotency.mdb	Standard Actions/Chronic cumulative exposure assessment PFAS	Variable
RelativePotencyFactors	PFAS-RelativePotencyFactorsProposal.mdb	Standard Actions/Chronic cumulative exposure assessment PFAS	Variable

4.4 Demo acute cumulative risk assessment

This standard action is of type: *Risks*

In this demo with fictitious data, acute cumulative risk assessments can be performed following various calculation methods (EFSA 2012 Optimistic and Pessimistic, EC 2018 Tier 1 and Tier 2). Here, also the effect of applying processing factors can be assessed.

Table 4.3: Datasources for Demo acute cumulative risk assessment.

Table Group	Name	Repository	Type
MaximumResidue-Limits	DemoConcentrationLim-its.mdb	Standard Actions/Demo Acute Cumulative Risk Assessment	Fixed
Concentrations	DemoConcentrations.mdb	Standard Actions/Demo Acute Cumulative Risk Assessment	Fixed
Survey	DemoConsumptions.mdb	Standard Actions/Demo Acute Cumulative Risk Assessment	Fixed
Effects	DemoEffects.mdb	Standard Actions/Demo Acute Cumulative Risk Assessment	Fixed
FoodExtrapolations	DemoFoodExtrapola-tions.mdb	Standard Actions/Demo Acute Cumulative Risk Assessment	Fixed
FoodTranslations	DemoFoodRecipes.mdb	Standard Actions/Demo Acute Cumulative Risk Assessment	Fixed
Foods	DemoFoods.mdb	Standard Actions/Demo Acute Cumulative Risk Assessment	Fixed
HazardDoses	DemoPointsOfDepar-ture.mdb	Standard Actions/Demo Acute Cumulative Risk Assessment	Fixed
Processing	DemoProcessingFac-tors.mdb	Standard Actions/Demo Acute Cumulative Risk Assessment	Vari-able
Processing	DemoProcessingfac-torOri.mdb	Standard Actions/Demo Acute Cumulative Risk Assessment	Vari-able
AuthorisedUses	DemoSubstanceAuthorisa-tions.mdb	Standard Actions/Demo Acute Cumulative Risk Assessment	Fixed
ResidueDefinitions	DemoSubstanceConver-sions.mdb	Standard Actions/Demo Acute Cumulative Risk Assessment	Fixed
Compounds	DemoSubstances.mdb	Standard Actions/Demo Acute Cumulative Risk Assessment	Fixed
UnitVariability	DemoUnitVar36.mdb	Standard Actions/Demo Acute Cumulative Risk Assessment	Vari-able
UnitVariability	DemoUnitVarPRIMo.mdb	Standard Actions/Demo Acute Cumulative Risk Assessment	Vari-able

4.5 Acute single substance dietary exposure assessment of carbofuran or chlorpyrifos

This standard action is of type: *Dietary exposures*

This standard action allows you to perform probabilistic acute dietary exposure assessments of single substances in food. This standard actions contains examples for the two pesticides carbofuran and chlorpyrifos. The assessment is based on consumption data for two-years-old from the Dutch National Food Consumption Survey (DNFCS)-Young children (Ocké et al. (2008)). The consumption data are coded according to EFSAs coding system raw primary commodity (RPC) (EFSA (2019)). The 2014 concentration data of carbofuran and chlorpyrifos were extracted from EFSAs Data Warehouse and are organised according to the SSD data format. The standard action is developed for training purposes and results should not be regarded as representative outcomes of full dietary exposure assessments of carbofuran or chlorpyrifos.

In this standard action you can:

- Select the substance; carbofuran or chlorpyrifos,
- Choose the method for handling left-censored concentration measurements (lower bound or upper bound approach),
- Choose whether to include *correction for processing*,
- Choose whether to account for *unit-to-unit variation* using the Beta model,

- Run with or without uncertainty.

Table 4.4: Datasources for Acute single substance dietary exposure assessment of carbofuran or chlorpyrifos.

Table Group	Name	Repository	Type
Compounds	Catalogues_WHO_Training_Acute_NL2.xlsx	Standard Actions/Acute single substance dietary exposure assessment of carbofuran or chlorpyrifos	Fixed
Food-Translations	Catalogues_WHO_Training_Acute_NL2.xlsx	Standard Actions/Acute single substance dietary exposure assessment of carbofuran or chlorpyrifos	Fixed
Foods	Catalogues_WHO_Training_Acute_NL2.xlsx	Standard Actions/Acute single substance dietary exposure assessment of carbofuran or chlorpyrifos	Fixed
Populations	Catalogues_WHO_Training_Acute_NL2.xlsx	Standard Actions/Acute single substance dietary exposure assessment of carbofuran or chlorpyrifos	Fixed
Processing	Catalogues_WHO_Training_Acute_NL2.xlsx	Standard Actions/Acute single substance dietary exposure assessment of carbofuran or chlorpyrifos	Fixed
UnitVariability	Catalogues_WHO_Training_Acute_NL2.xlsx	Standard Actions/Acute single substance dietary exposure assessment of carbofuran or chlorpyrifos	Fixed
Concentrations	Concentrations_WHO_Carbofuran_Chlorpyrifos.xlsx	Standard Actions/Acute single substance dietary exposure assessment of carbofuran or chlorpyrifos	Fixed
Survey	Consumptions_NL_2yr_VCP-2002-2006_RPC.xlsx	Standard Actions/Acute single substance dietary exposure assessment of carbofuran or chlorpyrifos	Fixed

4.6 Chronic single substance dietary exposure assessment of lead or atropine

This standard action is of type: *Dietary exposures*

This standard action enables you to perform probabilistic chronic dietary exposure assessments of single substances in food (Boon et al. (2017), Boon et al. (2022)). The assessment is based on consumption data of women aged 15-44 years from the FAO-WHO Gift survey of Bangladesh, coded according to EFSA's FoodEx2 classification. The atropine concentration data were extracted from the GEMS database. The lead concentration data for foods are artificial, manually generated around the means reported by EFSA in 2012 (EFSA 2012). The concentration data are organised according the SSD data format. The standard action is developed for training purposes and results should not be seen as a full dietary exposure assessment of atropine or lead

This example allows you to:

- Select the substance atropine or lead,
- Choose the method for handling left-censored concentration measurements (lower bound or upper bound approach),
- Select a method for modelling long term intakes (OIM/LNN),
- Run with or without uncertainty

Table 4.5: Datasources for Chronic single substance dietary exposure assessment of lead or atropine.

Table Group	Name	Repository	Type
Compounds	Catalogues_WHO_Training_chronic_BD.xlsx	Standard Actions/Chronic single substance dietary exposure assessment lead or atropine	Fixed
Food-Translations	Catalogues_WHO_Training_chronic_BD.xlsx	Standard Actions/Chronic single substance dietary exposure assessment lead or atropine	Fixed
Foods	Catalogues_WHO_Training_chronic_BD.xlsx	Standard Actions/Chronic single substance dietary exposure assessment lead or atropine	Fixed
Populations	Catalogues_WHO_Training_chronic_BD.xlsx	Standard Actions/Chronic single substance dietary exposure assessment lead or atropine	Fixed
Concentrations	Concentrations_WHO_Training_Pb_At.xlsx	Standard Actions/Chronic single substance dietary exposure assessment lead or atropine	Fixed
Survey	Consumptions_BD_GIFT_2022.xlsx	Standard Actions/Chronic single substance dietary exposure assessment lead or atropine	Fixed

4.7 Demo Human Monitoring Analysis bisphenols

This standard action is of type: *Biological matrix concentration comparisons*

Human BioMonitoring (HBM) is a primary instrument to measure real-life exposure to chemicals. Because of the associated high costs, chemical levels in body fluids such as blood or urine would ideally be predictable from estimated external exposure levels, such as in the diet and/or from other non-dietary sources. It is needed to convert external exposures to internal concentrations by use of a kinetic (PBK) model or the application of a simple absorption factor.

This standard action provides a simple demonstration of this approach. It is based on the EuroMix bio-monitoring study (Husøy et al. (2019), Karrer et al. (2019)), and considers three of the investigated chemical substances in this study, i.e. the bisphenols BPA, BPS and BPF, which could have adverse estrogenic effects and therefore require risk assessment. These substances were measured in the urine of 144 adult individuals on two days (in this demo, we only use the data from day 1). Measurement values below the limit of detection were imputed. The specific gravity of the urine samples was measured as well, which is used in the calculation to calculate adjusted urine concentrations as decribed in Husøy et al. (Husøy et al. (2019)). The study subjects also kept detailed diaries on their food consumption, and use of personal care products. In this demonstrator standard action, the dietary exposure and optionaly also the non-dietary exposure from personal care products and thermal paper is modelled. The dietary exposure is predicted from the consumption data for 226 modelled foods derived from the food diaries and food monitoring concentration data for BPA, BPS and BPF in these foods. Additionally, non-dietary exposure from personal care products and from handling thermal paper was modelled separately by Karrer et al. (Karrer et al. (2020)) and these exposures are available as data for three non-dietary routes, i.e. dermal, oral and inhalation. In the demonstrator these exposures are aggregated with the dietary exposures at the individual level. In this demonstrator the conversion from external exposures to urine concentration is based on a simple absorption factor approach.

Note that this standard action only shows the principles of comparing modelled and measured biomonitoring levels. After conversion to a full MCRA action it will be possible to extend the example in various ways. The simple absorption factor could be replaced by a kinetic model. In fact, such a model (Karrer et al. (2018), Karrer et al. (2019)) is available in MCRA, but it needs further optimisation in the current implementation, e.g. in terms of running times. See Karrer et al. (Karrer et al. (2020)) for more information.

In this demonstrator standard action, several choices can be made:

- Acute or chronic exposure. For acute, HBM data are taken per day and modelled dietary exposures are based on Monte Carlo integration of consumptions and food concentrations. For chronic, HBM data are averaged per individual (not relevant in this standard action because only one sample per individual is available) and dietary exposures are modelled using the observed individual means model.
- Derive modelled concentrations from dietary exposur only or aggregate with non-dietary exposure sources.

- For the modelled concentrations, replace censored values in the food concentration data (measurements below the limit of reporting LOR) with 0 or with 0.5 x LOR.
- Choose from example data files with substance-specific absorption factors.

This standard action summarises data available from the EuroMix HBM study for bisphenols BPA, BPS and BPF in human urine. After imputing missing values and correcting for urine specific gravity, the resulting distribution is compared to the distribution of bisphenol concentrations predicted from dietary exposure and exposure from personal care products and thermal paper.

Table 4.6: Datasources for Demo Human Monitoring Analysis bisphenols.

Table Group	Name	Repository	Type
Compounds	Demo_HBM_CataloguesAndSecondaryStandards.xlsx	Standard Actions/Demo Human Monitoring Analysis	Fixed
Effects	Demo_HBM_CataloguesAndSecondaryStandards.xlsx	Standard Actions/Demo Human Monitoring Analysis	Fixed
FoodTranslations	Demo_HBM_CataloguesAndSecondaryStandards.xlsx	Standard Actions/Demo Human Monitoring Analysis	Fixed
Foods	Demo_HBM_CataloguesAndSecondaryStandards.xlsx	Standard Actions/Demo Human Monitoring Analysis	Fixed
RelativePotencyFactors	Demo_HBM_CataloguesAndSecondaryStandards.xlsx	Standard Actions/Demo Human Monitoring Analysis	Fixed
Concentrations	Demo_HBM_ConcentrationsSSD.xlsx	Standard Actions/Demo Human Monitoring Analysis	Fixed
Survey	Demo_HBM_Consumptions.xlsx	Standard Actions/Demo Human Monitoring Analysis	Fixed
KineticModels	Demo_HBM_KineticModelsBisphenols_Standard.xlsx	Standard Actions/Demo Human Monitoring Analysis	Variable
KineticModels	Demo_HBM_KineticModelsBisphenols_Standard2.xlsx	Standard Actions/Demo Human Monitoring Analysis	Variable
NonDietary	Demo_HBM_NonDietaryExposures.xlsx	Standard Actions/Demo Human Monitoring Analysis	Fixed
HumanMonitoringData	Demo_HBM_PhenolsUrinePooledOneDay.xlsx	Standard Actions/Demo Human Monitoring Analysis	Fixed

4.8 TDS-based long term dietary exposure and risk assessment

This standard action is of type: *Risks*

This standard action provides an overview about probabilistic long-term dietary exposure assessments with TDS data in MCRA. Possibilities to calculate exposure for population (sub-)groups from different countries and for different substances are provided. In addition, options for regional, seasonal or production type specific exposure are demonstrated along with options to select for consumers of different foods. In order to highlight options of the MCRA TDS module, the standard actions provide a simple and condensed overview about possible settings in MCRA. Consult the documentation pages on *TDS-based exposure and risk assessment* if you want to learn more about running these types of assessments in MCRA.

4.9 Long-term dietary exposure and risk of nickel for the Belgian population

This standard action is of type: *Risks*

The risk assessment originally conducted by the European Food Safety Authority stated some concerns regarding the chronic exposure of the European population to nickel due to food intake. This study aimed to evaluate the extent of the Belgian population's exposure to nickel via intake of different foods/drinks available in their market.

This standard actions is using nickel concentration data and Belgian consumption data-2014 data, for training and demonstration purposes and is meant to demonstrate the average nickel exposure for total Belgian consumers, for differences Belgian sub-populations and specifying the exposure for different foods (LB and UB).

This standard action demonstrates the aggregated exposure assessment of the effects of nickel occurrence in food of Belgian population. It uses 2017 Nickel contamination data in foods available in Belgian market and Belgian consumption data 2014.

Table 4.7: Datasources for Long-term dietary exposure and risk of nickel for the Belgian population.

Table Group	Name	Repository	Type
Survey	2021-09-17_BelgianConsumptions.xlsx	Standard Actions/Long-term dietary exposure and risk of nickel for Belgian age groups	Fixed
Compounds	Catalogues_FNS_Case_Study_BE_Ni_MCRA_v9.xlsx	Standard Actions/Long-term dietary exposure and risk of nickel for Belgian age groups	Fixed
Effects	Catalogues_FNS_Case_Study_BE_Ni_MCRA_v9.xlsx	Standard Actions/Long-term dietary exposure and risk of nickel for Belgian age groups	Fixed
Foods	Catalogues_FNS_Case_Study_BE_Ni_MCRA_v9.xlsx	Standard Actions/Long-term dietary exposure and risk of nickel for Belgian age groups	Fixed
HazardCharacterisations	Catalogues_FNS_Case_Study_BE_Ni_MCRA_v9.xlsx	Standard Actions/Long-term dietary exposure and risk of nickel for Belgian age groups	Fixed
Populations	Catalogues_FNS_Case_Study_BE_Ni_MCRA_v9.xlsx	Standard Actions/Long-term dietary exposure and risk of nickel for Belgian age groups	Fixed
Concentrations	Contaminants_BE_2017-2019_INNIBEL_Ni_MCRA_FE2.xlsx	Standard Actions/Long-term dietary exposure and risk of nickel for Belgian age groups	Fixed
FoodTranslations	FoodTranslations_INNIBEL_Ni.xlsx	Standard Actions/Long-term dietary exposure and risk of nickel for Belgian age groups	Fixed

4.10 TDS-based long-term exposure and risk assessment of methylmercury for German children

This standard action is of type: *Risks*

Note: This standard action was developed within the FNS-Cloud project (https://www.fns-cloud.eu). Related consumption data and example conversion scripts for uploading data will be made available in the FNS Cloud (status 05/2022: in progress).

4.10.1 Introduction

This standard action performs *a TDS-based probabilistic risk assessment* regarding the exposure of German children to methylmercury. Total Diet Study (TDS) data are combined with food consumption data, and stratification of the calculations regarding region, season and production type (organic or conventional) can be considered. Total Diet Studies (TDS) are an appropriate method to representatively reflect background occurrence levels of substances in food in a population's diet. More than 90 % of the diet are covered, the foods are prepared as consumed and pooled to composite samples prior to chemical analysis. Next to the representation of almost the total diet in Germany, the German TDS (BfR MEAL Study) also allows to elaborate potential differences in occurrence of substances in different stratifications. The stratifications consider four different regions (North, East, South and West Germany), two different seasons (which are in this demo fictitious summer and winter season), and the production type (i.e. organic versus conventional production).

In this example, consumption data for children in Germany and fictive TDS occurrence data for methylmercury are used. The commonly used lower bound (LB) or upper bound (UB) scenarios are used for treatment of censored data. Exposure is calculated applying the Observed Individual Means model. The exposure is expressed as µg per kg body weight and day for the mean, and the 50th, 90th, 95th, 97.5th, 99th and 99.9th percentiles of the exposure distribution and the relative contributions from modelled foods or main food groups to total exposure is displayed. For risk assessment, the exposure distribution is compared to the tolerable weekly intake (TWI) of 1.3 µg per kg body weight and day set for methylmercury set by EFSA. The risk in indicated by the Hazard Index (HI) quantified by exposure relative to the TWI.

4.10.2 Standard action options

In this standard action, potential differences between stratification related exposure assessments are demonstrated. The user is able to select and combine the above-mentioned stratifications and compare exposure outcomes and risks.

- By selection of a certain region the occurrence data from this region will be combined with the respective consumptions data from children living in this region.
- By selection of a certain season the occurrence data from this season will be combined with consumption events during this season. Summer season is defined from April to September and Winter season from October till March.
- By selection of a type of production occurrence data from organically or conventionally produced foods will be combined with the overall consumption, since information on consumption of organically or conventionally produced foods is not included in the consumption data.
- Exposure from foods that are not stratified according to the above-mentioned characteristics is included on an unspecific basis into the assessments (e.g., the item "pasta" is not sampled in different regions, therefore regional exposure will be calculated by combining regional consumption with the concentration of the unspecified sample).
- The assessments can be done either for all individuals included in the consumption survey or can be restricted to consumers only. The restriction will consider consumers of one or more selected TDS (modelled) foods but includes all consumptions of these individuals.
- The option to calculate uncertainties analysis results in 95 % confidence intervals (CIs) by resampling the individuals.

4.10.3 Standard action data

Table 4.8: Datasources for TDS-based long-term exposure and risk assessment of methylmercury for German children.

Table Group	Name	Repository	Type
Compounds	Catalogues_FNS_DE_MeHg	Standard Actions/TDS-based long-term exposure and risk assessment of methylmercury for German children	Fixed
Effects	Catalogues_FNS_DE_MeHg	Standard Actions/TDS-based long-term exposure and risk assessment of methylmercury for German children	Fixed
Foods	Catalogues_FNS_DE_MeHg	Standard Actions/TDS-based long-term exposure and risk assessment of methylmercury for German children	Fixed
HazardCharacterisations	Catalogues_FNS_DE_MeHg	Standard Actions/TDS-based long-term exposure and risk assessment of methylmercury for German children	Fixed
TotalDietStudy	Catalogues_FNS_DE_MeHg	Standard Actions/TDS-based long-term exposure and risk assessment of methylmercury for German children	Fixed
Concentrations	Concentrations_FNS_DE_MeHg_SSD	Standard Actions/TDS-based long-term exposure and risk assessment of methylmercury for German children	Fixed
Survey	Consumptions_DE_VELS_2001-2002	Standard Actions/TDS-based long-term exposure and risk assessment of methylmercury for German children	Fixed

4.11 Acute Cumulative Risk Assessment Craniofacial Alterations (EFSA 2022)

This standard action is of type: *Single value risks*

This standard action will enable you to run acute cumulative dietary risk assessment of craniofacial alterations by pesticide residues using the same methods and data as used by EFSA in 2022.

Table 4.9: Datasources for Acute Cumulative Risk Assessment Craniofacial Alterations (EFSA 2022).

Table Group	Name	Repository
Concentrations	ConcentrationsSSD_DAC.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani
Concentrations	ConcentrationsSSD_DAH.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani
Survey	ConsumptionsBE.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani
Survey	ConsumptionsCZ.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani
Survey	ConsumptionsDE.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani
Survey	ConsumptionsDK.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani
Survey	ConsumptionsES.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani
Survey	ConsumptionsFI.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani
Survey	ConsumptionsFR.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani
Survey	ConsumptionsHU.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani
Survey	ConsumptionsIE.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani
Survey	ConsumptionsIT.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani
Survey	ConsumptionsLT.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani
Survey	ConsumptionsNL.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani
Survey	ConsumptionsRO.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani
Survey	ConsumptionsSE.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani
AssessmentGroupMemberships	SecondaryInputDataFPA3-v3.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani
AuthorisedUses	SecondaryInputDataFPA3-v3.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani
Compounds	SecondaryInputDataFPA3-v3.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani
Effects	SecondaryInputDataFPA3-v3.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani

Table 4.9 – continued from previous page

Table Group	Name	Repository
FoodTranslations	SecondaryInputDataFPA3-v3.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani
Foods	SecondaryInputDataFPA3-v3.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani
HazardDoses	SecondaryInputDataFPA3-v3.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani
MaximumResidueLimits	SecondaryInputDataFPA3-v3.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani
Populations	SecondaryInputDataFPA3-v3.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani
Processing	SecondaryInputDataFPA3-v3.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani
ResidueDefinitions	SecondaryInputDataFPA3-v3.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani
SubstanceApprovals	SecondaryInputDataFPA3-v3.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani
UnitVariability	UnitVar36.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani
UnitVariability	UnitVarPrimo.mdb	Standard Actions/Acute Cumulative Risk Assessment Crani

4.12 EU acute cumulative exposure assessment (2018) Tier I and Tier II

This standard action is of type: *Single value risks*

This standard action is based on work done in 2018 (van Klaveren et al. (2019a)). In the context of the second framework partnership agreement between the National Institute for Public Health and the Environment of the Netherlands (RIVM) and the European Food Safety Authority (EFSA) acute cumulative dietary exposure assessments were performed for two cumulative assessment groups (CAGs) of pesticides that affect the nervous system: pesticides causing brain and/or erythrocyte AChE inhibition (CAG-NAN, 47 pesticides) and pesticides causing functional alterations of the motor division (CAG-NAM, 100 pesticides). The exposure assessments used monitoring data collected by the Netherlands under their official monitoring programmes in 2014, 2015 and 2016 and individual Dutch food consumption data. Exposure estimates were obtained for each group of pesticides using the MCRA software. The Standing Committee on Plants, Animals, Food and Feed (SC PAFF) discussed the scope of the assessment in 2018 and agreed on the parameters to be used for the cumulative exposure assessment. Based on that discussion, a very conservative tier I modelling approach and a refined, but still conservative tier II modelling approach were used. In these assessments, common risk assessment practice was followed and the cumulative exposure was expressed as the total margin of exposure (MOET) at the 50th, 90th, 95th, 99th and 99.9th percentile of the exposure distribution.

Table 4.10: Datasources for EU acute cumulative exposure assessment (2018) Tier I and Tier II.

