MCRA, Release 6

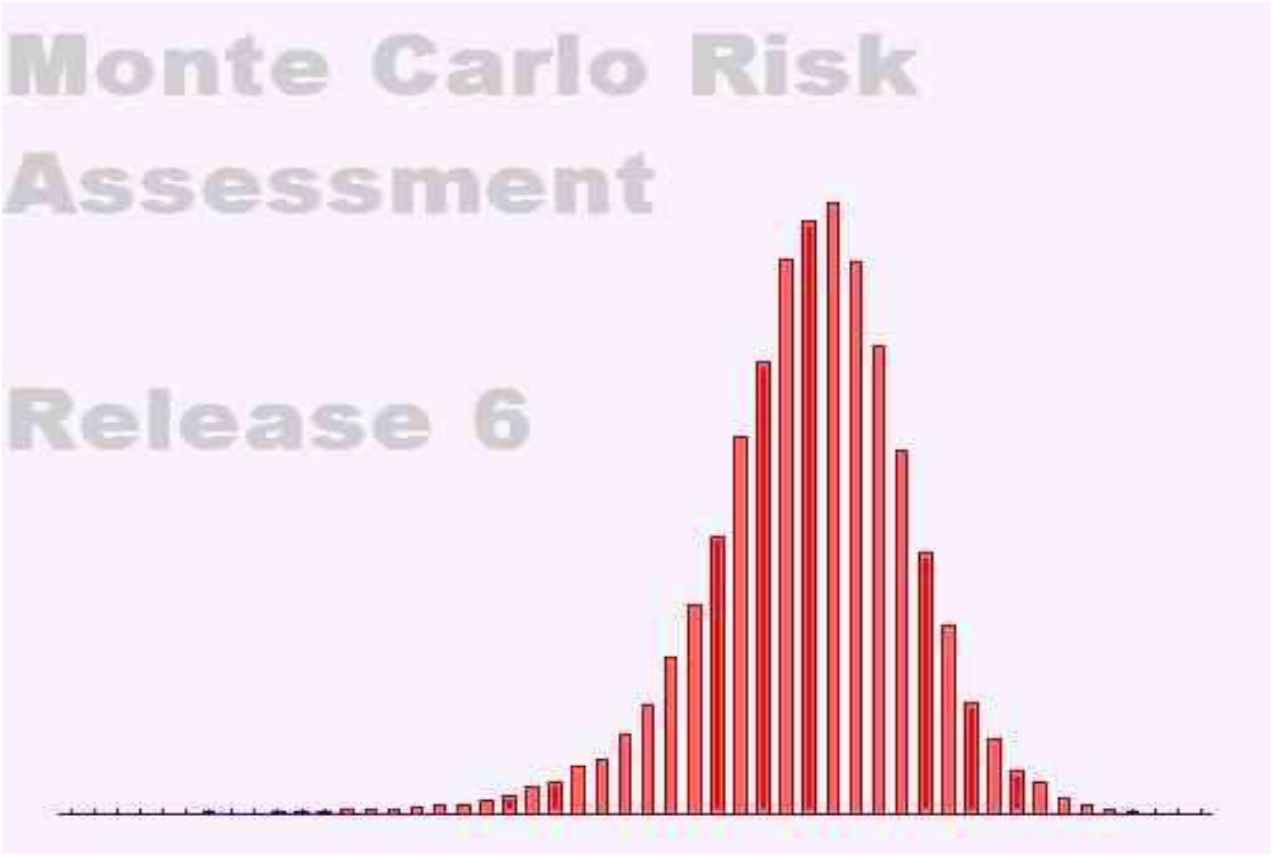
a web-based program for Monte Carlo Risk Assessment

On line Manual 2007-07-11

documenting MCRA Release 6.0

Waldo J. de Boer
Hilko van der Voet

With contributions by: Paul W. Goedhart, Jac T.N.M. Thissen



Biometris and RIKILT
Wageningen University and Research centre
RIVM
Centre for Substances and Integrated Risk Assessment,
National Institute for Public Health and the Environment

July 07

Biometris is the unit for Mathematical and Statistical Methods of Wageningen University & Research centre.

Post address
P.O. Box 100
6700 AC Wageningen
The Netherlands
Visiting address
Bornsesteeg 47, building no. 116
6708 PD Wageningen
Telephone: +31 (0)317 476925
Telefax: +31 (0)317 483554

RIKILT (Institute of Food Safety) is a part of Wageningen University & Research centre.

Post address
P.O. Box 230
6700 AE Wageningen
The Netherlands
Visiting address
Bornsesteeg 45, building no. 123
6708 PD Wageningen
Telephone: 0031 317 475400
Telefax: 0031 317 417717

RIVM National Institute for Public Health and the Environment

Post address
P.O. Box 1
3729 BA Bilthoven
The Netherlands
Visiting address
Antonie van Leeuwenhoeklaan 9
3721 MA Bilthoven
Telephone: 0031 30 2749111
Telefax: 0031 30 2742971

Contents

Contents 3

1 Foreword 7

2 Getting started with MCRA 8

2.1 Introduction 8

2.2 Registration 9

2.3 MCRA Main menu 9

2.4 Data needed 10

2.5 Manage input/output 11

2.6 Starting the MCRA program 13
2.6.1 Selection of the tables 14
2.6.2 Selection of food consumption survey and of compound 16
2.6.3 Selection of foods 17

3 Specifying MCRA options and running a job 19

3.1 Overview 19

3.2 Acute risk type 20
3.2.1 Concentration model options 21
3.2.1.1 Empirical *22*
3.2.1.2 Binomial/normal (no pooling) *22*
3.2.1.3 Binomial/normal with pooling *22*
3.2.2 Concentration data options 23
3.2.2.1 Replacement of non-detects *23*
3.2.2.2 Modelling processing effects *23*
3.2.3 Unit variability model options 24
3.2.3.1 Estimated parameters for unit variability *24*
3.2.3.2 Beta distribution *25*
3.2.3.3 Lognormal distribution *26*
3.2.3.4 Bernoulli distribution *26*
3.2.4 Additional options concerning IESTI and consumption days 27
3.2.5 Intake model options 28
3.2.5.1 Only empirical estimates *28*
3.2.5.2 Empirical estimates and betabinomial/normal *28*

3.3 Chronic risk type 29
3.3.1 Concentration model options 30
3.3.2 Intake model options 30
3.3.2.1 Betabinomial/normal (BBN) *30*
*3.3.2.2 Discrete/*semi-parametric *(ISUF)* *32*

3.4 Uncertainty analysis 33
3.4.1 Uncertainty analysis options for acute risks 33
3.4.2 Uncertainty analysis options for chronic risks: betabinomial/normal (BBN) 34
3.4.3 Uncertainty analysis options for chronic risks: discrete/semi-parametric (ISUF) 34

3.5 Output 34
3.5.1 Output options for acute risks: only empirical estimates 34

3.5.2 Output options for acute risks: empirical estimates + betabinomial/normal (BBN) 35
3.5.3 Output options for chronic risks: betabinomial/normal (BBN) 35
3.5.4 Output options for chronic risks: discrete/semi-parametric (ISUF) 36

3.6 Running an MCRA job 36

3.7 Checking the processing time 36

4 Example output 38

4.1 Acute risk assessment: basic analysis 38

4.2 Chronic risk assessment: betabinomial/normal (BBN) 46

4.3 Help on Charts 55
4.3.1 To display the Property Editor 56

5 Model description 58

5.1 Basic model 58

5.2 How to deal with limited information on compound concentration data 59
5.2.1 The choice between a parametric and non-parametric approach 59
5.2.2 Estimation based on histogram data 60
5.2.3 Estimation based on summary data 60
5.2.3.1 Moments and other characteristics 61
5.2.3.2 Estimation 61

5.3 Acute risk assessment 63
5.3.1 Concentration model: modelling of compound concentrations in consumed food 63
5.3.1.1 Distributional assumptions 63
5.3.1.1.1 Non-parametric modelling of concentrations (empirical) 63
5.3.1.1.2 Parametric modelling of concentrations (binomial/lognormal no pooling) 63
5.3.1.1.3 Parametric modelling of concentrations (binomial/lognormal with pooling) 64
5.3.1.2 Modelling of missing data and replacement of non-detects 64
5.3.1.3 Modelling of processing effects 65
5.3.1.3.1 No processing factor 65
5.3.1.3.2 Processing (fixed factors) 65
5.3.1.3.3 Processing (distribution based) 66
5.3.2 Modelling of unit variability 66
5.3.2.1 Introduction, variability in deterministic modelling 66
5.3.2.2 Approaches to unit variability in probabilistic modelling: specifying distributions 67
5.3.2.2.1 Beta model for unit variability 68
5.3.2.2.2 Lognormal model for unit variability 70
5.3.2.2.3 Bernoulli model for unit variability 70
5.3.2.2.4 Estimation of intake values using the concept of unit variability 71
5.3.2.3 Additional: Comparison of probabilistic with deterministic estimates of acute risk 71
5.3.3 Intake model 72
5.3.3.1 Only empirical estimates 72
5.3.3.2 Empirical estimates and betabinomial/normal 73
5.3.3.2.1 Monte Carlo data generation of intake 73
5.3.3.2.2 Intake frequency model 73
5.3.3.2.3 Intake amount model 73
5.3.3.2.4 Estimating the acute risk variability of positive intake amounts 73
5.3.3.2.5 Estimating the acute intake distribution 73

5.4 Chronic risk assessment 73
5.4.1 Introduction 74
5.4.2 Modelling the intake frequency distribution 74
5.4.3 Modelling the positive intake amounts 75
5.4.3.1 Power or log transformation 75
5.4.3.2 Model with between-consumer and within-consumer variance component 75

5.4.3.2.1 Log transformation 75
5.4.3.2.2 Power transformation 75
5.4.4 Modelling usual daily intake 76
5.4.4.1 Analytical integration *76*
5.4.4.2 Numerical integration *76*
5.4.5 Covariable and/or cofactor in the models 76
5.4.5.1 Cofactor *77*
5.4.5.1.1 Intake frequency model 77
5.4.5.1.2 Intake amount model 77
5.4.5.1.3 Usual intake model 77
5.4.5.2 Covariable *77*
5.4.5.2.1 Intake frequency model 77
5.4.5.2.2 Intake amount model 77
5.4.5.2.3 Usual intake model 77
5.4.5.3 Covariable and cofactor *77*
5.4.5.3.1 Intake frequency model 77
5.4.5.3.2 Intake amount model 78
5.4.5.3.3 Usual intake model 78
5.4.6 Discrete/semi-parametric (ISUF) 78
5.4.6.1 Power or log transformation *78*
5.4.6.2 Spline fit *79*
5.4.6.3 Estimation of the parameters of the usual intake distribution *79*
5.4.6.4 Back transformation and estimation of usual intake *79*

5.5 Uncertainty analysis: resampling data sets and resampling from distributions **80**
5.5.1 Resampling datasets 81
5.5.2 Resampling parametric distributions 81

5.6 Binning **82**

6 Selection of consumers and foods **83**

6.1 Subset selection: consumers **84**

6.2 Subset selection: consumption and concentration data **86**

7 On-line editing of data **88**

7.1 List or add concentrations **91**

7.2 List or add variability factors **93**

7.3 List or add processing factors **94**

7.4 List or edit ARfD or ADI **95**

8 References **97**

9 APPENDIX A: Preparing the data **99**

9.1 Overview **99**

9.2 Harmonised CODEX codes **100**
9.2.1 Definitions 100
9.2.2 Conversion of codes 101
9.2.3 Conversion rules 101
9.2.4 Scheme to link food consumption and compound concentration data 101
9.2.5 Example of use 103

9.3 Food tables **104**
9.3.1 Food (compulsory) 104

9.3.2 FoodProperties (optional, (for unit variability compulsory)) 104
9.3.3 FoodComposition (optional) 105
9.3.4 FoodMarketShare (optional) 105

9.4 Food consumption tables 105
9.4.1 FoodConsumption (compulsory) 105
9.4.2 Individual(compulsory) 106
9.4.3 FoodSurvey (optional) 106
9.4.4 ProcessingType (optional, (for processing compulsory)) 107
9.4.5 Processing (optional) 107

9.5 Compound concentration tables 109
9.5.1 Compound (compulsory) 109
9.5.2 Country (compulsory) 109
9.5.3 ConcentrationValues (optional) 109
9.5.4 ConcentrationSummaryStatistics (optional) 110
9.5.5 ConcentrationDiscreteValues (optional) 111
9.5.6 VariabilityProd (optional) 111
9.5.7 VariabilityCompProd (optional) 112
9.5.8 VariabilityProcCompProd (optional) 112
9.5.9 AgriculturalUse (optional) 113
9.5.10 ConcentrationWorstcaseValues (optional) 113

10 APPENDIX B: Example output 114

10.1 Acute risk assessment: processing fixed factors 114

10.2 Acute risk assessment: unit variability, Beta distribution 116

10.3 Acute risk assessment: IESTI 118

10.4 Acute risk assessment: uncertainty 119

10.5 Acute risk assessment: diagnostics 120

10.6 Acute risk assessment: betabinomial distribution 122

10.7 Acute risk assessment: binomial/lognormal with pooling 126

10.8 Chronic risk assessment: discrete/semi-parametric (ISUF) 129

10.9 Chronic risk assessment: betabinomial/normal with covariable age and uncertainty 134

10.10 Chronic risk assessment: betabinomial/normal with covariable age and cofactor sex 140

11 APPENDIX C: Errors in displaying the page and scripting errors 146

1 Foreword

This manual describes a stochastic (or Monte Carlo) model for dietary risk assessment of chemical compounds based on monitoring data concerning the quality of foods and agricultural products. Intake (exposure) assessment is an important step in risk assessment of chemical compounds, such as agricultural chemicals (pesticides, veterinary drugs), toxins (*e.g.* mycotoxins) and environmental contaminants (*e.g.* dioxins). Occasionally, we use the term residue when we refer to compound and the term individual when we refer to consumer. The use of multiple names is avoided as much as possible.

The methods for probabilistic modelling described here are implemented in the program Monte Carlo Risk Assessment (MCRA). MCRA is a computational tool for dietary risk assessment. MCRA can calculate intake distributions for both short-term (acute) and long-term (chronic) intakes. Basically, it simulates daily consumptions by sampling a food consumption database and combines these with a random sample from either a compound database (empirical distribution) or a parametric distribution of compound concentrations. The result is a full *distribution* of intakes, rather than traditional deterministic methods which only provide a point estimate. Percentiles of the intake distribution can be used to assess risks by relating them to *e.g.* an acute reference dose (ARfD). In a chronic risk assessment, MCRA calculates the distribution of the usual intakes over consumers based on the average concentration and the empirical distribution of intake between consumers and between different intake days of the same consumers. Percentiles of this usual intake distribution can then be related to *e.g.* the acceptable daily intake (ADI). Uncertainty of percentiles can be established by resampling methods. MCRA allows including processing factors (*e.g.* the effect of cooking on the concentration) and variability factors (to correct for the fact that monitoring data are obtained from composite samples, whereas consumers may eat individual units). Analyses can be done for a total population or for a subpopulation (*e.g.* children, males or females or consumption-days only). The effects of concentration below analytical reporting limits (LOR) can be assessed. Large portion consumption and the highest compound or median compound in case of bulking or blending in the composite sample is used in IESTI (International Estimated Short Term Intake) calculations.

The current release of MCRA is written in **Microsoft Visual C# .NET 2005**. Release 5 and earlier versions were written in the statistical package GenStat (2005). MCRA is internet-based and can be used by registered users at http://mcra.rikilt.wur.nl. It consists of a basic program to do the computations and of additional database selection possibilities implemented in HTML and Active Server Pages (ASP). MCRA runs with Component One Chart (1999) which offers the possibility to manipulate graphical output after it has been obtained.
An earlier version of the MCRA program, as well as an implementation of the Monte Carlo method in @Risk (1996), have been described in van der Voet *et al*. (1999), and further elaboration was given in de Boer & van der Voet (2000, 2001) and van der Voet *et al*. (2001).

This manual gives a complete description and justification of the statistical methods used in the program MCRA and offers an introduction to assist with the practical application of MCRA in dietary risk assessment. The documentation describes MCRA Release 6. It covers the current release 6.0 (release 6 version 0) and all future updates starting with the same release number. Major updates of the program, encompassing new or improved facilities will be released with an increased release number and a new manual.

MCRA is a result of an ongoing co-operation between RIKILT and Biometris since 1998. RIKILT co-ordinates the Dutch KAP programme (Quality of Agricultural Products) where results of monitoring programs for chemical compounds in food are gathered in a national database. RIKILT also has a recipe database to link food codes from the Dutch food consumption table to primary agricultural products. Biometris contributes statistical models and programs for quantitative risk analysis.
Since 2005, the program is extended in collaboration with RIVM to include models similar to those available in the STEM (Statistical Exposure Modelling) software.

2 Getting started with MCRA

2.1 Introduction

The MCRA system (Monte Carlo Risk Assessment) can be used for assessment of risks due to the intake of compounds on foods. MCRA provides the following options:

- Acute probabilistic risk assessment: MCRA will calculate the intake distribution (mg or microgram compound per kg body weight) from input data on consumption and compound concentrations in food.
- In addition, covariable and/or cofactor dependent percentiles and uncertainty intervals may be estimated using simulated acute intakes as input data.
- Percentiles: the intake distribution can be characterised by percentiles, *i.e.* compound concentration levels exceeded with only a small specified probability (for example the 99th percentile p99 is exceeded only in 1% of the cases).
- Uncertainty due to small samples: resampling of Monte Carlo (MC) variation, of consumers and of compound concentrations to assess the uncertainty of the percentiles in the form of an approximate confidence interval.
- Diagnostics on the amount of MC-variation and the amount of variability due to resampling consumption and compound concentration data.
- Calculation of point estimates (IESTI) and comparison with MC-results.
- Decomposition of foods into ingredients using the composition, *e.g.* convert pizza consumption to consumption of wheat, tomato, cheese, etc.
- Decomposition of foods into marketshares, *e.g.* for apple marketshares are specified for Jonagold, Granny Smith and Golden Delicious.
- Parametric or empirical modelling of concentrations: MCRA can resample the compound concentration data directly (empirical model), or it can sample from a binomial-lognormal model fitted to the concentration data (parametric model). Note: consumption data are always re-sampled empirically from the consumer data set.
- Modelling of processing effects: sometimes it is known that concentrations are reduced by food processing, *e.g.* cooking, and frying. MCRA can incorporate processing factors as fixed effects or by sampling from a processing factor distribution. The latter possibility requires the specification of a nominal and an upper value for the processing factor.
- Modelling of unit variability: compound concentrations are often measured in large composite samples, thus hiding part of the variability that exists between individual units. MCRA has extensive possibilities to model unit variability *e.g.* sampling from a Beta, Bernoulli or Lognormal distribution.
- Modelling of non-detects levels: compound concentrations are often only known above a certain limit, the Limit Of Reporting (LOR). In a worst-case analysis, all non-detect measurements may be replaced by the LOR value.
- Subset selection: extensive possibilities to select data on age, weight or sex of consumers, day of consumption, consumed and derived foods, year, country and sampling type of concentration data.
- Insertion of worst case values for foods without concentration measurement values.
- Calculate intake distribution for consumption-days only
- Chronic risk assessment: MCRA calculates the usual intake distribution when the total number of intake days per consumer is 2 or more. In MCRA, basically, two methods are implemented: the first method (parametric) is called the betabinomial/normal method. The betabinomial distribution is used to model the intake frequency and the normal distribution is used to model logarithmically or power transformed intake amounts. Both distributions are numerically integrated to obtain the usual intake distribution for the entire population. In an extended version of this method, the frequencies and the amounts can also be related to a covariable and/or cofactor to estimate covariable- and cofactor-dependent percentiles and uncertainty intervals. The second method (discrete/semi-parametric) follows an approach proposed by Nusser *et al.* (1996, 1997) and Dodd (1996). The chronic intake distribution is characterised by percentiles and uncertainty intervals on

these percentiles. Following Dodd *et al.* (2006) we refer to this method as the ISUF (Iowa State University Foods) model.

2.2 Registration

To use MCRA, navigate the web browser to http://mcra.rikilt.wur.nl. The opening screen gives some general information, as well as links to fixed versions of the manual of the latest release. Find the latest developments and most recent information in the On Line Manual.

As a potential new user, first fill in the registration form. Click registration form to get the form displayed in Figure 1. Here, specify your name, organisation, address and email address. Choose a username (no spaces allowed) and a password for use of the MCRA system. Click the 'OK' -button to send the request to the MCRA webmaster at RIKILT, and you will get a response by email as soon as possible.

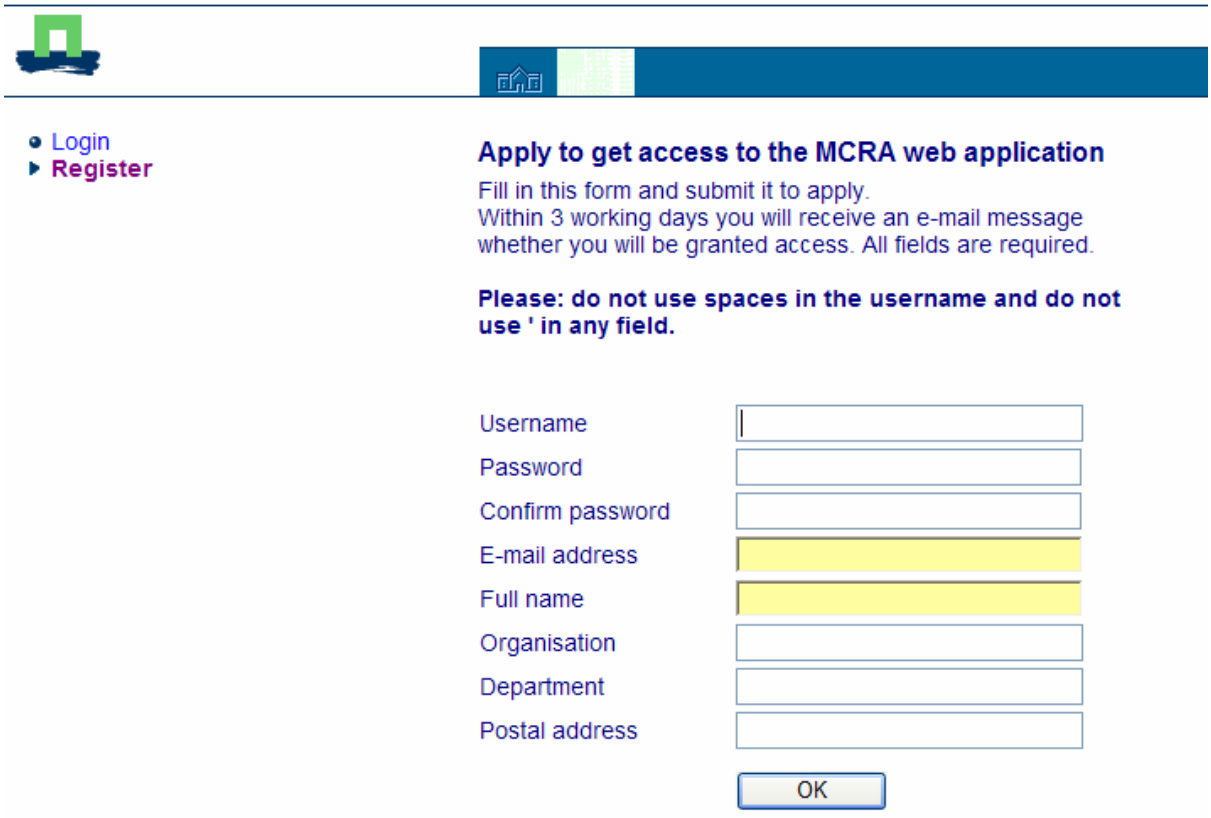


Figure 1: Registration form

Registered users enter the website by clicking Login to MCRA for registered users on the home page. Specify username and password, and the first screen of the MCRA website is shown: **Main Menu**.

2.3 MCRA Main menu

In Figure 2, options within the Main menu are shown:

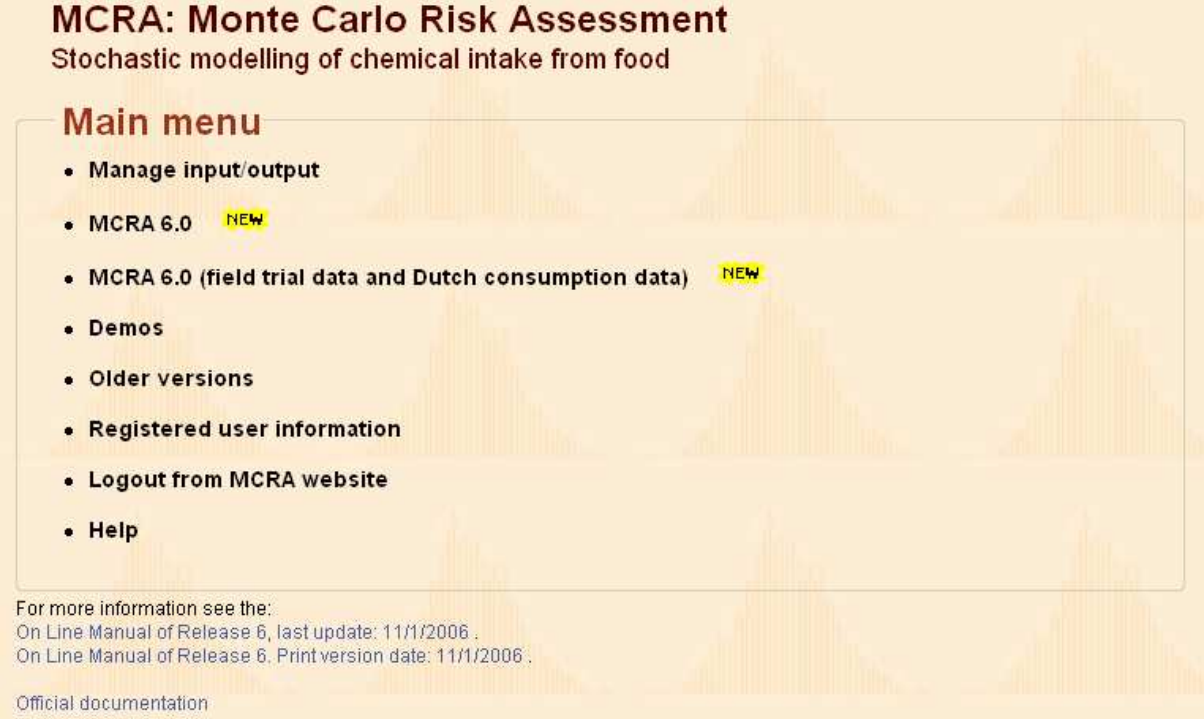


Figure 2: MCRA Main Menu

- Click 'MCRA 6.0' to start preparations needed to do an MCRA analysis.
- Click 'Manage input/output' to upload (and download) your own data (see 2.5) or to download output from a former MCRA analysis.
- Click 'Older versions' to run old releases of the program (older versions are not supported).
- Click 'MCRA 6.0 (field trial data and Dutch consumption data)' to run MCRA with features to edit your own compound concentration data.
- Click 'Registered user information' to view which information about you is stored in the user database.
- Click 'Help' for explanation about options in the Main menu.
- Click 'Logout from MCRA website' to leave the MCRA website. Your personal data files, latest output files and the latest input options remain stored for later use.

2.4 Data needed

What data are needed to run MCRA? All data for MCRA are stored in Microsoft Access database tables according to a predefined format. To use your own data, prepare your database off-line and upload it to your personal user area on the MCRA website. The 'MCRA 6.0 (field trial data and Dutch consumption data)' option in the Main menu (see 2.3) offers some possibilities to edit data on-line. However, on-line editing is restricted to compound concentration values, unit variability, processing factors, acute reference dose (ARfD) and average daily intake (ADI) (see Chapter 7).

In Chapter 9 , a full description of the format is given how data should be saved in a MS Access databases. Basically, input data for MCRA originate from two sources: food consumption surveys and monitoring programs on compound concentration data. Often, additional tables are needed to link consumption data to compound concentration data or to implement model options like unit variability. In Figure 3, a short outline is presented how tables are linked to each other: consumption data are linked directly to compound concentration data or in an indirect way, through the use of food composition data, food marketshare data, processing data or by the use of a supertype algorithm.

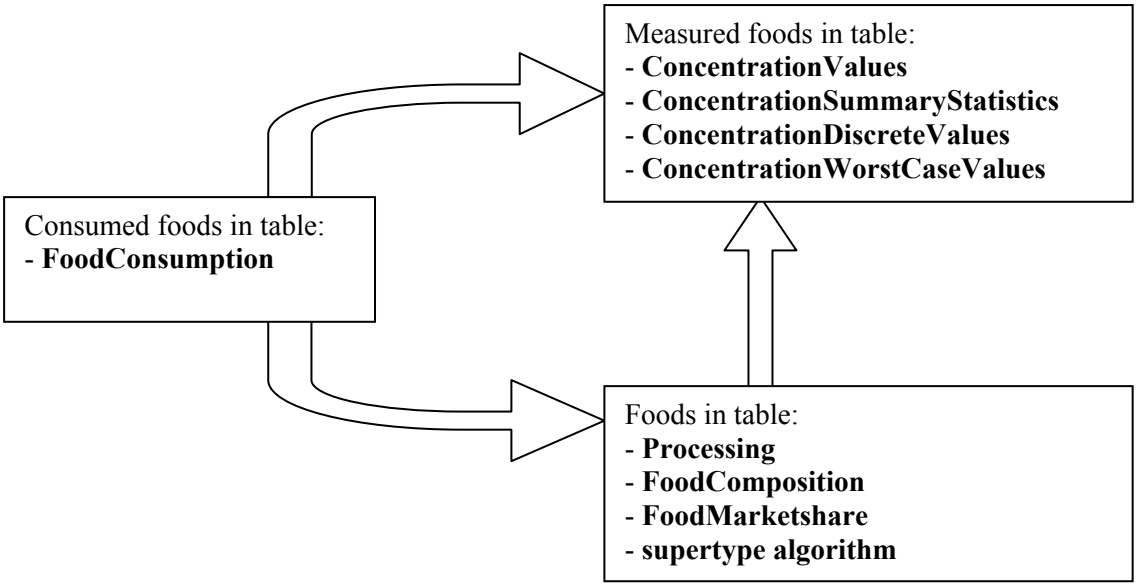


Figure 3: Links between consumption and compound concentration data

The MCRA system has a central database with example data. You may run example analyses without having data yourself. However, MCRA is primarily designed to work with user databases, or with a mixture of user data and centrally supplied data. For example, provide your own data on compound concentrations and combine these with the centrally supplied consumption data. Be careful when using tables from different databases: the codes of foods of the centrally supplied data and your own data should be consistent with each other.
Consumption data are consumed portions of food (consumed at different days) of consumers. To get standardized intakes, in any case the weight of each consumer should be supplied. Other characteristics of the consumers, like age and/or sex, may be used in further analyses. The second type of data, compound concentration data, are the amounts of compound found on monitoring samples of food. The third category, additional tables, provides information that links consumption data to compound concentration data or store information for more sophisticated analyses like unit variability (see Figure 3).
Food composition data specifies the composition of foods. So, speaking about pizza, the composition specifies proportions for *e.g.* wheat, tomato, cheese etc. Food marketshare data specifies the proportion of subtypes, so for apple, marketshares are *e.g.* Jonagold, Granny Smith, Golden Delicious etc. Processing data specify the unprocessed food, the processed food and the corresponding processing factors, *e.g.* for grapes raisins are specified. The supertype of a food is, if needed, automatically determined. So the supertype of *e.g.* Granny Smith is apple.
As a registered MCRA user you have complete control over the file management in your personal area by starting 'Manage input/output' in the Main menu (see 2.5).

2.5 Manage input/output

The 'Manage input/output' option in the Main menu brings you to a screen where you can upload your data files (see Figure 4). Each user has a personal data area with two subdirectories named 'IN' and 'OUT'. The 'IN'-directory is used to upload your own MS Access databases. Databases are uploaded directly or in zipped form. Other options in this menu are *e.g.* zip, rename or delete files. MS Access databases and zip-files can also be downloaded. Note: never delete subdirectories 'IN' and 'OUT'.

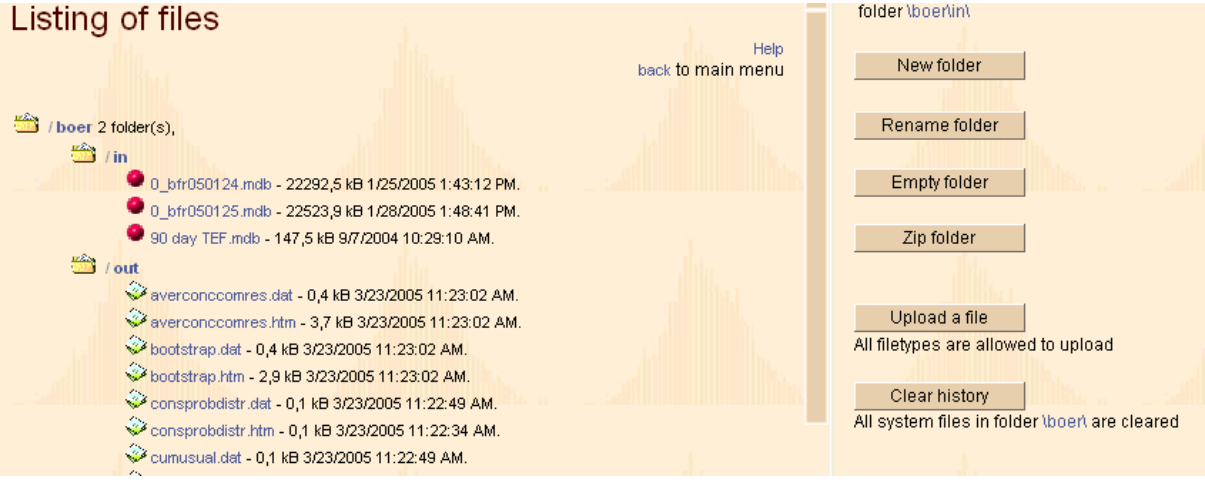


Figure 4: Manage input/output

Output from MCRA is written to the personal 'OUT'-directory. To download output for off-line viewing do the following. Go to the central menu (see Figure 7), click the 'View-output'-button and click DownloadOutput (see Figure 5). Output files are downloaded in a zipped format. The download includes a file 'viewoutput.htm' which gives the same options to study output as available on the website.

Occasionally, after pasting ComponentOne Charts into Word the chart is not displayed (at all) and instead, an icon appears. To our experience, pasting charts from the clipboard encounters no difficulties when the Word document is opened first, then press the 'Copy to Clipboard'-button and paste the contents of the clipboard into the Word document.

Occasionally, system faults occur due to errors like incorrect database contents, queries giving empty subsets, subset selections combining inconsistent levels or scripting errors. The best way to proceed is to log out and enter the website again (login). Then, click the 'Manage input/output' link in the main menu, click your personal directory link or the 'IN' or 'OUT'-subdirectory links (a number of buttons appear) and click the 'Clear history'-button: all system files (files created by MCRA, but not visible) are deleted from the personal directory. Files on the 'IN' or 'OUT'-subdirectory are not deleted (see also ch. 11).

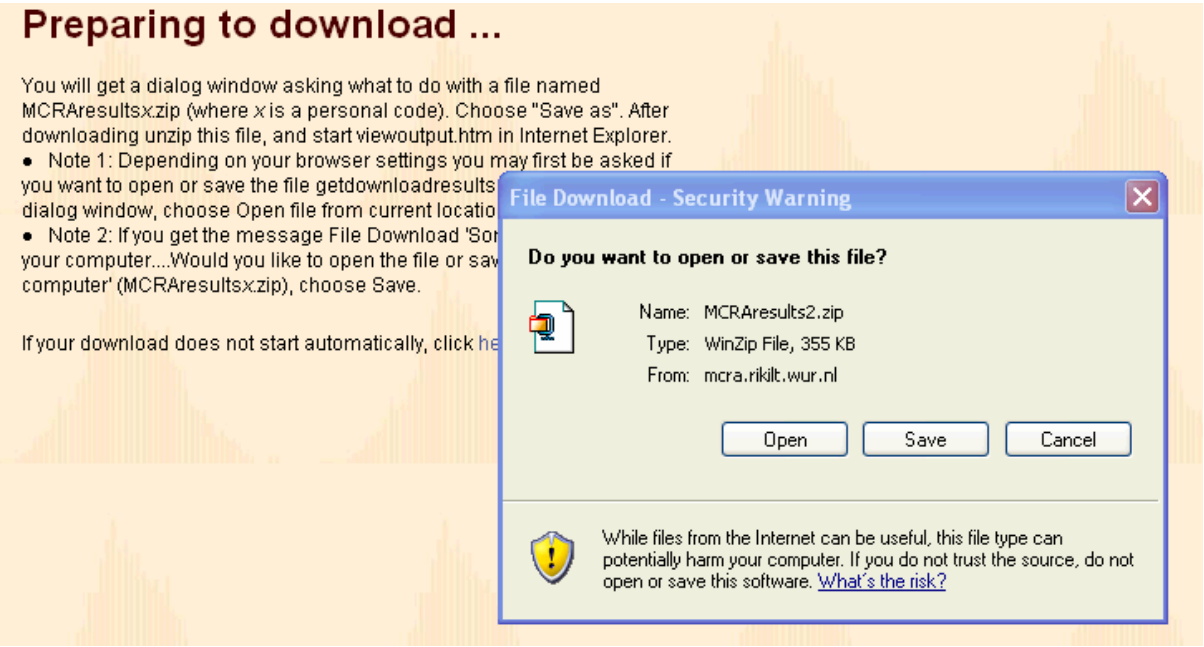


Figure 5: Download output for off-line viewing

2.6 Starting the MCRA program

To start MCRA, click the 'MCRA 6.0' option in the Main menu (see Figure 2). Then, depending on whether you are a new user or not, the screen in Figure 6 or Figure 7 is shown. Each activity is started from the MCRA central menu. After finishing the activity, the user returns to the MCRA central menu to start a new activity.
New users are automatically brought to the MCRA central menu in Figure 6. Click the 'go'-button to start selection of data.

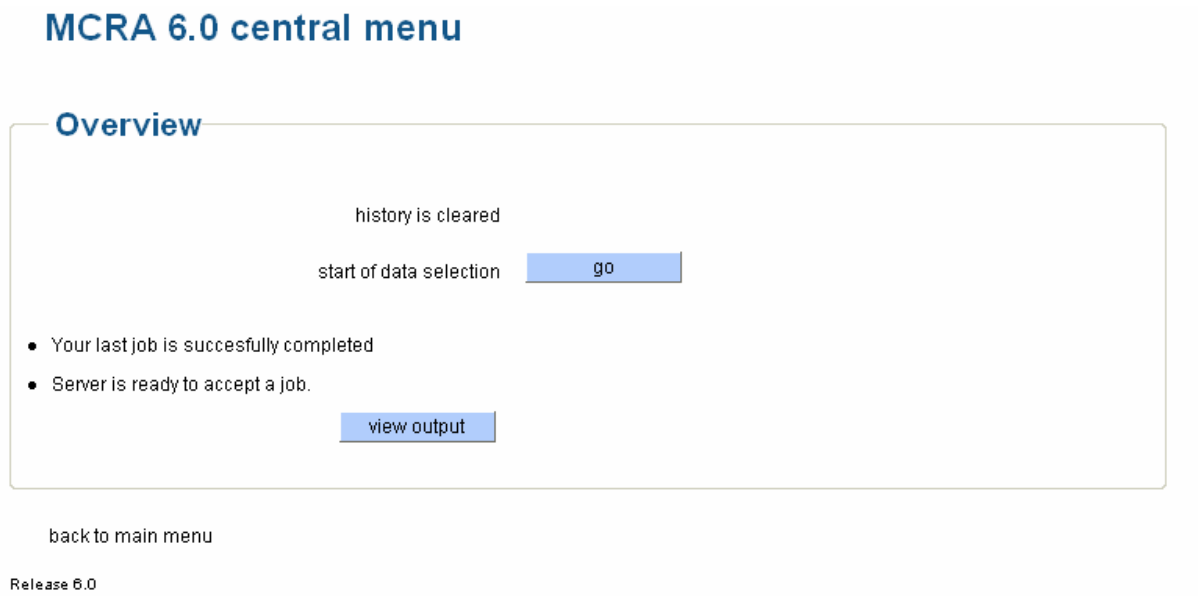


Figure 6: MCRA central menu, start of data selection

In all subsequent cases, the menu in Figure 7 is shown.

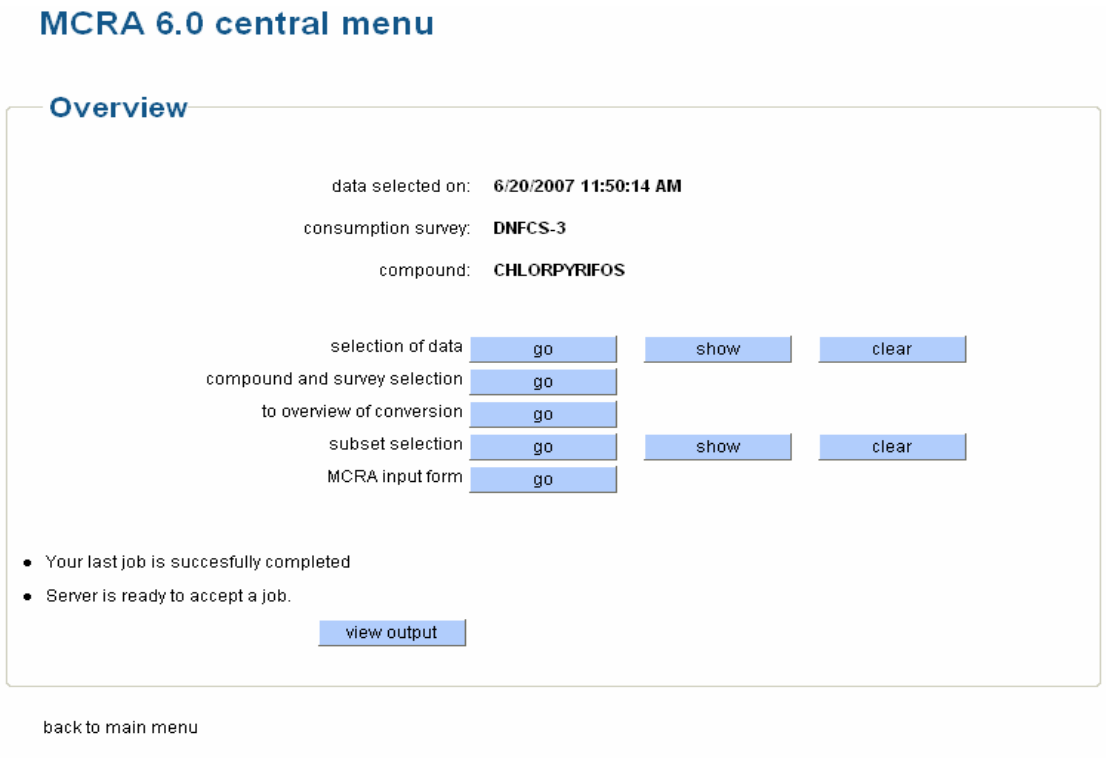


Figure 7: MCRA central menu

First, select the data. In Figure 6, there are no further options. When a selection already exist, continue with the same selections as before by clicking the MCRA-input-form 'go'-button (see Figure 7). For selecting new data, click the selection-of-data 'clear'-button and enter the screen in Figure 6. Click the to-overview-of-conversion 'go'-button to enter an overview of the latest food conversion.

2.6.1 Selection of the tables

Select tables from remote data servers or supply own data by clicking the selection-of-data 'go' or 'clear'-button in the MCRA central menu (see Figure 7). The screen in Figure 8 displays a list of data servers that are sharing data with you (according to your user credentials). Select one or more data servers and click the 'go'-button. In Figure 8, data server 'Biometris' is checked.