Table Group	Name	Repository	Type
AssessmentGroupMemberships	LeanCAGsAssessmentGroupMembership10%p999.xlsx	Standard Actions/EU 2018 Acute Cumulative Exposure Assessment	Fixed
AuthorisedUses	SecondaryInputData.mdb	Standard Actions/EU 2018 Acute Cumulative Exposure Assessment	Fixed
Compounds	SecondaryInputData.mdb	Standard Actions/EU 2018 Acute Cumulative Exposure Assessment	Fixed
Effects	SecondaryInputData.mdb	Standard Actions/EU 2018 Acute Cumulative Exposure Assessment	Fixed
FoodExtrapolations	SecondaryInputData.mdb	Standard Actions/EU 2018 Acute Cumulative Exposure Assessment	Fixed
FoodTranslations	SecondaryInputData.mdb	Standard Actions/EU 2018 Acute Cumulative Exposure Assessment	Fixed
Foods	SecondaryInputData.mdb	Standard Actions/EU 2018 Acute Cumulative Exposure Assessment	Fixed
HazardDoses	SecondaryInputData.mdb	Standard Actions/EU 2018 Acute Cumulative Exposure Assessment	Fixed
MaximumResidueLimits	SecondaryInputData.mdb	Standard Actions/EU 2018 Acute Cumulative Exposure Assessment	Fixed
Processing	SecondaryInputData.mdb	Standard Actions/EU 2018 Acute Cumulative Exposure Assessment	Fixed
ResidueDefinitions	SecondaryInputData.mdb	Standard Actions/EU 2018 Acute Cumulative Exposure Assessment	Fixed
UnitVariability	UnitVar36.mdb	Standard Actions/EU 2018 Acute Cumulative Exposure Assessment	Variable
UnitVariability	UnitVarPrimo.mdb	Standard Actions/EU 2018 Acute Cumulative Exposure Assessment	Variable
Concentrations	a_ConcentrationsSSD_NAM.mdb	Standard Actions/EU 2018 Acute Cumulative Exposure Assessment	Variable
Concentrations	a_ConcentrationsSSD_NAN.mdb	Standard Actions/EU 2018 Acute Cumulative Exposure Assessment	Variable
Survey	a_ConsumptionsNL2.mdb	Standard Actions/EU 2018 Acute Cumulative Exposure Assessment	Fixed
Survey	a_ConsumptionsNL3_6.mdb	Standard Actions/EU 2018 Acute Cumulative Exposure Assessment	Fixed

4.13 EU chronic cumulative exposure assessment (2018) Tier I and Tier II

This standard action is of type: *Risks*

This standard action is based on research done in 2018 van Klaveren et al. (2019b).

This standard action will enable you to reproduce the exposure assessment of chronic cumulative effects of pesticide residues in food affecting the thyroid. These are retrospective exposure assessments of the cumulative exposure for the thyroid using monitoring data from 2014, 2015 and 2016. In this standard action Dutch monitoring and consumption data are used. The results, data used and methodology are reported in a scientific report following published on the EFSA website in September 2019. The methodology fulfils the requirements set by the European Commission.

Table 4.11: Datasources for EU chronic cumulative exposure assessment (2018) Tier I and Tier II.

Table Group	Name	Repository	Type
AssessmentGroupMemberships	LeanCAGsAssessmentGroupMembership10%p999.xlsx	Standard Actions/EU 2018 Chronic Cumulative Exposure Assessment	Fixed
AuthorisedUses	SecondaryInputData.mdb	Standard Actions/EU 2018 Chronic Cumulative Exposure Assessment	Fixed
Compounds	SecondaryInputData.mdb	Standard Actions/EU 2018 Chronic Cumulative Exposure Assessment	Fixed
Effects	SecondaryInputData.mdb	Standard Actions/EU 2018 Chronic Cumulative Exposure Assessment	Fixed
FoodExtrapolations	SecondaryInputData.mdb	Standard Actions/EU 2018 Chronic Cumulative Exposure Assessment	Fixed
FoodTranslations	SecondaryInputData.mdb	Standard Actions/EU 2018 Chronic Cumulative Exposure Assessment	Fixed
Foods	SecondaryInputData.mdb	Standard Actions/EU 2018 Chronic Cumulative Exposure Assessment	Fixed
HazardDoses	SecondaryInputData.mdb	Standard Actions/EU 2018 Chronic Cumulative Exposure Assessment	Fixed
MaximumResidueLimits	SecondaryInputData.mdb	Standard Actions/EU 2018 Chronic Cumulative Exposure Assessment	Fixed
Processing	SecondaryInputData.mdb	Standard Actions/EU 2018 Chronic Cumulative Exposure Assessment	Fixed
ResidueDefinitions	SecondaryInputData.mdb	Standard Actions/EU 2018 Chronic Cumulative Exposure Assessment	Fixed
Concentrations	c_ConcentrationsSSD_TCF.mdb	Standard Actions/EU 2018 Chronic Cumulative Exposure Assessment	Variable
Concentrations	c_ConcentrationsSSD_TCP.mdb	Standard Actions/EU 2018 Chronic Cumulative Exposure Assessment	Variable
Survey	c_ConsumptionsNL2.mdb	Standard Actions/EU 2018 Chronic Cumulative Exposure Assessment	Fixed
Survey	c_ConsumptionsNL3_6.mdb	Standard Actions/EU 2018 Chronic Cumulative Exposure Assessment	Fixed

4.14 Risk steatosis from imazalil

This standard action is of type: *Risks*

Traditional risk assessment often uses animal data to evaluate toxicological limit values, such as the acceptable daily intake (ADI). However, in-vitro tests may help to reduce the use of test animals. In the standard action described here such a traditional risk calculation can be compared to a similar calculation based on in-vitro data. For example, a risk calculation of steatosis as health effect and imazalil as chemical substance can be compared to the toxicological hazard characterisation based on in-vitro data from AdipoRed measurements in human liver cells. In-vitro concentrations are assumed to represent internal liver concentrations. A human physiologically based kinetic model for imazalil is used to extrapolate from in-vitro to in-vivo doses.

In this standard action, risk for steatosis from exposure to imazalil is estimated from in-vivo or in-vitro based hazard characterisations.

Two parameter sets are available:

1. estimates based on QSAR models only, and
2. estimates based on QSAR models and in-vitro experiments.

The latter parameter set shows that in the long term concentration levels of imazalil in the liver are stationary.

Table 4.12: Datasources for Risk steatosis from imazalil.

Table Group	Name	Repository	Type
AdverseOutcomePathwayNetworks	AOPN-Effects-EffectRelations-181017.xlsx	Standard Actions/Risk steatosis from imazalil	Fixed
Effects	AOPN-Effects-EffectRelations-181017.xlsx	Standard Actions/Risk steatosis from imazalil	Fixed
DoseResponseData	BfR-HepaRG-AdipoRed-Single.xlsx	Standard Actions/Risk steatosis from imazalil	Fixed
Concentrations	ConcentrationsSSD_20190129.zip	Standard Actions/Risk steatosis from imazalil	Fixed
Compounds	EuroMix Substances Inventory (v8) (PPPs).zip	Standard Actions/Risk steatosis from imazalil	Fixed
KineticModels	EuroMix-Cosmos_Tebby et al ParamA_QSAR.xlsx	Standard Actions/Risk steatosis from imazalil	Variable
KineticModels	Euromix-Cosmos_Tebby et al ParamB_QSAR_vitro.xlsx	Standard Actions/Risk steatosis from imazalil	Variable
UnitVariability	Foods coded in FoodEx1.mdb	Standard Actions/Risk steatosis from imazalil	Fixed
HazardCharacterisations	HazardCharacterisation Imazalil.xlsx	Standard Actions/Risk steatosis from imazalil	Fixed
Survey	NL-VCP-RPC 2005-2006 2-6yr.mdb	Standard Actions/Risk steatosis from imazalil	Fixed
Survey	NL-VCP-RPC 2007-2010 7-69yr.mdb	Standard Actions/Risk steatosis from imazalil	Fixed
Survey	NL-VCP-RPC 2010-2012 70+yr.mdb	Standard Actions/Risk steatosis from imazalil	Fixed
FoodTranslations	NL-VCP-RPC Foods.mdb	Standard Actions/Risk steatosis from imazalil	Fixed
Foods	NL-VCP-RPC Foods.mdb	Standard Actions/Risk steatosis from imazalil	Fixed
EffectRepresentations	TestSystems-Responses-EffectRepresentations-181026.xlsx	Standard Actions/Risk steatosis from imazalil	Fixed
Responses	TestSystems-Responses-EffectRepresentations-181026.xlsx	Standard Actions/Risk steatosis from imazalil	Fixed
TestSystems	TestSystems-Responses-EffectRepresentations-181026.xlsx	Standard Actions/Risk steatosis from imazalil	Fixed

4.15 Training prospective risk assessment acute Tier II

This standard action is of type: *Single value risks*

This standard action allows you to run and compare the background and different Tier II approaches to probabilistic prospective risk assessment for a newly proposed use (focal substance/food) as defined by the EC working group. This is an acute risk assessment of a neurological effect (functional alternations of the motor division) for the Dutch toddlers, the Dutch other children, the Bulgarian other children and Italian adults populations. The focal food and substance combination that can be selected are emamectin with peach, acrinathrin with wheat (fictitious data) or pirimicarb with lettuces (fictitious data). The focal exposure can be based on a MRL, GAP or actual exposure scenario as defined by the working group, or it can be omitted (for comparison) so that only the background exposure is calculated. The 99.9th percentile of the exposure has been adjusted using uncertain adjustment factors from the EFSA cumulative risk assessment report 2020. The exposure is calculated probabilistically and the risk is expressed as margin of exposure (MOE) and is compared to a threshold value of 100.

Table 4.13: Datasources for Training prospective risk assessment acute Tier II.

Table Group	Name	Repository	Type
FocalFoods	FocalConcentrations-ActualScenario.mdb	Standard Actions/Training Acute Prospective Risk Assessment Tier II	Variable
FocalFoods	FocalConcentrations-GapScenario.mdb	Standard Actions/Training Acute Prospective Risk Assessment Tier II	Variable
FocalFoods	FocalConcentrations-MrlScenario.mdb	Standard Actions/Training Acute Prospective Risk Assessment Tier II	Variable
AssessmentGroupMemberships	SecondaryInput-DataMRL2016.mdb	Standard Actions/Training Acute Prospective Risk Assessment Tier II	Fixed
AuthorisedUses	SecondaryInput-DataMRL2016.mdb	Standard Actions/Training Acute Prospective Risk Assessment Tier II	Fixed
Compounds	SecondaryInput-DataMRL2016.mdb	Standard Actions/Training Acute Prospective Risk Assessment Tier II	Fixed
Effects	SecondaryInput-DataMRL2016.mdb	Standard Actions/Training Acute Prospective Risk Assessment Tier II	Fixed
FoodExtrapolations	SecondaryInput-DataMRL2016.mdb	Standard Actions/Training Acute Prospective Risk Assessment Tier II	Fixed
FoodTranslations	SecondaryInput-DataMRL2016.mdb	Standard Actions/Training Acute Prospective Risk Assessment Tier II	Fixed
Foods	SecondaryInput-DataMRL2016.mdb	Standard Actions/Training Acute Prospective Risk Assessment Tier II	Fixed
HazardDoses	SecondaryInput-DataMRL2016.mdb	Standard Actions/Training Acute Prospective Risk Assessment Tier II	Fixed
MaximumResidueLimits	SecondaryInput-DataMRL2016.mdb	Standard Actions/Training Acute Prospective Risk Assessment Tier II	Variable
Processing	SecondaryInput-DataMRL2016.mdb	Standard Actions/Training Acute Prospective Risk Assessment Tier II	Fixed
ResidueDefinitions	SecondaryInput-DataMRL2016.mdb	Standard Actions/Training Acute Prospective Risk Assessment Tier II	Fixed
MaximumResidueLimits	SecondaryInput-DataMRL2019.mdb	Standard Actions/Training Acute Prospective Risk Assessment Tier II	Variable
DeterministicSubstanceConversionFactors	SingleValueCalculations.mdb	Standard Actions/Training Acute Prospective Risk Assessment Tier II	Fixed
UnitVariability	UnitVar36.mdb	Standard Actions/Training Acute Prospective Risk Assessment Tier II	Fixed
Concentrations	a_ConcentrationsSSD_NAM.mdb	Standard Actions/Training Acute Prospective Risk Assessment Tier II	Fixed
Survey	a_ConsumptionsBU.mdb	Standard Actions/Training Acute Prospective Risk Assessment Tier II	Fixed
Survey	a_ConsumptionsIT.mdb	Standard Actions/Training Acute Prospective Risk Assessment Tier II	Fixed
Survey	a_ConsumptionsNL2.mdb	Standard Actions/Training Acute Prospective Risk Assessment Tier II	Fixed
Survey	a_ConsumptionsNL3_6.mdb	Standard Actions/Training Acute Prospective Risk Assessment Tier II	Fixed

4.16 Training prospective risk assessment chronic Tier II

This standard action is of type: *Single value risks*

This standard action allows you to run and compare the background and different Tier II approaches to probabilistic chronic prospective risk assessment for a newly proposed use (focal substance/food) as defined by the EC working group. This is a chronic risk assessment of an effect (hypothyroidism) for the Dutch toddlers, Dutch other children and German adults populations. The focal food and substance combination that can be selected are cyprodinil with apples (fictitious data), valifenalate with lettuce or valifenalate with aubergines. The focal exposure can be based on a GAP or actual exposure scenario as defined by the working group, or it can be omitted (for comparison) so that only the background exposure is calculated. The 99.9th percentile of the exposure has been adjusted using uncertain adjustment factors from the EFSA cumulative risk assessment report 2020. The exposure is calculated probabilistically and. the risk is expressed as margin of exposure (MOE) and is compared to a threshold value of 100.

Table 4.14: Datasources for Training prospective risk assessment chronic Tier II.

Table Group	Name	Repository	Type
FocalFoods	FocalConcentrations-ActualScenario.mdb	Standard Actions/Training Chronic Prospective Risk Assessment Tier II	Variable
FocalFoods	FocalConcentrations-GAPScenario.mdb	Standard Actions/Training Chronic Prospective Risk Assessment Tier II	Fixed
FocalFoods	FocalConcentrations-GapScenario.mdb	Standard Actions/Training Chronic Prospective Risk Assessment Tier II	Variable
AssessmentGroupMemberships	SecondaryInput-DataMRL2016.mdb	Standard Actions/Training Chronic Prospective Risk Assessment Tier II	Fixed
AuthorisedUses	SecondaryInput-DataMRL2016.mdb	Standard Actions/Training Chronic Prospective Risk Assessment Tier II	Fixed
Compounds	SecondaryInput-DataMRL2016.mdb	Standard Actions/Training Chronic Prospective Risk Assessment Tier II	Fixed
Effects	SecondaryInput-DataMRL2016.mdb	Standard Actions/Training Chronic Prospective Risk Assessment Tier II	Fixed
FoodExtrapolations	SecondaryInput-DataMRL2016.mdb	Standard Actions/Training Chronic Prospective Risk Assessment Tier II	Fixed
FoodTranslations	SecondaryInput-DataMRL2016.mdb	Standard Actions/Training Chronic Prospective Risk Assessment Tier II	Fixed
Foods	SecondaryInput-DataMRL2016.mdb	Standard Actions/Training Chronic Prospective Risk Assessment Tier II	Fixed
HazardDoses	SecondaryInput-DataMRL2016.mdb	Standard Actions/Training Chronic Prospective Risk Assessment Tier II	Fixed
MaximumResidueLimits	SecondaryInput-DataMRL2016.mdb	Standard Actions/Training Chronic Prospective Risk Assessment Tier II	Variable
Processing	SecondaryInput-DataMRL2016.mdb	Standard Actions/Training Chronic Prospective Risk Assessment Tier II	Fixed
ResidueDefinitions	SecondaryInput-DataMRL2016.mdb	Standard Actions/Training Chronic Prospective Risk Assessment Tier II	Fixed
MaximumResidueLimits	SecondaryInput-DataMRL2019.mdb	Standard Actions/Training Chronic Prospective Risk Assessment Tier II	Variable
DeterministicSubstanceConversionFactors	SingleValueCalculations.mdb	Standard Actions/Training Chronic Prospective Risk Assessment Tier II	Fixed
Concentrations	c_ConcentrationsSSD_TCF.mdb	Standard Actions/Training Chronic Prospective Risk Assessment Tier II	Fixed
Survey	c_ConsumptionsDE.mdb	Standard Actions/Training Chronic Prospective Risk Assessment Tier II	Fixed
Survey	c_ConsumptionsNL2.mdb	Standard Actions/Training Chronic Prospective Risk Assessment Tier II	Fixed
Survey	c_ConsumptionsNL3_6.mdb	Standard Actions/Training Chronic Prospective Risk Assessment Tier II	Fixed

4.17 Training substance prioritisation acute neuro

This standard action is of type: *Risks*

This standard action allows to prioritise substances to be included in a cumulative risk assessment by inspecting a list of (semi-)probabilistic hazard quotients, ranked from high to low. This is a risk assessment for either all neurotoxicological effects in the Dutch toddler population (level 1), or one of the effect subgroups (NAM or NAN) (level 2). The hazard quotient (termed hazard index in MCRA) compares a percentile from the dietary exposure distribution (P99 or P99.9 can be selected) to the ARfD, which a fixed value for the hazard characterisation (hence

the designation as semi-probabilistic).

Table 4.15: Datasources for Training substance prioritisation acute neuro.

Table Group	Name	Repository	Type
AssessmentGroupMemberships	ActiveSubstances.xlsx	Standard Actions/Training Substance Prioritisation Acute Neuro	Fixed
MaximumResidueLimits	ConcentrationLimits.xlsx	Standard Actions/Training Substance Prioritisation Acute Neuro	Fixed
Effects	Effects.xlsx	Standard Actions/Training Substance Prioritisation Acute Neuro	Fixed
FoodExtrapolations	FoodExtrapolations.xlsx	Standard Actions/Training Substance Prioritisation Acute Neuro	Fixed
FoodTranslations	FoodRecipes.xlsx	Standard Actions/Training Substance Prioritisation Acute Neuro	Fixed
Foods	Foods.xlsx	Standard Actions/Training Substance Prioritisation Acute Neuro	Fixed
HazardCharacterisations	HazardCharacterisations.xlsx	Standard Actions/Training Substance Prioritisation Acute Neuro	Fixed
Processing	ProcessingFactors.xlsx	Standard Actions/Training Substance Prioritisation Acute Neuro	Fixed
AuthorisedUses	SubstanceAuthorisations.xlsx	Standard Actions/Training Substance Prioritisation Acute Neuro	Fixed
ResidueDefinitions	SubstanceConversions.xlsx	Standard Actions/Training Substance Prioritisation Acute Neuro	Fixed
Compounds	Substances.xlsx	Standard Actions/Training Substance Prioritisation Acute Neuro	Fixed
UnitVariability	UnitVarPrimo.xlsx	Standard Actions/Training Substance Prioritisation Acute Neuro	Fixed
Concentrations	a_ConcentrationsSSD_Neurotadd	Standard Actions/Training Substance Prioritisation Acute Neuro	Fixed
Survey	a_ConsumptionsNL2.mdb	Standard Actions/Training Substance Prioritisation Acute Neuro	Fixed
Survey	a_ConsumptionsNL3_6.mdb	Standard Actions/Training Substance Prioritisation Acute Neuro	Fixed

- *Chronic mixture risk assessment of metals relevant for chronic kidney disease (nephrotoxicity)*
- *Chronic cumulative exposure assessment PA*
- *Chronic cumulative exposure assessment PFAS*
- *Demo acute cumulative risk assessment*
- *Acute single substance dietary exposure assessment of carbofuran or chlorpyrifos*
- *Chronic single substance dietary exposure assessment of lead or atropine*
- *Demo Human Monitoring Analysis bisphenols*
- *TDS-based long term dietary exposure and risk assessment*
- *Long-term dietary exposure and risk of nickel for the Belgian population*
- *TDS-based long-term exposure and risk assessment of methylmercury for German children*
- *Acute Cumulative Risk Assessment Craniofacial Alterations (EFSA 2022)*
- *EU acute cumulative exposure assessment (2018) Tier I and Tier II*
- *EU chronic cumulative exposure assessment (2018) Tier I and Tier II*
- *Risk steatosis from imazalil*
- *Training prospective risk assessment acute Tier II*

- *Training prospective risk assessment chronic Tier II*
- *Training substance prioritisation acute neuro*

CHAPTER

FIVE

TYPE AND UNIT DEFINITIONS

5.1 Adjustment factor distribution method types

Accepted justment factor distribution method types. Controlled terminology.

Table 5.1: Unit definition for Adjustment factor distribution method types.

Name	Short name	Aliases	Description
No adjustment factor	None		No adjustment factor.
Fixed adjustment factor	Fixed		Fixed adjustment factor.
LogNormal	LogNormal		Lognormal distribution with parameters a and b and offset c (default c = 0).
LogStudents_t	LogStudents_t		Log Students-t distribution with parameters a, b and c and offset d (default d = 0).
Beta	Beta		Beta distribution with shape parameters a and b on interval [c, d], (default = 0, 1).
Gamma	Gamma		Gamma distribution with shape parameter a and rate parameter b with offset = c (default = 0).

5.2 Assessment group membership calculation methods

Accepted Assessment group membership calculation methods. Controlled terminology.

Table 5.2: Unit definition for Assessment group membership calculation methods.

Name	Short name	Aliases	Description
Any (crisp)	Any (crisp)		Assign the highest membership value as membership. For crisp memberships, assign positive substance membership if any model indicates positive membership, and negative membership otherwise.
Majority (crisp)	Majority (crisp)		Assign positive substance membership if the majority of the membership models indicates positive membership, otherwise, the substance is considered not to be in the assessment group.
Ratio (probabilistic)	Ratio (probabilistic)		Express substance membership as a probability ranging from zero (certainly out) to one (certainly in), computed as the average membership score.
Bayesian (probabilistic)	Bayesian (probabilistic)		Express substance memberships as a probability with values ranging from zero (certainly out) to one (certainly in) computed using a Bayesian approach.

5.3 Benchmark response type

Accepted benchmark response types. Controlled terminology.

Table 5.3: Unit definition for Benchmark response type.

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.

5.4 Biological organisation type

Accepted biological organisation types. Controlled terminology.

Table 5.4: Unit definition for Biological organisation type.

Name	Short name	Aliases	Description
Molecular	Molecular	Molecular	Molecular level
Cellular	Cellular	Cellular	Cellular level
Tissue	Tissue	Tissue	Tissue level
Organ	Organ	Organelle, Organ	Organ level
Individual	Individual	Individual	Whole body level
Population	Population	Population	Population level

5.5 Body weight unit

Accepted units for person body weights. Controlled terminology.

Table 5.5: Unit definition for Body weight unit.

Name	Short name	Aliases
kilogram	kg	kg, kilograms, kilogr, 3, G167A
gram	g	g, grams, gr, 0, G148A

5.6 Boolean type

Accepted boolean types. Controlled terminology.

Table 5.6: Unit definition for Boolean type.

Name	Short name	Aliases
True	True	True, Yes, T, Y
False	False	False, F, No, N

5.7 Cluster method type

Accepted cluster method types. Controlled terminology.

Table 5.7: Unit definition for Cluster method type.

Name	Short name	Aliases	Description
Component selection (SNMU)	NoClustering		Only component selection is performed.
Component selection and population subgrouping (SNMU + k-means clustering)	Kmeans		Component selection followed by K-Means clustering of individuals based on their component exposure. K-means classifies individuals in multiple groups (i.e., clusters), such that individuals within the same cluster are as similar as possible (i.e., high intra-class similarity), whereas individuals from different clusters are as dissimilar as possible (i.e., low inter-class similarity). In k-means clustering, each cluster is represented by its center (i.e, centroid) which corresponds to the mean of points assigned to the cluster.
Component selection and population subgrouping (SNMU + hierarchical clustering)	Hierarchical		Component selection followed by hierarchical (Ward's) clustering of individuals based on their component exposure. Hierachical clustering builds a hierarchy from the bottom-up, and doesn't require to specify the number of clusters beforehand. Hierarchical clustering produces a tree-based representation of the observations known as a dendrogram.

5.8 Combination method membership info and PoD presence types

Accepted Combination method membership info and PoD presence types. Controlled terminology.

Table 5.8: Unit definition for Combination method membership info and PoD presence types.

Name	Short name	Aliases	Description
Consider active if POD/HC AND (in-silico) memberships indicate active	Intersection		Consider active if POD/HC AND (in-silico) memberships indicate active.
Consider active if POD/HC OR (in-silico) memberships indicates active	Union		Consider active if POD/HC OR (in-silico) memberships indicates active.

5.9 Concentration limit value type

Accepted concentration limit value types. Controlled terminology.

Table 5.9: Unit definition for Concentration limit value type.

Name	Short name	Aliases
Maximum residue limit	MRL	MRL, MaximumResidueLimit
Proposed maximum residue limit	Proposed-MRL	ProposedMrl, ProposedMaximumResidueLimit

5.10 Concentration model choice types

Accepted Concentration model choice types. Controlled terminology.

Table 5.10: Unit definition for Concentration model choice types.

Name	Short name	Aliases	Description
Custom	Custom		By setting this tier to custom, the exposure model can be configured in any way desirable (without tier specific presets). The model is fully specified by the user. Both EFSA 2012 Optimistic and EFSA 2012 Pessimistic can be specified using the custom model choice. This choice allows for a sensitivity analysis where each factor is varied, one at the time.
EFSA 2012 Optimistic	EfsaOptimistic		EFSA 2012 Optimistic.
EFSA 2012 Pessimistic	EfsaPessimistic		EFSA 2012 Pessimistic.
EC 2018 Tier 1	ComTier1		EC 2018 Tier 1.
EC 2018 Tier 2	ComTier2		EC 2018 Tier 2.
EFSA 2012 Pessimistic - Acute	EfsaPessimisticAcute		EFSA 2012 Pessimistic - Acute.
EFSA 2012 Pessimistic - Chronic	EfsaPessimistic-Chronic		EFSA 2012 Pessimistic - Chronic.

5.11 Concentration model types

Accepted Concentration model types. Controlled terminology.

Table 5.11: Unit definition for Concentration model types.

Name	Short name	Aliases	Description
Empirical	Empirical		Residues are sampled from the empirical distribution. Fallback: zero.
Censored value Spike LogNormal	CVSpike-LogN		A lognormal model (logarithmic transformed values, with parameters mu and sigma^2) is fitted to the positive residues values. LOR information is not used. Fallback (if number of positives less than 2): Empirical, but Maximum Residu Limit for pessimistic assessments.
Censored Spike Truncated LogNormal	CVSpike-TruncLogN		A truncated lognormal model (with parameters mu and sigma^2) is fitted to the positive residues values. The LOR is used to estimate the truncated left tail of the distribution. Fallback: Lognormal non-detect spike.
Censored LogNormal	CensLogN		Advanced. A censored lognormal model (with parameters mu and sigma^2) is fitted to the censored and positives residue values. Note, this model is not available when agricultural use information is used. Fallback: Lognormal non-detect spike.
Zero Spike Censored LogNormal	ZeroSpike-CensLogN		Advanced. A mixture model with zero spike (p0) and censored lognormal model (with parameters mu and sigma^2) is fitted to the censored and positives residue values. Note, this model is not available when agricultural use information is used. Fallback: Censored lognormal.
Censored Spike Maximum Residue Limit	CVSpike-MRL		Censored Spike Maximum Residue Limit.
Summary statistics	Summary statistic		Summary statistics.
LogNormal	LogN		Lognormal model.

5.12 Concentration unit

Accepted units for substance concentrations. Controlled terminology.

Table 5.12: Unit definition for Concentration unit.

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
microgram/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
microgram/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
milligram/deciliter	mg/dL	mg/dl, milligram/deciliter, milligr/dL, milligr/dl

5.13 Concentration value type

Accepted concentration value type. Controlled terminology.

Table 5.13: Unit definition for Concentration value type.

Name	Short name	Aliases	Description
Mean concentration	MC	MeanConcentration, ConcentrationMean, MC	Mean value from the residue trials.
Median concentration	MR	MedianConcentration, MR, STMR, SupervisedTrialMedianResidue	Median concentration / residue value of the positive measurements of the residue trials.
Highest concentration	HR	HighestConcentration, HighestResidue, HR	Highest measured residue / concentration value.
Concentration percentile	CP	Percentile	
Limit of quantification	LOQ	LOQ	
Maximum residue limit	MRL	MRL	

5.14 Concentrations tier types

Accepted Concentrations tier types. Controlled terminology.

Table 5.14: Unit definition for Concentrations tier types.

Name	Short name	Aliases	Description
Custom	Custom		By setting this tier to custom, the exposure model can be configured in any way desirable (without tier specific presets). The model is fully specified by the user. Both EFSA 2012 Optimistic and EFSA 2012 Pessimistic can be specified using the custom model choice. This choice allows for a sensitivity analysis where each factor is varied, one at the time.
EC 2018 Tier 1	ComTier1		EC 2018 Tier 1.
EC 2018 Tier 2	ComTier2		EC 2018 Tier 2.

5.15 Consumption intake unit

Accepted units for consumption intake amounts. Controlled terminology.

Table 5.15: Unit definition for Consumption intake unit.

Name	Short name	Aliases
gram/kilogram bodyweight/day	g/kg bw/day	g/kg bw, gram/kg bw, g/kg bw/day, gram/kg bw/day, gr/kg bw/day, G212A
gram/day	g/day	gram, grams, g/day, g/day, gram/day, gr/day

5.16 Consumption unit

Accepted units for consumption amounts. Controlled terminology.

Table 5.16: Unit definition for Consumption unit.

Name	Short name	Aliases
kilogram	kg	kg, kilograms, kilogr, 3, G167A
gram	g	g, grams, gr, 0, G148A

5.17 Consumption value type

Accepted consumption value types. Controlled terminology.

Table 5.17: Unit definition for Consumption value type.

Name	Short name	Aliases
Large portion	LP	LP, LargePortion
Mean consumption	MC	MC, MeanConsumption
Percentile	Percentile	Percentile, P

5.18 Covariate model types

Accepted Covariate model types. Controlled terminology.

Table 5.18: Unit definition for Covariate model types.

Name	Short name	Aliases	Description
Only constant	Constant		No relation between exposure and e.g. age or gender.
Only covariable	Covariable		Exposure depends on the covariable, e.g. age.
Only cofactor	Cofactor		Exposure depends on the level of the cofactor, e.g. gender.
Both covariable and cofactor	CovariableCofactor		Exposure depends on both covariable and cofactor (additive model).
Both covariable and cofactor and interaction	CovariableCofactorInteraction		Exposure depends on both covariable and cofactor and the effect of the covariable differs for different levels of the cofactor (multiplicative model).

5.19 Dietary exposures details level types

Accepted ietary exposures details level types. Controlled terminology.

Table 5.19: Unit definition for Dietary exposures details level types.

Name	Short name	Aliases	Description
Full	Full		Show all details.
Restrict to risk-drivers (dietary exposures screening)	OnlyRiskDrivers		Restrict to detailed output for risk-drivers identified by dietary exposures screening.
Omit foods-as-eaten details	OmitFoodsAsEaten		Restrict to detailed output for modelled foods and substances. Omit foods-as-eaten details.

5.20 Dietary intake calculation tier types

Accepted Dietary intake calculation tier types. Controlled terminology.

Table 5.20: Unit definition for Dietary intake calculation tier types.

Name	Short name	Aliases	Description
Custom	Custom		By setting this tier to custom, the exposure model can be configured in any way desirable (without tier specific presets).
EFSA 2012 Optimistic	EfsaOptimistic		EFSA 2012 Optimistic.
EFSA 2012 Pessimistic	EfsaPessimistic		EFSA 2012 Pessimistic.
EC 2018 Tier 1	ComTier1		EC 2018 Tier 1.
EC 2018 Tier 2	ComTier2		EC 2018 Tier 2.
EFSA 2012 Pessimistic - Acute	EfsaPessimisticAcute		EFSA 2012 Pessimistic - Acute.
EFSA 2012 Pessimistic - Chronic	EfsaPessimistic-Chronic		EFSA 2012 Pessimistic - Chronic.

5.21 Dose response model type

Accepted dose response model types. Controlled terminology.

Table 5.21: Unit definition for Dose response model type.

Name	Short name	Aliases	Description
Exp-m1	Exp-m1	Expm1	Exponential model 1
Exp-m2	Exp-m2	Expm2	Exponential model 2
Exp-m3	Exp-m3	Expm3	Exponential model 3
Exp-m4	Exp-m4	Expm4	Exponential model 4
Exp-m5	Exp-m5	Expm5	Exponential model 5
Hill-m1	Hill-m1	Hillm1	Hill model 1
Hill-m2	Hill-m2	Hillm2	Hill model 2
Hill-m3	Hill-m3	Hillm3	Hill model 3
Hill-m4	Hill-m4	Hillm4	Hill model 4
Hill-m5	Hill-m5	Hillm5	Hill model 5
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	

5.22 Dose unit

Accepted units for substance doses. Controlled terminology.

Table 5.22: Unit definition for Dose unit.

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
microgram/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
femtogram/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
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, G225A
microgram/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

continues on next page

Table 5.22 – continued from previous page

Name	Short name	Aliases
moles	moles	moles, Moles
millimoles	mmoles	mmoles, mMoles
micromoles	µmoles	umoles, uMoles
nanomoles	nmoles	nmoles, nMoles
gram/kilogram bodyweight/week	g/kg bw/week	g/kg bw/week, gram/kg bw/week, gr/kg bw/week, G218A
milligram/kilogram bodyweight/week	mg/kg bw/week	mg/kg bw/week, milligram/kg bw/week, milligr/kg bw/week, G217A
microgram/kilogram bodyweight/week	µg/kg bw/week	µg/kg bw/week, microgram/kg bw/week, microgr/kg bw/week, G216A
nanogram/kilogram bodyweight/week	ng/kg bw/week	ng/kg bw/week, nanogram/kg bw/week, nanogr/kg bw/week, G215A
picogram/kilogram bodyweight/week	pg/kg bw/week	pg/kg bw/week, picogram/kg bw/week, picogr/kg bw/week
femtogram/kilogram bodyweight/week	fg/kg bw/week	fg/kg bw/week, femtogram/kg bw/week, femtogr/kg bw/week
gram/gram bodyweight/week	g/g bw/week	g/g bw/week, gram/g bw/week, gr/g bw/week
milligram/gram bodyweight/week	mg/g bw/week	mg/g bw/week, milligram/g bw/week, milligr/g bw/week
microgram/gram bodyweight/week	µg/g bw/week	µg/g bw/week, microgram/g bw/week, microgr/g bw/week
nanogram/gram bodyweight/week	ng/g bw/week	ng/g bw/week, nanogram/g bw/week, nanogr/g bw/week
picogram/gram bodyweight/week	pg/g bw/week	pg/g bw/week, picogram/g bw/week, picogr/g bw/week
femtogram/gram bodyweight/week	fg/g bw/week	fg/g bw/week, femtogram/g bw/week, femtogr/g bw/week
kilogram/week	kg/week	kg/week, kilogram/week, kilogr/week
gram/week	g/week	g/week, gram/week, gr/week
milligram/week	mg/week	mg/week, milligram/week, milligr/week
microgram/week	µg/week	µg/week, microgram/week, microgr/week
nanogram/week	ng/week	ng/week, nanogram/week, nanogr/week
picogram/week	pg/week	pg/week, picogram/week, picogr/week
femtogram/week	fg/week	fg/week, femtogram/week, femtogr/week

5.23 Estimates nature types

Accepted Estimates nature types. Controlled terminology.

Table 5.23: Unit definition for Estimates nature types.

Name	Short name	Aliases	Description
Realistic	Realistic		For lognormal: no censoring at the value of the composite sample concentration, no upper limit to the unit concentration. For Beta: no censoring at the value of the composite sample concentration, unit values are never higher than the number of units in composite sample * value of composite sample concentration.
Conservative	Conservative		For lognormal: unit values will be left-censored at the value of the composite sample concentration, no upper limit to the unit concentration. For Beta: unit values will be left-censored at the value of the value of composite sample concentration, unit are values never higher than the number of units in composite sample * value of composite sample concentration.

5.24 Exposure approach types

Accepted Exposure approach types. Controlled terminology.

Table 5.24: Unit definition for Exposure approach types.

Name	Short name	Aliases	Description
Risk based (RPFs)	Risk based (RPFs)		Exposures are multiplied by the RPF and thus exposures to different substances are on the same and comparable scale.
Standardised	Standardised		All substances are standardised to equal variance (selection of the components will work on patterns of correlation only).
Unweighted (RPFs = 1)	ExposuresUW		Exposures as such are taken (unweighted). This is equivalent to RPFs equal to 1. Thus exposures to different substances are not on the same and comparable scale anymore.

5.25 Exposure method types

Accepted Exposure method types. Controlled terminology.

Table 5.25: Unit definition for Exposure method types.

Name	Short name	Aliases	Description
Manual	Manual		Exposure levels are determined by explicit specification.
Automatic	Automatic		Exposure levels are generated automatically based on the estimated exposure distribution.

5.26 Exposure route type

Accepted exposure routes how an individual is exposed to substance concentrations. Controlled terminology.

Table 5.26: Unit definition for Exposure route type.

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).

5.27 Exposure type

Accepted exposure types. Controlled terminology.

Table 5.27: Unit definition for Exposure type.

Name	Short name	Aliases	Description
Acute	Acute	Acute	Acute exposure.
Chronic	Chronic	Chronic	Chronic exposure.

5.28 Exposure unit

Accepted units for substance exposures. Controlled terminology.

Table 5.28: Unit definition for Exposure unit.

Name	Short name	Aliases
gram/kilogram bodyweight/day	g/kg bw/day	g/kg bw/day, g/kg/day, gram/kg bw/day, gr/kg bw/day, G212A
milligram/kilogram bodyweight/day	mg/kg bw/day	mg/kg bw/day, mg/kg/day, milligram/kg bw/day, milligr/kg bw/day, G211A
microgram/kilogram bodyweight/day	µg/kg bw/day	µg/kg bw/day, µg/kg/day, microgram/kg bw/day, microgr/kg bw/day, G210A
nanogram/kilogram bodyweight/day	ng/kg bw/day	ng/kg bw/day, ng/kg/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
femtogram/kilogram bodyweight/day	fg/kg bw/day	fg/kg bw/day, fg/kg/day, femtogram/kg bw/day, femtogr/kg bw/day
gram/gram bodyweight/day	g/g bw/day	g/g bw/day, g/g/day, gram/g bw/day, gr/g bw/day
milligram/gram bodyweight/day	mg/g bw/day	mg/g bw/day, mg/g/day, milligram/g bw/day, milligr/g bw/day
microgram/gram bodyweight/day	µg/g bw/day	µg/g bw/day, µg/g/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, pg/g/day, picogram/g bw/day, picogr/g bw/day
femtogram/gram bodyweight/day	fg/g bw/day	fg/g bw/day, fg/g/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
gram/kilogram	g/kg	g/kg, gram/kg, gr/kg, G015A
milligram/kilogram	mg/kg	mg/kg, milligram/kg, milligr/kg, G061A
microgram/kilogram	µg/kg	µg/kg, microgram/kg, microgr/kg, G050A
nanogram/kilogram	ng/kg	ng/kg, nanogram/kg, nanogr/kg, G077A
picogram/kilogram	pg/kg	pg/kg, picogram/kg, picogr/kg, G081A
femtogram/kilogram	fg/kg	fg/kg, femtogram/kg, femtogr/kg
gram	g	g, gram, gr, G148A
milligram	mg	mg, milligram, milligr, G155A
microgram	µg	µg, microgram, microgr
nanogram	ng	ng, nanogram, nanogr, G120A
picogram	pg	pg, picogram, picogr, G125A
femtogram	fg	fg, femtogram, femtogr

5.29 Focal commodity replacement method types

Accepted Focal commodity replacement method types. Controlled terminology.

Table 5.29: Unit definition for Focal commodity replacement method types.

Name	Short name	Aliases	Description
Replace samples with focal commodity samples	Replace samples with focal commodity samples		Replace all samples of the selected focal commodity/commodities.
Append focal commodity samples	Append focal commodity samples		Add the samples of the focal commodity/commodities to the background concentration data.
Replace measurements of focal food/substance combinations with measurements from focal commodity samples	Replace measurements of focal food/substance combinations with measurements from focal commodity samples		Replace the substance concentrations of the background concentrations by substance concentrations from the focal commodity concentration data.
Remove measurements of focal food/substance combinations	Remove measurements of focal food/substance combinations		Remove substance measurements for the selected focal food/substance combinations.
Replace measurements of focal food/substance combinations with concentration limit value	Replace measurements of focal food/substance combinations with concentration limit value		Replace the substance concentrations of the background concentrations by a concentration limit value.

5.30 Function types

Accepted Function types. Controlled terminology.

Table 5.30: Unit definition for Function types.

Name	Short name	Aliases	Description
Polynomial	Polynomial		A polynomial regression fits a nonlinear relationship between the value of the independent variable (e.g. age) and the corresponding conditional mean of y (here the exposure). A polynomial with a degree of 0 is simply a constant function; with a degree of 1 is a line; with a degree of 2 is a quadratic; with a degree of 3 is a cubic, and so on.
Spline	Spline		A spline fits a nonlinear relationship between the value of the independent variable (e.g. age) and the corresponding conditional mean of y (here the exposure). A spline with a degree of 0 is simply a constant function; with a degree of 1 is a line; with a degree of 2 is a quadratic; with a degree of 3 is a cubic, and so on.

5.31 Gender type

Accepted gender types. Controlled terminology.

Table 5.31: Unit definition for Gender type.

Name	Short name	Aliases	Description
Female	F	Female, F	Female
Male	M	Male, M	Male

5.32 Harvest application type

Accepted harvest application types. Controlled terminology.

Table 5.32: Unit definition for Harvest application type.

Name	Short name	Aliases	Description
Pre-harvest application	Pre-harvest	PreHarvest	Pre-harvest application
Post-harvest application	Post-harvest	PostHarvest	Post-harvest application

5.33 Hazard characterisation type

Accepted hazard characterisation types. Controlled terminology.

Table 5.33: Unit definition for Hazard characterisation type.

Name	Short name	Aliases
Benchmark dose	BMD	BMD
No observed adverse effect level	NOAEL	NOAEL
Lowest observed adverse effect level	LOAEL	LOAEL
Acceptable daily intake	ADI	ADI
Acute reference dose	ARfD	ARfD
No observed effect level	NOEL	NOEL
Tolerable daily intake	TDI	TDI
Tolerable weekly intake	TWI	TWI
Benchmark dose lower confidence limit of 1%	BMDL01	BMDL01
Benchmark dose lower confidence limit of 10%	BMDL10	BMDL10
Human biomonitoring guidance values	HBMGV	HBMGV
Other	Other	Other

5.34 Hazard dose imputation method types

Accepted Hazard dose imputation method types. Controlled terminology.

Table 5.34: Unit definition for Hazard dose imputation method types.

Name	Short name	Aliases	Description
Munro P5 (TTC approach)	Munro P5 (TTC approach)		Use the P5 of the Munro NOEL collection.
Munro central value	Munro central value		Use an unbiased nominal value from the Munro NOEL collection; draw randomly from this collection in the uncertainty runs.
Available hazard characterisations distribution P5	Available hazard characterisations distribution P5		Use the P5 of the available points of departure.
Available hazard characterisations distribution central value	Available hazard characterisations distribution central value		Use an unbiased nominal value from the collection of available points of departure; draw randomly from this collection in the uncertainty runs.

5.35 Health effect types

Accepted Health effect types. Controlled terminology.

Table 5.35: Unit definition for Health effect types.

Name	Short name	Aliases	Description
Risk	Risk		Health effect is negative (risk).
Benefit	Benefit		Health effect is positive (benefit).

5.36 Individual property type

Accepted individual property types. Controlled terminology.

Table 5.36: Unit definition for Individual property type.

Name	Short name	Aliases	Description
Categorical	Categorical	Categorical	Categorical e.g. blood type A, B, AB, O or region East, West, North, South.
Boolean	Boolean	Boolean	Boolean e.g. yes, no, true, false. See Boolean types unit definitions.
Numeric	Numeric	Numeric	Numeric, real numbers.
Nonnegative	Nonnegative	Nonnegative	Nonnegative real numbers, positive or zero.
Integer	Integer	Integer	Integer, integer numbers.
NonnegativeInteger	NonnegativeInteger	NonnegativeInteger	NonnegativeInteger integer numbers, positive or zero.
Month	Month	Month	Month. See Month types unit definitions.
DateTime	DateTime	DateTime	DateTime, period.
Gender	Gender	Gender, Sex	Gender, sex or sexuality. See Gender types unit definitions.
Location	Location	Location	Location, country.

5.37 Individual subset types

Methods for selecting/matching survey individuals with a specified/scoped population.

Table 5.37: Unit definition for Individual subset types.

Name	Short name	Aliases	Description
Match individuals selection to population definition	Match to population definition		Match individuals selection to population definition.
Ignore population definition (use all individuals in survey)	Ignore population definition		Ignore population definition (use all individuals in survey).
Match individuals selection to population definition using selected properties only	Match using selected properties		Match individuals selection to population definition using selected properties only.

5.38 Intake model types

Accepted Intake model types. Controlled terminology.

Table 5.38: Unit definition for Intake model types.