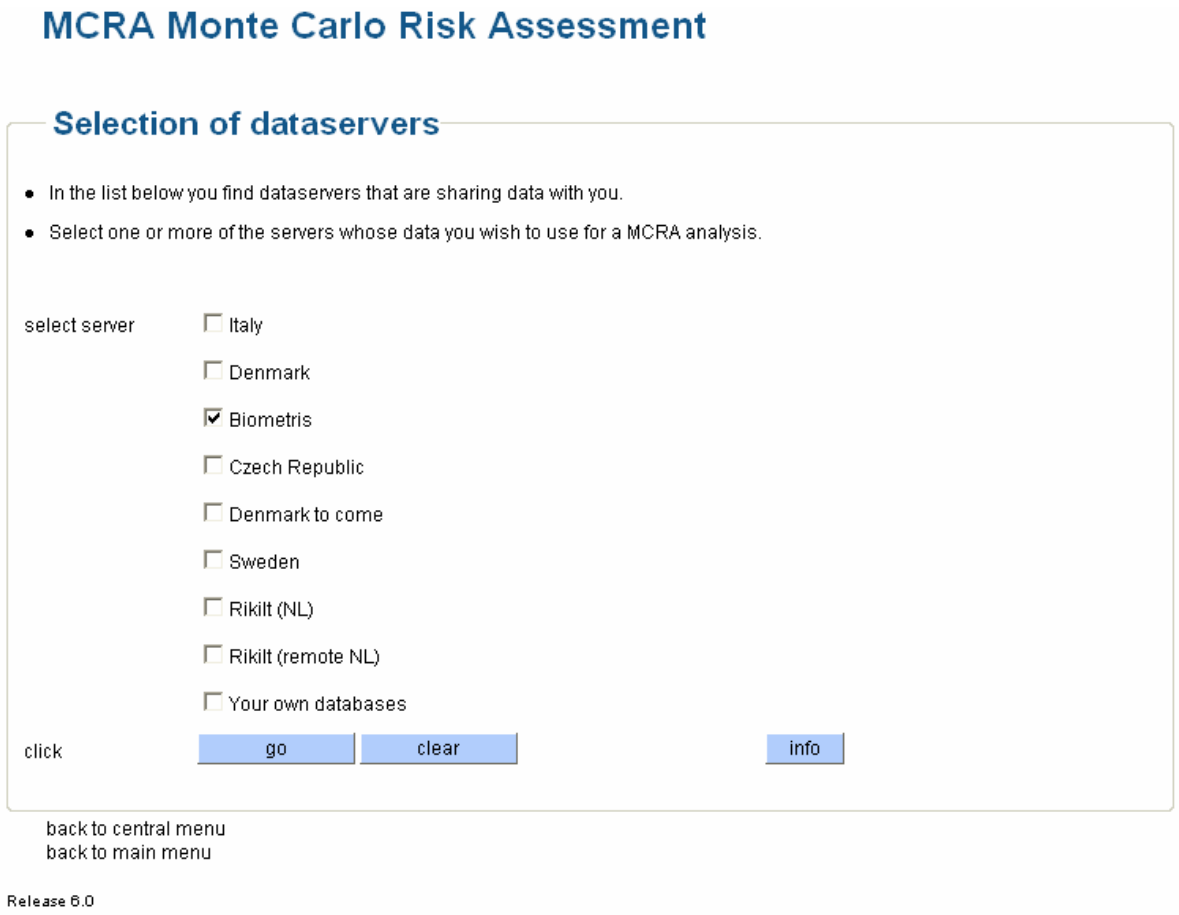


Figure 8: Selection of data servers

In Figure 9 all databases that are available to you are shown. From here select databases for further use in the MCRA analysis. Click the compound-and-survey-selection 'go'-button for selection of a new compound or survey if data are retrieved on an earlier occasion.
As mentioned before, the coding used in various tables should be consistent. **Therefore, the safest way to select data is using data from one and only one source**. If not, convince yourself that data coming from different sources are consistent and suited for your purposes.
On each of these sources, find some information by clicking the 'info'-button and, next screen (not shown), by clicking the buttons with country names.

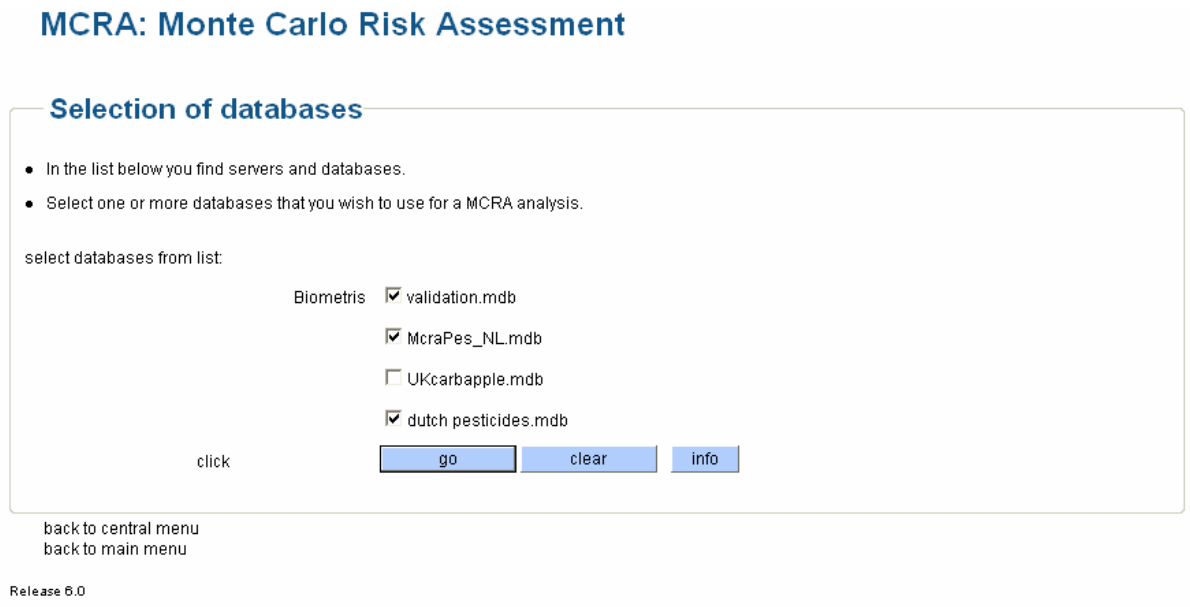


Figure 9: Selection of databases

Figure 10 displays the selected databases and shows which tables are available. Select whole databases at once (check 'All Tables') or make combinations of tables from different databases as done in Figure 10.

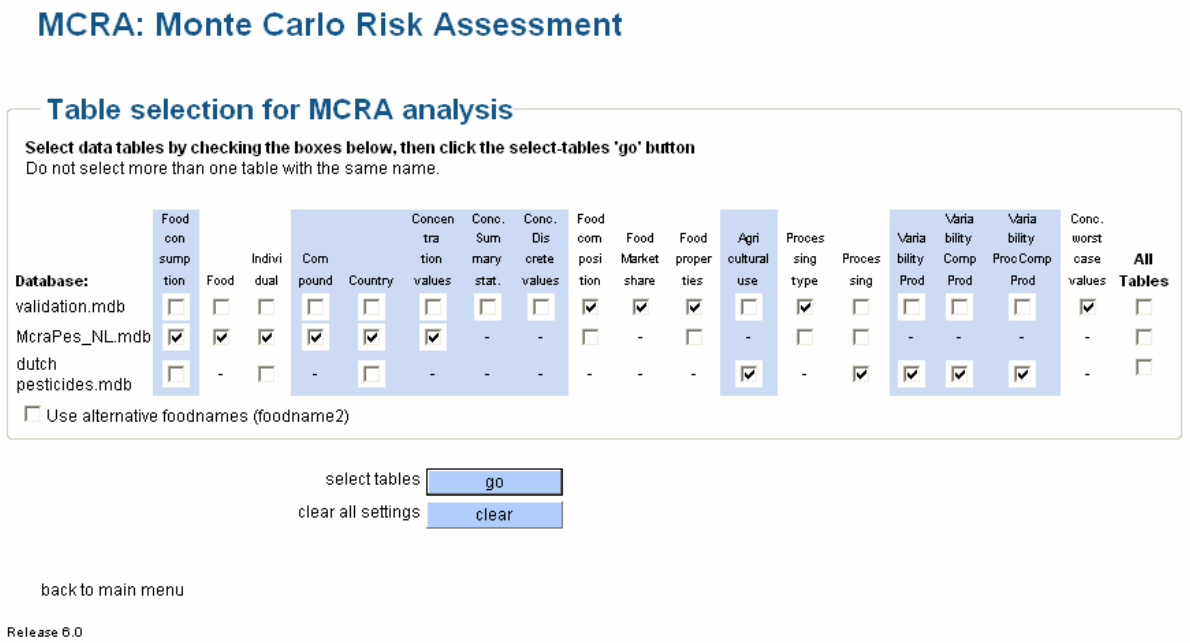


Figure 10: Table selection for MCRA analysis

Scroll the mouse over the checkboxes and tablenames are displayed in red, green or grey textboxes.

- It is compulsory to select all tables for which red textboxes appear (Foodconsumption, Foods, Individual, Compounds and Country).
- It is compulsory to select one of the tables containing compound concentration data (ConcentrationValues, ConcentrationSummaryStatistics and ConcentrationDiscreteValues). These tables have their name displayed in green textboxes (see also 3.2.1).

- All tables with grey textboxes are optional (Foodproperties, Processing, Processingtype, ConcentrationWorstCaseValues, AgriculturalUse, VariabilityProd, VariabilityCompProd, VariabilityProcCompProd).

Make sure to select all compulsory tables otherwise the menu keeps returning with warnings about not selecting the required tables. Make also sure that each type of table is represented once.
To use an alternative language for labelling foods, check 'Use alternative foodnames (foodname2))'.
Check your tabels and click the 'go'-button..

2.6.2 Selection of food consumption survey and of compound

After selecting tables, the screen of Figure 11 is shown. Checkboxes for 'allow marketshares not summing to 100% (step 4)' and 'allow worst-case concentrations (step 7)' are only shown when table FoodMarketshare and ConcentrationWorstCaseValues are selected. Here, table ConcentrationWorstCaseValues is not selected (checkbox not shown). When checkbox 'allow marketshares not summing to 100% (step 4)' is not checked, all derived foods not summing up to 100% are ignored in the analysis.
Select a survey name from the scroll-down menu. MCRA works on single compounds. Select a compound name from the scroll-down menu (see Figure 11).

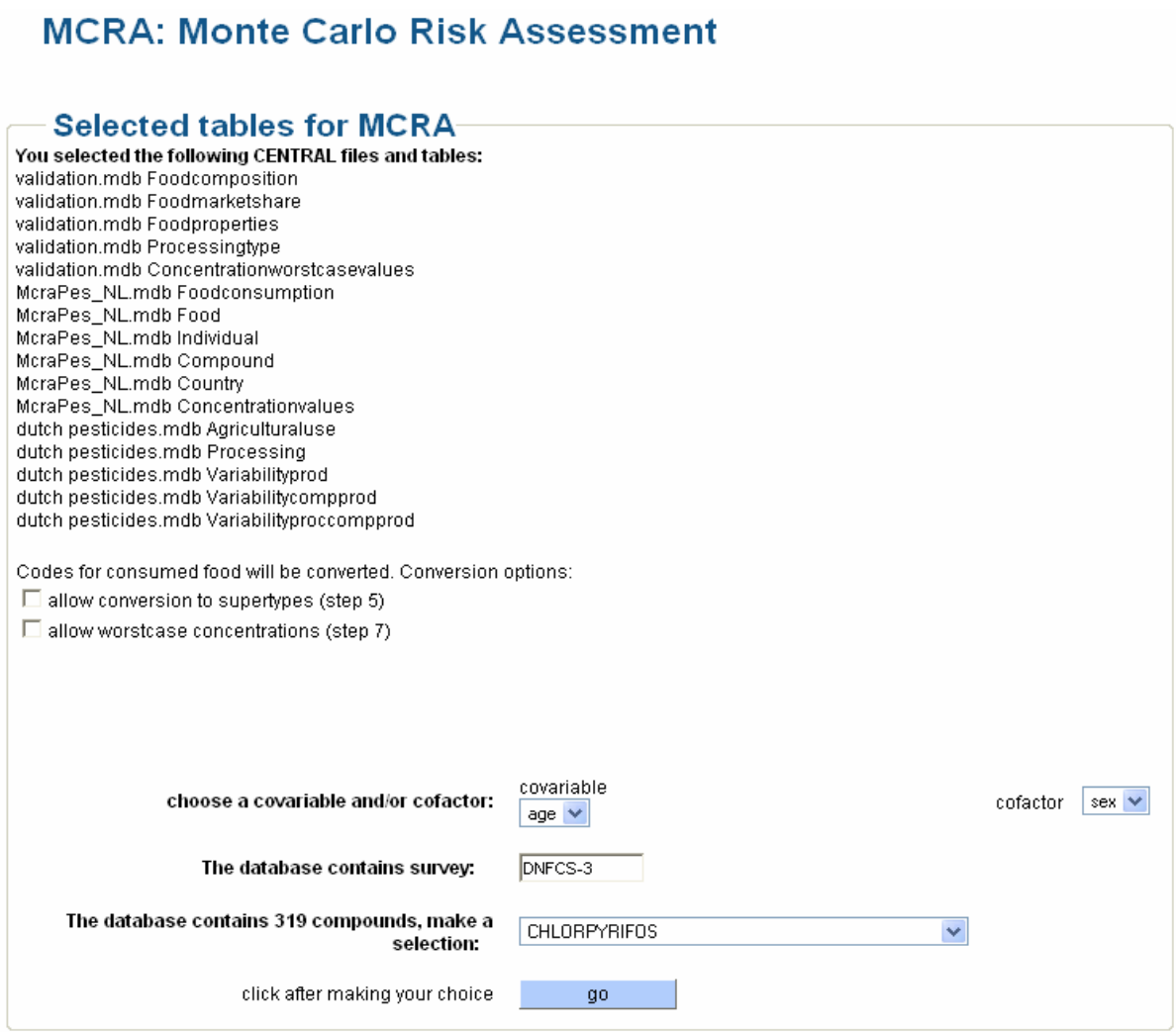


Figure 11: Selection of food consumption survey and compound

MCRA performs a check whether all columns are available in the selected tables. If not, a report of errors is produced, see Figure 12

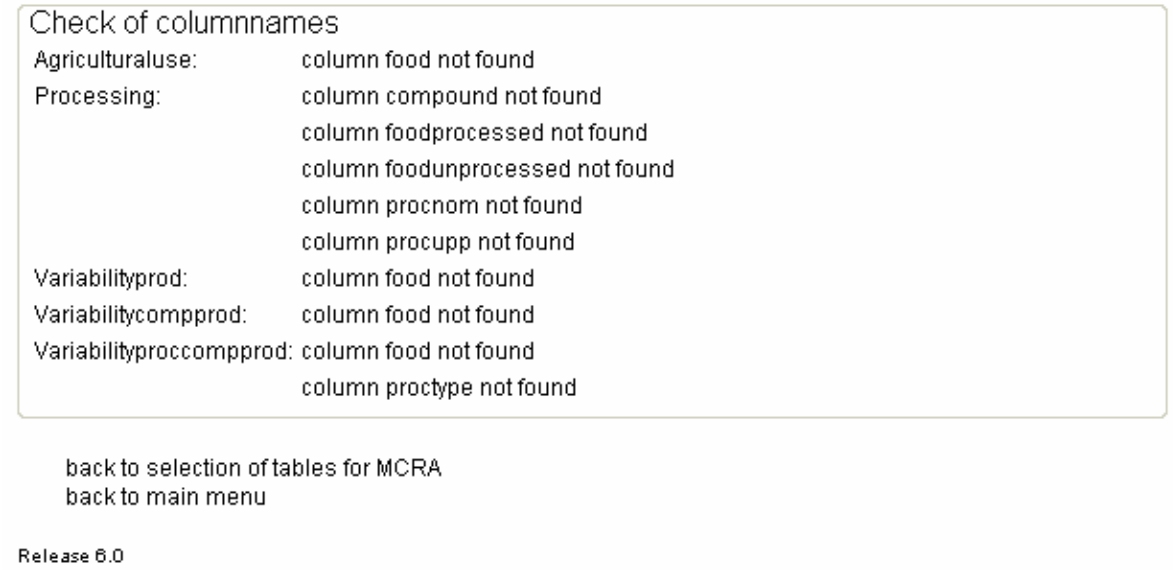


Figure 12: Error report on missing columns

In all examples, the centrally supplied database 'Validation.mdb' is used**.**

2.6.3 Selection of foods

After table selection, information on the number of consumed and of derived foods is displayed.

For consumed foods, three situations may occur:
1. on some foods only positive concentration values are measured,
2. on some foods only non-detects are found,
3. on some foods only worst-case values are found.

Further output is:
4. on some derived foods no information is found,
5. on some consumed foods no information is found.

Now, three options are available (see Figure 13):
1. continue with only foods for which positive concentration measurements are available,
2. continue with foods for which concentration data are available *i.e.* positive concentrations and/or non-detects,
3. continue with all foods *i.e.* also foods for which worst-case values are found (not shown).

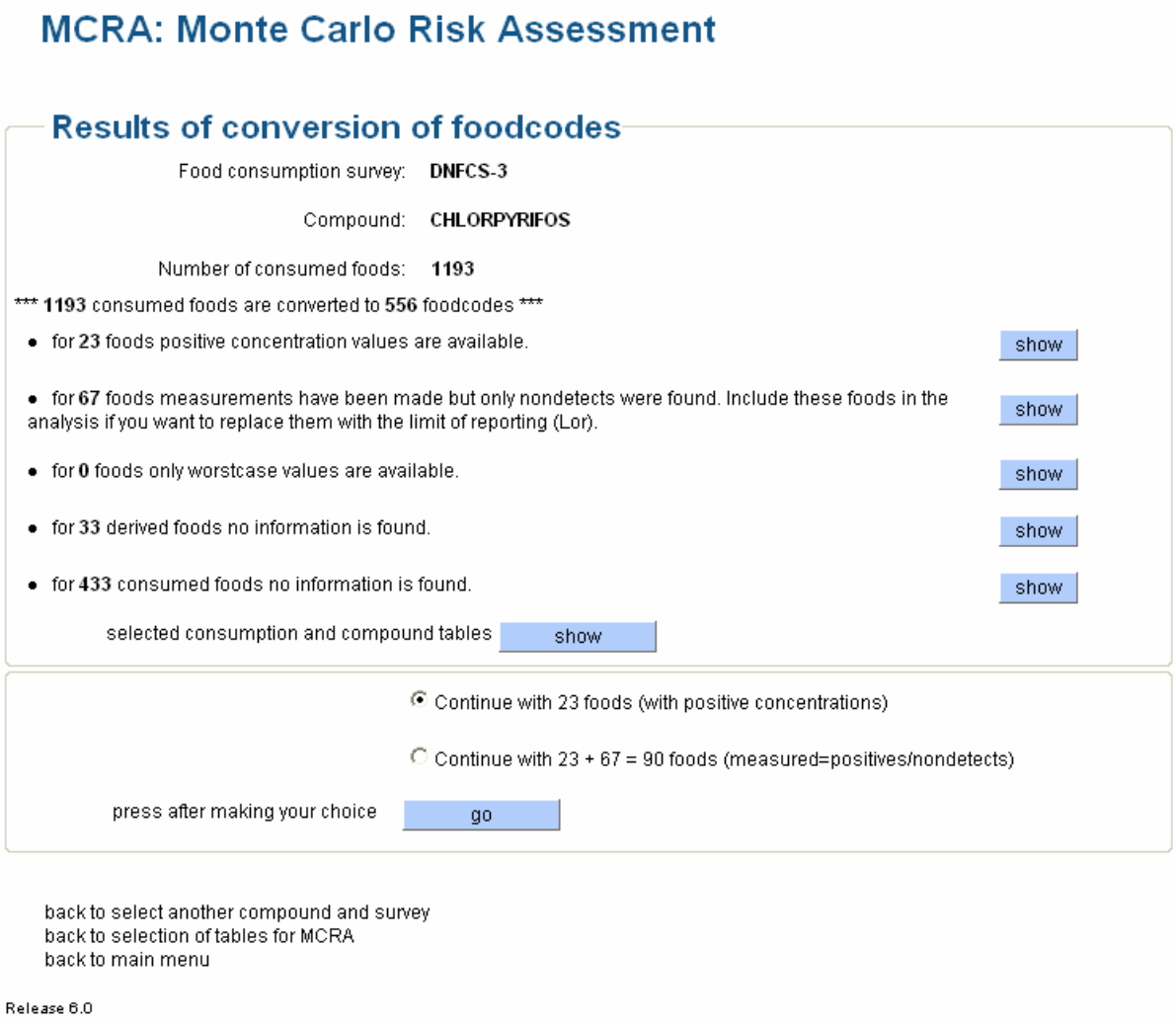


Figure 13: Food selection for MCRA analysis

Check one of the radio buttons and click the ‘go’-button to enter the MCRA central menu (see Figure 7). Click the MCRA-input-form ‘go’-button to enter the MCRA options menu treated in Chapter 3 .

Information about subset selection of the consumer population and foods is found in Chapter 6 .

3 Specifying MCRA options and running a job

3.1 Overview

In this chapter, all options of the input form are discussed with references to paragraphs in Chapter 5 for more theoretical background.
Click the MCRA-input-form 'go'-button of the central menu (see Figure 7). When MCRA is entered for the first time, the screen in Figure 14 (default options) is entered.

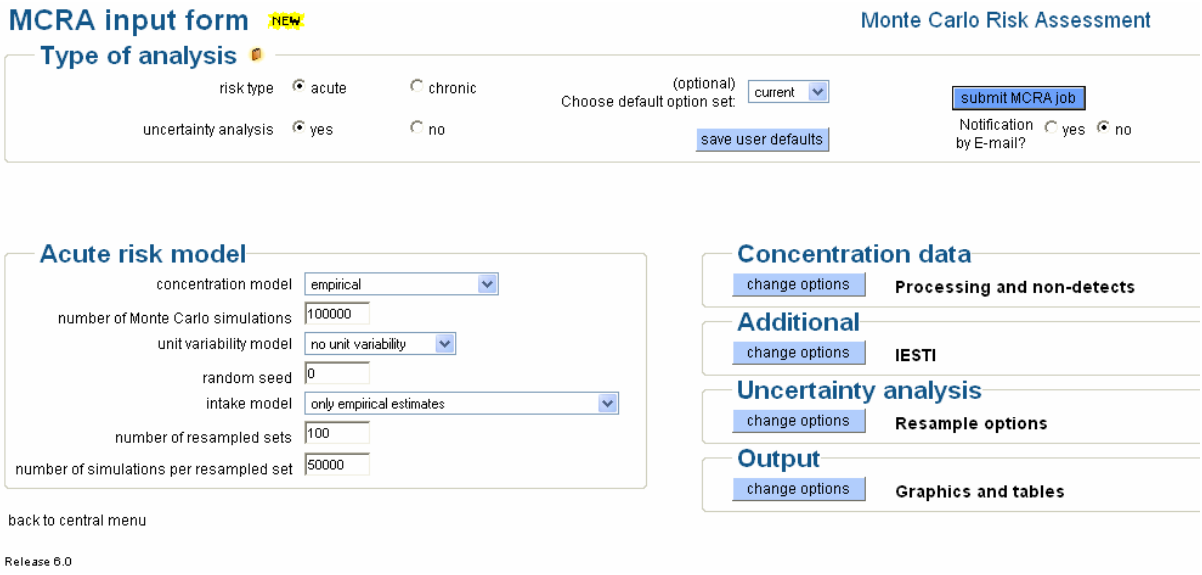


Figure 14: MCRA input form.

The input form is divided into a top, left and right section in which model options are specified:

- In the top section, choose the risk type, acute or chronic, and whether to perform an uncertainty analysis.
- In the left section, a number of options related to the choices made in the top section are specified. For an acute risk assessment, three more specifications are made concerning 1) the concentration model; 2) the unit variability model for the concentrations; 3) and the intake model which may be extended with an analysis to relate the acute intake to a covariable (*e.g.* age) and/or cofactor (*e.g.* sex). For a chronic risk assessment, two models are specified 1) the concentration model; 2) and the intake model. Note that the unit variability model is not relevant for chronic risks.
- The right part of the screen displays a number of special option blocks which depend on the choices made in the top and left section. For an acute risk model at least 3 option blocks are displayed 1) Concentration data; 2) Additional; 3) and Output. For a chronic risk model at least 3 option blocks appear 1) Concentration data; 2) Intake model; 3) and Output. In Figure 14 the option block Uncertainty analysis is shown because in the top section uncertainty analysis is checked. Click a change-options-button to reveal the options (see Figure 15, default options).