Name	Short name	Aliases	Description
Observed Individual Means	OIM		Observed Individual Means: just the empirical means over the observed days.
BetaBinomial Normal	BBN		BetaBinomial distribution for frequency of exposure + (transformed) Normal distribution for amounts (de Boer et al. 2009).
Logistic-Normal Normal	LNN0		Logistic-Normal distribution for frequency of exposure + (transformed) Normal distribution for amounts.
Logistic-Normal Normal with correlation	LNN		Logistic-Normal distribution for frequency of exposure + (transformed) Normal distribution for amounts. Both models are estimated taking into account the correlation between exposure frequency and amounts.
Iowa State University Foods model	ISUF		Iowa State University Foods model: semiparametric distribution for frequency of exposure + (transformed) Normal distribution for amounts (de Boer et al. 2009, Dodd (1996)).

5.39 Internal concentration types

Accepted Internal concentration types. Controlled terminology.

Table 5.39: Unit definition for Internal concentration types.

Name	Short name	Aliases	Description
Internal modelled concentrations	Internal modelled concentrations		Internal modelled concentrations from dietary and/or non-dietary routes, aggregated.
Human monitoring concentrations	Human monitoring concentrations		Human monitoring concentrations as measured in blood, urine or other compartments.

5.40 Internal model type

Accepted internal model types. PBK model or absorption factor model Controlled terminology.

Table 5.40: Unit definition for Internal model type.

Name	Short name	Aliases	Description
Absorption Factor Model	AbsorptionFactorModel	AbsorptionFactorModel	Use absorption factor model.
PBK Model	PBKModel	PBKModel	Use PBK model.

5.41 Mean value correction types

Accepted Mean value correction types. Controlled terminology.

Table 5.41: Unit definition for Mean value correction types.

Name	Short name	Aliases	Description
Unbiased	Unbiased		The mean of the lognormal is unbiased (bias correction).
Biased	Biased		The mean of the lognormal is biased (no bias correction).

5.42 Measurement result type

Specifies the type of a measurement result. E.g., a positive value, a non-detect, or missing value.

Table 5.42: Unit definition for Measurement result type.

Name	Short name	Aliases	Description
VAL	VAL	VAL	Positive measurement greater than zero.
LOD	LOD	LOD	Measurement below the limit of detection (LOD).
LOQ	LOQ	LOQ	Measurement below the limit of quantification (LOQ).
MV	MV	MV	Missing value (MV).

5.43 Missing value imputation method types

Accepted Missing value imputation method types. Controlled terminology.

Table 5.43: Unit definition for Missing value imputation method types.

Name	Short name	Aliases	Description
Set zero	Set zero		Set missing values to zero.
Impute from data	Impute from data		Replace missing measurements by random other measurements of the same substance, biological matrix and sampling type.
No missing value imputation	No missing value imputation		No missing value imputation, all missing values remain in the data set and samples with missing values will be removed before analysis.

5.44 Modelled foods calculation source types

Accepted Modelled foods calculation source types. Controlled terminology.

Table 5.44: Unit definition for Modelled foods calculation source types.

Name	Short name	Aliases	Description
Derive modelled foods from concentrations	DeriveModelledFoodsFromSampleBasedConcentrations		Derive modelled foods from sample based concentration data.
Derive modelled foods from single value concentrations	DeriveModelledFoodsFromSingleValueConcentrations		Derive modelled foods from single value concentrations.
Derive modelled foods from concentration limits	UseWorstCaseValues		Derive modelled foods from concentration limits.

5.45 Month type

Accepted months types. Controlled terminology.

Table 5.45: Unit definition for Month type.

Name	Short name	Aliases
January	Jan	Jan, Januari, 1
February	Feb	Feb, Februari, 2
March	Mar	Mar, 3
April	Apr	Apr, 4
May	May	May, 5
June	Jun	Jun, June, 6
July	Jul	Jul, July, 7
August	Aug	Aug, 8
September	Sep	Sep, Sept, 9
October	Oct	Oct, 10
November	Nov	Nov, 11
December	Dec	Dec, 12

5.46 Multiple substance handling method types

Accepted Multiple substance handling method types. Controlled terminology.

Table 5.46: Unit definition for Multiple substance handling method types.

Name	Short name	Aliases	Description
Combined assessment of selected substances	Combined		Combined assessment of selected substances.
Loop over selected substances	Loop		Loop over selected substances.

5.47 Network analysis type

Accepted Network analysis types. Controlled terminology.

Table 5.47: Unit definition for Network analysis type.

Name	Short name	Aliases	Description
No network analysis	No network analysis		No network analysis is applied.
Apply network analysis	Apply network analysis		Network analysis is applied on the substance x component (U) matrix.

5.48 Non-quantifications handling method types

Accepted Non-quantifications handling method types. Controlled terminology.

Table 5.48: Unit definition for Non-quantifications handling method types.

Name	Short name	Aliases	Description
By zero	By zero		Non-quantifications are assumed to be zero's (set to 0).
By f * LOR	By f * LOR		Non-quantifications are replaced by f * LOR where f is a constant.
By f * LOD or by LOD + f * (LOQ - LOD)	By f * LOD or by LOD + f * (LOQ - LOD)		Left censored are replaced by f * LOD; Non-quantifications are replaced by LOD + f * (LOQ - LOD), where f is a constant.

5.49 Nondetect imputation method types

Accepted nondetect imputation method types. Controlled terminology.

Table 5.49: Unit definition for Nondetect imputation method types.

Name	Short name	Aliases	Description
Replace by LOR/LOQ/LOD	ReplaceLimit		Non-quantifications are replaced by f * LOR or f * LOD or by LOD + f * (LOQ - LOD) where f is a constant.
Impute from censored lognormal distribution	Impute from censoredln		Replace nondetect measurements by a random draw from the lower (left) tail of the censored lognormal distribution.

5.50 Occurrence patterns tier types

Accepted Occurrence patterns tier types. Controlled terminology.

Table 5.50: Unit definition for Occurrence patterns tier types.

Name	Short name	Aliases	Description
Custom	Custom		By setting this tier to custom, the occurence model can be configured in any way desirable (without tier specific presets). The model is fully specified by the user. Both EFSA 2012 Optimistic and EFSA 2012 Pessimistic can be specified using the custom model choice. This choice allows for a sensitivity analysis where each factor is varied, one at the time.
EC 2018 Tier 1	ComTier1		A test tier with realistic model settings.
EC 2018 Tier 2	ComTier2		A test tier with realistic model settings.

5.51 Point of departure type

Accepted point of departure types. Controlled terminology.

Table 5.51: Unit definition for Point of departure type.

Name	Short name	Aliases
Benchmark dose	BMD	BMD
No observed adverse effect level	NOAEL	NOAEL
Lowest observed adverse effect level	LOAEL	LOAEL
No observed effect level	NOEL	NOEL
Median lethal dose	LD50	LD50
Benchmark dose lower confidence limit of 1%	BMDL01	BMDL01
Benchmark dose lower confidence limit of 10%	BMDL10	BMDL10

5.52 Point of departure types

Accepted Point of departure types. Controlled terminology.

Table 5.52: Unit definition for Point of departure types.

Name	Short name	Aliases	Description
Unspecified (no conversion to common expression type)	FromReference		Do not convert non-standard point of departures.
BMD (convert all hazard characterisations as BMDs)	BMD		Convert all point of departures to bench mark doses.
NOAEL (convert all hazard characterisations as NOAELs)	NOAEL		Convert all point of departures to NOAELs.

5.53 Probability distribution type

Probability distribution types.

Table 5.53: Unit definition for Probability distribution type.

Name	Short name	Aliases	Description
LogNormal	LogNormal	LogNormal	Lognormal distribution.
Normal	Normal	Normal	Normal distribution.
LogisticNormal	LogisticNormal	LogisticNormal	Logisticnormal distribution.
Deterministic	Deterministic	Deterministic	Deterministic distribution.

5.54 Processing distribution type

Accepted processing distribution types. Controlled terminology.

Table 5.54: Unit definition for Processing distribution type.

Name	Short name	Aliases	Description
Logistic Normal distribution	LogisticNormal	LogisticNormal, 1	Logisticnormal distribution.
Log Normal distribution	LogNormal	LogNormal, 2	Lognormal distribution.

5.55 Property level type

Accepted property level types. Controlled terminology.

Table 5.55: Unit definition for Property level type.

Name	Short name	Aliases	Description
Individual	Individual	Individual	Individual level.
IndividualDay	IndividualDay	IndividualDay	IndividualDay.

5.56 Response type

Accepted response types. Controlled terminology.

Table 5.56: Unit definition for Response type.

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.

5.57 Risk metric types

Accepted Risk metric types. Controlled terminology.

Table 5.57: Unit definition for Risk metric types.

Name	Short name	Aliases	Description
Margin of exposure	MOE		Margin of exposure (MOE).
Hazard index	HI		Hazard index (HI).

5.58 Riskmetric calculation types

Accepted Riskmetric calculation types. Controlled terminology.

Table 5.58: Unit definition for Riskmetric calculation types.

Name	Short name	Aliases	Description
RPF weighted	RPF Weighted		Calculates risk as a single ratio involving cumulative RPF-weighted exposures.
Sum of risk ratios	Sum of ratios		Calculate risk as sum of ratios of exposure and hazard.

5.59 Single value dietary exposures calculation method types

Accepted Single value dietary exposures calculation method types. Controlled terminology.

Table 5.59: Unit definition for Single value dietary exposures calculation method types.

Name	Short name	Aliases	Description
IESTI	IESTI		IESTI.
IESTI new	IESTI new		IESTI new.
TMDI	TMDI		Theoretical Maximum Daily Intake.
IEDI	IEDI		International Estimated Daily Intake.
Rees–Day model (I)	Rees–Day(I)		Rees–Day model (I).
Rees–Day model (II)	Rees–Day (II)		Rees–Day model (II).

5.60 Single value risk calculation method types

Accepted Single value risk calculation method types. Controlled terminology.

Table 5.60: Unit definition for Single value risk calculation method types.

Name	Short name	Aliases	Description
From single value dietary exposures	From single value dietary exposures		From single value dietary exposures and hazard characterisations.
As percentile from risks distribution	As percentile		As percentile from risks distribution.

5.61 Standardise blood methods

Accepted Standardise blood methods. Controlled terminology.

Table 5.61: Unit definition for Standardise blood methods.

Name	Short name	Aliases	Description
Standardise by total lipid measured via gravimetric analysis	Gravimetric		Standardise by total lipid measured via gravimetric analysis.
Standardise by total lipid measured via enzymatic summation	Enzymatic		Standardise by total lipid measured via enzymatic summation.
Standardise by derived total lipid content of Triglyc-erides/Cholesterol (Bernert et al. 2007)	Bernert		Standardise by derived total lipid content (Bernert et al. 2007).

5.62 Standardise/normalise urine concentration method

Accepted Standardise urine concentration methods based on creatinine. Controlled terminology.

Table 5.62: Unit definition for Standardise/normalise urine concentration method.

Name	Short name	Aliases	Description
Normalise by specific gravity	SGNorm		Normalise by specific gravity.
Standardise by creatinine concentration	CreatStand		Standardise by creatinine concentration.

5.63 Substance group selection method types

Accepted Substance group selection method types. Controlled terminology.

Table 5.63: Unit definition for Substance group selection method types.

Name	Short name	Aliases	Description
All substances	IncludeAll		Include all substances of the substances table and use hazard characterisation imputation for missing hazard data.
Restrict to available hazard data	RestrictHazardDoseRpf		Restrict to the substances with available hazard data (either in the form of dose response models or RPFs).
Restrict to available hazard data and possible membership	RestrictHazardDoseRpfAndProbableMembership		Consider only the substances with available hazard data and non-zero membership (i.e., P(AG) > 0).
Restrict to available hazard data and certain membership	RestrictHazardDoseRpfAndCertainMembership		Consider only substances with certain assessment group membership (i.e., P(AG) = 1) and for which a hazard characterisation is available.
Restrict to non-zero membership	RestrictProbableMembership		Consider all substances, use TTC based on the Cramer class for the substances for which no limit dose or RPF is defined.
Restrict to certain membership	RestrictCertainMembership		Consider only the substances with certain assessment group membership (i.e., P(AG) = 1).

5.64 Substance translation allocation method types

Accepted Substance translation allocation method types. Controlled terminology.

Table 5.64: Unit definition for Substance translation allocation method types.

Name	Short name	Aliases	Description
Random allocation	Random allocation		Random allocation.
Allocate most potent	Allocate most potent		Allocate most potent active substance.
Nominal estimate	Nominal estimate		Allocate nominal estimate (weighted average allocation).
Allocate to all	Allocate to all		Allocate for each active substance independently as if all concentrations were allocated to this active substance.

5.65 Target dose selection method types

Accepted Target dose selection method types. Controlled terminology.

Table 5.65: Unit definition for Target dose selection method types.

Name	Short name	Aliases	Description
Select most toxic	MostToxic		Choose the most toxic (default).
Take aggregate	Aggregate		Choose an aggregated hazard characterisation when there there are multiple available candidates in nominal runs.
Random draw	Draw		Draw a random hazard characterisation.

5.66 Target doses calculation method types

Accepted Target doses calculation method types. Controlled terminology.

Table 5.66: Unit definition for Target doses calculation method types.

Name	Short name	Aliases	Description
In-vivo PoDs (BMDs, NOAELs, etc.)	InVivoPods		In-vivo Points of Departures (BMDs, NOAELs, etc.).
In-vitro BMDs	InVitroBmds		In-vitro Bench Mark Doses
In-vivo PoDs for index substance, others using RPFs from in-vitro dose response models	CombineInVivoPodInVitroDrms		In-vivo Points of Departures for index substance, others using RPFs from in-vitro dose response models

5.67 Target level type

Accepted units whether a dose is assumed to be an internal or external dose. Controlled terminology.

Table 5.67: Unit definition for Target level type.

Name	Short name	Aliases	Description
External	Ext	Ext	External exposure.
Internal	Int	Int	Internal exposure.

5.68 Test system type

Accepted test system types. Controlled terminology.

Table 5.68: Unit definition for Test system type.

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

5.69 Testing method types

Accepted xxx types. Controlled terminology.

Table 5.69: Unit definition for Testing method types.

Name	Short name	Aliases	Description
Backward	Backward		Backward selection starts with selecting a model with a function of the highest degree. Then, the degree of the function is decreased by one and the model is tested again. This process is repeated until decreasing the degree does not improve the model fit anymore.
Forward	Forward		Forward selections starts with selecting a model with a function of the lowest degree. Then, the degree of the function is increased by one and the model is tested again. This process is repeated until increasing the degree does not improve the model fit anymore.

5.70 Time unit

Accepted time units. Controlled terminology.

Table 5.70: Unit definition for Time unit.

Name	Short name	Aliases	Description
hours	h	Hours, h	In hours
minutes	min	Minutes, min	In minutes

5.71 Transform types

Accepted Transform types. Controlled terminology.

Table 5.71: Unit definition for Transform types.

Name	Short name	Aliases	Description
Logarithmic	Logarithmic		Exposure amounts are transformed to normality using a logarithmic transformation.
No transformation	No transformation		Exposure amounts are not transformed.
Power	Power		Exposure amounts are transformed to normality using a Box-Cox power transformation.

5.72 Uncertainty types

Accepted Uncertainty types. Controlled terminology.

Table 5.72: Unit definition for Uncertainty types.

Name	Short name	Aliases	Description
Empirical	Empirical		Data are taken as such.
Parametric	Parametric		A parametric model is fitted to the data.

5.73 Unit variability correlation types

Accepted Unit variability correlation types. Controlled terminology.

Table 5.73: Unit definition for Unit variability correlation types.

Name	Short name	Aliases	Description
No correlation	NoCorrelation		The unit residue values for unit portions (consumption amount/unitweight) are randomly drawn, explicitly ignoring any correlation between unit residues.
Full correlation	FullCorrelation		The unit residue values for unit portions (consumption amount/unitweight) are randomly drawn, explicitly introducing correlation between unit residues, e.g. high (small) values occur more frequently together.

5.74 Unit variability model types

Accepted Unit variability model types. Controlled terminology.

Table 5.74: Unit definition for Unit variability model types.

Name	Short name	Aliases	Description
Beta distribution	Beta distribution		Requires knowledge of the number of units in a composite sample, and of the variability between units (realistic or conservative estimates). Under the beta model, the simulated unit values are drawn from a bounded distribution on the interval.
Lognormal distribution	Lognormal distribution		Requires only knowledge of the variability between units (realistic or conservative estimates). The lognormal distribution is considered as an appropriate model for many empirical positive concentration distributions (unbounded distribution).
Bernoulli distribution	Bernoulli distribution		Requires only knowledge of the number of units in a composite sample (results are always conservative). 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.

5.75 Unit variability types

Accepted Unit variability types. Controlled terminology.

Table 5.75: Unit definition for Unit variability types.

Name	Short name	Aliases	Description
Variation coefficient	Variation coefficient		Standard deviation divided by the mean.
Variability factor	Variability factor		Defined as 97.5th percentile divided by the mean.

5.76 Unit weight value type

Accepted unit weight types. Controlled terminology.

Table 5.76: Unit definition for Unit weight value type.

Name	Short name	Aliases
Unit weight raw agricultural commodity	RAC	RAC, UnitWeightRAC, UnitWeightRawAgriculturalCommodity
Unit Weight edible portion	EP	EP, UnitWeightEP, UnitWeightEdiblePortion

5.77 Value qualifier

Supported value qualifiers.

Table 5.77: Unit definition for Value qualifier.

Name	Short name	Aliases
=	=	=, Equals
<	<	lt, LessThan, <

CHAPTER

SIX

APPLICATION PROGRAMMING INTERFACE (API)

The entire API interface is described using the Swagger application.

CHAPTER

SEVEN

COMMAND LINE INTERFACE (CLI)

7.1 Introduction

The MCRA core library comes with a command line interface (CLI) utility to run MCRA actions using input files and producing output files. Data files and specification of sequences of modules and settings are supplied in an action template. An action template is a local disk folder with a defined structure which is described in *Action template structure*. A new empty action template can be created by the CLI as demonstrated in *Create a new action template*. Alternatively, the template can also be a compressed (zip) file format. Running an action via the CLI is described in *Run an action*. This generates results in local disk output files that can be accessed via, for instance, a web browser as described in *Action template structure*.

The latest release of the CLI utility of MCRA Core is available as a download from GitHub (https://github.com/rivm-syso/mcra-core/releases) or can be built from the sources.

7.2 Action template structure

The CLI can process either folders or zip files as input. This paragraph explains the structure of the files and subfolders inside the input folder or zip file which the CLI needs to run an MCRA action correctly. This structure is the same for both the contents of a zip file and the contents of a folder. The term base folder will be used to refer to either the zip file or the folder containing the settings and data of the action. The base folder must contain the following XML files:

- _ActionData.xml: contains the configuration of the input data for the action
- _ActionSettings.xml: contains the configuration for the action

The data for the action may be included as CSV files containing MCRA table data directly in the base folder together with the settings files. In this case the _ActionData.xml file is optional. The other option is to put the input data files in a subfolder named Data. In this case the following file formats are accepted:

- Excel files: files with an XLS or XLSX extension;
- Access files: files with an ACCDB or MDB extension;
- Zipped CSV: zip files containing the MCRA data in one CSV file per table.

The files mentioned above must adhere to the data formats of MCRA available in the documentation.

Action settings XML file structure

The model settings XML file must be named '_ActionSettings.xml' (with a leading underscore '_' character). A simple example, configuring a 'Foods' action, is:

```
<Project>
  <Name>ExampleFoodsAction</Name>
  <ActionType>Foods</ActionType>
</Project>
```

The Project tag is the root element of the XML file; it is named Project for historic reasons. This configuration file is an example of a minimalistic configuration in which all other settings have been omitted. By using the *Create a new action template* option, a configuration file will be automatically generated that contains all possible options for the action.

Data configuration

As mentioned before, MCRA accepts Excel, Access and zipped CSV data files as input when they are included in the Data subfolder. For simple actions there is also the option to include CSV files directly in the root of the folder or zip file. In this case the folder or zip file itself becomes the single input file for the action. The next section lists all possible options. Data for running an action with the CLI is accepted in the following ways: 1. As CSV (comma separated values) files directly in the root of the folder or zip file. In this case, the zip file or folder is itself the single data file, containing the MCRA tables. A data configuration XML file is optional, but it can still be used to filter the table groups that are loaded during the run. 2. A Data subfolder containing one or more of the accepted MCRA file types. In this case a data configuration XML file is necessary to link the MCRA table groups and data files to use.

Data configuration XML file structure

The data configuration XML file must be named _ActionData.xml (with the leading underscore '_' character) and must be placed in the root of the zip file or folder. It describes the links between the modules and data. The link is made based on a so-called MCRA table group definition. A data file can contain data for one or more table groups. An example of a simple data configuration file is as follows:

```
<DataSourceConfiguration xmlns:xsd=http://www.w3.org/2001/XMLSchema
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <DataSourceMappingRecords>
    <DataSourceMappingRecord>
      <SourceTableGroup>Foods</SourceTableGroup>
      <Name>InputData.xls</Name>
    </DataSourceMappingRecord>
    <DataSourceMappingRecord>
      <SourceTableGroup>Compounds</SourceTableGroup>
      <Name>InputData.xls</Name>
    </DataSourceMappingRecord>
    <DataSourceMappingRecord>
      <SourceTableGroup>Survey</SourceTableGroup>
      <Name>Consumptions.mdb</Name>
    </DataSourceMappingRecord>
    <DataSourceMappingRecord>
      <SourceTableGroup>FoodConsumptions</SourceTableGroup>
      <Name>Consumptions.mdb</Name>
    </DataSourceMappingRecord>
  </DataSourceMappingRecords>
</DataSourceConfiguration>
```

The data source configuration contains one collection of DataSourceMappingRecords. Each DataSourceMappingRecord must contain at least the SourceTableGroup. The Name element refers to the file name in the Data subfolder when data files are used. If the zip file or folder contains CSV table files in the root itself, the name is optional. In this case the name of the zip file itself is used.

7.3 Create a new action template

The CLI can generate an empty action template for any type of action that is supported by MCRA. This generated template consists of the correct files and folder structure for the action, with all default settings, and Excel data files with empty tables as a starting point to fill in your own input data.

Use the following command to create a new action of a given type:

```
mcra.exe create '<name_of_new_action>' -a <action type>
```

where the option -a specifies the type of action to be created. For example, to create a basic skeleton for analysing your own single value risk estimate with the name MySingleValueRisk, run the command,

```
mcra.exe create MySingleValueRisk  -a SingleValueRisk
```

The name of the action type, SingleValueRisk in the example above, is a reserved keyword and must match with one of the supported action types. To get a list of all supported action types, run the command,

```
mcra.exe create -u
```

This will print out a list of reserved names of all action types. After creating this template, you can start filling in the input data in the Excel files and adjusting the calculation settings in the _ActionSettings.xml file.