After clicking any (radio) button or any item in a scroll-down menu, the screen is rebuilt to implement the choice. After changing a value in a text box, leave the field (move the mouse/replace pointer) to implement the value.
Each description (like 'risk type', 'uncertainty analysis', 'concentration model' etc. in Figure 14) has its own mouse-over function revealing a blue screen-tip with short information on the topic. By double-clicking the description name, the On Line Manual is opened for more information.

Save your settings by clicking the save-user-defaults-button in the top section.
The scroll-down menu Choose-default-option-setting has 4 settings: current, previous, system and user.

- 'current' is shown when one of the input options is changed.
- 'previous' recalls all program settings from the last performed MCRA analysis.
- 'system' sets all default settings.
- 'user' implements all user defaults.

When no MCRA analysis has been performed in the past, setting 'previous', 'user' and 'system' are identical.

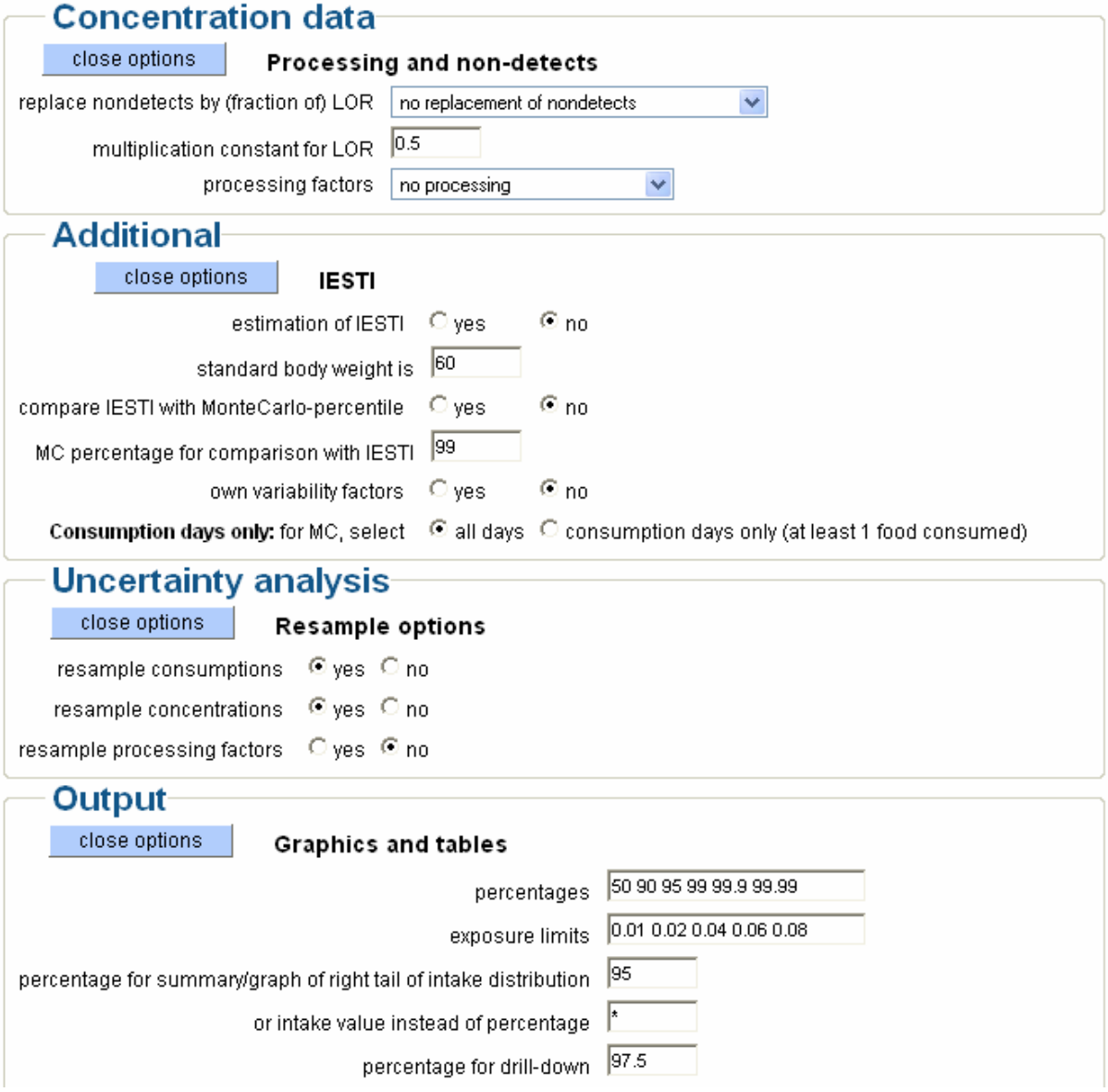


Figure 15: Right section after clicking the 'change options'-buttons in the input form.

Click the 'submit MCRA job'-button in the top section to start MCRA (see 3.6)

3.2 Acute risk type

Specify risk type is acute and find the specifications in the left section of the screen (see Figure 16, default options). Scroll-down menus for the concentration model (see 3.2.1), the unit variability

model (see 3.2.3) and the intake model (see 3.2.5) are shown together with some numerical specifications.

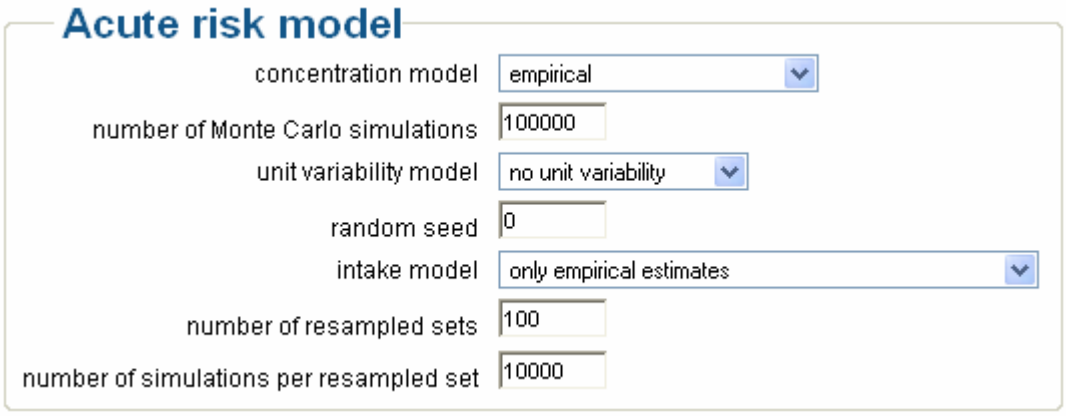


Figure 16: Left section of the MCRA input form if risk type is acute

Number of Monte Carlo simulations

Days of consumption are randomly sampled from the consumer database. Each time a consumption day is sampled, it contributes to the probability distribution of intakes. Each individual contribution is called a simulation. In Figure 16, the total number of MC-simulations is 100,000.

Random seed

The MC-simulation uses a pseudo-random number generator that is initialised by setting the seed. To get time-based values, set seed to zero. For repeatable sequences, set the seed to an integer number.

Number of resampled sets

Specify the number of resampled sets to assess the uncertainty distrubution. In Figure 16, the specified number is 100.

Number of simulations per resampled set

Specify the number of simulations per resampled set. In Figure 16, the specified number of simulations per resampled set is 10,000.

3.2.1 Concentration model options

Settings are:

- **empirical** (only shown for full data) [default for full data]
- **binomial/normal (no pooling)**
- **binomial/normal with pooling**

Compound concentration data are present as full data (a list of compound concentrations is available), summary data (only some summary statistics, for example means, percentiles or maxima are available) or histogram data (only numbers of observations classified in intervals are available). Select the type of concentration data in the selection of tables menu in 2.6.1 .

In the probabilistic model, a distribution of compound concentration data is used to sample from. A choice is made between a parametric and a non-parametric (empirical) approach. Compound concentration data may be used as such (empirical modelling, only with full data) or fitted by binomial/normal distributions *e.g.* parametric modelling, based on full data, summary data or histogram data.

Parametric modelling becomes important in data-scarce situations. The normal distribution with parameters μ and σ has been selected as being both theoretically sensible and practically useful (Shimizu & Crow 1988, Van der Voet *et al*. 1999). The non-parametric approach requires more data to obtain a satisfying representation of the full distribution.

Summarizing, for concentration data we have:

Type of concentration data	full data	summary data	histogram data
Option settings in concentration model			
Non-parametric: empirical	x	-	-
Parametric: binomial/normal (with or without pooling)	x	x	x

Table 1: Possible combinations of option settings with type of concentration data.

See also: How to deal with limited information on compound concentration data

3.2.1.1 Empirical

In the non-parametric approach, choose 'empirical': concentration values are sampled at random from the available compound concentration data and combined with food consumption data to generate the intake distribution of intake values.
See also: Distributional assumptions

3.2.1.2 Binomial/normal (no pooling)

In the parametric approach, compound concentration values per food are sampled from parametric distributions based on full, histogram or summary data. Parameters μ and σ of the normal distribution are estimated using the log-transformed non-zero compound concentrations (full data) or condensed data (summary or histogram data). Choose this setting only, when enough data are available to estimate μ's and σ's for all foods.
Estimation of the variance and/or mean may fail because compound concentration data on specific foods are sparse or even missing. In case of missing parameters, a warning message is printed. Re-run MCRA with setting 'binomial/normal with pooling'.
A related question is the reliability of estimates based on a few degrees of freedom. To overcome these problems, basically, compound concentration data on other foods are used to give sufficient data to base estimates upon. Foods are classified into groups of similar foods and missing or unreliable parameters are estimated using all compound concentration data in a group. This process of using concentration data on similar foods to base estimates for μ and σ upon is called pooling (see also 5.3.1.1.3)
See also: Distributional assumptions

3.2.1.3 Binomial/normal with pooling

Specifying pooling means that foods are automatically assigned into groups and pooled. The identification of groups is based on the CODEX system (see 9.2), using the first 2 characters of the code. When the code contains information on supertypes *e.g.* indicated by the presence of symbol '$', then the supertype is used to form groups. Foods with equal first 2 characters or equal supertypes are placed into the same group.

Pooling is performed in a two step procedure following the next scheme:

1. **Test homogeneity of variances within the groups**
   if variances are homogeneous,
   pool variances.
   test homogeneity of means within the groups
   if means are homogeneous,
   pool means.
2. **Test homogeneity of variances of foods with df < 10 against overall-variance**
   if variances are homogeneous,
   replace variances with overall-variance.

Results of step 1 and 2 are (sub) groups with:
a) pooled variances and pooled means,
b) pooled variances and the original (unpooled, heterogene) means,
c) the original (unpooled, heterogene) variances and original means.

An example of pooling is given in 10.7 .
See also: Distributional assumptions

3.2.2 Concentration data options

Find in the right section of the input form, option block Concentration data (see Figure 17, default options). Specify settings for missing data and non-detects and processing effects.

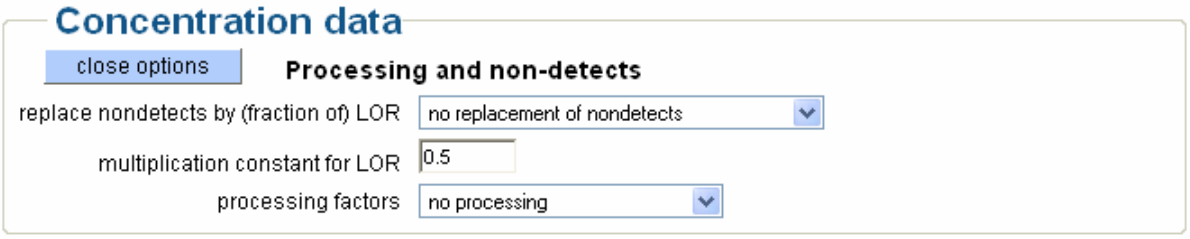


Figure 17: Option block Concentration data

3.2.2.1 Replacement of non-detects

Settings are:

- **no replacement of non-detects** [default]
- **replace all non-detects**
- **replace non-detects based on crop treated**

In many cases of compound risk assessment (*e.g.* pesticides) the majority of the monitoring measurements are non-detects, *i.e.* no quantitative measurement is reported. Only values higher than the Limit Of Reporting (LOR) are reported. When a compound enters the food chain only via crop treatment and the percentage crop treated is (approximately) known, this knowledge is used to infer that some of the monitoring measurements should be real zeroes, contributing nothing to the intake, whereas other non-detects in the monitoring data could have any value below the limit of reporting.
Non-detects (all non-detects or a specified fraction) are replaced with 0 or the LOR multiplied by a multiplication constant. If percent crop treated data are available (see table AgriculturalUse, 9.5.9), then replacement by LOR is restricted to an appropriate fraction of the non-detects by specifying 'Replace non-detects based on crop treated'.
See also: Modelling of missing data and replacement of non-detects

3.2.2.2 Modelling processing effects

Settings are:

- **no processing** [default]
- **processing (fixed factors)**
- **processing (distribution based)**

Concentrations in the consumed food may be different from the monitoring compound due to processing such as peeling, washing and cooking. Usually, processing lowers the concentration in the consumed food compared to the concentration in the unprocessed food. The effect of processing is modelled by multiplying the monitoring compound by a factor f_k which will typically be between 0 and 1. Occasionally, the processing factor may also be > 1, *e.g.* for drying. Often, processing factors may be variable across situations. This variability may be entered into the model by specifying two values:

- $f_{k,nom}$, the nominal value, typically some sort of central value;
- and, $f_{k,upp}$, an upper 95% confidence limit.

Distribution based processing factors require both values whereas for fixed factors only $f_{k,nom}$ or $f_{k,upp}$ needs to be specified (when both are specified, the highest value will be used; worst case scenario). No processing implies that $f_k = 1$.

To use processing factors f_k , choose processing (fixed factors) or processing (distribution based). Processing factors are read from table 'Processing' ($f_{k,nom}$ = procnom, $f_{k,upp}$ = procupp) and processing codes and labels from table 'ProcessingType'. Note that specifying no processing is a worst case scenario ($f_k = f_{k,upp} = 1$).
The program multiplies concentrations with fixed processing factors (in which case the conservative value $f_k = f_{k,upp}$ is used), or with random values sampled from a normal distribution with parameters μ and σ. The mean and standard deviation are based on transformed values of $f_{k,upp}$ and $f_{k,nom}$. The type of transformation for each processing type is specified in the last column of table ProcessingType. Choose disttype = 1 for a logistic-normal distribution or disttype = 2 for a lognormal distribution. To process simultaneously some foods using fixed factors and others distribution based, choose 'processing (distribution based)'. Now, fixed factors f_k are obtained by providing only $f_{k,upp}$ whereas random factors f_k are sampled when both $f_{k,upp}$ and $f_{k,nom}$ are given.
It is not necessary to fill out a complete list of processing factors for all foods. Missing values of $f_{k,nom}$ and $f_{k,upp}$ are, by default, replaced by the value 1.
See also: Modelling of processing effects

3.2.3 Unit variability model options

Settings are:

- **no unit variability** [default]
- **beta distribution**
- **lognormal distribution**
- **bernoulli distribution**

Monitoring measurements are typically made on homogenised composite samples. Each sample is composed of nu_k units with nominal unit weight wu_k each. The weight of a composite sample is often larger than a daily consumer portion. This implies that the mean level of the monitoring compound may be a fair estimate of the mean level of the food, but the variability of the monitoring measurements is certainly not appropriate to estimate the variance. Therefore, acute risks may be higher than would follow from a direct use of the composite sample data. This problem has been addressed by modelling unit variability.
In MCRA, the following three models for unit variability are available:

1. **Beta model**, requires knowledge of the number of units in a composite sample, and of the variability between units (realistic or conservative estimates);
2. **Lognormal model** requires only knowledge of the variability between units (realistic or conservative estimates).
3. **Bernoulli model**, requires only knowledge of the number of units in a composite sample (results are always conservative);

See also: Modelling of unit variability

3.2.3.1 Estimated parameters for unit variability

When parameters for unit variability based on empirical studies are available, these are used to simulate concentrations for a unit, assuming a parametric form for the unit-to-unit variability within a batch *e.g.* the beta or lognormal distribution.
Table 2 describes the four options when a parametric form for unit variability is specified. Compounds are simulated for a new unit in the batch using a lognormal distribution or for a unit belonging to the composite sample leading to the use of the beta distribution.

	Simulate for new unit in batch (lognormal distribution)	Simulate for unit belonging to composite sample (beta distribution)
Estimates of unit variability are realistic (R)	• no censoring at cm_k • no upper limit to the unit concentration	• no censoring at cm_k • unit values never higher than $nu_k \cdot cm_k$
Estimates of unit variability are conservative (C)	• unit values will be left-censored at cm_k • no upper limit to the unit concentration	• unit values will be left-censored at cm_k • unit values never higher than $nu_k \cdot cm_k$

Table 2: Choices for estimated variability factors. cm_k = value of composite sample concentration, nu_k = number of units in composite sample.

See also: Approaches to unit variability in probabilistic modelling: specifying distributions

3.2.3.2 Beta distribution

Find in the right section of the input form, option block Unit variability: Beta distribution (see Figure 18, default options).

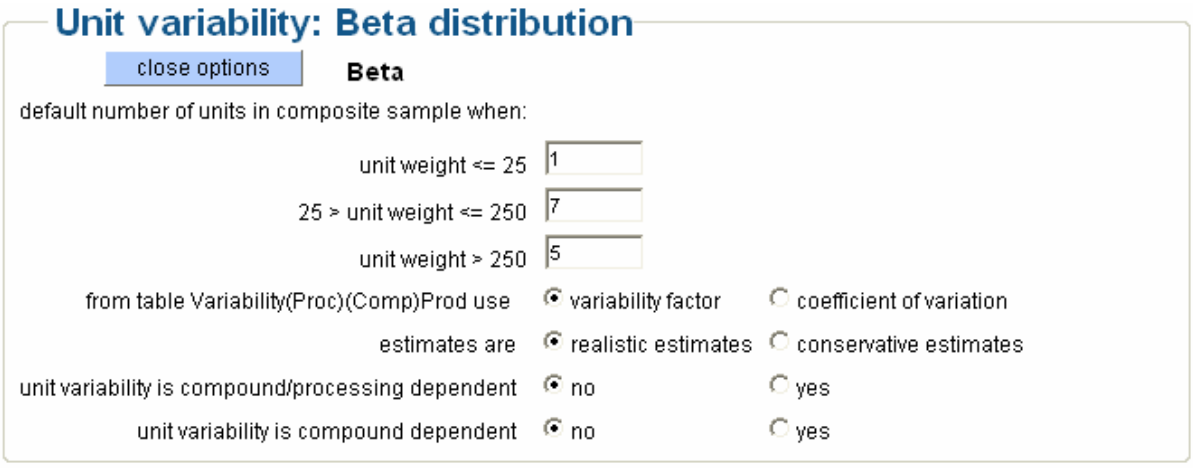


Figure 18: Option block Unit variability: Beta distribution

The parameter for unit variability is specified as a variability factor *v* or as a coefficient of variation *cv* of the unit values in the composite sample. Variability factors *v* (97.5th percentile divided by mean), coefficient of variation *cv* (standard deviation divided by mean) and number of units *nu* in the composite sample are retrieved from table VariabilityProd when unit variability is independent of the compound and processing type. If the variability factor is dependent on compound and/or processing type, data are expected in tables VariabilityCompProd or VariabilityProcCompProd, respectively.

The following variability factors *v* are recommended:

- for large crops (wu_k > 250g) value $v = 5$;
- for medium crops (wu_k 25- 250g) $v = 7$;
- for small crops ($wu_k \leq$ 25g) $v = 1$ (FAO/WHO, 1997).
- and for foods which are processed in large batches, *e.g.* juicing, marmalade/jam, sauce/puree, $v = 1$.

The latter information is specified in field bulkingblending of table ProcessingType (see 9.4.4). If the parameter for variability is missing, zero variability is assumed, and the unit concentrations are equal to the sampled composite sample concentrations.

Specify whether the supplied values for variability are realistic or conservative estimates. In the latter case, unit values are left-censored at the value of the mean (composite sample concentration). If there are no user-defined values for the number of units in the composite sample these are taken using a default scheme of nominal unit weights. This scheme follows in principle the definition of FAO/WHO (1997), as illustrated in Figure 18, but can be modified by the user.
See also: Beta model for unit variability

3.2.3.3 Lognormal distribution

Find in the right section of the input form, option block Unit variability: Lognormal distribution (see Figure 19, default options).

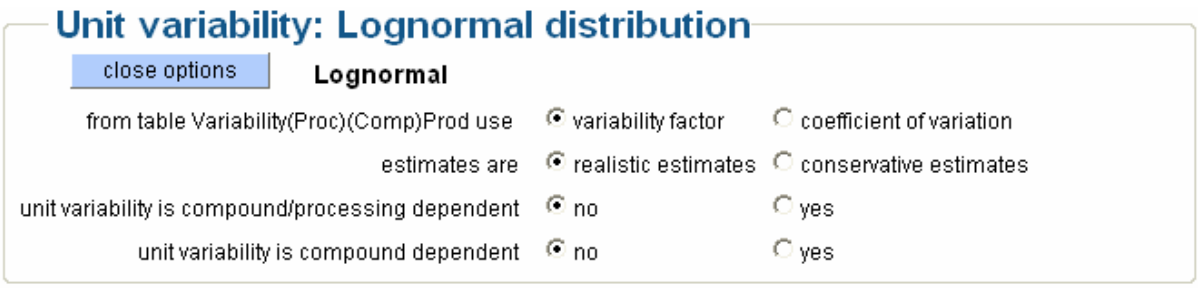


Figure 19: Option block: Unit variability: Lognormal distribution

In Figure 19 a parametric form for the unit-to-unit variability is specified. Concentrations are simulated for new units in the batch leading to the lognormal distribution.
The parameter for unit variability is specified as a variability factor v or a coefficient of variation cv.

The conversion of a variability factor into parameters of the lognormal distribution requires an exact definition of what is meant. Here, 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. Specify whether estimates are realistic or conservative.

In the conservative approach, unit concentrations of the composite sample are left-censored at the value of the monitoring compound. The realistic approach implies that the unit value may be lower than the value of the monitoring compound.
See also: Lognormal model for unit variability

3.2.3.4 Bernoulli distribution

Find in the right section of the input form, option block Unit variability: Bernoulli distribution, (see Figure 20, default options).

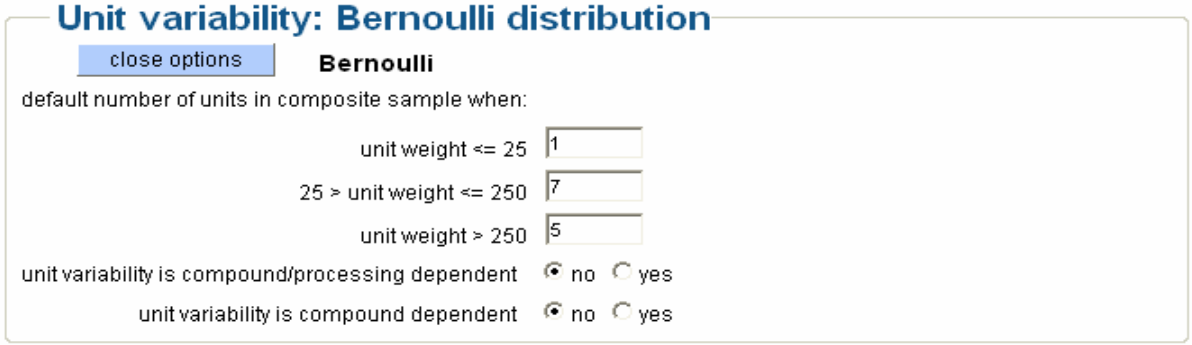


Figure 20: Option block Unit variability: Bernoulli distribution

In practice, measurements on individual units to obtain a measure for unit variability are not very common. Therefore, the number of units nu_k in the composite sample is used to define the parameter for unit variability (see van der Voet *et al*. 2001). 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.
The following variability factors v are recommended:

- for large crops (wu_k > 250g) value v = 5;
- for medium crops (wu_k 25- 250g) v = 7;
- for small crops ($wu_k \leq$ 25g) v = 1 (FAO/WHO, 1997).
- and for foods which are processed in large batches, *e.g.* juicing, marmalade/jam, sauce/puree, v = 1.

The latter information is specified in field bulkingblending of table Processing (see 9.4.4). The number of units within a consumption is calculated and for each unit a Bernoulli distribution is used to sample the monitoring compound with probability (v-1)/v or a multiple v of it with probability 1/v (see Figure 20).
See also: Bernoulli model for unit variability

3.2.4 Additional options concerning IESTI and consumption days

Find in the right section of the input form, option block Additional options that handles IESTI and consumption days only (see Figure 21, default options).

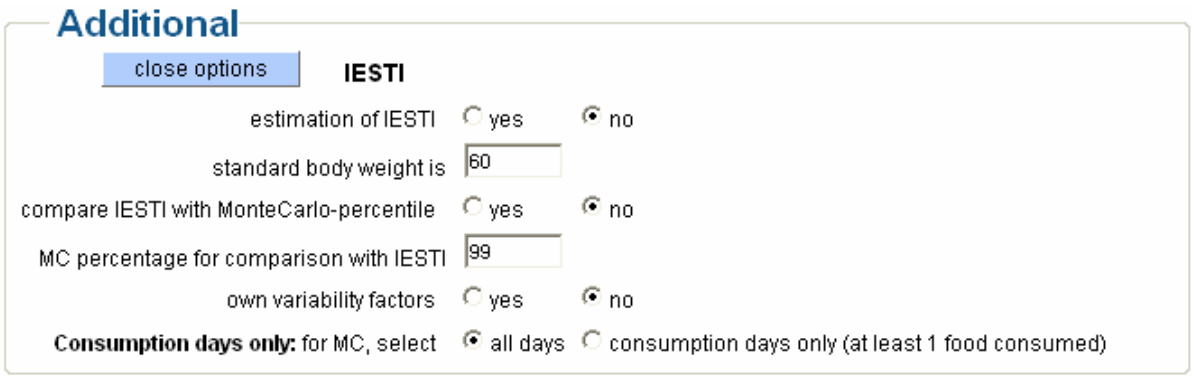


Figure 21: Option block Additional in the third section of the MCRA input form screen

The IESTI (International Estimated Short-Term Intake) is a prediction of the short-term intake of a compound on the basis of the assumptions of high daily food consumption per consumer and highest compounds and, in case of blending and bulking, the median compound from supervised trials. The IESTI is expressed in microgram/kg body weight/day and estimated per food.
IESTI estimates are requested when estimation of IESTI is set to yes. Standard unit variability factors and a standard body weight of 60 kg are specified. The IESTI is compared with estimates of a specified percentile (per food) of the MC-simulation.
In the output (not shown) two kinds of estimates of the MC-percentile are given: one for “All days” and one for “Consumption days only”. Be aware that specification of option ‘consumption days only’ may alter the interpretation (and estimate) of the percentile for “All days”. In the latter case the estimate refers to a smaller subset containing consumption days only. However, note that still not every food is eaten on every consumption day. The interpretation and estimate of the percentile for “Consumption days only” is not affected by setting option ‘consumption days only’.
The IESTI calculations correspond to the definition of FAO/WHO (2002) that may be considered as the deterministic counterpart of the probabilistic approach used in MCRA.
In 10.3 output of IESTI calculations are shown.
See also: Additional: Comparison of probabilistic with deterministic estimates of acute risk

3.2.5 Intake model options

Settings are:

- **only empirical estimates** [default]
- **empirical estimates and betabinomial/normal**

3.2.5.1 Only empirical estimates

This is a straightforward acute risk analysis.

3.2.5.2 Empirical estimates and betabinomial/normal

Find in the right section of the input form, option block Intake model (see Figure 22, default options).

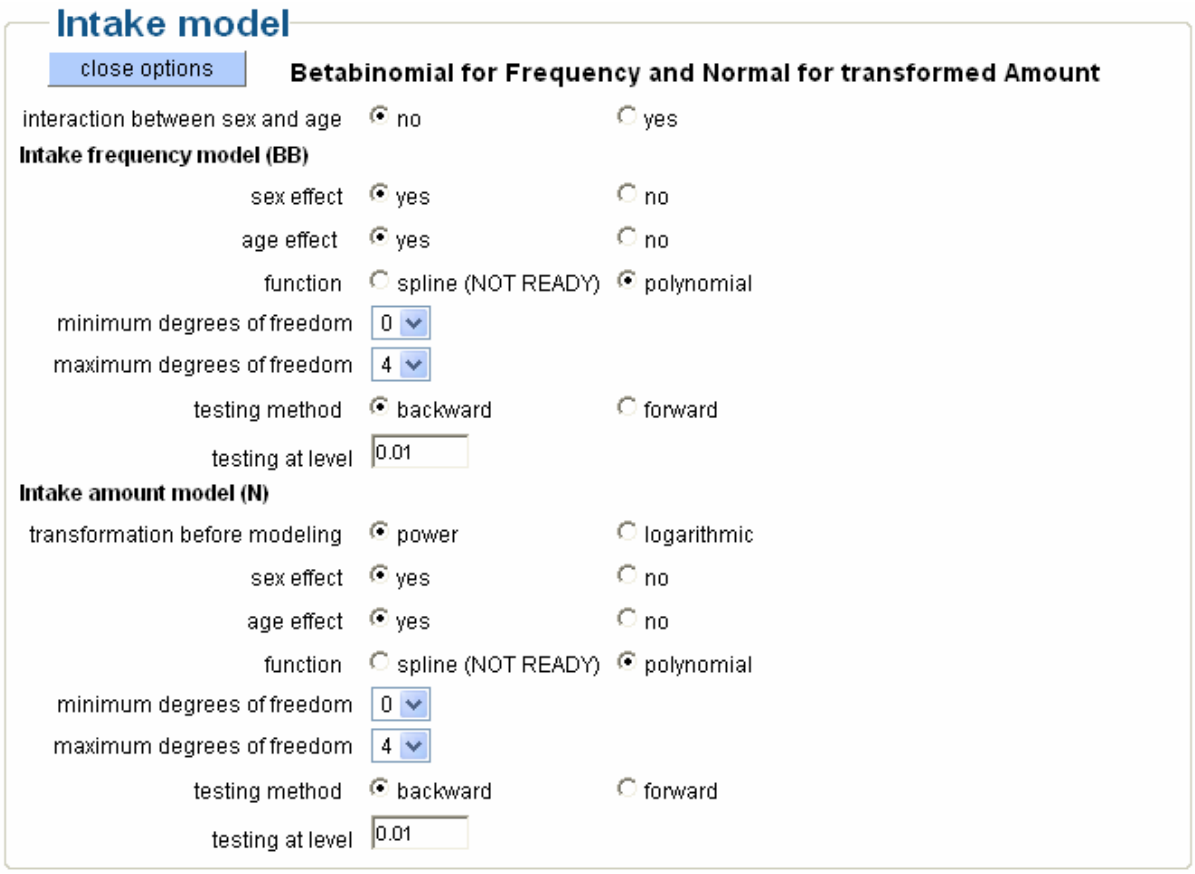


Figure 22: Option block Intake model in the right section of the MCRA input form screen

Note that estimation of a covariable and/or cofactor dependent acute intake distribution is additional to the standard analysis. The simulated intake values of an acute analysis are used as input. Then the analysis proceeds:

- the betabinomial distribution is used to model the intake frequency. Frequencies may be related to a covariable using a spline or polynomial function and/or to a cofactor;
- non-zero intakes are transformed using a logarithmic or power transformation. The normal distribution is used to model the transformed intake amounts and a spline or polynomial function may be used to describe the effect of a covariable;
- both distributions are numerically integrated to obtain the acute intake distribution. Covariable- and cofactor-dependent percentiles can be derived from the corresponding acute intake distributions.

See also: Empirical estimates and betabinomial/normal

Intake frequency model

The intake frequency function models the probability of consumers of having an intake. Depending on the consumption pattern, we have regular, less regular and incidental consumers. So each consumer has his own probability of having an intake. For many foods, there may be a relation with age (or sex or, if available, other demographic data): the probability of having an intake may be related to *e.g.* the age of the consumer. The betabinomial distribution is very suited to sample consumer probabilities using *e.g.* age and/or sex as explanatory variables.
In Figure 22, decide on modelling the effect of a covariable. Choose a spline function or a polynomial. A smoothing spline is a complicated function, constructed from segments of cubic polynomials with constraints to ensure smoothness. A polynomial function is based on orthogonal linear, quadratic, cubic or quartic curves. The degree of smoothness of the spline or polynomial function is controlled by increasing or decreasing the degrees of freedom. A spline with the maximum degrees of freedom is less smooth than a spline with the minimal degrees of freedom. Decide on the method of testing: backward selection means that testing starts with a spline or polynomial of the highest degree. In each elimination round the number of degrees of freedom is decreased one at a time, and the process is stopped when the resulting decrease in fit is significant at the specified significance level as judged on the basis of a deviance test. Forward selection means that the evaluation of the degree of the spline or polynomial is started with a function of the lowest degree. In all evaluations the testing level is 0.01.
See also: Modelling the intake frequency distribution

Intake amount model
Choose whether the non-zero intakes are transformed by a logarithmic or power transformation and whether the amounts should be related to *e.g.* age and/or sex. The transformed amounts are analysed with maximum likelihood using a spline or polynomial function to describe the effect of a covariable. The analysis provides covariable- and cofactor-dependent mean intakes and the standard deviation of the transformed intake distribution, see Figure 22.
See also: Modelling transformed intake amounts

Estimation of covariable- and/or cofactor-dependent percentiles
For each combination of levels of the covariable and cofactor the acute intake distribution is obtained by numerical integration. The parametric intake frequency and intake amount model are used to derive distributions of intake frequency and intake amount values through the use of MC-sampling. For each level of the covariable and cofactor, the corresponding distributions are multiplied, and an acute intake distribution is obtained which represents the acute intake for that specific combination of levels of the explanatory variable(s). The number of MC simulations is specified in the MCRA input form, see Figure 16.
See also: Estimating the acute intake distribution

3.3 Chronic risk type

Specify in the top section (see Figure 14) risk type is chronic and uncertainty. Specifications in the left section (see Figure 23, default options) are:

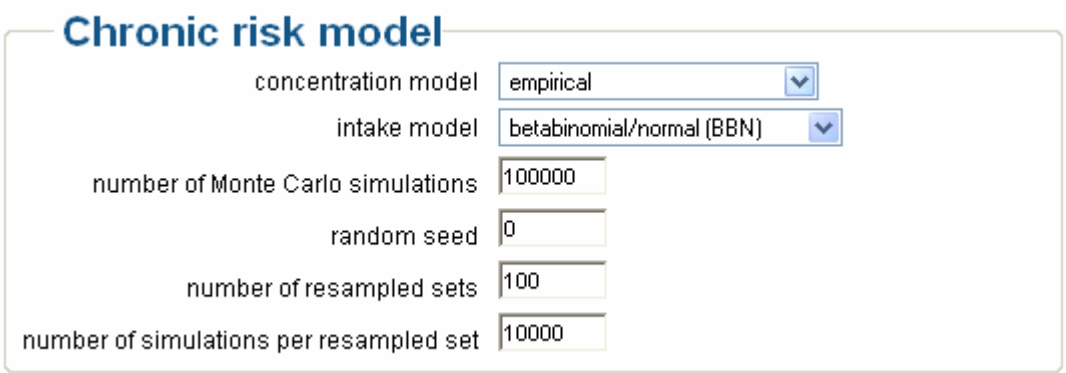


Figure 23: Left section of the MCRA input form if type of analysis is Chronic

In dietary risk assessment, usual intake is defined as the long-run average of daily intakes of a dietary component by an individual consumer. In the MCRA program, for chronic risk assessment, the intake is calculated as the consumption on each day of each consumer multiplied by the average value of the compound concentration levels (non-detects and detects) divided by body weight and, if specified, applying processing and/or replacing zeros with the LOR (based on percent crop treated). Note, unit variability is not relevant in chronic risk assessment. Compound data for a chronic risk assessment may be present as full, summary or histogram data. For full data, a choice can be made between a parametric and a non-parametric (empirical) approach. For summary or histogram data a parametric approach is obligatory. Note, option 'consumption-days only' is not relevant for chronic risks and chronic risk assessment is only performed when the total number of days per consumer is 2 or more.
See also: Chronic risk assessment

Number of Monte Carlo simulations
The usual intake distribution is obtained by a numerical integration procedure. Specify the number of MC-simulations to estimate the distribution. In Figure 16, the total number of MC-simulations is 100,000.

Random seed
The numerical integration is based on MC-simulation and requires a pseudo-random number generator that is initialised by setting the seed. To get time-based values, set seed to zero. For repeatable sequences, set the seed to an integer number. See 3.2

Number of resampled sets
Specify the number of resampled sets to assess the uncertainty. Settings are the same as for the acute risk model. See 3.2

Number of simulations per resampled set
Specify the number of simulations per resampled set. In Figure 16, the specified number of simulations per resampled set is 10,000.

3.3.1 Concentration model options
Settings are:

- **empirical** [default] (see 3.2.1.1)
- **binomial/normal (no pooling)** (see 3.2.1.2)
- **binomial/normal with pooling** (see 3.2.1.3)

For options of Concentration data (see 3.2.2)

3.3.2 Intake model options
Settings are:

- **betabinomial/normal (BBN)** [default]
- **discrete/semi-parametric (ISUF)**

The covariable and/or cofactor are selected in the screen shown in Figure 11**.**

3.3.2.1 Betabinomial/normal (BBN)
Find in the right section of the input form, option block Intake model (Figure 24, default options).

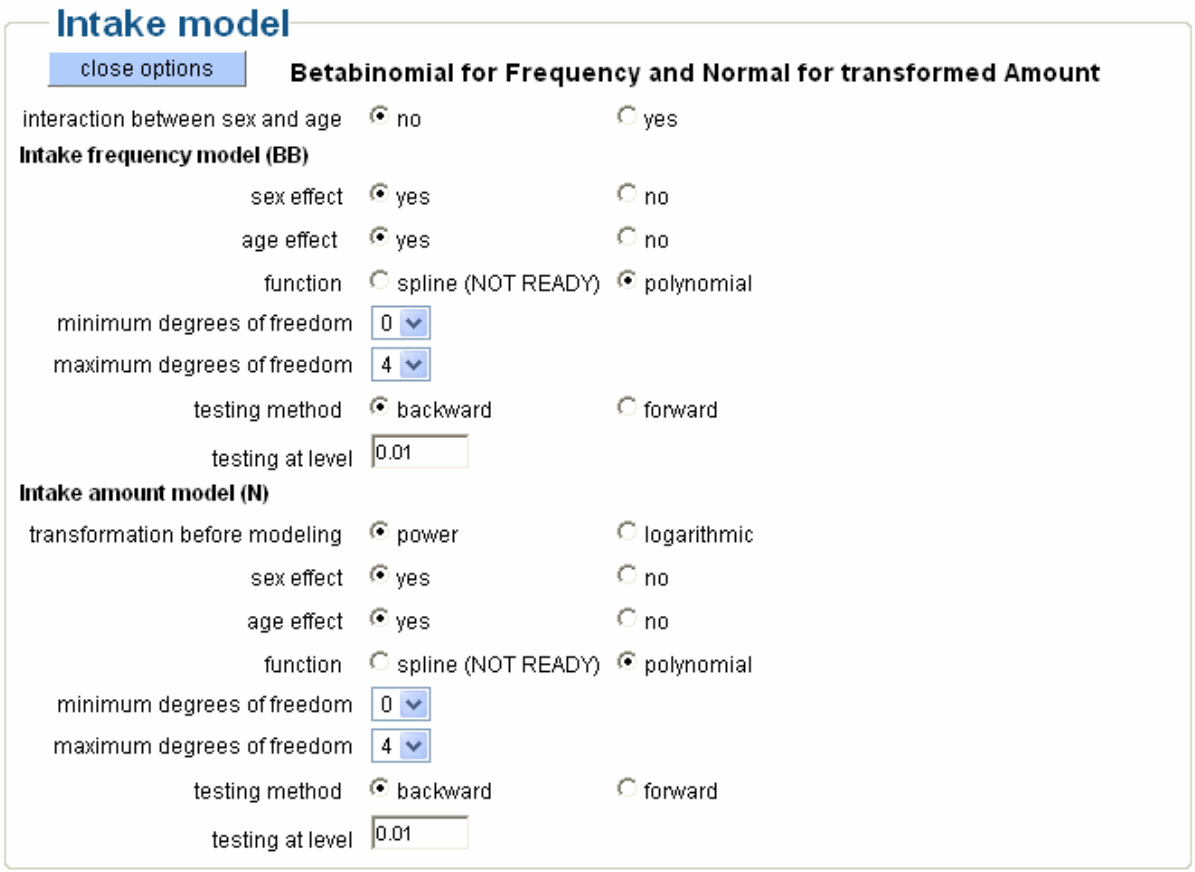


Figure 24: Option block Intake model if intake model is betabinomial/normal (BBN)

For a chronic risk assessment, the intake is calculated as the consumption on each day of each consumer multiplied by the average value of the compound concentration levels divided by body weight (see 3.3). Apply the betabinomial/normal model:

- the betabinomial distribution is used to model the intake frequency. A spline or polynomial function are available to describe the effect of a covariable. Choose whether an interaction is included when both cofactor and covariable effect are modelled.
- non-zero intakes are transformed to normality and a maximum likelihood algorithm is used to estimate the effects of explanatory variables and the standard deviation parameter of the usual intake distribution (variance between individuals).
- both distributions are numerically integrated to obtain the usual intake distribution. For each combination of levels of the explanatory variables, an intake frequency distribution and a transformed intake amount dsitribution is simulated through MC-sampling. Both distributions are multiplied to obtain the usual intake distribution. The number of MC-simulations is specified in the MCRA input form, see Figure 23. Covariable- and cofactor-dependent percentiles can be derived from the corresponding usual intake distributions.

Note that the interaction option refers to both models (no separate modelling).
See also: Betabinomial/normal

Intake frequency model

The intake frequency model describes the probability of having an intake. Depending on the consumption pattern, we have regular, less regular and incidental consumers, each consumer has his own intake frequency.
For many foods, there may be a relation with explanatory variables like *e.g.* age and/or sex. The betabinomial distribution is very suited to model the intake frequency as function of explanatory variables.

In Figure 24, choose whether to model *e.g.* age and/or sex effects. Choose a spline or a polynomial function. A smoothing spline is a complicated function, constructed from segments of cubic polynomials with constraints to ensure smoothness. A polynomial function is based on orthogonal linear, quadratic, cubic or quartic curves. The degree of smoothness of the spline or polynomial function is controlled by increasing or decreasing the degrees of freedom. A spline with the maximum degrees of freedom is less smooth than a spline with the minimal degrees of freedom. Decide on the method of testing: backward selection means that testing starts with a spline or polynomial of the highest degree. In each elimination round the number of degrees of freedom is decreased one at a time, and the process is stopped when the resulting decrease in fit is significant at the specified significance level as judged on the basis of a deviance test. Forward selection means that the evaluation of the degree of the spline or polynomial is started with a function of the lowest degree. Choose your testing level. If both cofactor and covariable are included, decide on modelling the interaction. For a polynomial, the interaction means that curves are no longer parallel and intercepts may differ.
See also: Modelling the intake frequency distribution
See also: Intake frequency model with covariable and/or cofactor

Intake amount model
The non-zero intake amounts are transformed to approximately normality by a logarithmic or power transformation. The transformed amounts are analysed with maximum likelihood using a spline or polynomial function to model the effect of a covariable. The analysis provides mean intakes of the transformed intake distribution dependent on explanatory variables. The total variance of the non-zero transformed intake amounts is divided into a between individuals and a between days within individuals a variance component. The between-individuals component is the basis for the estimation of the distribution of the usual intake.
See also: Modelling the positive intake amounts
See also: Intake amount model with covariable and/or cofactor

Estimation of the usual intake distribution by numerical integration
The *e.g.* age and/or sex dependent usual intake distribution is derived by a numerical integration. The parametric intake frequency model and parametric intake amount model are used to derive for each combination of levels of age and sex, a distribution of intake frequency values and intake amount values. This is done by MC-sampling. By multiplying both distributions, a distribution is obtained which represents the usual daily intake for a specific combination of levels of age and/or sex. The number of MC simulations is specified in the MCRA input form, see Figure 16.
See also: Estimating usual intake distributions

*3.3.2.2 Discrete/*semi-parametric *(ISUF)*
Find in the right section of the input form option block Intake model (see Figure 25, default options).