7.4 Run an action

The most basic use of the CLI is to specify a folder name or a zip file containing the settings and data for the action that you want to run (see section *Action template structure* for a detailed description). In a terminal window. e.g. PowerShell or the command prompt, run the command:

```
mcra.exe run '<path to action folder>' [options]
or:
mcra.exe run '<name_of_action>.zip' [options]
```

As an example, assume that an action folder has been composed for doing a calculation on the retrospective Tier 2 of cumulative exposure to pesticides. This model can be run by the command:

```
mcra.exe run 'EU acute cumulative exposure assessment (2018) Tier 2'
```

The name of the action in this example is equal to the name of the folder that contains the data and settings as input for the CLI, so in this example there is a folder with the name 'EU acute cumulative exposure assessment (2018) Tier 2'. To get an overview of the available options for running the command line interface, run the following command:

```
mcra.exe help run
```

The output of this command lists the options for the run command as shown below (exact details maybe different, depending on the version of the CLI utility).

Table 7.1: command-line parameters for running an action

Parameter	Description
task input name	(pos. 0, required) Name of the input zip file or base folder containing the simulation task to be processed.
-o, --output	Base folder for output. Project output will be written to a subfolder, using the project name.
--overwrite	(Default: false) Overwrite existing output. If set to false, the output will be written to a uniquely named folder, otherwise the project name is used as output folder
--skipreport	(Default: false) Don't render the full output report.
--skiptables	(Default: false) Don't generate CSV output tables.
--skipcharts	(Default: false) Don't generate SVG charts.
--keeptempfiles	(Default: false) Keep temporary (intermediate) files.
-r, --randomseed	Use this value as the Monte Carlo random seed for the project.
-i, --interactive	(Default: false) Set to true to run in interactive mode.
-s, --silent	(Default: false) Set to true to run in silent mode.
--dbType	(Default: Csv) Database type. Possible options: - csv: Intermediate data will be written to CSV files during the run, - binary: Intermediate data will be written as binary files
--help	Display this help screen.
--version	Display version information.

Output

By default the CLI utility creates the output files in a subfolder in the location where the CLI command is run, use the '--output' command line option to specify a different output folder.

7.5 Output files and folder structure

The output of an action performed with the CLI is saved to a folder, which by default has the name of the action suffixed with date and time to create a unique folder name. The output of the action consists of metadata, data and image (chart) files. The CLI options '--skipreport', '--skiptables' and/or '--skipcharts' are available to limit the output if desired. The contents of the output folder are as follows.

Table 7.2: Output file and folders of an action run

File	Description
Metadata files	
_CsvFileIndex.txt	Tab delimited text file containing lookup data for the output CSV data files. It contains the name of the file and the path to the file in the full MCRA output.
_MCRAVersion.txt	Contains detailed MCRA version information of the CLI that was used.
_TOC.txt	Contains a detailed internal table of contents of references to report sections.
_TOC-Charts.txt	Contains a detailed table of contents of the output charts (if any).
_TOC-CsvData.txt	Contains a detailed table of contents of the output CSV files.
_TOC-XmlData.txt	Contains a detailed table of contents of any output data in XML format.
ProjectOriginalSettings.xml	A copy of the original action settings file (_ActionSettings.xml)
ProjectSimulatedSettings.xml	A fully populated XML file containing all settings that were used in the run. This includes the defaults of all other settings that are available in MCRA.
Data files	
FoodsTable.csv	Foods output data table
SubstancesTable.csv	Substances output data table
Reports	
_Report.html	The MCRA report in HTML format. This report contains all output sections with tables and charts in one page.
Tables/charts	The tables and charts in the HTML file are included in this folder and are referenced from the HTML file.

CHAPTER

EIGHT

APPENDICES

8.1 NOAEL collection of Munro et al.

This collection is from Munro et al. (1996) and can be downloaded here.

8.2 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 $y_i^{(p)}$, see Box and Cox (1964).

8.3 Gauss-Hermite Integration

8.3.1 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). N-point integration is exact for all polynomials $f(x)$ of degree 2N-1, see Dahlquist and Bjorck (1974). 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:

$$\int_{-\infty}^{\infty} F(x)\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)\mathrm{d}y$$

$$= \int_{-\infty}^{\infty} F(\mu + \sqrt{2}\sigma x)\frac{1}{\sqrt{\pi}}\exp(-x^2)\mathrm{d}x$$

$$= \frac{1}{\sqrt{\pi}}\sum_{j=1}^{N} w_j F(\mu + \sqrt{2}\sigma x_j)$$

8.3.2 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) 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) .

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 $w_i w_j$.

8.3.3 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

$$P(Y_{ij} = 0|b_i, v_i) = 1\text{–}H(\lambda + v_i).$$

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 covariables, such as sex 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.

CHAPTER

NINE

GLOSSARY

ADI Acceptable daily intake. The ADI is an estimate of the amount of a substance in food or drinking water that can be consumed daily over a lifetime without presenting an appreciable risk to health. It is usually expressed as milligrams of the substance per kilogram of body weight and day and applies to chemical substances such as food additives, pesticide residues and veterinary drugs.

ADME An abbreviation for "absorption, distribution, metabolism and excretion", the four key processes which describe how drugs and chemicals get into the body, what happens to them while they are there, and how they are eliminated

AIC Akaike Information Criterion.

AOP Adverse Outcome Pathways. An AOP is a structured representation of biological events leading to adverse effects and is considered relevant to risk assessment.

ARfD Acute reference dose. Estimate of the amount of a substance in food and/or drinking water, normally expressed on a body weight basis, that can be ingested in a period of 24 h or less without appreciable health risk to the consumer on the basis of all known facts at the time of the evaluation.

AU Agricultural Use.

BBN Beta binomial normal model.

BMD Benchmark dose. A dose or concentration that produces a predetermined change in response rate of an adverse effect (called the benchmark response or *BMR*) compared to background.

BMDL Benchmark dose lower confidence limit.

BMDU Benchmark dose upper confidence limit.

BMI The body mass index is a measurement that expresses the relationship between an individual's weight and height. BMI is calculated by dividing weight in kilograms by height in metres squared (i.e. height x height). Used to assess whether someone's weight is appropriate.

BMR Benchmark response.

BREAM Bystander and resident exposure assessment model.

BROWSE Bystanders, residents, operators and workers exposure models.

BW Body weight.

CA Concentration Addition. This model is based on a dilution principle, and was designed for chemicals with a similar mechanism of action.

CAG Cumulative assessment group. A group of chemicals that could plausibly act by a common mode of action, not all of which will necessarily do so. Membership of a CAG can usually be refined (reduced) by application of successively higher tiers of assessment.

ConsExpo Consumer exposure model.

CRA Cumulative risk assessment. Risk assessment for combined exposure to two or more chemicals by all relevant pathways and routes.

DA Dose addition. A process to establish the response of organisms to a mixture of chemicals with similar toxicity. This involves adding up their individual effects to predict the likely impact of the overall mixture.

DR Dose response. The relationship between the amount of a substance to which an individual organism, population or ecosystem is exposed and the way in which it responds (e.g. in terms of toxicity).

FA Food additive. A substance deliberately added to foods or beverages for beneficial technological reasons (e.g. to preserve, flavour, colour or ensure a particular texture). Food additives are not normally consumed by themselves nor used as typical ingredients in food.

FC Focal commodity. A commodity for which an *MRL* is to be set or for which a high residue event has been monitored, and which is therefore the focus of an exposure assessment.

GAP Good agricultural practice. GAP is a certification system for agriculture, specifying procedures (and attendant documentation) that must be implemented to create food for consumers or further processing that is safe and wholesome, using sustainable methods.

HBM Human biomonitoring. A direct measurement of the level of toxic chemical compounds present in the body. Often, these measurements are made using blood and urine.

HBGV Health based guidance value. HBGV is a science-based recommendation for the maximum (oral) exposure to a substance that is not expected to result in an appreciable health risk, taking into account current safety data, uncertainties in these data, and the likely duration of consumption.

HC Hazard characterisations is a generic term for any reference value for a substances at a chosen biological target level (external or internal) beyond which exposure is associated with potential adverse health effects. Hazard characterisations can be specified as external values (e.g., a human based guidance value, such as an *ADI* or *ARfD*) or are based on a point of departure (*POD*), such as BMDs from dose-response models or externally specified points of departure (*NOAEL*, *LOAEL*, MDS). The computation may involve assessment factors, e.g., for inter-species conversion, intra-species variation or additional sources of uncertainty. The calculation may also use kinetic models or absorption factors to convert external doses to internal doses or vice versa.

HI The hazard index is the sum of all hazard quotiens (*HQ*) of the substances that are associated with the same potential adverse health effect.

HQ The hazard quotient is the ratio of the exposure to a substance and the reference level at which no adverse effects are expected (i.e., exposure divided by the reference level). A HQ smaller than 1 is associated with no expected adverse health effect and a HQ larger than 1 is associated with possible adverse health effects. The HQ is closely related to the *MOE*, which can be seen as the inverse metric.

HR Highest residue. The HR is the highest residue level (expressed as mg/kg) in a composite sample of the edible portion of a food commodity when a pesticide has been used according to maximum GAP conditions. The HR is estimated as the highest of the residue values (one from each trial) from supervised trials conducted according to maximum GAP conditions, and includes residue components defined by the JMPR for estimation of dietary intake.

ICED Individual critical effect dose.

In silico Research theoretical method, particularly involving computer models, to predict the likely toxicological, or other, effects of substances.

In vitro Research method which involves testing cells or tissues extracted from living organisms.

In vivo Research method which involves testing individual live animals or populations of live animals.

IVIVE In vitro to in vivo extrapolation. Refers to the qualitative or quantitative transposition of experimental results or observations made in vitro to predict phenomena in vivo, biological organisms.

JRC Joint Research Centre

Lipid Fat and fat-like substance.

LNN Logistic normal normal model.

LOAEL Lowest observed adverse effect level.

LOD Limit of detection. Lowest concentration of a pesticide residue in a defined matrix where positive identification can be achieved using a specified method (IUPAC, 2006).

LOQ Limit of quantification. Lowest concentration of a pesticide residue in a defined matrix where positive identification and quantification measurement can be achieved using a specified analytical method (IUPAC, 2006).

LOR Limit of reporting. Practical limit of residue quantification at or above the *LOQ*. The conservative limit of quantification for a defined matrix and method which may vary between laboratories or within the one laboratory from time to time because of different equipment, techniques, and reagents. Commonly either the lower limit of the calibrated range of the method or the lowest level at which quantitative recovery of the analyse has been demonstrated (IUPAC, 2006).

MCR Maximum cumulative ratio.

MCRA Monte Carlo Risk Assessment.

MIE Molecular initiating event.

MoA Mode of Action.

MOE The margin of exposure (MOE) is the ratio between the reference level at which no adverse effects are expected to the exposure to a substance (i.e., reference level divided by exposure). Commonly, the reference level is assumed to be a point of departure (*POD*) based on animal studies that does not incorporate all factors to translate to a human reference value. A MOE is therefore typically compared to a uncertainty/safety factor (UF) composed of the product of the uncertainty factors. An MOE is smaller than the UF is associated with risk. A MOE larger than the UF is associated with no expected adverse health effects.

MOET The harmonic sum of all individual MOEs.

MRA Mixture risk assessment.

MRL Maximum residue level. Maximum concentration of a residue that is legally permitted or recognized as acceptable in,or on, a food, agricultural commodity, or animal feedstuff as set by Codex or a national regulatory authority(IUPAC, 2006).

MV Missing value

NAMs New approach methodologies. The term NAM has been emerged as a descriptive reference to any non-animal-based approaches that can be used to provide information in the context of chemical hazard and risk assessment

NMF Non-negative Matrix Factorization

NOAEL No observed adverse effect level is the greatest concentration or amount of a substance at which no detectable adverse effects occur in an exposed population.

NOEC No observed effect concentration.

OIM Observed Individual Means approach. An approach for estimating longer term exposures by taking each individual's observed mean consumption over the duration of a dietary survey.

OP Occurence pattern

PARC Partnership for the Assessment of the Risk of Chemicals

PBPK Physiologically based pharmacokinetic/toxicokinetic models.

PCPs Personal care products

POCE The probability of critical exposure (PoCE) is the proportion of the *HI* distribution above the threshold (or of the generalised *MOE* below the threshold) is the probability of critical exposure in the particular (sub)population. The threshold value can be 1 if all assessment factors have already been accounted for in the calculation of HI or MOE.

POD A point of departure is defined as a point on a toxicological dose-response curve obtained from a dose dose-response experiment in the region at which the curve transitions from no effects to effects. It is used as the base value for deriving toxicological reference values, or hazard characterisations. Common PODs are the no-observed adverse effect level (*NOAEL*) and benchmark dose (*BMD*).

PPP Plant protection products. Products used to protect, preserve or influence the growth of desirable plants or to destroy or control the growth of unwanted plants or parts of plants.

PRIMo EFSA pesticide residue intake model.

QSAR Quantitative structure activity relationship. The quantitative/qualitative structure activity relationships are a set of methods by which the effects of different compounds are related to their molecular structures. It allows the likely adverse or beneficial effects of a particular chemical to be predicted by comparing it with others which have similar structures.

RA Response Addition. An approach to the risk assessment of mixtures of substances in which responses to each of the individual components are determined and added together in order to predict the response to the mixture as a whole. This approach is only valid if the individual components do not interact with each other, i.e. their effects are completely independent.

RAC Raw agricultural commodity. Part of a crop used as a food or feed commodity directly from the harvested crop without processing.

RIVM Rijksinstituut voor Volksgesondheid en Milieu (Dutch National Institute for Public Health and the Environment). (Dutch) National Institute for Public Health and the Environment.

RPF Relative potency factor. The ratio of the toxic potency of a given chemical to that of an index chemical in the Cumulative Assessment Group (*CAG*). Relative potency factors are used to convert exposures of all chemicals in the CAG into their exposure equivalents of the index chemical.

SA Standard action

SG Specific gravity of urine

SNMU Sparse Nonnegative Matrix Underapproximation

SRA Standard Regulatory Action

SSC Source/Substance Combination

SSD EFSA Standard sample description.

TDS Total diet study. A study designed to estimate the likely consumption of harmful or beneficial substances in the diet. When undertaking such a study, commonly-consumed foods are purchased from shops in a particular country before being analysed.

TDI Tolerable daily intake. Is an estimate of the amount of a substance in food or drinking water which is not added deliberately (e.g contaminants) and which can be consumed over a lifetime without presenting an appreciable risk to health.

TEF Toxic equivalency factor.

TK Toxicokinetics. The study of the processes by which potentially toxic substances are handled in the body. This involves an understanding of the absorption, distribution, metabolism and excretion of such substances *ADME*.

TP Thermal paper.

TTC Threshold of toxicological concern. A screening tool that provides conservative exposure limits in the absence of sufficient chemical-specific toxicological data. It is a science-based approach for prioritising chemicals with low-level exposures that require more data over those that can be presumed to present no appreciable human health risk.

The tolerable weekly intake is the maximum intake of substances in food, such as nutrients or contaminants, that can be consumed weekly over a lifetime without risking adverse health effects.

Part III

Bibliography

CHAPTER

TEN

PUBLICATIONS USING MCRA

2021

- Boon, P.E., te Biesebeek, J.D., Bokkers, B.G.H., and Bulder, A.S. Herziening van de risicobeoordeling van genx en pfoa in moestuingewassen in dordrecht, papendrecht en sliedrecht. 2021. doi:10.21945/RIVM-2021-0064.
- Boon, P.E., te Biesebeek, J.D., Bokkers, B.G.H., and Bulder, A.S. Herziening van de risicobeoordeling van pfas in moestuingewassen in helmond. 2021. doi:10.21945/RIVM-2021-0071.
- Ioannidou, S., Cascio, C., and Gilsenan, M. B. European food safety authority open access tools to estimate dietary exposure to food chemicals. *Environment International*, 149:106357, 2021. URL: https://doi.org/10.1016/j.envint.2020.106357.
- Kruisselbrink, J. W., van Lenthe, M. S., van der Voet, H., de Boer, W. J., and van Klaveren, J. D. Feasibility study open mcra. *EFSA Supporting Publications*, 18:6515E, 2021. URL: https://doi.org/10.2903/sp.efsa.2021.EN-6515.
- Pustjens, A.M., Castenmiller, J.J.M., te Biesebeek, J.D., de Rijk, T.C., van Dam, R.C.J., and Boon, P.E. Dietary exposure to mycotoxins of 1- and 2-year-old children from a dutch total diet study. *World Mycotoxin Journal*, 2021. doi:https://doi.org/10.3920/WMJ2020.2676.
- Pustjens, A.M., Castenmiller, J.J.M., te Biesebeek, J.D., and Boon, P.E. Dietary intake of protein and fat of 12- to 36-month-old children in a dutch total diet study. *European Journal of Nutrition*, 2021. doi:https://doi.org/10.1007/s00394-021-02653-6.
- te Biesebeek, J., Sam, M., Sprong, R., van Donkersgoed, G., Kruisselbrink, J., de Boer, W., van Lenthe, M., van der Voet, H., and van Klaveren, J. Potential impact of prioritisation methods on the outcome of cumulative exposure assessments of pesticides. *EFSA Supporting Publications*, 18:6559E, 2021. URL: https://doi.org/10.2903/sp.efsa.2021.EN-6559.
- van Klaveren, J. D., van den Brand, A. D., van Donkersgoed, G., van der Velde-Koerts, T., van der Voet, H., Kruisselbrink, J. W., de Boer, W. J., van Lenthe, M., and Sprong, C. Proposed prospective scenarios for cumulative risk assessment of pesticide residues. *EFSA Supporting Publications*, 18:6811E, 2021. URL: https://doi.org/10.2903/sp.efsa.2021.EN-6811.
- European Food Safety Authority (EFSA), Anastassiadou, M., Choi, J., Coja, T., Dujardin, B., Hart, A., Hernandez-Jerrez, A.F., Jarrah, S., Lostia, A., Machera, K., Mangas, I., Mienne, A., Schepens, M., Widenfalk, A., and Mohimont, L. Cumulative dietary risk assessment of chronic acetylcholinesterase inhibition by residues of pesticides. *EFSA Journal*, 19(2):e06392, 2021. URL: https://doi.org/10.2903/j.efsa.2021.6392.

2020

- Beronius, A., Zilliacus, J., Hanberg, A., Luijten, M., van der Voet, H., and van Klaveren, J. Methodology for health risk assessment of combined exposures to multiple chemicals. *Food and Chemical Toxicology*, pages 111520, July 2020. URL: https://doi.org/10.1016/j.fct.2020.111520.
- Cotterill, J., Price, N., Rorije, E., and Peijnenburg, A. Development of a QSAR model to predict hepatic steatosis using freely available machine learning tools. *Food and Chemical Toxicology*, 142:111494, August 2020. URL: https://doi.org/10.1016/j.fct.2020.111494.
- Fischer, B.C., Rotter, S., Schubert, J., Marx-Stoelting, P., and Solecki, R. Recommendations for international harmonisation, implementation and further development of suitable scientific approaches regarding the assessment of mixture effects. *Food and Chemical Toxicology*, 141:111388, July 2020. URL: https://doi.org/10.1016/j.fct.2020.111388.
- Karrer, C., Andreassen, M., von Goetz, N., Sonnet, F., Sakhi, A.K., Hungerbühler, K., Dirven, H., and Husøy, T. The EuroMix human biomonitoring study: source-to-dose modeling of cumulative and aggregate exposure for the bisphenols BPA, BPS, and BPF and comparison with measured urinary levels. *Environment International*, 136:105397, March 2020. URL: https://doi.org/10.1016/j.envint.2019.105397.
- Kennedy, M.C., Hart, A.D.M., Kruisselbrink, J.W., van Lenthe, M., de Boer, W.J., van der Voet, H., Rorije, E., Sprong, C., and van Klaveren, J. A retain and refine approach to cumulative risk assessment. *Food and Chemical Toxicology*, 138:111223, April 2020. URL: https://doi.org/10.1016/j.fct.2020.111223.
- Sprong, C., Crépet, A., Metruccio, F., Blaznik, U., Anagnostopoulos, C., Christodoulou, D.L., Jensen, B.H., Kennedy, M., González, N., Rehurkova, I., Ruprich, J., te Biesebeek, J.D., Vanacker, M., Moretto, A., and van Klaveren, J. Cumulative dietary risk assessment overarching different regulatory silos using a margin of exposure approach: a case study with three chemical silos. *Food and Chemical Toxicology*, 142:111416, August 2020. URL: https://doi.org/10.1016/j.fct.2020.111416.
- Tebby, C., van der Voet, H., de Sousa, G., Rorije, E., Kumar, V., de Boer, W.J., Kruisselbrink, J.W., Bois, F.Y., Faniband, M., Moretto, A., and Brochot, C. A generic PBTK model implemented in the MCRA platform: predictive performance and uses in risk assessment of chemicals. *Food and Chemical Toxicology*, 142:111440, August 2020. URL: https://doi.org/10.1016/j.fct.2020.111440.
- van den Brand, A.D., Beukers, M., Niekerk, M., van Donkersgoed, G., van der Aa, M., van de Ven, B., Bulder, A., van der Voet, H., and Sprong, C.R. Assessment of the combined nitrate and nitrite exposure from food and drinking water: application of uncertainty around the nitrate to nitrite conversion factor. *Food Additives & Contaminants: Part A*, 37(4):568–582, January 2020. URL: https://doi.org/10.1080/19440049.2019.1707294.
- van der Voet, H., Kruisselbrink, J.W., de Boer, W.J., van Lenthe, M.S., van den Heuvel, J.J.B., Crépet, A., Kennedy, M.C., Zilliacus, J., Beronius, A., Tebby, C., Brochot, C., Luckert, C., Lampen, A., Rorije, E., Sprong, C., and van Klaveren, J.D. The MCRA toolbox of models and data to support chemical mixture risk assessment. *Food and Chemical Toxicology*, 138:111185, April 2020. URL: https://doi.org/10.1016/j.fct.2020.111185.
- Vanacker, M., Quindroit, P., Angeli, K., Mandin, C., Glorennec, P., Brochot, C., and Crépet, A. Aggregate and cumulative chronic risk assessment for pyrethroids in the French adult population. *Food and Chemical Toxicology*, 143:111519, September 2020. URL: https://doi.org/10.1016/j.fct.2020.111519.
- Vlachou, C., Hofstädter, D., Rauscher-Gabernig, E., Griesbacher, A., Fuchs, K., and König, J. Risk assessment of nitrites for the Austrian adult population with probabilistic modelling of the dietary exposure. *Food and Chemical Toxicology*, 143:111480, September 2020. URL: https://doi.org/10.1016/j.fct.2020.111480.
- European Food Safety Authority (EFSA), Craig, P.S., Dujardin, B., Hart, A., Hernández-Jerez, A.F., Hougaard Bennekou, S., Kneuer, C., Ossendorp, B., Pedersen, R., Wolterink, G., and Mohimont, L. Cumulative dietary risk characterisation of pesticides that have acute effects on the nervous system. *EFSA Journal*, 18(4):e06087, 2020. URL: https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/j.efsa.2020.6087.
- European Food Safety Authority (EFSA), Craig, P.S., Dujardin, B., Hart, A., Hernandez-Jerez, A.F., Hougaard Bennekou, S., Kneuer, C., Ossendorp, B., Pedersen, R., Wolterink, G., and Mohimont, L. Cu-

mulative dietary risk characterisation of pesticides that have chronic effects on the thyroid. *EFSA Journal*, 18(4):e06088, 2020. URL: https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/j.efsa.2020.6088.