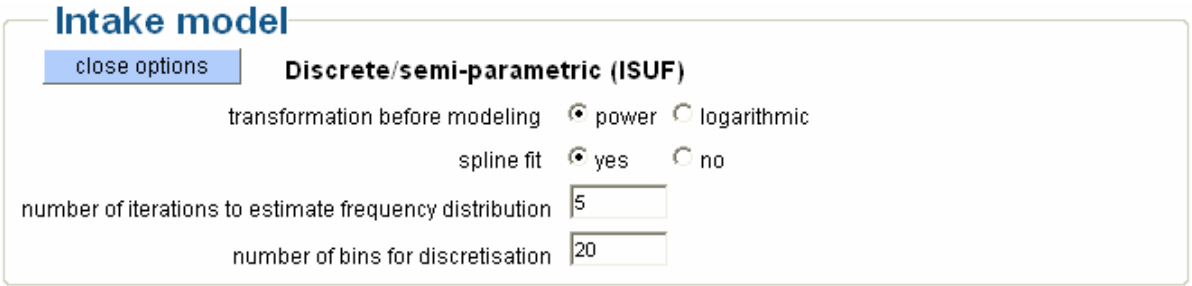


Figure 25: Option block Intake model if intake model setting is discrete/semi-parametric (ISUF)

For a chronic risk assessment the intake is calculated as the consumption on each day of each consumer multiplied by the average value of the compound concentration values divided by body

weight (see 3.3). Before estimating the chronic percentiles of the distribution applying the discrete/semi-parametric (ISUF, Iowa State University Foods) model, non-normal intake data are transformed to approximate normality following an approach proposed by Nusser *et al*. (1996, 1997) and Dodd (1996).
See also: Discrete/semi-parametric (ISUF)

Transformation: power or logarithmic
Specify a power or logarithmic transformation. Usually, a power transformation is satisfactory.
See also: Power or log transformation

Spline fit or not
The transformation to normality is improved by fitting a spline to the transformed intakes. A smoothing spline is a complicated function, constructed from segments of cubic polynomials with constraints to ensure smoothness.
See also: Spline fit

Number of iterations to estimate the intake frequency distribution
The intake frequency distribution is estimated in an iterative process. Specify the number (x 1000); here 5 is specified, giving a total of 5000 iterations, see Figure 25:
See also: Back transformation and estimation of usual intake

Number of bins for discretisation
The intake frequency is estimated on a discretised probability grid. Here, the resolution of the grid is equal to 20. So, the probability mass is discretised at a grid ranging from 0 to 1 with step length equal to 0.05.
See also: Back transformation and estimation of usual intake

3.4 Uncertainty analysis

Specify in the top section (see Figure 14) an uncertainty analysis.
See also: Uncertainty of risk assessments: resampling data

3.4.1 Uncertainty analysis options for acute risks

Find in the right section of the input form, option block Uncertainty analysis (see Figure 26, default options).

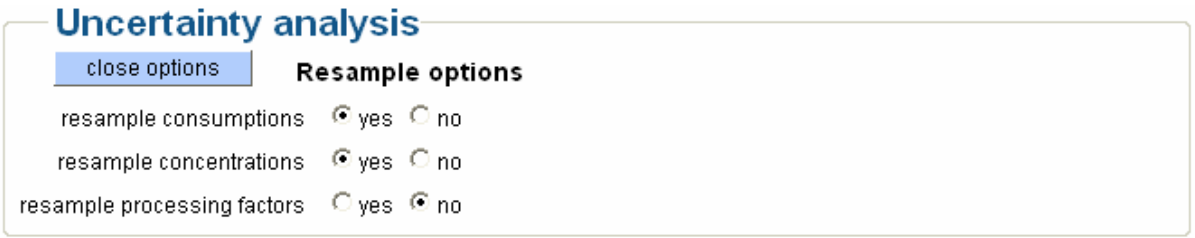


Figure 26: Option block Uncertainty analysis if an acute (or chronic BBN) risk model is chosen

The uncertainty of output statistics (*e.g.* mean or percentiles of the intake distribution) is assessed by resampling datasets (consumptions, concentrations) or distributions (processing factors). Resampling can be applied on the level of fresh MC-samples, on the level of consumers and on the level of the compound concentration data. To examine the uncertainty due to MC-variability in each analysis only, specify 'no' for all three options. Then data are resampled from the original data. Apply resample consumptions and resample concentration data: from each dataset, data are resampled (with replacement) to construct a so-called bootstrap sample. Setting resample procesing factors to yes will generate new processing factors from a parametric uncertainty distribution. Parameters of this uncertainty distribution have to be entered in the Processing table (see 9.4.5). From the resampled

data sets and parameters an intake distribution is simulated. Each resampled set provides a mean, maximum and percentiles according to the specified percentages and all replicates together contain the information to make inferences from the data, *e.g.* to establish the uncertainty of mean, maximum and percentiles.
The number of resampled sets and the number of simulations per set is specified in the left section of the MCRA input form (see Figure 16). Here, 100 resampled sets are specified and each set is resampled 10,000 times. The number of values within a set restricts which percentiles are displayed. Here, the highest possible percentage for which uncertainty information can be calculated is the 99.99^{th} percentile, for a set containing 1000 simulations this is the 99.9^{th} percentile.

3.4.2 Uncertainty analysis options for chronic risks: betabinomial/normal (BBN)

Options are the same as for an uncertainty analysis for acute risks, described in 3.4.1 .

3.4.3 Uncertainty analysis options for chronic risks: discrete/semi-parametric (ISUF)

Find in the right section of the input form, option block Uncertainty analysis (see Figure 27, default options).

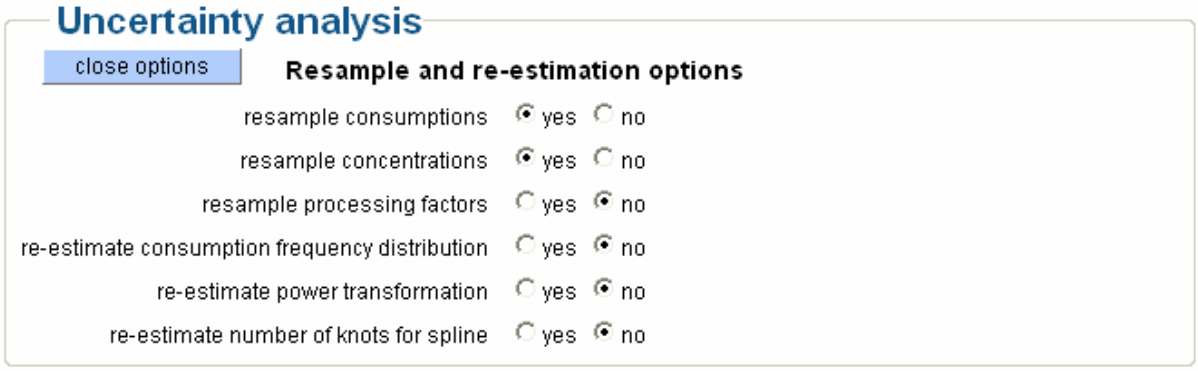


Figure 27: Option block Uncertainty analysis if chronic risk model ISUF is chosen

Uncertainty is assessed by resampling the consumption and compound concentration data. Option 're-estimate consumption frequency distribution' is only relevant when consumptions are resampled. Note that estimation of the frequency distribution is time consuming. Options 're-estimate power transformation' and 're-estimate number of knots for spline' are used in the transformation to normality of the intake distribution.

3.5 Output

Depending on the type of analysis there are different output option blocks to specify graphics and tables.

3.5.1 Output options for acute risks: only empirical estimates

In Figure 28 (default options), percentages and exposure limits are specified separated by a space.

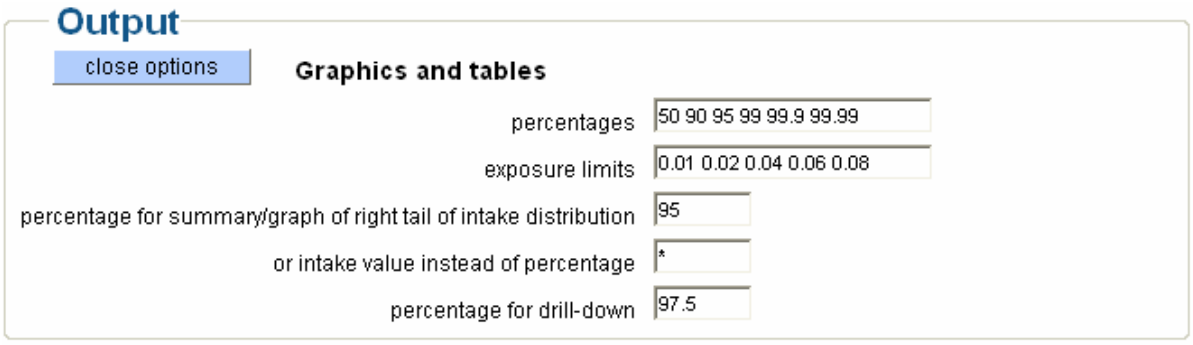


Figure 28: Option block Output if an acute risk model is chosen with only empirical estimates

The next option specifies the percentage that is used for summarising the contribution of foods to the right tail of the intake distribution and to display a graph of the upper tail. A percentage may be specified, but specifying an intake value in the option below overrules the percentage. The drill-down percentage is used to select the nine consumers around the percentile of the intake distribution that corresponds to the specified percentage. For these consumers the age and weight are displayed, the intake, the consumption and the compound concentration contents found on each consumption. To display information of consumers with the highest intake, specify percentage 100.

3.5.2 Output options for acute risks: empirical estimates + betabinomial/normal (BBN)

The upper part of Figure 29 (default options), see 3.5.1 . In the lower part, options related to the covariable are shown, here age. The minimum and maximum age are retrieved from the database, but are overruled by specifying own values. The step length is automatically determined but can be overruled. Default, approximately 20 steps are taken and the calculated step length is rounded to the nearest integer. To get predictions for specific ages, specify your extra age's space separated. Extra ages may lay outside the specified range determined by the minimum and maximum age.

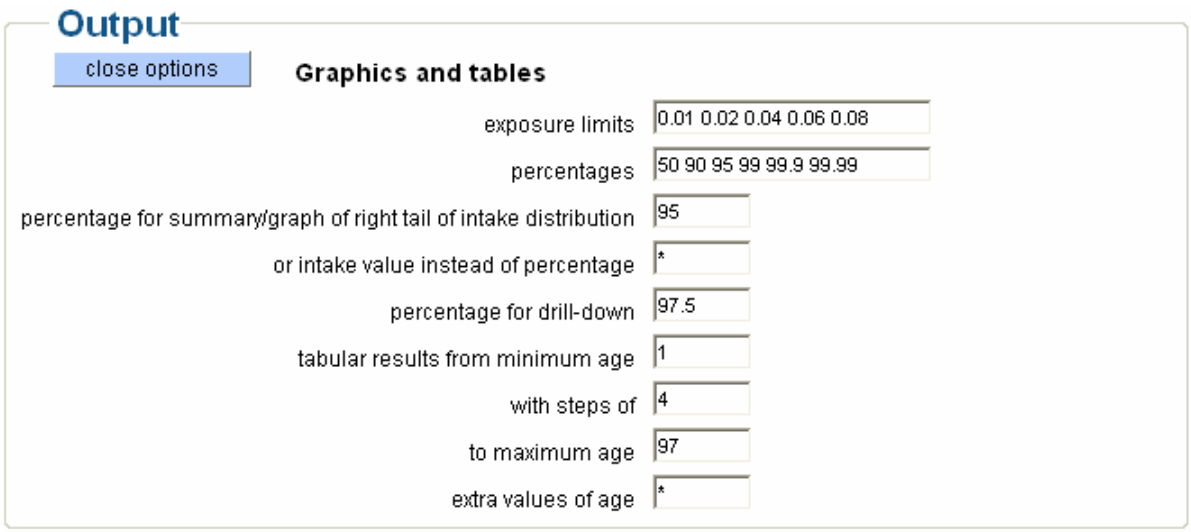


Figure 29: Option block Output if an acute risk model is chosen with empirical estimates + betabinomial/normal (BBN)

3.5.3 Output options for chronic risks: betabinomial/normal (BBN)

See Figure 30 (default options) and 3.5.1 and 3.5.2 .

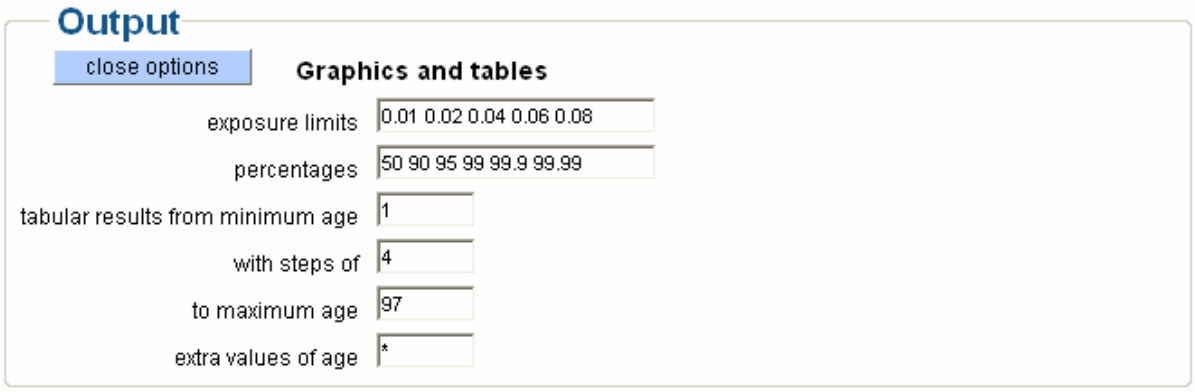


Figure 30: Option block Output if chronic risk model betabinomial/normal (BBN) is chosen

The last 4 options are only relevant when a covariable is modelled (see Figure 24).

3.5.4 Output options for chronic risks: discrete/semi-parametric (ISUF)

See Figure 31 (default options) and 3.5.1 .

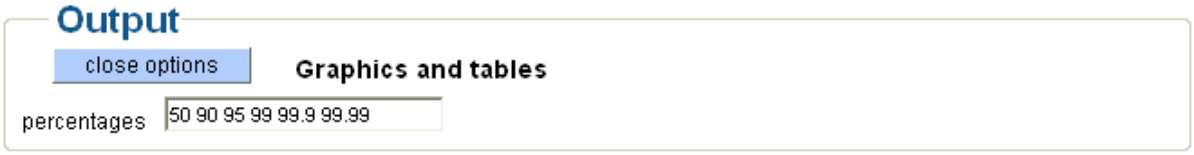


Figure 31: Option block Output if chronic risk model discrete/semi-parametric (ISUF) is chosen

3.6 Running an MCRA job

Click the 'MCRA-submit-job'-button in the top section to run a MCRA analysis. Check the radio button if you wish to be notified when the analysis is completed. After submitting the form, enter the MCRA central menu (see Figure 32).

3.7 Checking the processing time

After a submit, all model specifications are passed to the server and the analysis is initiated. First, all data are exported to system files located on the user directory. Export of data takes is only done when data are changed, that is after selection of new tables or subset selection.
Click the 'show progress'-button to view the progress of the analysis (see Figure 32). Here, also information is given about the estimated CPU time. Performing a risk assessment may be time consuming when data files are large and/or the number of simulations is high. Click the 'update window'-button to refresh the screen. Click the 'kill job'-button to end a job (with a fatal error status, not shown).

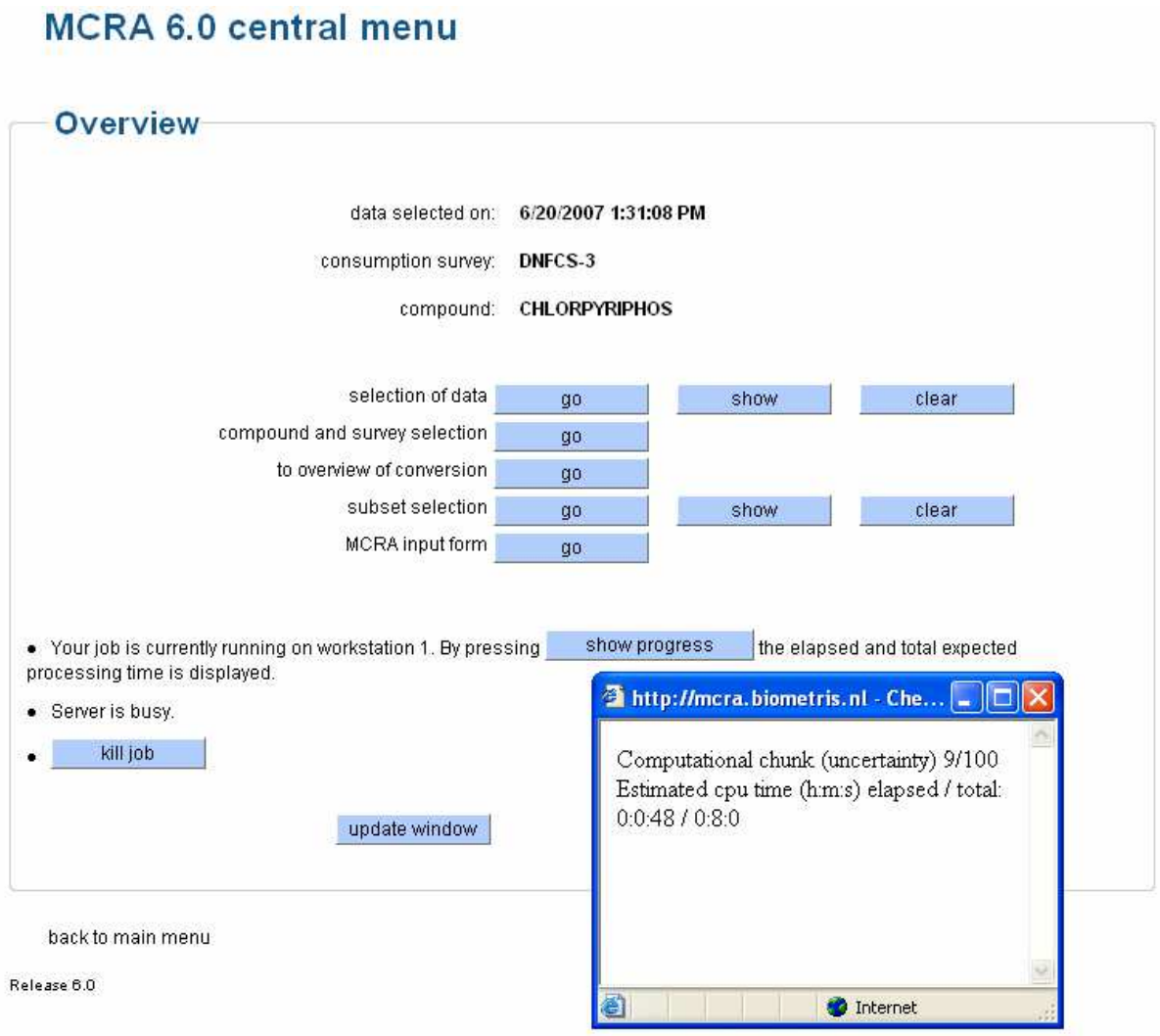


Figure 32: Central menu, show progress

After completion (click the 'update window' button) a new button 'view output' appears (see Figure 7). Click the 'view output'-button to get the screen in Figure 34 (see 4.1).

4 Example output

Click the 'view output'-button in Figure 7. Depending on the type of analysis, several output windows may appear (for example Figure 34 and Figure 39).
In 4.1 , the output is shown of an acute risk analysis. In 4.2 an example is given of a chronic risk analysis. Other examples can be found in Ch. 10 . For downloading results for off-line viewing, see 2.5 .

4.1 Acute risk assessment: basic analysis

This example shows output of an acute risk assessment for organo phosphate pesticide Chlorpyrifos which has neurotoxic effects. The Dutch Validation database from RIKILT (NL) is used. Select 'All tables', survey 'DNFCS-3' and compound 'CHLORPYRIFOS'. Then, after selection of (see 6.2): Table_grapes, Grapefruit, Lemon, Mandarin, Orange, Apple, Peach, Peppers and Potato, the input form is reached (see Figure 33).

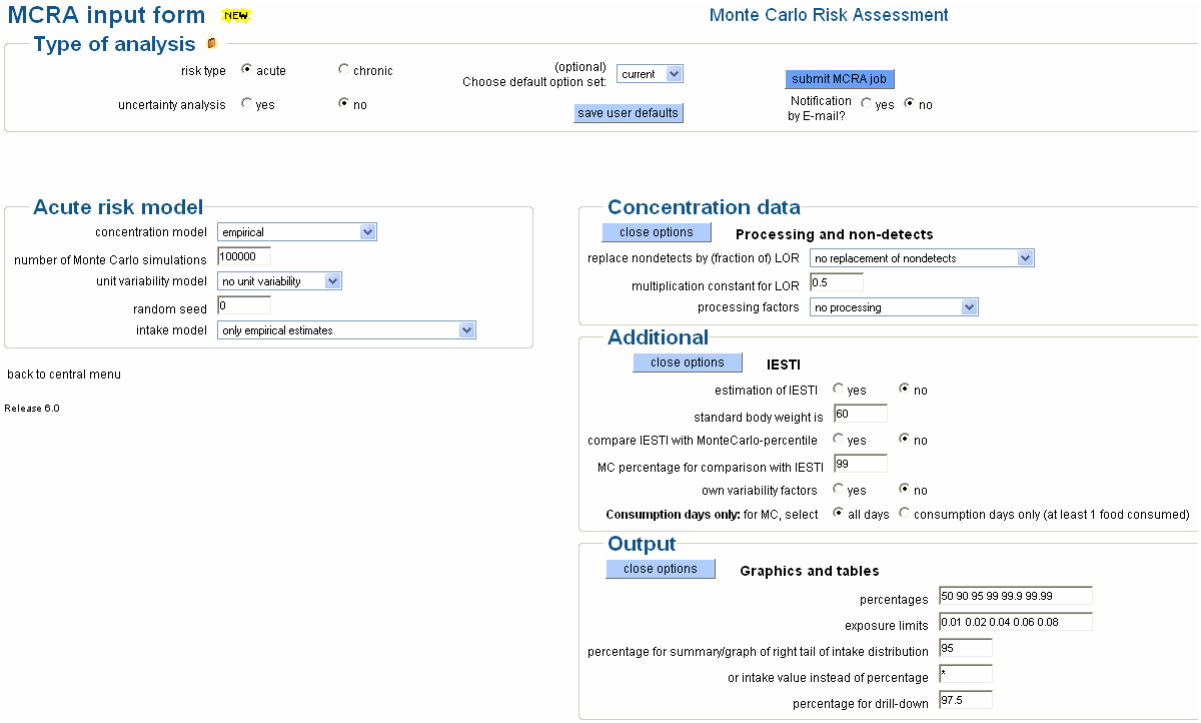


Figure 33: MCRA input form for an acute risk analysis

Table 3 lists the main options:

Input form	
risk type	acute
uncertainty analysis	no
concentration model	empirical
number of Monte Carlo simulations	100000
unit variability model	no unit variability
random seed	0
intake model	only empirical estimates
concentration data	system defaults
Additional	system defaults
output:	
percentage for summary/graph	95

Table 3: Input form options: basic acute risk analysis

Click the 'submit MCRA job' in Figure 33 to start the analysis. When the analysis is finished, view your results using the 'view output'-button in the central menu. This gives Figure 34:

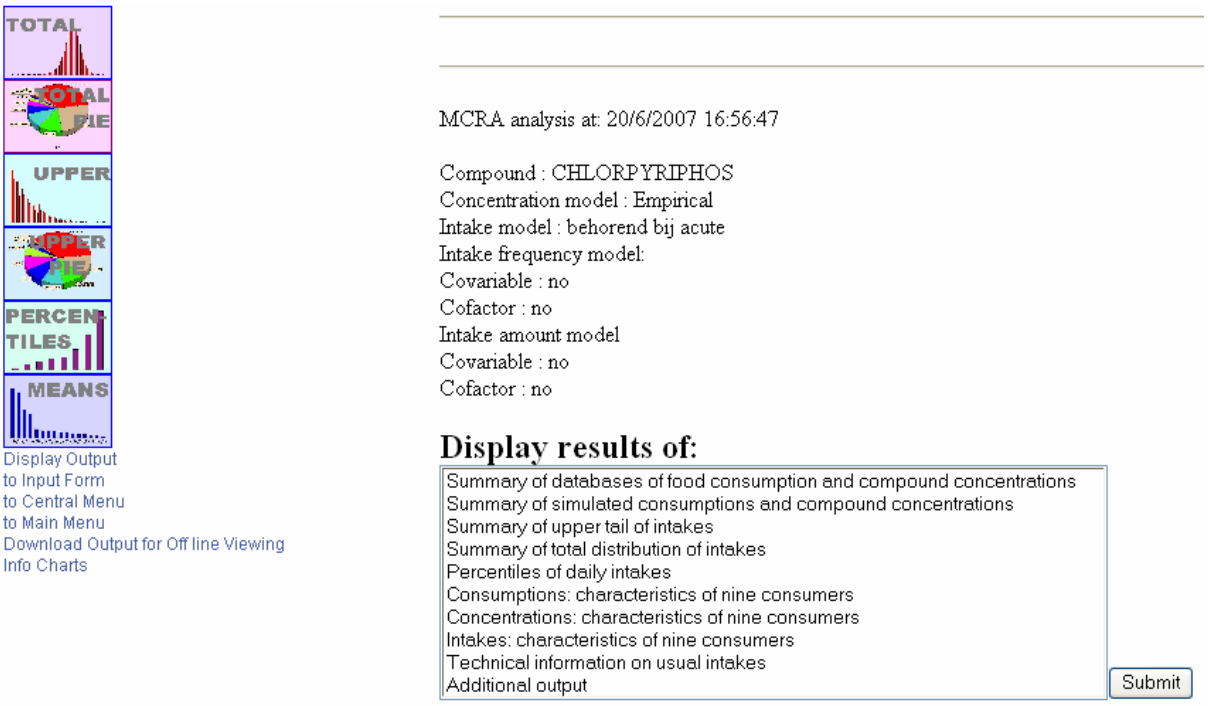


Figure 34: View output window for an acute risk analysis

Choose 'Additional output' and click the 'Submit'-button in Figure 34. In Table 4 you find the main characteristics concerning this analysis taken from the 'Additional output' file.

'Additional output' file	
Survey name	DNFCS-3
Compound code	CHLORPYRIFOS
Number of foods	9
Acute reference dose (ARfD)	100
Acceptable daily intake (ADI)	10
Number of detects	655
Number of non-detects	3267
No of consumers	6250
Population characteristics,	
minimum age	1
maximum age	97
minimum weight	8
maximum weight	150
Sex	female, male
Total no of consumption days	12132

Table 4: Information taken from the 'Additional output' file

Choose 'Summary of databases of food consumption and compound concentrations' and click the 'Submit'-button in Figure 34: a summary is given from the data stored in your databases (see Table 5). The upper part of the table displays all information concerning consumption data and the lower part all information concerning compound concentrations. For each food you will find the code and label.

In the upper part the average consumption over all consumers and all days (MeanConsum) together with the average consumption on consumption days only (MeanConsDays) is given. In addition, the number of consumption days (NconsDays) and total number of days (Ndays) are displayed, also expressed as percentage consumption days (%ConsDays). So, there are 12500 days (2 days for 6250 consumers). Grapefruit is consumed on 1785 days (14.3%). The average consumption of Grapefruit on these days is 4.06 x 12500/1785 = 28.5 g.

```
--------------------------------------------------------------------------------
Summary of databases of consumptions and compound concentrations with respect to:

Compound:  CHLORPYRIPHOS

--------------------------------------------------------------------------------

Code         : food code
Food         : food label
--------------------------------------------------------------------------------
MeanConsum   : average consumption, all consumers, all days
MeanConsDays : average consumption, consumption days only
NConsDays    : number of consumption days in the data set
NDays        : total number of days
%ConsDays    : percentage consumption days
--------------------------------------------------------------------------------
MeanConcen   : mean concentration of all samples
MeanPosConc  : mean concentration of samples with positive concentrations
NSamplPos    : number of samples with positive concentrations
NSamples     : total number of samples with concentration measurements
%SamplPos    : percentage samples with positive concentrations
--------------------------------------------------------------------------------


FOOD CONSUMPTION DATA

        Code         Food MeanConsum MeanConsDays NConsDays    NDays   %ConsDays
                               (g)          (g)                            (%)
      FB1235 TABLE-GRAPES      13.23         34.2      4836    12500        38.7
      FC0203 GRAPEFRUIT,        4.06         28.5      1785    12500        14.3
      FC0204 LEMON, SEE A       1.54          4.2      4589    12500        36.7
      FC0206 MANDARIN, SE       8.83         41.2      2677    12500        21.4
      FC0208 ORANGE, SWEE      57.06         99.6      7163    12500        57.3
      FP0226 APPLE             61.50         99.4      7737    12500        61.9
      FS0247 PEACH              2.08          7.3      3538    12500        28.3
      VO0445 PEPPERS, SWE       3.33         16.5      2533    12500        20.3
      VR0589 POTATO           138.52        173.1     10005    12500        80.0

COMPOUND CONCENTRATION DATA

        Code Food         MeanConcen  MeanPosConc NSamplPos NSamples   %SamplPos
                             (mg/kg)      (mg/kg)                          (%)
      FB1235  TABLE-GRAPES    0.0159       0.1178       136     1007        13.5
      FC0203  GRAPEFRUIT,     0.0693       0.1591        47      108        43.5
      FC0204  LEMON, SEE A    0.0128       0.0575        16       72        22.2
      FC0206  MANDARIN, SE    0.1073       0.1467       166      227        73.1
      FC0208  ORANGE, SWEE    0.0522       0.1078       222      459        48.4
      FP0226  APPLE           0.0030       0.0554        37      680         5.4
      FS0247  PEACH           0.0076       0.0562        21      156        13.5
      VO0445  PEPPERS, SWE    0.0011       0.1156         9      942         1.0
      VR0589  POTATO          0.0002       0.0600         1      271         0.4
                                                        ----    ----
                              number of detects =        655     3922 = total
                                                                number of samples

                    the number of non-detects =        3267 ( 3922 - 655  )
```

Table 5: Summary of the database, consumptions and compounds

The lower part of Table 5 displays information concerning compound concentrations on each food. The mean concentrations of all samples (MeanConcen) and of positives only (MeanPosConc) are calculated. The number of positives (NSamplPos) and the total number of concentration measurements (NSamples) are given as well as the percentage of positive concentrations (%SamplPos). In this example for Grapefruit 47 positive concentrations are found out of 108 samples (43.5%). The mean concentration of the positive samples is 0.0693 x 108/47 = 0.1591 mg/kg.

Choose 'Summary of simulated intakes'. In Figure 34 you find essentially the same kind of information but all statistics are calculated using simulated data (see Table 6). The simulation is performed by sampling N (in Table 6 100,000) times a consumption pattern from the food consumption database (with replacement). For each consumed food a concentration value is sampled from the distribution of concentrations of that food. Multiplying consumption and concentration gives the intake per food. Summing up the intakes over foods divided by the body weight gives the total intake of the compound (here CHLORPYRIFOS) expressed in microgram per kg body weight per day. For both consumption and concentration find additional statistics DeltaC and DeltaR. These statistics displays the difference expressed as percentage between the mean (MeanConsum and MeanConcen, respectively) of the simulated data and the mean of the data as found in your database, respectively. The better the simulation mimics the database, the lower the percentages indicating the deviation between simulated and input data.

```
-------------------------------------------------------------------------------
Summary of simulations of consumptions and compound concentrations with respect to:

Compound:  CHLORPYRIPHOS

-------------------------------------------------------------------------------
Code         : food code
Food         : food label
-------------------------------------------------------------------------------
MeanConsum   : average consumption, all consumers, all days
DeltaC       : difference (%) compared to average consumption
               in database
MeanConsDays : average consumption, consumption days only
NConsDays    : number of consumption days in the data set
NDays        : total number of days
%ConsDays    : percentage consumption days
-------------------------------------------------------------------------------
MeanConcen   : mean concentration in simulations with positive amount consumed
DeltaR       : difference (%) compared to average concentration
               in database
NSamplPos    : number of positive concentrations in simulations with positive
               amount consumed
NSamples     : total number of concentration measurements
               (detects and non-detects) in simulations with positive
               amount consumed.
%SamplPos    : percentage positive concentrations
ProcFact     : mean processing factor
-------------------------------------------------------------------------------

Food consumption data

   Code  Food        MeanConsum  DeltaC  MeanConsDays   NConsDays   NDays   %ConsDays
                          (g)     (%)           (g)                             (%)
 FB1235  TABLE-GRAPES     13.14    -0.7         33.93       38713  100000        38.7
 FC0203  GRAPEFRUIT,       4.02    -1.2         28.54       14075  100000        14.1
 FC0204  LEMON, SEE A      1.48    -4.0          4.04       36635  100000        36.6
 FC0206  MANDARIN, SE      8.99     1.9         41.86       21484  100000        21.5
 FC0208  ORANGE, SWEE     57.08     0.0         99.53       57345  100000        57.3
 FP0226  APPLE            61.56     0.1         99.53       61853  100000        61.9
 FS0247  PEACH             2.12     2.3          7.54       28146  100000        28.1
 VO0445  PEPPERS, SWE      3.39     1.7         16.62       20392  100000        20.4
```

VR0589	POTATO	38.03	-0.4	172.74	79907	100000	79.9

Compound concentration data

Code	Food	MeanConc mg/kg	DeltaR (%)	NSamplPos	NSamples	%SamplPos (%)	ProcFact
FB1235	TABLE-GRAPES	0.0154	-3.3	5229	38713	13.5	1.00
FC0203	GRAPEFRUIT,	0.0684	-1.3	6058	14075	43.0	1.00
FC0204	LEMON, SEE A	0.0127	-0.3	8147	36635	22.2	1.00
FC0206	MANDARIN, SE	0.1069	-0.3	15705	21484	73.1	1.00
FC0208	ORANGE, SWEE	0.0520	-0.4	27701	57345	48.3	1.00
FP0226	APPLE	0.0030	-1.8	3376	61853	5.5	1.00
FS0247	PEACH	0.0076	0.1	3828	28146	13.6	1.00
VO0445	PEPPERS, SWE	0.0010	-6.5	186	20392	0.9	1.00
VR0589	POTATO	0.0002	7.8	318	79907	0.4	1.00

Table 6: Summary of simulated intakes

In the left part of Figure 34, a number of icons are displayed. Click the icon 'Total' to display a histogram of the simulated total intakes (see the left plot in Figure 35). Click the icon 'Upper' to display the upper tail of the intake distribution (see the right plot of Figure 35). The specified percentage for the upper tail is 5%, corresponding with an intake of 0.44 microgram/kg bw/day. The acute reference dose (ARfD) for Chlorpyrifos is equal to 100 microgram/kg bw/day.

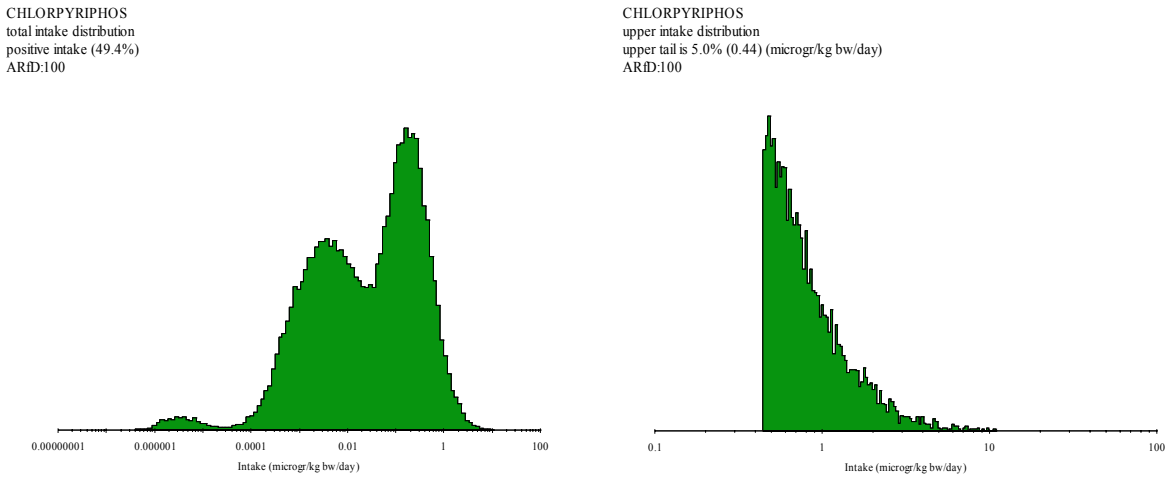


Figure 35: Total intake distribution and upper tail (5%)

The part of the intake distribution exceeding the ARfD is displayed in red (not happening in Figure 35).

Choose 'Summary of total distribution of intakes' and see Table 7. Here the contributions are expressed as percentages (RelContr) of each food to the total intake distribution. A pie chart (see Figure 36) of the foods with the 9 largest contributions is displayed by clicking the icon 'Pie' in Figure 34. Also the mean, median and the percentiles of the 2.5% and 97.5% (p2.5% and p97.5%) point of the total intake distribution per food are given in Table 7. The last column (%Zeros) shows the number of zero intakes per food. If %Zeros is greater than 97.5% the p97.5% is lower than the mean!
Let's take Orange as an example. It contributes 61.8% to the total intake distribution and its average concentration is 0.052 microgram/kg bw/day. The p2.5%, median, p97.5% and %Zeros of the total intake distribution of Orange are: 0.000, 0.000, 0.542 microgram/kg bw/day and 72.3% respectively.

```
-------------------------------------------------------------------------------------------------
Summary of the characteristics of the total intake distribution

Compound:  CHLORPYRIPHOS

-------------------------------------------------------------------------------------------------
Contribution (%) per food to the total distribution,
RelContr: relative contribution (%) per food
p2.5%   : 2.5% perc. of intake distr. per food(microgram/kg bw/day)
Mean    : mean of intake distr.per food(microgram/kg bw/day)
Median  : median of intake distr.per food(microgram/kg bw/day)
p97.5%  : 97.5% perc. of intake distr. per food(microgram/kg bw/day)
%Zeros  : percentage zeros per food
-------------------------------------------------------------------------------------------------
         Food  RelContr       Mean         Median         p2.5%         p97.5%     %Zeros
                      %   (microgr/     (microgr/     (microgr/      (microgr/          %
                          kg bw/day)    kg bw/day)    kg bw/day)     kg bw/day)
 TABLE-GRAPES       4.4        0.004         0.000         0.000          0.014       94.8
 GRAPEFRUIT,        5.3        0.004         0.000         0.000          0.008       93.9
 LEMON, SEE A       0.4        0.000         0.000         0.000          0.002       91.9
 MANDARIN, SE      22.1        0.019         0.000         0.000          0.215       84.3
 ORANGE, SWEE      61.8        0.052         0.000         0.000          0.542       72.3
 APPLE              4.8        0.004         0.000         0.000          0.004       96.6
 PEACH              0.4        0.000         0.000         0.000          0.001       96.2
 PEPPERS, SWE       0.1        0.000         0.000         0.000          0.000       99.8
 POTATO             0.7        0.001         0.000         0.000          0.000       99.7
```

Table 7: Contribution to the total intake distribution

Choose 'Summary of upper tail of intakes' and see Table 8. Find the contributions expressed as percentages (RelContr) of each food to the upper tail according to the specified percentage (here 5%). A pie chart (right plot of Figure 36) of the foods with the 9 largest contributions to the upper 5% of the intake distribution is displayed by clicking the icon 'Pie' in Figure 34. Also the mean, median and the percentiles of the 2.5% and 97.5% (p2.5% and p97.5%) point of the upper 1% of the intake distribution per food are given in Table 8. The table shows the same statistics for that part of the intakes per food that correspond with the upper 5% of the intake distribution. See also Figure 37, where the mean intake for each food in the upper tail is graphically displayed. The last column (%Zeros) shows the number of zero intakes per food.
Let's take Orange as an example. It contributes 68.5% to the upper 5% of the intakes and its average concentration is 0.670 microgram/kg bw/day. The p2.5%, median, p97.5% and %Zeros of the upper 5% of the intakes are: 0.000, 0.542, 2.624 microgram/kg bw/day and 22.3% respectively.

```
-------------------------------------------------------------------------------------------------
Summary of characteristics of the upper intake distribution

Compound:  CHLORPYRIPHOS

-------------------------------------------------------------------------------------------------
Characteristics per food of the upper 5.0% of the distribution,
corresponding with a total intake higher than 0.4438 (microgram/kg bw/day)
RelContr: relative contribution (%) per food
p2.5%   : 2.5% perc. of intake distr. per food(microgram/kg bw/day)
Mean    : mean of intake distr.per food(microgram/kg bw/day)
Median  : median of intake distr.per food(microgram/kg bw/day)
p97.5%  : 97.5% perc. of intake distr. per food(microgram/kg bw/day)
%Zeros  : percentage zeros per food
-------------------------------------------------------------------------------------------------
         Food  RelContr       Mean         Median         p2.5%         p97.5%     %Zeros
                      %   (microgr/     (microgr/     (microgr/      (microgr/          %
                          kg bw/day)    kg bw/day)    kg bw/day)     kg bw/day)
 TABLE-GRAPES       3.8        0.037         0.000         0.000          0.537       90.0
 GRAPEFRUIT,        5.8        0.057         0.000         0.000          0.723       86.7
 LEMON, SEE A       0.1        0.001         0.000         0.000          0.004       90.0
 MANDARIN, SE      16.9        0.165         0.000         0.000          1.174       65.8
 ORANGE, SWEE      68.5        0.670         0.542         0.000          2.624       22.3
 APPLE              4.5        0.044         0.000         0.000          0.553       92.5
 PEACH              0.1        0.001         0.000         0.000          0.001       95.6
```

PEPPERS, SWE	0.0	0.000	0.000	0.000	0.000	99.7
POTATO	0.3	0.003	0.000	0.000	0.000	99.3

Table 8: Contribution to the upper tail (1%) of the intake distribution

In Figure 36, find a graphical display of the figures in Table 7 and Table 8.

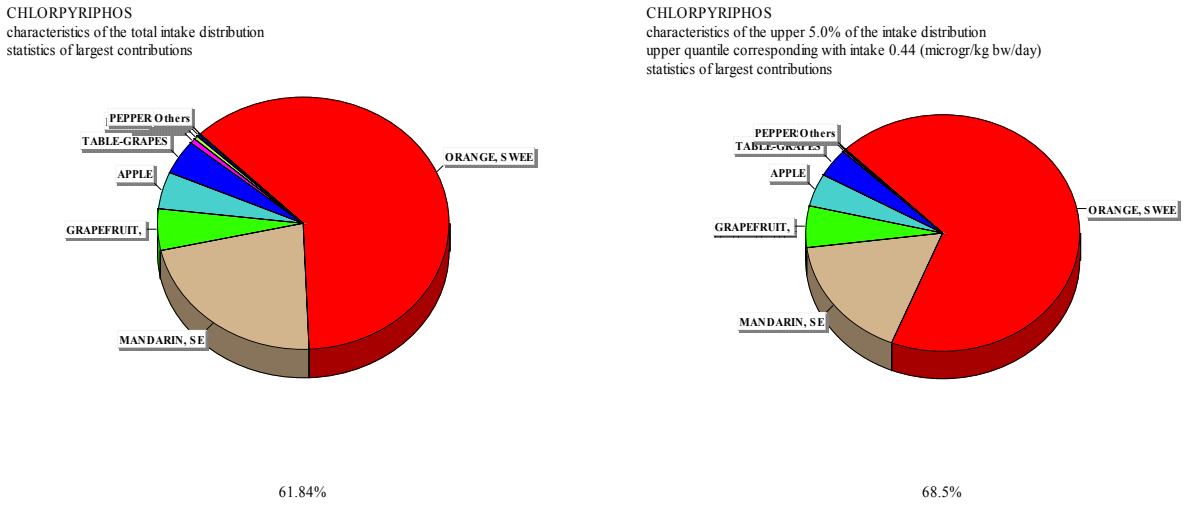


Figure 36: Relative contribution of foods to the total intake distribution and upper tail (5%)

Scroll the mouse over the pie chart to find the contribution of each food. Here, the relative contribution for Orange is displayed.

Choose 'Percentiles of daily intakes' and see Table 9. The percentiles of the acute intake distribution for the percentages specified in the first line of the Output options in the MCRA input form (see Figure 33) are shown. In this example, 0.01% of the population has an intake higher than 7.35849 microgram/kg bw/day. Also the mean and maximum are given.

```
--------------------------------------------------------------------------------
Random sampling is based on seed :         0
Number of simulations (consumers):    100000  out of    6250
CHLORPYRIPHOS (microgr/kg bw/day)  consumption:     49412 out of 100000

Compound:  CHLORPYRIPHOS

--------------------------------------------------------------------------------
     Percentiles, maximum and average intake
--------------------------------------------------------------------------------

   Percentage     Percentiles of CHLORPYRIPHOS  (microgr/kg bw/day)

        50.00         0.00000
        90.00         0.23915
        95.00         0.44365
        99.00         1.21429
        99.90         3.41250
        99.99         7.35849
         mean         0.08382
      maximum        10.77044
```

Table 9: Percentiles for the acute intake distribution

In the left plot of Figure 37 find a graphical display of the figures in Table 9. The right plot of Figure 37 shows the mean intake per food in the upper tail (the third column of Table 8).