2019

- Bois, F.Y., Tebby, C., and Brochot, C. EuroMix PBPK model for combined exposures. 2019. URL: https://zenodo.org/record/2532334.
- Boobis, A. Report of EuroMix workshops on international harmonisation on the risk assessment of combined exposure to multiple chemicals. 2019. URL: https://zenodo.org/record/3479150.
- Boon, P.E., Van Der Aa, M., Dusseldorp, A., Janssen, P., Zeilmaker, M.J., and Schulpen, S. Loodinname via kraanwater: blootstellingsschatting en risicobeoordeling voor diverse risicogroepen. RIVM Letter report 2019-0090, 2019. URL: https://rivm.openrepository.com/handle/10029/623516.
- Boon, P.E., Zeilmaker, M.J., and Mengelers, M.J.B. Risicobeoordeling van GenX en PFOA in moestuingewassen in helmond. RIVM Letter report 2019-0024, 2019. URL: https://rivm.openrepository.com/handle/10029/622988.
- Boon, P.E., Van Donkersgoed, G., Van Der Vossen, W., Sam, M., Noordam, M.Y., and Van Der Schee, H. Tussenevaluatie van de nota 'gezonde groei, duurzame oogst'. RIVM Letter report 2018-0127, 2019. URL: https://rivm.openrepository.com/handle/10029/623125.
- Crépet, A., Vanacker, M., Sprong, C., de Boer, W.J., Blaznik, U., Kennedy, M., Anagnostopoulos, C., Christodoulou, D.L., Ruprich, J., Rehurkova, I., Domingo, J.L., Jensen, B.H., Metruccio, F., Moretto, A., Jacxsens, L., Spanoghe, P., Senaeve, D., van der Voet, H., and van Klaveren, J. Selecting mixtures on the basis of dietary exposure and hazard data: application to pesticide exposure in the European population in relation to steatosis. *International Journal of Hygiene and Environmental Health*, 222(2):291–306, March 2019. URL: https://doi.org/10.1016/j.ijheh.2018.12.002.
- de Rop, J., Senaeve, D., Jacxsens, L., Houbraken, M., van Klaveren, J., and Spanoghe, P. Cumulative probabilistic risk assessment of triazole pesticides in Belgium from 2011-2014. *Food Additives & Contaminants: Part A*, 36(6):911–921, April 2019. URL: https://doi.org/10.1080/19440049.2019.1606943.
- Fischer, B., Schubert, J., Rotter, S., and Solecki, R. Specific recommendations regarding implementation of mechanism-based test strategy for harmonised cumulative risk assessment according oecd, who, efsa and EuroMix guidance. 2019. URL: https://zenodo.org/record/3490547.
- Heinemeyer, G., Jantunen, M., and Hakkinen, P. *The Practice of Consumer Exposure Assessment*. Springer International Publishing, 2019. URL: https://doi.org/10.1007/978-3-319-96148-4.
- Karrer, C., de Boer, W.J., Delmaar, C., Cai, Y., Crépet, A., Hungerbühler, K., and von Goetz, N. Linking probabilistic exposure and pharmacokinetic modeling to assess the cumulative risk from the bisphenols BPA, BPS, BPF, and BPAF for Europeans. *Environmental Science & Technology*, 53(15):9181–9191, July 2019. URL: https://doi.org/10.1021/acs.est.9b01749.
- Kennedy, M., Hart, A., Kruisselbrink, J.W., van Lenthe, M., de Boer, W., van der Voet, H., Rorije, E., Sprong, C., and van Klaveren, J. Methodology and results of the retain and refine approach. 2019. URL: https://zenodo.org/record/3465690.
- Kennedy, M.C., Garthwaite, D.G., de Boer, W.J., and Kruisselbrink, J.W. Modelling aggregate exposure to pesticides from dietary and crop spray sources in UK residents. *Environmental Science and Pollution Research*, 26(10):9892–9907, February 2019. URL: https://doi.org/10.1007/s11356-019-04440-7.
- Kolbaum, A.E., Berg, K., Müller, F., Kappenstein, O., and Lindtner, O. Dietary exposure to elements from the German pilot total diet study (TDS). *Food Additives & Contaminants: Part A*, 36(12):1822–1836, October 2019. URL: https://doi.org/10.1080/19440049.2019.1668967.
- Sachse, B., Kolbaum, A.E., Ziegenhagen, R., Andres, S., Berg, K., Dusemund, B., Hirsch-Ernst, K.I., Kappenstein, O., Müller, F., Röhl, C., Lindtner, O., Lampen, A., and Schäfer, B. Dietary manganese exposure in the adult population in Germany—what does it mean in relation to health risks? *Molecular Nutrition & Food Research*, 63(16):1900065, July 2019. URL: https://doi.org/10.1002/mnfr.201900065.

- Tietz, T., Lenzner, A., Kolbaum, A.E., Zellmer, S., Riebeling, C., Gürtler, R., Jung, C., Kappenstein, O., Tentschert, J., Giulbudagian, M., Merkel, S., Pirow, R., Lindtner, O., Tralau, T., Schäfer, B., Laux, P., Greiner, M., Lampen, A., Luch, A., Wittkowski, R., and Hensel, A. Aggregated aluminium exposure: risk assessment for the general population. *Archives of Toxicology*, 93(12):3503–3521, October 2019. URL: https://doi.org/10.1007/s00204-019-02599-z.
- van der Voet, H., Kruisselbrink, J.W., de Boer, W.J., van Lenthe, M.S., van den Heuvel, J.J.B., Crépet, A., Kennedy, M.C., Zilliacus, J., Beronius, A., Tebby, C., Brochot, C., Rorije, E., Sprong, C., and van Klaveren, J.D. Draft paper on the EuroMix toolbox of models and data to support chemical mixture risk assessment. 2019. URL: https://zenodo.org/record/3474943.
- van der Voet, H., Kruisselbrink, J.W., de Boer, W.J., van Lenthe, M.S., van den Heuvel, J.J.B., Crépet, A., Kennedy, M.C., Zilliacus, J., Beronius, A., Rorije, E., Sprong, C., and van Klaveren, J.D. The EuroMix model toolbox MCRA 9. 2019. URL: https://zenodo.org/record/3462181.
- van Klaveren, J.D., Kruisselbrink, J.W., de Boer, W.J., van Donkersgoed, G., Biesebeek, J.D. t., Sam, M., and van der Voet, H. Cumulative dietary exposure assessment of pesticides that have acute effects on the nervous system using MCRA software. *EFSA Supporting Publications*, 16(9):1708E, 2019. URL: https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/sp.efsa.2019.EN-1708.
- van Klaveren, J.D., Kruisselbrink, J.W., de Boer, W.J., van Donkersgoed, G., Biesebeek, J.D. t., Sam, M., and van der Voet, H. Cumulative dietary exposure assessment of pesticides that have chronic effects on the thyroid using MCRA software. *EFSA Supporting Publications*, 16(9):1707E, 2019. URL: https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/sp.efsa.2019.EN-1707.
- van Lenthe, M.S., de Boer, W.J., Kruisselbrink, J.W., van der Voet, H., Crépet, A., Vanacker, M., and Trocellier, L. Validation of the EuroMix model toolbox and comparison with us software. 2019. URL: https://zenodo.org/record/3467409.
- Zilliacus, J., Rorije, E., Kennedy, M., and van Klaveren, J. Proceedings and training material from second training session for stakeholders. 2019. URL: https://zenodo.org/record/3560731.
- Zilliacus, J., Beronius, A., Hanberg, A., Luijten, M., van Klaveren, J., and van der Voet, H. EuroMix handbook for mixture risk assessment. 2019. URL: https://zenodo.org/record/3560719.
- French Agency for Food, Environmental and Occupational Health & Safety (ANSES), France, Regulated Products Assessment Department, Residues and Food Safety Unit , Chatzidimitriou, E., Mienne, A., Pierlot, S., Noel, L., and Sarda, X. Assessment of combined risk to pesticide residues through dietary exposure. *EFSA Journal*, 17(S2):e170910, 2019. URL: https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/j.efsa.2019.e170910.

2018

- Boon, P.E., Van Donkersgoed, G., Te Biesebeek, J.D., Wolterink, G., and Rietveld, A.G. Cumulative exposure to residues of plant protection products via food in the Netherlands. RIVM Letter report 2017-0018, 2018. URL: http://rivm.openrepository.com/rivm/handle/10029/622169.
- Boon, P.E., Te Biesebeek, J.D., Brants, H., Bouwmeester, M.C., and Hessel, E.V.S. Dietary sources of exposure to bisphenol A in the Netherlands. RIVM Letter report 2017-0187, 2018. URL: http://rivm.openrepository.com/rivm/handle/10029/621792.
- Jardim, A.N.O, Mello, D.C., Brito, A.P., van Donkersgoed, G., Boon, P.E., and Caldas, E.D. Dietary cumulative acute risk assessment of organophosphorus, carbamates and pyrethroids insecticides for the Brazilian population. *Food and Chemical Toxicology*, 112:108–117, February 2018. URL: https://doi.org/10.1016/j.fct.2017.12.010.
- Jardim, A.N.O, Mello, D.C., Brito, A.P., van der Voet, H., Boon, P.E., and Caldas, E.D. Probabilistic dietary risk assessment of triazole and dithiocarbamate fungicides for the Brazilian population. *Food and Chemical Toxicology*, 118:317–327, August 2018. URL: https://doi.org/10.1016/j.fct.2018.05.002.

- Mengelers, M., Te Biesebeek, J.D., Schipper, M., Slob, W., and Boon, P.E. Risicobeoordeling van GenX en PFOA in moestuingewassen in Dordrecht, Papendrecht en Sliedrecht. RIVM Letter report 2017-0017, 2018. URL: http://rivm.openrepository.com/rivm/handle/10029/621785.
- Rotter, S., Beronius, A., Boobis, A.R., Hanberg, A., van Klaveren, J., Luijten, M., Machera, K., Nikolopoulou, D., van der Voet, H., Zilliacus, J., and Solecki, R. Overview on legislation and scientific approaches for risk assessment of combined exposure to multiple chemicals: the potential EuroMix contribution. *Critical Reviews in Toxicology*, 48(9):796–814, October 2018. URL: https://doi.org/10.1080/10408444.2018.1541964.
- Suomi, J., Tuominen, P., Niinistö, S., Virtanen, S.M., and Savela, K. Dietary heavy metal exposure of Finnish children of 3 to 6 years. *Food Additives & Contaminants: Part A*, 35(7):1305–1315, June 2018. URL: https://doi.org/10.1080/19440049.2018.1480065.
- van De Ven, B.M., Fragki, S., te Biesebeek, J.D., Rietveld, A.G., and Boon, P.E. Mineral oils in food; a review of toxicological data and an assessment of the dietary exposure in the Netherlands. RIVM Letter report 2017-0018, 2018. URL: http://rivm.openrepository.com/rivm/handle/10029/622044.

2017

- Boon, P.E., te Biesebeek, J.D., and van Donkersgoed, G. Dietary exposure to lead in the Netherlands. RIVM Letter report 2016-0206, 2017. URL: https://www.rivm.nl/bibliotheek/rapporten/2016-0206.pdf.
- Presser, K., Zoom, C., Szymanek, J., and Zappa, G. Development of a pilot service for the electronic infrastructure of METROFOOD-RI. In *Proceedings of 3rd IMEKOFOODS Conference: Metrology Promoting Harmonization and Standardization in Food and Nutrition*. International Measurement Confederation, 2017. URL: https://imeko.org/publications/tc23-2017/IMEKO-TC23-2017-045.pdf.
- Sieke, C., Michalski, B., and Kuhl, T. Probabilistic dietary risk assessment of pesticide residues in foods for the German population based on food monitoring data from 2009 to 2014. *Journal of Exposure Science & Environmental Epidemiology*, 28(1):46–54, July 2017. URL: https://doi.org/10.1038/jes.2017.7.
- Sprong, R.C., Niekerk, E.M., and Beukers, M.H. Intake assessment of the food additives nitrites (e 249 and e 250) and nitrates (e 251 and e 252). RIVM Letter report 2016-0208, 2017. URL: https://www.rivm.nl/bibliotheek/rapporten/2016-0208.pdf.

2016

- Boon, P.E. and te Biesebeek, J.D. Preliminary assessment of dietary exposure to 3-MCPD in the Netherlands. RIVM Letter report 2015-0199, 2016. URL: https://www.rivm.nl/bibliotheek/rapporten/2015-0199.pdf.
- Boon, P.E., te Biesebeek, J.D., van Leeuwen, S.P.J., Zeilmaker, M.J., and Hoogenboom, L.A.P. Dietary exposure to polybrominated diphenyl ethers in the Netherlands. RIVM Letter report 2016-0037, 2016. URL: https://www.rivm.nl/bibliotheek/rapporten/2016-0037.pdf.
- Rompelberg, C., Heringa, M.B., van Donkersgoed, G., Drijvers, J., Roos, A., Westenbrink, S., Peters, R., van Bemmel, G., Brand, W., and Oomen, A.G. Oral intake of added titanium dioxide and its nanofraction from food products, food supplements and toothpaste by the Dutch population. *Nanotoxicology*, 10(10):1404–1414, September 2016. URL: https://doi.org/10.1080/17435390.2016.1222457.
- Sprong, R.C., de Wit-Bos, L., te Biesebeek, J.D., Alewijn, M., Lopez, P., and Mengelers, M.J.B. A mycotoxin-dedicated total diet study in the Netherlands in 2013: part III – exposure and risk assessment. *World Mycotoxin Journal*, 9(1):109–128, February 2016. URL: https://doi.org/10.3920/wmj2015.1905.
- Stephenson, C.L. and Harris, C.A. An assessment of dietary exposure to glyphosate using refined deterministic and probabilistic methods. *Food and Chemical Toxicology*, 95:28–41, September 2016. URL: https://doi.org/10.1016/j.fct.2016.06.026.

- van der Voet, H., de Boer, W.J., Kruisselbrink, J.W., van Donkersgoed, G., and van Klaveren, J.D. MCRA made scalable for large cumulative assessment groups. *EFSA Supporting Publications*, 13(1):910E, 2016. URL: https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/sp.efsa.2016.EN-910.

2015

- Akhandaf, Y., van Klaveren, J., de Henauw, S., van Donkersgoed, G., van Gorcum, T., Papadopoulos, A., Sirot, V., Kennedy, M., Pinchen, H., Ruprich, J., Rehurkova, I., Perelló, G., and Sioen, I. Exposure assessment within a total diet study: a comparison of the use of the pan-European classification system FoodEx-1 with national food classification systems. *Food and Chemical Toxicology*, 78:221–229, April 2015. URL: https://doi.org/10.1016/j.fct.2015.01.019.
- Blaznik, U., Yngve, A., Eržen, I., and Ribič, C.H. Consumption of fruits and vegetables and probabilistic assessment of the cumulative acute exposure to organophosphorus and carbamate pesticides of schoolchildren in Slovenia. *Public Health Nutrition*, 19(3):557–563, May 2015. URL: https://doi.org/10.1017/s1368980015001494.
- Boon, P.E. and Van der Voet, H. Probabilistic dietary exposure models relevant for acute and chronic exposure assessment of adverse chemicals via food. RIVM Letter report 2015-0191, 2015. URL: https://www.rivm.nl/bibliotheek/rapporten/2015-0191.pdf.
- Boon, P.E., van Donkersgoed, G., Christodoulou, D., Crépet, A., D'Addezio, L., Desvignes, V., Ericsson, B., Galimberti, F., Ioannou-Kakouri, E., Jensen, B.H., Rehurkova, I., Rety, J., Ruprich, J., Sand, S., Stephenson, C., Strömberg, A., Turrini, A., van der Voet, H., Ziegler, P., Hamey, P., and van Klaveren, J.D. Cumulative dietary exposure to a selected group of pesticides of the triazole group in different European countries according to the EFSA guidance on probabilistic modelling. *Food and Chemical Toxicology*, 79:13–31, May 2015. URL: https://doi.org/10.1016/j.fct.2014.08.004.
- He, D., Ye, X., Xiao, Y., Zhao, N., Long, J., Zhang, P., Fan, Y., Ding, S., Jin, X., Tian, C., Xu, S., and Ying, C. Dietary exposure to endocrine disrupting chemicals in metropolitan population from China: a risk assessment based on probabilistic approach. *Chemosphere*, 139:2–8, November 2015. URL: https://doi.org/10.1016/j.chemosphere.2015.05.036.
- Jacobs, R., van der Voet, H., and ter Braak, C.J.F. Integrated probabilistic risk assessment for nanoparticles: the case of nanosilica in food. *Journal of Nanoparticle Research*, June 2015. URL: https://doi.org/10.1007/s11051-015-2911-y.
- Kennedy, M.C., Glass, C.R., Bokkers, B., Hart, A.D.M., Hamey, P.Y., Kruisselbrink, J.W., de Boer, W.J., van der Voet, H., Garthwaite, D.G., and van Klaveren, J.D. A European model and case studies for aggregate exposure assessment of pesticides. *Food and Chemical Toxicology*, 79:32–44, May 2015. URL: https://doi.org/10.1016/j.fct.2014.09.009.
- Kennedy, M.C., van der Voet, H., Roelofs, V.J., Roelofs, W., Glass, C.R., de Boer, W.J., Kruisselbrink, J.W., and Hart, A.D.M. New approaches to uncertainty analysis for use in aggregate and cumulative risk assessment of pesticides. *Food and Chemical Toxicology*, 79:54–64, May 2015. URL: https://doi.org/10.1016/j.fct.2015.02.008.
- Kennedy, M.C., Glass, C.R., Fustinoni, S., Moretto, A., Mandic-Rajcevic, S., Riso, P., Turrini, A., van der Voet, H., Hetmanski, M.T., Fussell, R.J., and van Klaveren, J.D. Testing a cumulative and aggregate exposure model using biomonitoring studies and dietary records for Italian vineyard spray operators. *Food and Chemical Toxicology*, 79:45–53, May 2015. URL: https://doi.org/10.1016/j.fct.2014.12.012.
- Mancini, F.R., Sirot, V., Busani, L., Volatier, J.L., and Hulin, M. Use and impact of usual intake models on dietary exposure estimate and risk assessment of chemical substances: a practical example for cadmium, acrylamide and sulphites. *Food Additives & Contaminants: Part A*, 32(7):1065–1074, May 2015. URL: https://doi.org/10.1080/19440049.2015.1041428.
- Sprong, R.C. and Boon, P.E. Dietary exposure to cadmium in the Netherlands. RIVM Letter report 2015-0085, 2015. URL: https://www.rivm.nl/bibliotheek/rapporten/2015-0085.pdf.

- Suomi, J., Ranta, J., Tuominen, P., Putkonen, T., Bäckman, C., Ovaskainen, M.L., Virtanen, S.M., and Savela, K. Quantitative risk assessment on the dietary exposure of Finnish children and adults to nitrite. *Food Additives & Contaminants: Part A*, 33(1):41–53, November 2015. URL: https://doi.org/10.1080/19440049.2015.1117145.
- van der Voet, H., de Boer, W.J., Kruisselbrink, J.W., Goedhart, P.W., van der Heijden, G.W.A.M., Kennedy, M.C., Boon, P.E., and van Klaveren, J.D. The MCRA model for probabilistic single-compound and cumulative risk assessment of pesticides. *Food and Chemical Toxicology*, 79:5–12, May 2015. URL: https://doi.org/10.1016/j.fct.2014.10.014.
- van Klaveren, J.D., Kennedy, M.C., Moretto, A., Verbeke, W., van der Voet, H., and Boon, P.E. The ACROPOLIS project: its aims, achievements, and way forward. *Food and Chemical Toxicology*, 79:1–4, May 2015. URL: https://doi.org/10.1016/j.fct.2015.03.006.

2014

- Boon, P.E. Estimation of the acute dietary exposure to pesticides using the probabilistic approach and the point estimate methodology. *European Journal of Nutrition & Food Safety*, 4(1):1–3, January 2014. URL: https://doi.org/10.9734/ejnfs/2014/6899.
- Boon, P.E., van der Voet, H., Ruprich, J., Turrini, A., Sand, S., and van Klaveren, J.D. Computational tool for usual intake modelling workable at the European level. *Food and Chemical Toxicology*, 74:279–288, December 2014. URL: https://doi.org/10.1016/j.fct.2014.10.019.
- Boon, P.E., te Biesebeek, J.D., de Wit, L., and van Donkersgoed, G. Dietary exposure to dioxins in the Netherlands. RIVM Letter report 2014-0001, 2014. URL: https://www.rivm.nl/bibliotheek/rapporten/2014-0001.pdf.
- van der Voet, H., Kruisselbrink, J.W., de Boer, W.J., and Boon, P.E. Model-then-add: usual intake modelling of multimodal intake distributions. RIVM Letter report 090133001/2014, 2014. URL: http://hdl.handle.net/10029/314361.

2013

- Roodenburg, A.J.C., van Ballegooijen, A.J., Dötsch-Klerk, M., van der Voet, H., and Seidell, J.C. Modelling of usual nutrient intakes: potential impact of the Choices programme on nutrient intakes in young Dutch adults. *PLoS ONE*, 8(8):e72378, August 2013. URL: https://doi.org/10.1371/journal.pone.0072378.
- Temme, E.H.M., van der Voet, H., Thissen, J.T.N.M., Verkaik-Kloosterman, J., van Donkersgoed, G., and Nonhebel, S. Replacement of meat and dairy by plant-derived foods: estimated effects on land use, iron and SFA intakes in young Dutch adult females. *Public Health Nutrition*, 16(10):1900–1907, February 2013. URL: https://doi.org/10.1017/s1368980013000232.

2012

- Boon, P.E., te Biesebeek, J.D., Sioen, I., Huybrechts, I., Moschandreas, J., Ruprich, J., Turrini, A., Azpiri, M., Busk, L., Christensen, T., Kersting, M., Lafay, L., Liukkonen, K.-H., Papoutsou, S., Serra-Majem, L., Traczyk, I., de Henauw, S., and van Klaveren, J.D. Long-term dietary exposure to lead in young European children: comparing a pan-European approach with a national exposure assessment. *Food Additives & Contaminants: Part A*, 29(11):1701–1715, November 2012. URL: https://doi.org/10.1080/19440049.2012.709544.
- Goedhart, P.W., van der Voet, H., Knüppel, S., Dekkers, A.L.M., Dodd, K.W., Boeing, H., and van Klaveren, J. A comparison by simulation of different methods to estimate the usual intake distribution for episodically consumed foods. *EFSA Supporting Publications*, 9(6):299E, 2012. URL: https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/sp.efsa.2012.EN-299.

- Sioen, I., Fierens, T., van Holderbeke, M., Geerts, L., Bellemans, M., de Maeyer, M., Servaes, K., Vanermen, G., Boon, P.E., and de Henauw, S. Phthalates dietary exposure and food sources for Belgian preschool children and adults. *Environment International*, 48:102–108, November 2012. URL: https://doi.org/10.1016/j.envint.2012.07.004.
- van Klaveren, J.D., Goedhart, P.W., Wapperom, D., and van der Voet, H. A European tool for usual intake distribution estimation in relation to data collection by EFSA. *EFSA Supporting Publications*, 9(6):300E, 2012. URL: https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/sp.efsa.2012.EN-300.
- EFSA Panel on Plant Protection Products and their Residues (PPR). Guidance on the use of probabilistic methodology for modelling dietary exposure to pesticide residues. *EFSA Journal*, 10(10):2839, 2012. URL: https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/j.efsa.2012.2839.

2011

- Boon, P.E., Bonthuis, M., van der Voet, H., and van Klaveren, J.D. Comparison of different exposure assessment methods to estimate the long-term dietary exposure to dioxins and ochratoxin A. *Food and Chemical Toxicology*, 49(9):1979–1988, September 2011. URL: https://doi.org/10.1016/j.fct.2011.05.009.
- Noorlander, C.W., van Leeuwen, S.P.J., te Biesebeek, J.D., Mengelers, M.J.B., and Zeilmaker, M.J. Levels of perfluorinated compounds in food and dietary intake of PFOS and PFOA in the Netherlands. *Journal of Agricultural and Food Chemistry*, 59(13):7496–7505, July 2011. URL: https://doi.org/10.1021/jf104943p.
- Souverein, O.W., de Boer, W.J., Geelen, A., van der Voet, H., de Vries, J.H., Feinberg, M., and van 't Veer, P. Uncertainty in intake due to portion size estimation in 24-hour recalls varies between food groups. *The Journal of Nutrition*, 141(7):1396–1401, May 2011. URL: https://doi.org/10.3945/jn.111.139220.

2010

- Boon, P.E., te Biesebeek, J.D., Sioen, I., Huybrechts, I., de Neve, M., Amiano, P., Arganini, C., Azpiri, M., Busk, L., Christensen, T., Hilbig, A., Hirvonen, T., Koulouridaki, S., Lafay, L., Liukkonen, K.-H., Moschandreas, J., Papoutsouk, S., Ribas-Barba, L., Ruprich, J., Serra-Majem, L., Tornaritis, M., Turrini, A., Urtizberea, M., Verger, E., Westerlund, A., Kersting, M., de Henauw, S., and van Klaveren, J.D. Long-term dietary exposure to chromium in young children living in different European countries. *EFSA Supporting Publications*, 7(5):54E, 2010. URL: https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/sp.efsa.2010.EN-54.
- Boon, P.E., Sioen, I., van der Voet, H., Huybrechts, I., de Neve, M., Amiano, P., Azpiri, M., Busk, L., Christensen, T., Hilbig, A., Hirvonen, T., Koulouridaki, S., Lafay, L., Liukkonen, K.-H., Moschandreas, J., Papoutsou, S., Ribas-Barba, L., Ruprich, J., Serra-Majem, L., Tornaritis, M., Turrini, A., Urtizberea, M., Verger, E., Westerlund, A., Kersting, M., de Henauw, S., and van Klaveren, J.D. Long-term dietary exposure to lead in young children living in different European countries. *EFSA Supporting Publications*, 7(5):51E, 2010. URL: https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/sp.efsa.2010.EN-51.
- Huybrechts, I., Sioen, I., Boon, P.E., de Neve, M., Amiano, P., Arganini, C., Bower, E., Busk, L., Christensen, T., Hilbig, A., Hirvonen, T., Kafatos, A., Koulouridaki, S., Lafay, L., Liukkonen, K.-H., Papoutsou, S., Ribas-Barba, L., Ruprich, J., Rehurkova, I., Kersting, M., Serra-Majem, L., Turrini, A., Verger, E., Westerlund, A., Tornaritis, M., van Klaveren, J.D., and de Henauw, S. Long-term dietary exposure to different food colours in young children living in different European countries. *EFSA Supporting Publications*, 7(5):53E, 2010. URL: https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/sp.efsa.2010.EN-53.
- König, A., Kuiper, A.H., Marvin, H.J.P., Boon, P.E., Busk, L., Cnudde, F., Cope, S., Davies, H.V., Dreyer, M., Frewer, L.J., Kaiser, M., Kleter, G.A., Knudsen, I., Pascal, G., Prandini, A., Renn, O., Smith, M.R., Traill, B.W., van der Voet, H., van Trijp, H., Vos, E., and Wentholt, M.T.A. The SAFE FOODS framework for improved risk analysis of foods. *Food Control*, 21(12):1566–1587, December 2010. URL: https://doi.org/10.1016/j.foodcont.2010.02.012.
- on Contaminants in the Food Chain (CONTAM), E. P. Scientific opinion on lead in food. *EFSA Journal*, 8(4):1570, 2010. URL: https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/j.efsa.2010.1570.

- Sioen, I., Boon, P.E., Huybrechts, I., de Neve, M., Amiano, P., Arganini, C., Busk, L., Chadjigeorgiou, C., Christensen, T., Hilbig, A., Hirvonen, T., Koulouridaki, S., Lafay, L., Liukkonen, K.-H., Moschandreas, J., Papoutsou, S., Ribas-Barba, L., Ruprich, J., Serra-Majem, L., Turrini, A., Urtizberea, M., Kersting, M., Verger, E., Westerlund, A., van Klaveren, J.D., and de Henauw, S. Long-term dietary exposure to selenium in young children living in different European countries. *EFSA Supporting Publications*, 7(5):56E, 2010. URL: https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/sp.efsa.2010.EN-56.
- Slob, W., de Boer, W.J., and van der Voet, H. 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, January 2010. URL: https://doi.org/10.1016/j.fct.2009.09.035.
- Temme, E.H.M., van der Voet, H., Roodenburg, A.J.C., Bulder, A., van Donkersgoed, G., and van Klaveren, J. Impact of foods with health logo on saturated fat, sodium and sugar intake of young Dutch adults. *Public Health Nutrition*, 14(4):635–644, September 2010. URL: https://doi.org/10.1017/s1368980010002089.
- van Klaveren, J.D., van Donkersgoed, G., van der Voet, H., Stephenson, C., and Boon, P.E. Cumulative exposure assessment of triazole pesticides. *EFSA Supporting Publications*, 7(2):40E, 2010. URL: https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/sp.efsa.2010.EN-40.

2009

- Bokkers, B.G.H., Bakker, M.I., Boon, P.E., Bosgra, S., van der Heijden, G.W.A.M., Janer, G., Slob, W., and van der Voet, H. The practicability of the integrated probabilistic risk assessment (IPRA) approach for substances in food. RIVM Report 320121001/2009, 2009. URL: http://hdl.handle.net/10029/260367.
- Boon, P.E., Svensson, K., Moussavian, S., van der Voet, H., Petersen, A., Ruprich, J., Debegnach, F., de Boer, W.J., van Donkersgoed, G., Brera, C., van Klaveren, J.D., and Busk, L. Probabilistic acute dietary exposure assessments to captan and tolylfluanid using several European food consumption and pesticide concentration databases. *Food and Chemical Toxicology*, 47(12):2890–2898, December 2009. URL: https://doi.org/10.1016/j.fct.2009.01.040.
- Boon, P.E., Bakker, M.I., van Klaveren, J.D., and van Rossum, C.T.M. Risk assessment of the dietary exposure to contaminants and pesticide residues in young children in the Netherlands. RIVM report 35007000, 2009. URL: http://www.rivm.nl/bibliotheek/rapporten/350070002.pdf.
- Boon, P.E., van Asselt, E.D., Bakker, M.I., Kruizinga, A.G., and Jansen, M.C.J.F. Trends in diet and exposure to chemicals in Dutch children. Report 2009.002, RIKILT, Wageningen, 2009. URL: http://edepot.wur.nl/7507.
- Bos, P.M.J., Boon, P.E., van der Voet, H., Janer, G., Piersma, A.H., Brüschweiler, B.J., Nielsen, E., and Slob, W. A semi-quantitative model for risk appreciation and risk weighing. *Food and Chemical Toxicology*, 47(12):2941–2950, December 2009. URL: https://doi.org/10.1016/j.fct.2009.03.009.
- Bosgra, S., van der Voet, H., Boon, P.E., and Slob, W. An integrated probabilistic framework for cumulative risk assessment of common mechanism chemicals in food: an example with organophosphorus pesticides. *Regulatory Toxicology and Pharmacology*, 54(2):124–133, July 2009. URL: https://doi.org/10.1016/j.yrtph.2009.03.004.
- de Boer, W.J., van der Voet, H., Bokkers, B.G.H., Bakker, M.I., and Boon, P.E. Comparison of two models for the estimation of usual intake addressing zero consumption and non-normality. *Food Additives & Contaminants: Part A*, 26(11):1433–1449, November 2009. URL: https://doi.org/10.1080/02652030903161606.
- Jensen, B.H., Petersen, A., and Christensen, T. Probabilistic assessment of the cumulative dietary acute exposure of the population of Denmark to organophosphorus and carbamate pesticides. *Food Additives & Contaminants: Part A*, 26(7):1038–1048, July 2009. URL: https://doi.org/10.1080/02652030902859754.
- Müller, A.K., Bosgra, S., Boon, P.E., van der Voet, H., Nielsen, E., and Ladefoged, O. Probabilistic cumulative risk assessment of anti-androgenic pesticides in food. *Food and Chemical Toxicology*, 47(12):2951–2962, December 2009. URL: https://doi.org/10.1016/j.fct.2009.07.039.

- Muri, S.D., van der Voet, H., Boon, P.E., van Klaveren, J.D., and Brüschweiler, B.J. Comparison of human health risks resulting from exposure to fungicides and mycotoxins via food. *Food and Chemical Toxicology*, 47(12):2963–2974, December 2009. URL: https://doi.org/10.1016/j.fct.2009.03.035.
- Roodenburg, A. J. C., Temme, E. H. M., Howell Davies, O., and Seidell, J. C. Potential impact of the Choices programme on nutrient intakes in the Dutch population. *Nutrition Bulletin*, 34(3):318–323, September 2009. URL: https://doi.org/10.1111/j.1467-3010.2009.01767.x.
- Ruprich, J., Rehurkova, I., Boon, P.E., Svensson, K., Moussavian, S., van der Voet, H., Bosgra, S., van Klaveren, J.D., and Busk, L. Probabilistic modelling of exposure doses and implications for health risk characterization: glycoalkaloids from potatoes. *Food and Chemical Toxicology*, 47(12):2899–2905, December 2009. URL: https://doi.org/10.1016/j.fct.2009.03.008.
- van der Voet, H., van der Heijden, G.W.A.M., Bos, P.M.J., Bosgra, S., Boon, P.E., Muri, S.D., and Brüschweiler, B.J. A model for probabilistic health impact assessment of exposure to food chemicals. *Food and Chemical Toxicology*, 47(12):2926–2940, December 2009. URL: https://doi.org/10.1016/j.fct.2008.12.027.
- van Ooijen, H.J., van der Voet, H., and Bakker, M.I. Identification and handling of uncertainties in dietary exposure assessment. RIVM Report 320103004, 2009. URL: http://hdl.handle.net/10029/261706.
- EFSA Panel on Contaminants in the Food Chain (CONTAM). Scientific opinion on arsenic in food. *EFSA Journal*, 7(10):1351, 2009. URL: https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/j.efsa.2009.1351.
- EFSA Panel on Plant Protection Products and their Residues (PPR Panel). Scientific opinion on risk assessment for a selected group of pesticides from the triazole group to test possible methodologies to assess cumulative effects from exposure through food from these pesticides on human health. *EFSA Journal*, 7(9):1167, 2009. URL: https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/j.efsa.2009.1167.

2008

- Boon, P.E., der Voet, H. V., Raaij, M.T.M. V., and Klaveren, J.D. V. Cumulative risk assessment of the exposure to organophosphorus and carbamate insecticides in the Dutch diet. *Food and Chemical Toxicology*, 46(9):3090–3098, September 2008. URL: https://doi.org/10.1016/j.fct.2008.06.083.
- Brantsæter, A.L., Haugen, M., de Mul, A., Bjellaas, T., Becher, G., van Klaveren, J., Alexander, J., and Meltzer, H.M. Exploration of different methods to assess dietary acrylamide exposure in pregnant women participating in the Norwegian mother and child cohort study (MoBa). *Food and Chemical Toxicology*, 46(8):2808–2814, August 2008. URL: https://doi.org/10.1016/j.fct.2008.05.020.
- de Mul, A., Bakker, M.I., Zeilmaker, M.J., Traag, W.A., van Leeuwen, S.P.J., Hoogenboom, R.L.A.P., Boon, P.E., and van Klaveren, J.D. Dietary exposure to dioxins and dioxin-like PCBs in the Netherlands anno 2004. *Regulatory Toxicology and Pharmacology*, 51(3):278–287, August 2008. URL: https://doi.org/10.1016/j.yrtph.2008.04.010.
- Jensen, B.H., Andersen, J.H., Petersen, A., and Christensen, T. Dietary exposure assessment of Danish consumers to dithiocarbamate residues in food: a comparison of the deterministic and probabilistic approach. *Food Additives & Contaminants: Part A*, 25(6):714–721, June 2008. URL: https://doi.org/10.1080/02652030701858262.
- Seal, C.J., de Mul, A., Eisenbrand, G., Haverkort, A.J., Franke, K., Lalljie, S.P.D., Mykkänen, H., Reimerdes, E., Scholz, G., Somoza, V., Tuijtelaars, S., van Boekel, M., van Klaveren, J., Wilcockson, S.J., and Wilms, L. Risk-benefit considerations of mitigation measures on acrylamide content of foods – a case study on potatoes, cereals and coffee. *British Journal of Nutrition*, 99(S2):S1–S46, April 2008. URL: https://doi.org/10.1017/s0007114508965314.

2007

- Bakker, M.I., de Winter-Sorkina, R., de Mul, A., Boon, P.E., van Donkersgoed, G., van Klaveren, J.D., Baumann, B.A., Hijman, W.C., van Leeuwen, S.P.J., de Boer, W., and Zeilmaker, M.J. Dietary intake and risk evaluation of polybrominated diphenyl ethers in the Netherlands. *Molecular Nutrition & Food Research*, 52(2):204–216, December 2007. URL: https://doi.org/10.1002/mnfr.200700112.
- Boon, P.E., Ragas., A.M.J., and van Klaveren, J.D. Exploration of aggregate exposure to compounds present in food. Report 2007.016, RIKILT, Wageningen, 2007. URL: http://www.rikilt.wur.nl/NL/publicaties/Rapporten.
- de Winter-Sorkina, R., Bakker, M.I., Wolterink, G., and Zeilmaker, M.J. Brominated flame retardants: occurrence, dietary intake and risk assessment. RIVM report 320100002/2006, 2007. URL: http://rivm.openrepository.com/rivm/handle/10029/7303.
- van der Voet, H. and Slob, W. Integration of probabilistic exposure assessment and probabilistic hazard characterization. *Risk Analysis*, 27(2):351–371, April 2007. URL: https://doi.org/10.1111/j.1539-6924.2007.00887.x.
- European Food Safety Authority (EFSA). Opinion of the scientific panel on plant protection products and their residues on acute dietary intake assessment of pesticide residues in fruit and vegetables. *EFSA Journal*, 5(8):538, 2007. URL: https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/j.efsa.2007.538.

2006

- Caldas, E.D., Tressou, J., and Boon, P.E. Dietary exposure of Brazilian consumers to dithiocarbamate pesticides—a probabilistic approach. *Food and Chemical Toxicology*, 44(9):1562–1571, September 2006. URL: https://doi.org/10.1016/j.fct.2006.04.014.
- Caldas, E.D., Boon, P.E., and Tressou, J. Probabilistic assessment of the cumulative acute exposure to organophosphorus and carbamate insecticides in the Brazilian diet. *Toxicology*, 222(1-2):132–142, May 2006. URL: https://doi.org/10.1016/j.tox.2006.02.006.
- van Klaveren, J.D., Noordam, M.Y., Boon, P.E., van Donkersgoed, G., Ossendorp, B.C., van Raaij, M.T.M., and van der Roest, J.G. Trends in normoverschrijdigen, overschrijdingen van de acute referentiewaarde en gesommeerde blootstelling - tussenevaluatie nota duurzame gewasbescherming - deelrapport voedselveiligheid. Report 2006.011, RIKILT, Wageningen, 2006. URL: http://edepot.wur.nl/24544.

2005

- Boon, P.E., de Mul, A., van der Voet, H., van Donkersgoed, G., Brette, M., and van Klaveren, J.D. Calculations of dietary exposure to acrylamide. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 580(1-2):143–155, February 2005. URL: https://doi.org/10.1016/j.mrgentox.2004.10.014.
- de Mul, A., de Winter-Sorkina, R., Boon, P.E., van Donkersgoed, G., Bakker, M.I., and van Klaveren, J.D. Dietary intake of brominated diphenyl ether congeners by the Dutch population. Report 2005.006, RIKILT, Wageningen, 2005. URL: http://edepot.wur.nl/26982.
- Paulo, M.J., van der Voet, H., Jansen, M.J.W., ter Braak, C.J.F., and van Klaveren, J.D. Risk assessment of dietary exposure to pesticides using a Bayesian method. *Pest Management Science*, 61(8):759–766, 2005. URL: https://doi.org/10.1002/ps.1060.
- Schothorst, R.C., van Egmond, H.P., de Mul, A., Boon, P.E., van Klaveren, J.D., and Speijers, G.J.A. Trichothecenes in baby food. RIVM Report 310301002, 2005. URL: http://www.rivm.nl/bibliotheek/rapporten/310301002.pdf.

2004

- Boon, P.E., Tjoe Nij, E.I.M., Koopman, N., and van Klaveren, J.D. Dietary habits and exposure to pesticides in Dutch infants. Report 2004.017, RIKILT, Wageningen, 2004. URL: http://edepot.wur.nl/44408.
- Boon, P.E., Lignell, S., van Klaveren, J.D., and Tjoe Nij, E.I.M. Estimation of the acute dietary exposure to pesticides using the probabilistic approach and the point estimate methodology - the generation of work examples using food consumption data from the Netherlands and Sweden. Report 2004.008, RIKILT, Wageningen, 2004. URL: http://edepot.wur.nl/28647.
- Boon, P.E., Tjoe Nij, E.I.M., van Donkersgoed, G., and van Klaveren, J.D. Probabilistic intake calculations performed for the codex committee on pesticide residues. Report 2004.005, RIKILT, Wageningen, 2004. URL: http://edepot.wur.nl/36066.
- van der Voet, H. and Paulo, M.J. Some explorations into Bayesian modelling of risks due to pesticide intake from food. In van Boekel, M.A.J.S., Stein, A., and van Bruggen, A.H.C., editors, *Bayesian statistics and quality modelling in the agro-food production chain*, pages 145–162. Kluwer, Dordrecht, 2004. URL: http://library.wur.nl/frontis/bayes/13_van_der_voet.pdf.

2003

- Boon, P.E. and van Klaveren, J.D. Cumulative exposure to acetylcholineterase inhibiting compounds in the Dutch population and young childeren. Report 2003.003, RIKILT, Wageningen, 2003. URL: http://edepot.wur.nl/30057.
- Boon, P.E. and van Klaveren, J.D. Dietary exposure to pesticides - relevant variables and probabilistic modelling. Report 2003.008, RIKILT, Wageningen, 2003. URL: http://edepot.wur.nl/23045.
- Boon, P.E., van der Voet, H., and van Klaveren, J.D. Validation of a probabilistic model of dietary exposure to selected pesticides in Dutch infants. *Food Additives and Contaminants*, 20(sup001):S36–S49, October 2003. URL: https://doi.org/10.1080/0265203031000134956.
- de Winter-Sorkina, R., Bakker, M.I., van Donkersgoed, G., and van Klaveren, J.D. Dietary intake of brominated flame retardants by the Dutch population. Report 2003.019, RIKILT, Wageningen, 2003. URL: http://hdl.handle.net/10029/7303.
- de Winter-Sorkina, R., van Donkersgoed, G., Bakker, M.I., and van Klaveren, J.D. Dietary intake of heavy metals (cadmium, lead and mercury) by the Dutch population. Report 2003.016, RIKILT, Wageningen, 2003. URL: http://edepot.wur.nl/41597.
- Gibney, M.J. and van der Voet, H. Introduction to the Monte Carlo project and the approach to the validation of probabilistic models of dietary exposure to selected food chemicals. *Food Additives and Contaminants*, 20(sup001):S1–S7, October 2003. URL: https://doi.org/10.1080/0265203031000134947.
- van der Voet, H., Boon, P.E., and van Klaveren, J.D. Validation of Monte Carlo models for estimating pesticide intake of Dutch infants. Report 2003.002, RIKILT, Wageningen, 2003. URL: http://edepot.wur.nl/39363.

2002

- Boon, P.E., van Donkersgoed, G., and van Klaveren, J.D. Human acute exposure assessment of pesticides in fruits and vegetables. Report 2002.002, RIKILT, Wageningen, 2002. URL: https://library.wur.nl/WebQuery/wurpubs/reports/320297.

CHAPTER

ELEVEN

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, version 9.2.8, build 9.2.8.0 for Prod-RIVM on Tuesday April 25, 2023, 11:25:02.

11.1 Contributors to MCRA

MCRA development team:

Waldo de Boer, Johannes Kruisselbrink, Marco van Lenthe, Hans van den Heuvel, Hilko van der Voet, Tijmen van Voorthuijsen, Jasper Engel

Many people contributed to the MCRA code over the years:

Frits van Evert, Jack van Galen, Paul Goedhart, Gerie van der Heijden, Paul Keizer, Marcel Koenders, Jaap Kokorian, Sanne Korzec, Helen Owen, Gerrit Polder, Pim Reijersen, Willem Roelofs, Gert-Jan Swinkels, Jac Thissen.

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, Annick van den Brand, Gerda van Donkersgoed, Corinne Sprong, Trijntje van der Velde-Koerts, Jacob van Klaveren.

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

CHAPTER

TWELVE

CHANGE LOG

12.1 Version 9.2.8 (2023-04-25)

12.1.1 Added

- Update HTML report in zip download with toc-functionality to make it easier to browse through (#1409)
- Risks: Implement cumulative as sum of single-substance ratios (Exposure/Hazard) (#1528)
- Add documentation core and web (#1448)

12.1.2 Fixed

- Data sources view is not updated/refreshed after upload new data source version (#1512)
- User name input check for reset password is not correct (#1594)
- Setting Multiple effects gives error (#1595)

12.2 Version 9.2.7 (2023-04-03)

12.2.1 Added

- Separate substance-weighing option for MCR and mixtures analysis and add checkbox to compute or not compute MCR in HBM analysis actions (#1499)
- Show 'Created' date in admin users panel (#1523)
- Add standardisation methods for blood and urine (#1359, #1477)
- Implement initial version of SRA acute CRA of craniofacial alterations EFSA 2022 (#1464)

12.2.2 Changed

- Remove MCRA version select page, MCRA 9 is now the default
- Update exposure mixtures module with option to log-transform data before network analysis (#1531)

12.2.3 Fixed

- Hide option for reference substance equivalents when cumulative does not apply for the action (#1502)

12.3 Version 9.2.6 (2023-03-10)

12.3.1 Added

- Add compartments including cumulative amounts like urine (#1272)
- Allow hbm as input for risks (#1394)
- Add documentation about allocation of substances (#1473)
- Add new model substance approvals (#1461)

12.3.2 Changed

- Remove service worker functionality from Angular app
- Change MCRA URLs to point to static Documentation URL (#889)
- Move and update documentation section on HBM4EU/PARC HBM data format (#1427)
- Update Angular version to v15.0 (#1468)
- Unique constraint of Hazard Characterisations table is too strict (#1491)
- Keep report name tabs visible when scrolling through report comparison view

12.3.3 Fixed

- File upload failed: max upload file size was not configured correctly
- Fixed subset range editor
- Transaction deadlocked on lock resources in Job scheduler and simulation worker (#1487)
- Job scheduler does not assign jobs correctly (#1489)
- Null reference exception when changing data source (#1501)
- Exception when trying to view task settings in admin tasks panel
- Report comparison fails to load
- Can't compare outputs of a loop task in tabbed output view, TOC doesn't work (#1509, #1510)
- Error messages are not displayed correctly in web application
- Missing user account when e-mail verification fails (#1520)

12.4 Version 9.2.5 (2023-02-10)

12.4.1 Added

- Documentation: added section on CLI (#1450)

12.4.2 Changed

- Update third party packages (NuGet) for MCRA Web

12.4.3 Fixed

- Failed dataset upload still creates a data source record in the repository (#1483)

12.5 Version 9.2.4 (2023-02-03)

12.5.1 Added

- Add documentation of HBM4EU/PARC HBM data format (#1427)
- Add maximum percentage missing value percentage imputation hbm data (#1470)
- Add population characteristics Real Life Mixtures (#1475)

12.5.2 Changed

- Update HBM analysis module to include imputation of missing data method established within HBM4EU/PARC RLM (#1397)
- Update HBM data/analysis module to allow for analysis of concentrations of multiple sampling methods/matrices (#1398)
- Implement changed rules for sending automatic mail from WUR and RIVM (#1466)
- Move internalconcentrationtype to assessment settings

12.5.3 Fixed

- Download action + data (data as zipped csv) fails (#1451)
- Loop task output doesn't show the comparison between the looped actions (#1459)
- New user registration in MCRA returns 400 Bad Request (#1465)
- Load correct png mime type for safety chart
- Upload of Demo HBM bisphenols standard action data fails (#1472)

12.6 Version 9.2.3 (2023-01-16)

12.6.1 Added

- Implement chlorpyrifos kinetic model with metabolites (#1285)

12.6.2 Changed

- Removed unit test report results from admin page (#1441)
- Allow links to MCRA documentation on other servers (#1457)

12.6.3 Fixed

- Documentation shows no version of MCRA in colophon
- Documentation links to web API documentation and web application
- Can't view info in data browser of a shared data source (#1458)

12.7 Version 9.2.2 (2022-12-20)

12.7.1 Added

- Documentation: Add PARC Real-life mixture guidance documents (pdf) and related data to the User guide - Examples section (#1366)

12.7.2 Changed

- Update HBM analysis module with non-detects imputation method according to method established within HBM4EU/PARC RLM (#1396)

12.7.3 Fixed

- Fix reference to dietary exposures section in risks output section (#1331)
- MCRA Build date-time not displayed correctly (#1410)
- Settings in overview only show headers and no settings (#1442)
- Mixtures crashes when no concentrations for HBM data are available (#1443)
- Output collection of task with subtask fails (#1444)
- MCRA settings loading fails: allow floating point literals (NaN, Inf etc) in Json serialization

BIBLIOGRAPHY

[Abramowitz and Stegun (1972)] Abramowitz, M. and Stegun, I. A. Handbook of mathematical functions. *National Bureau of Standards Applied Mathematics Series*, 55:589–626, 1972.

[Bernert et al. (2007)] Bernert, J. T., Turner, W. E., Patterson, D. G., and Needham, L. L. Calculation of serum "total lipid" concentrations for the adjustment of persistent organohalogen toxicant measurements in human samples. *Chemosphere*, 68(5):824–831, 2007. URL: https://www.sciencedirect.com/science/article/pii/S0045653507002664.

[Boon et al. (2017)] Boon, P.E, van Donkersgoed, G., and te Biesebeek, J.D. Dietary exposure to lead in the netherlands. *RIVM Letter report*, 2017. URL: https://www.rivm.nl/bibliotheek/rapporten/2016-0206.pdf.

[Boon et al. (2022)] Boon, P.E., Pustjens, A.M., te Biesebeek, J.D., Brust, G.M.H., and Castenmiller, J.J.M. Dietary intake and risk assessment of elements for 1- and 2-year-old children in the netherlands. *Food and Chemical Toxicology*, 161:112810, 2022. URL: https://www.sciencedirect.com/science/article/pii/S0278691522000047.

[Bopp et al. (2015)] Bopp, S., Berggren, E., Kienzler, A., van der Linden, S., and Worth, A. 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.

[Box and Cox (1964)] Box, G. E. and Cox, D. R. An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2):211–243, 1964.

[Butler Ellis et al. (2018)] Butler Ellis, M.C., Kennedy, M. C., Kuster, C.J., Alanis, R., and Tuck, C.R. 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.

[Béchaux et al. (2013)] Béchaux, C., Zetlaoui, M., Tressou, J., Leblanc, J.-C., Héraud, F., and Crépet, A. Identification of pesticide mixtures and connection between combined exposure and diet. *Food and chemical toxicology*, 59:191–198, 2013. doi:10.1016/j.fct.2013.06.006.

[Cramer et al. (1976)] Cramer, G.M., Ford, R.A., and Hall, R.L. 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.

[Crépet et al. (2022)] Crépet, A., Vasseur, P., Jean, J., Badot, P.-M., Nesslany, F., Vernoux, J.-P., Feidt, C., and Mhaouty-Kodja, S. Integrating selection and risk assessment of chemical mixtures: a novel approach applied to a breast milk survey. *Environmental Health Perspectives*, 130(3):035001, 2022. doi:10.1289/EHP8262.

[Dahlquist and Bjorck (1974)] Dahlquist, G. and Bjorck, A. Numerical methods (transl. by n. anderson). 1974.

[de Boer and van der Voet (2011)] de Boer, W. J. 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.

[de Boer et al. (2009)] de Boer, W. J., van der Voet, H., Bokkers, B. G., Bakker, M. I., and Boon, P. E. 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.

[Dodd (1996)] Dodd, K. A technical guide to c-side. *Ames, Iowa: Department of Statistics and Center for Agricultural and Rural Development, Iowa State University*, 1996.

[EC (2018)] European Commission Standing Committee on Plants Animals Food and Feed. European commission working document sante-2015-10216 rev. 7. 2018.

[Efron (1979)] Efron, B. Bootstrap methods: another look at the jackknife annals of statistics 7: 1–26. *View Article PubMed/NCBI Google Scholar*, 1979.

[Efron and Tibshirani (1993)] Efron, B. and Tibshirani, R. J. An introduction to the bootstrap chapman & hall. *New York*, 1993.

[EFSA (2011a)] 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.

[EFSA (2011b)] European Food Safety Authority (EFSA). The food classification and description system foodex 2 (draft-revision 1). *EFSA Journal*, pages 438, 2011.

[EFSA (2012)] 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:10.2903/j.efsa.2012.2839.

[EFSA (2014)] 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.

[EFSA (2017a)] European Food Safety Authority (EFSA), Buist, H., Craig, P., Dewhurst, I., Hougaard Bennekou, S., Kneuer, C., Machera, K., Pieper, C., Court Marques, D., Guillot, G., Ruffo, F., and Chiusolo, A. Guidance on dermal absorption. *EFSA Journal*, 15(6):e04873, 2017. doi:10.2903/j.efsa.2017.4873.

[EFSA (2017b)] EFSA Panel on Contaminants in the Food Chain (CONTAM), Knutsen, H. K., Alexander, J., Barregård, L., Bignami, M., Brüschweiler, B., Ceccatelli, S., Cottrill, B., Dinovi, M., Edler, L., Grasl-Kraupp, B., Hogstrand, C., Hoogenboom, L. (Ron), Nebbia, C. S., Oswald, I. P., Petersen, A., Rose, M., Roudot, Alain-Claude, Schwerdtle, T., Vleminckx, C., Vollmer, G., Wallace, H., Ruiz, J. A. G., and Binaglia, M. Risks for human health related to the presence of pyrrolizidine alkaloids in honey, tea, herbal infusions and food supplements. *EFSA Journal*, 7 2017. doi:10.2903/j.efsa.2017.4908.

[EFSA (2018)] European Food Safety Authority (EFSA), Brancato, A., Brocca, D., Ferreira, L., Greco, L., Jarrah, S., Leuschner, R., Medina, P., Miron, I., Nougadere, A., Pedersen, R., Reich, H., Santos, M., Stanek, A., Tarazona, J., Theobald, A., and Villamar-Bouza, L. Use of efsa pesticide residue intake model (efsa primo revision 3). *EFSA Journal*, 16(1):e05147, 2018. URL: https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/j.efsa.2018.5147.

[EFSA (2019)] European Food Safety Authority (EFSA), Dujardin, B., and Kirwan, L. The raw primary commodity (RPC) model: strengthening EFSA's capacity to assess dietary exposure at different levels of the food chain, from raw primary commodities to foods as consumed. *EFSA Supporting Publications*, 16(1):1532E, 2019. URL: https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/sp.efsa.2019.EN-1532.

[EFSA (2020a)] European Food Safety Authority (EFSA), Craig, P. S., Dujardin, B., Hart, A., Hernández-Jerez, A. F., Hougaard Bennekou, S., Kneuer, C., Ossendorp, B., Pedersen, R., Wolterink, G., and Mohimont, L. Cumulative dietary risk characterisation of pesticides that have acute effects on the nervous system. *EFSA Journal*, 18(4):e06087, 2020. URL: https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/j.efsa.2020.6087.

[EFSA (2020b)] European Food Safety Authority (EFSA), Craig, P. S., Dujardin, B., Hart, A., Hernandez-Jerez, A. F., Hougaard Bennekou, S., Kneuer, C., Ossendorp, B., Pedersen, R., Wolterink, G., and Mohimont, L. Cumulative dietary risk characterisation of pesticides that have chronic effects on the thyroid. *EFSA Journal*, 18(4):e06088, 2020. URL: https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/j.efsa.2020.6088.

[Friedman et al. (2008)] Friedman, J., Hastie, T., and Tibshirani, R. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

[Gillis and Plemmons (2013)] Gillis, N. and Plemmons, R. J. Sparse nonnegative matrix underapproximation and its application to hyperspectral image analysis. *Linear Algebra and its Applications*, 438(10):3991–4007, 2013.

[Goedhart et al. (2012)] Goedhart, P. W., van der Voet, H., Knüppel, S., Dekkers, A. L.M., Dodd, K. W., Boeing, H., and van Klaveren, J. D. 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.

[Goodhardt et al. (1984)] Goodhardt, G. J., Ehrenberg, A. S., and Chatfield, C. The dirichlet: a comprehensive model of buying behaviour. *Journal of the Royal Statistical Society. Series A (General)*, pages 621–655, 1984.

[Hoyer (2004)] Hoyer, P. O. Non-negative matrix factorization with sparseness constraints. *Journal of machine learning research*, 5(Nov):1457–1469, 2004.

[Husøy et al. (2019)] Husøy, T., Andreassen, M., Hjertholm, H., Carlsen, M.H., Norberg, N., Sprong, C., Papadopoulou, E., Sakhi, A.K., Sabaredzovic, A., and Dirven, H.A.A.M. The norwegian biomonitoring study from the eu project euromix: levels of phenols and phthalates in 24-hour urine samples and exposure sources from food and personal care products. *Environment International*, 132:105103, 2019. URL: https://www.sciencedirect.com/science/article/pii/S0160412019306944.

[Jäckel (2005)] Jäckel, P. A note on multivariate gauss-hermite quadrature. *London: ABN-Amro. Re*, 2005.

[Karrer et al. (2018)] Karrer, C., Roiss, T., von Goetz, N., Skledar, D. G., Mašič, L. P., and Hungerbühler, K. Physiologically based pharmacokinetic (pbpk) modeling of the bisphenols bpa, bps, bpf, and bpaf with new experimental metabolic parameters: comparing the pharmacokinetic behavior of bpa with its substitutes. *Environmental Health Perspectives*, 126(7):077002, 2018. URL: https://ehp.niehs.nih.gov/doi/abs/10.1289/EHP2739.

[Karrer et al. (2019)] Karrer, Cecile, Boer, W. d., Delmaar, C., Cai, Y., Crépet, A., Hungerbühler, K., and Goetz, N. v. Linking probabilistic exposure and pharmacokinetic modeling to assess the cumulative risk from the bisphenols bpa, bps, bpf, and bpaf for europeans. *Environmental science & technology*, 53(15):9181–9191, 2019. doi:10.1021/acs.est.9b01749.

[Karrer et al. (2020)] Karrer, C., Andreassen, M., von Goetz, N., Sonnet, F., Sakhi, A. K., Hungerbühler, K., Dirven, H., and Husøy, T. The euromix human biomonitoring study: source-to-dose modeling of cumulative and aggregate exposure for the bisphenols bpa, bps, and bpf and comparison with measured urinary levels. *Environment International*, 136:105397, 2020. URL: https://www.sciencedirect.com/science/article/pii/S0160412019324080.

[Kennedy and Butler Ellis (2017)] Kennedy, M. C. and Butler Ellis, M.C. 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.

[Kennedy et al. (2012)] Kennedy, M. C., Butler Ellis, C. M.J., and Miller, P. C.H. 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.

[Kennedy et al. (2015a)] Kennedy, M. C., Glass, C. R., Bokkers, B., Hart, A. D., Hamey, P. Y., Kruisselbrink, J. W., de Boer, W. J., van der Voet, H., Garthwaite, D. G., and van Klaveren, J. D. A european model and case studies for aggregate exposure assessment of pesticides. *Food and Chemical Toxicology*, 79:32–44, 2015.

[Kennedy et al. (2015b)] Kennedy, M. C., van der Voet, H., Roelofs, V. J., Roelofs, W., Glass, C. R., de Boer, W. J., Kruisselbrink, J. W., and Hart, A. D.M. New approaches to uncertainty analysis for use in aggregate and cumulative risk assessment of pesticides. *Food and Chemical Toxicology*, 79:54–64, 2015.

[Kennedy et al. (2020)] Kennedy, M. C., Hart, A. D.M., Kruisselbrink, J. W., van Lenthe, M., de Boer, W. J., van der Voet, H., Rorije, E., Sprong, C., and van Klaveren, J. A retain and refine approach to cumulative risk assessment. *Food and Chemical Toxicology*, April 2020. doi:10.1016/j.fct.2020.111223.

[Kipnis et al. (2009)] Kipnis, V., Midthune, D., Buckman, D. W., Dodd, K. W., Guenther, P. M., Krebs-Smith, S. M., Subar, A. F., Tooze, J. A., Carroll, R. J., and Freedman, L. S. 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.

[Lee and Seung (1999)] Lee, D. D. and Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788, 1999.

[Lee et al. (2015)] Lee, J.-G., Kim, S.-H., Kim, H.-J., and Yoon, H.-J. Total diet studies as a tool for ensuring food safety. *Toxicological research*, 31(3):221–226, 2015. doi:doi:10.5487/TR.2015.31.3.221.

[Merz and Schrenk (2016)] Merz, K.-H. and Schrenk, D. Interim relative potency factors for the toxicological risk assessment of pyrrolizidine alkaloids in food and herbal medicines. *Toxicology Letters*, 263:44–57, 2016. doi:10.1016/j.toxlet.2016.05.002.

[Mood et al. (1974)] Mood, A. M., Graybill, F. A., and Boes, D. C. *Introduction to the Theory of Statistics 1974*. McGraw-Hill Kogakusha, 1974.

[Mulder et al. (2015)] Mulder, P. P.J., Sánchez, P. L., These, A., Preiss-Weigert, A., and Castellari, M. Occurrence of pyrrolizidine alkaloids in food. *EFSA Supporting Publications*, 12(8):859E, 2015. doi:10.2903/sp.efsa.2015.EN-859.

[Munro et al. (1996)] Munro, I. C., Ford, R. A., Kennepohl, E., and Sprenger, J. G. 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.

[Murtagh and Legendre (2014)] Murtagh, F. and Legendre, P. Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion? *Journal of Classification*, 31(3):274–295, 2014. doi:10.1007/s00357-014-9161-z.

[Nusser et al. (1996)] Nusser, S. M., Carriquiry, A. L., Dodd, K. W., and Fuller, W. A. A semiparametric transformation approach to estimating usual daily intake distributions. *Journal of the American Statistical Association*, 91(436):1440–1449, 1996.

[Nusser et al. (1997)] Nusser, S. M., Fuller, W. A., Guenther, P. M., 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.

[Ocké et al. (2008)] Ocké, M.C., van Rossum, C.T.M., Fransen, H.P., Buurma, E.J.M., de Boer, E.J., Brants, H.A.M., Niekerk, E.M., van der Laan, J.D., Drijvers, J.J.M.M., and Ghameshlou, Z. Dutch national food consumption survey - young children 2005/2006. *RIVM Letter report*, 2008. URL: https://www.rivm.nl/publicaties/dutch-national-food-consumption-survey-young-children-20052006.

[Price and Han (2011)] Price, P. S. and Han, X. 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.

[Saul and Lee (2002)] Saul, L. K. and Lee, D. D. Multiplicative updates for classification by mixture models. In *Advances in Neural Information Processing Systems*, 897–904. 2002.

[Slob (2002)] Slob, W. Dose-Response Modeling of Continuous Endpoints. *Toxicological Sciences*, 66(2):298–312, 04 2002. URL: https://doi.org/10.1093/toxsci/66.2.298.

[Slob (2006)] Slob, W. Probabilistic dietary exposure assessment taking into account variability in both amount and frequency of consumption. *Food and Chemical Toxicology*, 44(7):933–951, 2006.

[Slob and Setzer (2013)] Slob, W. and Setzer, R. Shape and steepness of toxicological dose–response relationships of continuous endpoints. *Critical reviews in toxicology*, 44:, 11 2013. doi:10.3109/10408444.2013.853726.

[Slob et al. (2010)] Slob, W., de Boer, W. J., and van der Voet, H. 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.

[Souverein et al. (2011)] Souverein, O. W., de Boer, W. J., Geelen, A., van der Voet, H., de Vries, J. H., Feinberg, M., and van't Veer, P. 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.

[Tebby et al. (2020)] Tebby, C., van der Voet, H., de Sousa, G., Rorije, E., Kumar, V., de Boer, W., Kruisselbrink, J. W., Bois, F. Y., Faniband, M., Moretto, A., and Brochot, C. A generic pbtk model implemented in the mcra platform: predictive performance and uses in risk assessment of chemicals. *Food and Chemical Toxicology*, 142:111440, 2020. URL: https://www.sciencedirect.com/science/article/pii/S0278691520303306.

[Tooze et al. (2006)] Tooze, J. A., Midthune, D., Dodd, K. W., Freedman, L. S., Krebs-Smith, S. M., Subar, A. F., Guenther, P. M., Carroll, R. J., and Kipnis, V. 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.

[van den Berg et al. (2016)] van den Berg, F., Jacobs, C.M.J., Butler Ellis, M.C., Spanoghe, P., Doan Ngoc, K., and Fragkoulis, G. 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.

[van der Voet and Slob (2007)] van der Voet, H. and Slob, W. 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.

[van der Voet et al. (2009)] van der Voet, H., van der Heijden, G. W.A.M., Bos, P. M.J., Bosgra, S., Boon, P. E., Muri, S. D., and Brüschweiler, B. J. 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.

[van der Voet et al. (2014)] van der Voet, H., Kruisselbrink, J.W., Boer, W.J., and Boon, P.E. Model-then-add: usual intake modelling of multimodal intake distributions. *RIVM Letter report*, 2014. URL: https://rivm.openrepository.com/handle/10029/314361.

[van Klaveren et al. (2012)] van Klaveren, J. D., Goedhart, P. W., Wapperom, D., and van der Voet, H. A european tool for usual intake distribution estimation in relation to data collection by efsa. *EFSA Supporting Publications*, 9(6):300E, 2012. URL: https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/sp.efsa.2012.EN-300.

[van Klaveren et al. (2019a)] van Klaveren, J.D., Kruisselbrink, J.W., de Boer, W.J., van Donkersgoed, G., Biesebeek, J.D. t., Sam, M., and van der Voet, H. Cumulative dietary exposure assessment of pesticides that have acute effects on the nervous system using mcra software. *EFSA Supporting Publications*, 16(9):1708E, 2019. URL: https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/sp.efsa.2019.EN-1708.

[van Klaveren et al. (2019b)] van Klaveren, J.D., Kruisselbrink, J.W., de Boer, W.J., van Donkersgoed, G., Biesebeek, J.D. t., Sam, M., and van der Voet, H. Cumulative dietary exposure assessment of pesticides that have chronic effects on the thyroid using mcra software. *EFSA Supporting Publications*, 16(9):1707E, 2019. URL: https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/sp.efsa.2019.EN-1707.

[Verkaik-Kloosterman et al. (2011)] Verkaik-Kloosterman, J., Dodd, K. W., Dekkers, A. L., van't Veer, P., and Ocké, M. C. 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.

[WHO (2018)] World Health Organization (WHO). *Guidance document on evaluating and expressing uncertainty in hazard characterization*. World Health Organization, 2018.

[Zetlaoui et al. (2011)] Zetlaoui, M., Feinberg, M., Verger, P., and Clémençon, S. Extraction of food consumption systems by nonnegative matrix factorization (nmf) for the assessment of food choices. *Biometrics*, 67(4):1647–1658, 2011.

INDEX

Q

R

S

T