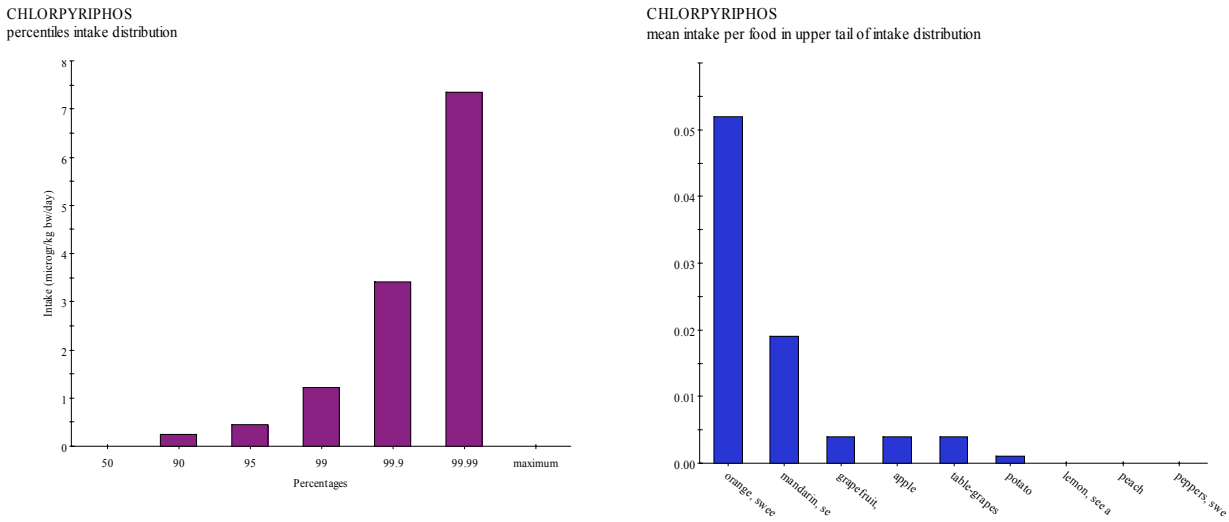


Figure 37: Percentiles and mean intake per food of the acute intake distribution

Choose ‘Consumptions: characteristics of nine consumers’, ‘Concentrations: characteristics of nine consumers’ or ‘Intakes: characteristics of nine consumers’. Summaries are displayed in Table 10, Table 11 and Table 12. A drill down of nine consumers according to a specified percentage (here 97.5%) is displayed: the sampled consumption, compound concentration and intake (as well as the consumer number, weight, age and Total for the intake in Table 12). So for consumer number 323461 (52 kg, 61 years), the total intake is 0.715 microgram/kg bw/day. This consumer consumed 2 foods (Potatoes 100.0 gr, Orange 120.0 gr). Only on Orange a compound was found (0.310 mg/kg), so the total intake for this consumer is 120.0 x 0.310/52 = 0.715 microgram/kg bw/day.

Drilldown consumption

Compound: CHLORPYRIPHOS

Drill down: consumption (g/day) per food of the 9 consumers
around the specified percentage (97.50%) for the intake distribution

Food	-4	-3	-2	-1	0	+1	+2	+3	+4
TABLE-GRAPES	.	0.5	.	8.4	1.0	.	.	.	.
GRAPEFRUIT,	.	.	75.2	0.8	.	.	.	.	.
LEMON, SEE A	.	.	.	0.6	3.0	.	.	.	.
MANDARIN, SE	110.0	83.0	.	29.0	1.5	.	43.6	290.0	55.0
ORANGE, SWEE	3.3	60.3	120.3	206.8	321.1	120.0	260.0	.	.
APPLE	303.3	0.4	434.0	1.7	125.4	.	.	.	.
PEACH	3.3	0.2	.	0.8	.	.	.	.	4.0
PEPPERS, SWE	.	.	.	2.8	.	.	.	.	.
POTATO	207.1	7.7	90.0	2.6	251.5	100.0	210.0	20.0	.

Table 10: Drill down of consumers for the acute intake distribution: consumptions

Drilldown concentrations

Compound: CHLORPYRIPHOS

Drill down: compound concentrations (mg/kg) per food of the 9 consumers
around the specified percentage (97.50%) for the intake distribution

	-4	-3	-2	-1	0	+1	+2	+3	+4
Food									
TABLE-GRAPES	.	0.040	.	0.000	0.000	.	.	.	.
GRAPEFRUIT,	.	.	0.300	0.000	.	.	.	.	.
LEMON, SEE A	.	.	0.000	0.000	0.000	.	.	.	.
MANDARIN, SE	0.320	0.130	.	0.100	0.090	.	0.200	0.180	0.260
ORANGE, SWEE	0.200	0.070	0.210	0.100	0.160	0.310	0.170	.	.
APPLE	0.040	0.000	0.030	0.000	0.000	.	.	.	.
PEACH	0.000	0.030	.	0.040	.	.	.	.	0.000
PEPPERS, SWE	.	.	.	0.000	.	.	.	.	.
POTATO	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.

Table 11: Drill down of consumers for the acute intake distribution:concentrations

Drilldown intake

Compound: CHLORPYRIPHOS

Drill down: the 9 consumers around the specified percentage (97.50%) for the intake distribution (microgr/kg bw/day)

	-4	-3	-2	-1	0	+1	+2	+3	+4
Individual	95243	361383	705152	263774	375142	323461	709242	328622	322893
Body Weight	67	21	85	33	72	52	74	73	20
age	21	7	79	7	23	61	41	38	6
Total	0.716	0.716	0.716	0.716	0.715	0.715	0.715	0.715	0.715
Food									
TABLE-GRAPES	.	0.001	.	0.000	0.000	.	.	.	.
GRAPEFRUIT,	.	.	0.265	0.000	.	.	.	.	.
LEMON, SEE A	.	.	0.000	0.000	0.000	.	.	.	.
MANDARIN, SE	0.525	0.514	.	0.088	0.002	.	0.118	0.715	0.715
ORANGE, SWEE	0.010	0.201	0.297	0.627	0.714	0.715	0.597	.	.
APPLE	0.181	0.000	0.153	0.000	0.000	.	.	.	.
PEACH	0.000	0.000	.	0.001	.	.	.	.	0.000
PEPPERS, SWE	.	.	.	0.000	.	.	.	.	.
POTATO	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.

Table 12: Drill down of consumers for the acute intake distribution: intake

4.2 Chronic risk assessment: betabinomial/normal (BBN)

In this example, output is shown for a chronic risk assessment for aspartaam. This database is not centrally supplied. This example is also used in 10.8 10.9 and 10.10 .
Find in Figure 38 the MCRA input form:

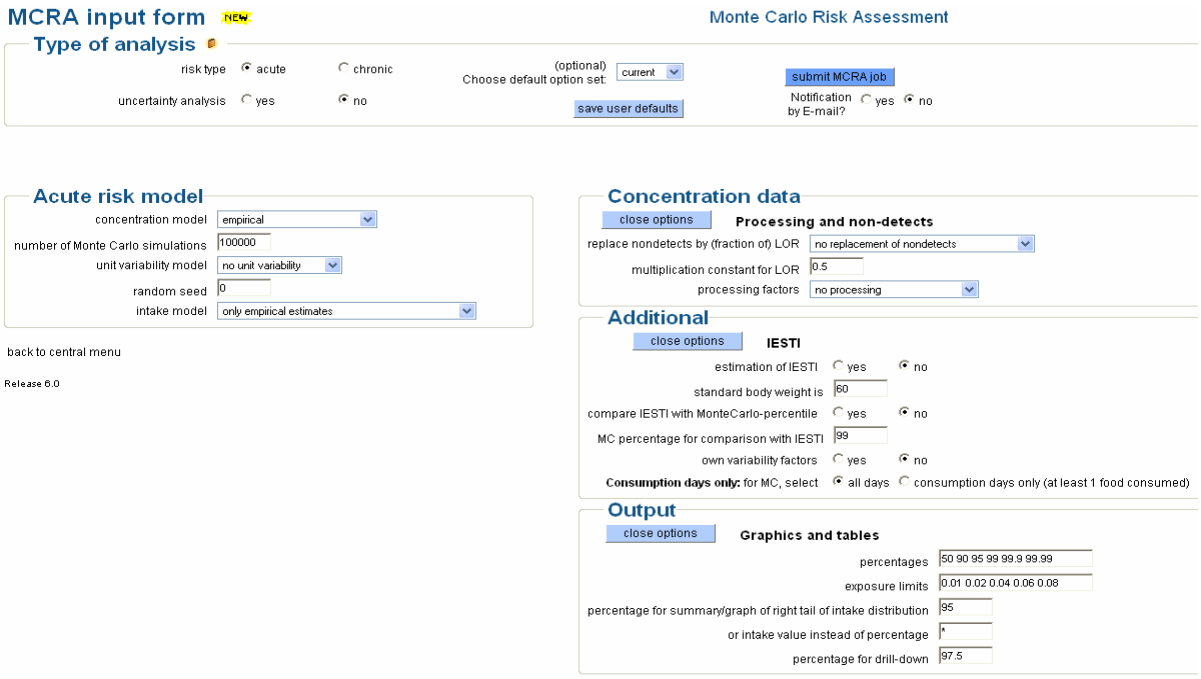


Figure 38: MCRA input form for chronic risk analysis betabinomial/normal

Table 13 lists the main options:

Input form	
risk type	chronic
uncertainty analysis	no
concentration model	empirical
number of MC simulations	100000
intake model	betabinomial/normal without covariable and cofactor
concentration data	system defaults
Output	Percentages at percentiles: 40 500 2000 40000

Table 13: Input form options: betabinomial/normal without covariable and cofactor

Click the ‘submit MCRA job’ to start the analysis. View your results using the ‘view output’-button in the central menu to get the screen of Figure 39:

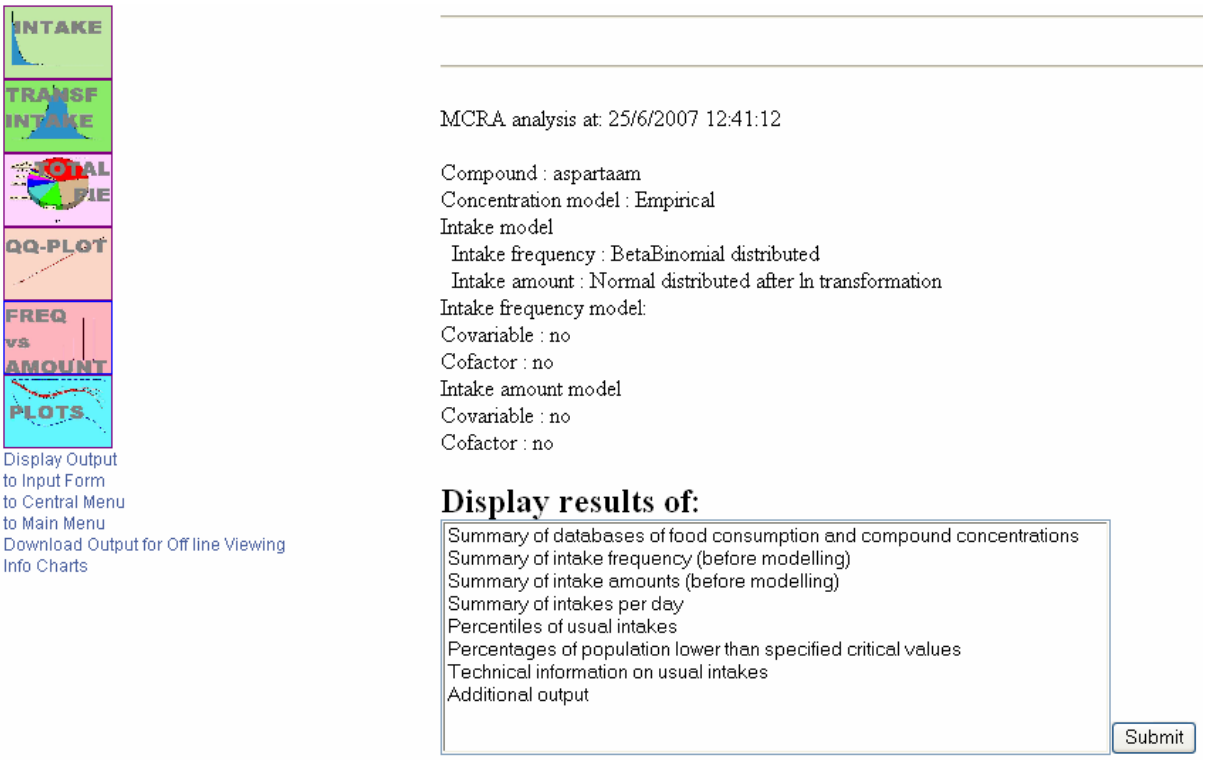


Figure 39: View output window for chronic risk analysis betabinomial/normal

Click the icon 'Intake' to display the empirical distribution of the non-zero daily intake values, calculated as daily intake values divided by body weight (see left plot Figure 40). Note the spike near 0 (zero intakes are already excluded here). The distribution is skewed. A logarithmic transformation is chosen to approximate normality.
Click the icon 'Transf Intake' to display the logarithmic transformed intake distribution (see right plot Figure 40). There is a better symmetry, but clearly the non-normal character of the data is not removed by a simple logarithmic transformation. Compare this figure also with the power transformed distribution in the right plot of Figure 78.

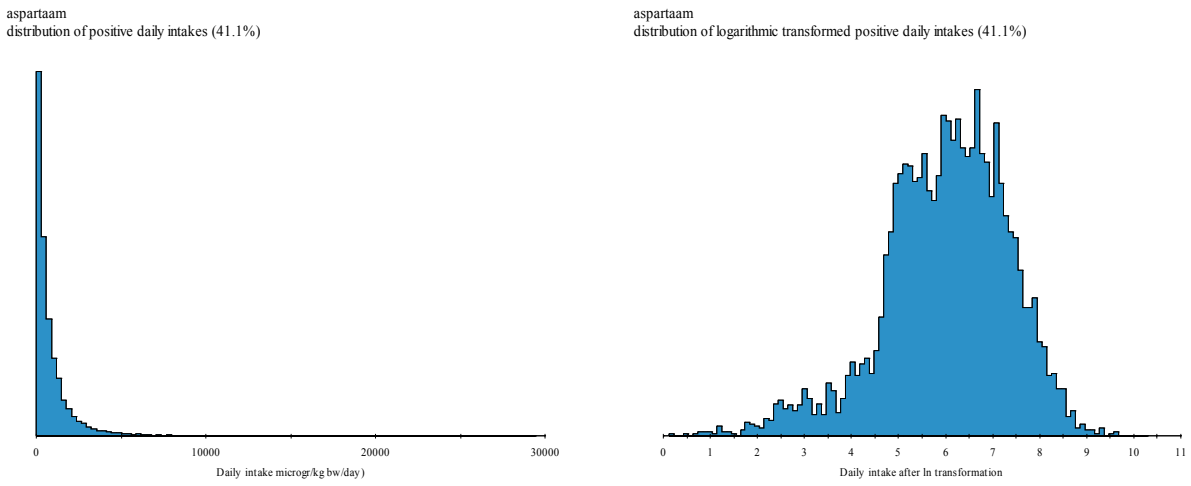


Figure 40: Untransformed and ln transformed intake distribution

The Dutch National Food Consumption Survey is used, which has 6250 consumers x 2 days is 12500 days in total.

Choose ‘Summary of databases of food consumption and compound concentrations’ to get Table 14. Aspartaam measurements were made in 113 samples. All of them had a positive value.

```
------------------------------------------------------------------------------
Summary of databases of consumptions and compound concentrations with respect to:

Compound:  aspartaam

------------------------------------------------------------------------------
Compound:  aspartaam
Code         : food code
Food         : food label
------------------------------------------------------------------------------
MeanConsum   : average consumption, all consumers, all days
MeanConsDays : average consumption, consumption days only
NConsDays    : number of consumption days in the data set
NDays        : total number of days
%ConsDays    : percentage consumption days
------------------------------------------------------------------------------
MeanConcen   : mean concentration of all samples
MeanPosConc  : mean concentration of samples with positive concentrations
NSamplPos    : number of samples with positive concentrations
NSamples     : total number of samples with concentration measurements
%SamplPos    : percentage samples with positive concentrations
------------------------------------------------------------------------------
Food consumption data
        Code Food         MeanConsum MeanConsDays NConsDays    NDays   %ConsDays
                                (g)          (g)                          (%)
        1002 IJs consumpt       0.02        110.0         2    12500         0.0
        1135 Siroop limon       0.09         19.3        61    12500         0.5
        1463 Dubbeldrank        4.13        270.2       191    12500         1.5
        1521 Limonadevruc       4.07        371.7       137    12500         1.1
        1522 Frisdranklig      10.79        385.4       350    12500         2.8
        1523 Frisdranklig      22.56        470.7       599    12500         4.8
        1591 Zoetstoftabl       0.20          7.1       357    12500         2.9
        1596 Zoetstofpoed       0.10          5.2       227    12500         1.8
        1813 Yakult drank       0.23         66.5        43    12500         0.3
        1833 Yoghurt mage       0.35        196.3        22    12500         0.2
        1834 Yoghurtdrank       0.34        305.6        14    12500         0.1
         305 Kwark magere       1.18        109.5       135    12500         1.1
         400 Frisdrank         38.96        362.9      1342    12500        10.7
         417 Limonade vru      19.65        309.4       794    12500         6.4
         447 Kauwgom z su       0.06          3.1       253    12500         2.0
         463 Siroop vruch       4.99         56.4      1107    12500         8.9
         657 Yoghurtdrank      24.46        294.6      1038    12500         8.3

Compound concentration data
        Code Food         MeanConcen  MeanPosConc NSamplPos NSamples   %SamplPos
                             (mg/kg)      (mg/kg)                          (%)
        1002 IJs consumpt   238.4000     238.4000         5        5       100.0
        1135 Siroop limon    42.0000      42.0000         2        2       100.0
        1463 Dubbeldrank     68.7500      68.7500         4        4       100.0
        1521 Limonadevruc   139.9286     139.9286        14       14       100.0
        1522 Frisdranklig   265.5000     265.5000         4        4       100.0
        1523 Frisdranklig   214.2222     214.2222        18       18       100.0
        1591 Zoetstoftabl 12033.3333   12033.3333         6        6       100.0
        1596 Zoetstofpoed 30333.3333   30333.3333         6        6       100.0
        1813 Yakult drank   164.0000     164.0000         2        2       100.0
        1833 Yoghurt mage    91.7500      91.7500         8        8       100.0
        1834 Yoghurtdrank    92.0000      92.0000         6        6       100.0
         305 Kwark magere    96.0000      96.0000         2        2       100.0
         400 Frisdrank       51.8000      51.8000         5        5       100.0
         417 Limonade vru    65.3571      65.3571        14       14       100.0
         447 Kauwgom z su   704.2500     704.2500         4        4       100.0
         463 Siroop vruch    68.3333      68.3333         6        6       100.0
         657 Yoghurtdrank    69.1429      69.1429         7        7       100.0
```

```
                                                    ----     ----
                              number of detects =   113      113 = total
                                                             number of samples

                          number of non-detects =     0   (  113  -   113 )
```

Table 14: Summary of the database, consumptions and compounds

On average, everyone consumes *e.g.* 39 g of Frisdrank per day. This is an average consumption of 362.9 g on 1342 consumption days and 0 g on 12500 - 1342 = 11158 non-consumption days. The average concentration of the 5 positive aspartaam samples of Frisdrank is 51.8 mg/kg.
Choose 'Summary of intake frequency (before modelling)' and see Table 15.

```
-----------------------------------------------------------------------------
Summary of intake frequencies (before modeling) with respect to:

Compound:  aspartaam

-----------------------------------------------------------------------------
Number and percentage of consumers with positive intakes

                nPosDays    %PosDays  nConsumers  %Consumers
                       0         0.0        3034        48.5
                       1        50.0        1296        20.7
                       2       100.0        1920        30.7

                                 All    Positive   %Positive
                             intakes     intakes     intakes
  Number of observations       12500        5136        41.1
     Number of consumers        6250        3216        51.5

For a graphical presentation of the distribution of the positive intakes:
click the first icon from the top

For a graphical presentation of the distribution of the transformed
positive intakes:
click the second icon from the top
```

Table 15: Summary of intake frequency

Choose 'Summary of intake amounts (before modelling)' and see Table 16.

```
-----------------------------------------------------------------------------
Summary of daily intakes (before modeling) with respect to:

Compound:  aspartaam

-----------------------------------------------------------------------------
Summary of all intake data (zeros included)

     Number of observations =      12500
                       Mean =    344.204
                     Median =      0.000
                    Minimum =      0.000
                    Maximum =  29446.326
             Lower quartile =      0.000
             Upper quartile =    315.822

Summary of positive intake data (without zeros)
```

```
     Number of observations =       5136
                       Mean =    837.723
                     Median =    454.944
                    Minimum =      1.120
                    Maximum =  29446.326
              Lower quartile =    183.425
              Upper quartile =   1055.158

For a graphical presentation of the distribution of the positive intakes:
click the first icon from the top

For a graphical presentation of the distribution of the transformed
positive intakes:
click the second icon from the top
```

Table 16: Summary statistics of the calculated daily intake amounts

In Table 17 you find the main characteristics concerning this analysis. The intake frequency function is estimated with a betabinomial model. The intake probability is 0.411 and the overdispersion parameter is 0.572. The model for the logarithmically transformed intake amounts is based on ML. The estimate of the 'between person' variance is 1.25 and of the 'within person' variance 0.54.

```
'Additional output' file
Number of foods                       1
Acute reference dose (ARfD)           *
Acceptable daily intake (ADI)         5
Number of detects                     113
Number of non-detects                 0
No of consumers                       6250
Population characteristics,
      minimum age                     1
      maximum age                     97
      minimum weight                  8
      maximum weight                  150
      Sex                             female, male
Total no of consumption days          5136
Replace all non-detects
Multiplication factor for LOR         0.5


Intake Frequency: BetaBinomial model (no cofactor and no covariable)
================================================================================

Estimates of parameters (autoscaled covariable)
-----------------------------------------------

                     Parameter       estimate      s.e.      t(*)
                      Constant      -0.360296  0.022751    -15.84


            Overdispersion phi         0.5717   0.01053     54.30
               _2Loglikelihood       12995.52
            Degrees of freedom           6248



Transformed Intake Amounts: ML model (no cofactor and no covariable)
================================================================================
```

```
Estimates of parameters (autoscaled covariable)
-----------------------------------------------
                     Parameter       estimate      s.e.      t(*)
                      Constant       5.929186  0.022446    264.15

       Between person variance         1.2526
        Within person variance         0.5390
                _2LogLikelihood       16281.80
            Degrees of freedom           5134


***** Usual intake percentiles/percentages are calculated from:
---------------------------------------------------------------
            sex       age   Intake probability   Transformed intake amount
              -         -               0.4109                        5.9292
```

Table 17: Information in 'Additional output' file

Click the icon 'Plots' in Figure 39 to display the screen of Figure 41. Here 3 more plots can be chosen: the density of the Beta distribution of the Frequency model, the cumulative distribution of the Amount model and the cumulative usual intake, see Figure 42.

Display plots of:

Frequency model
Density of beta distribution

Amount model
Cumulative amount distribution

Usual intake distribution
Cumulative usual intake distribution

Figure 41: Display of more plots

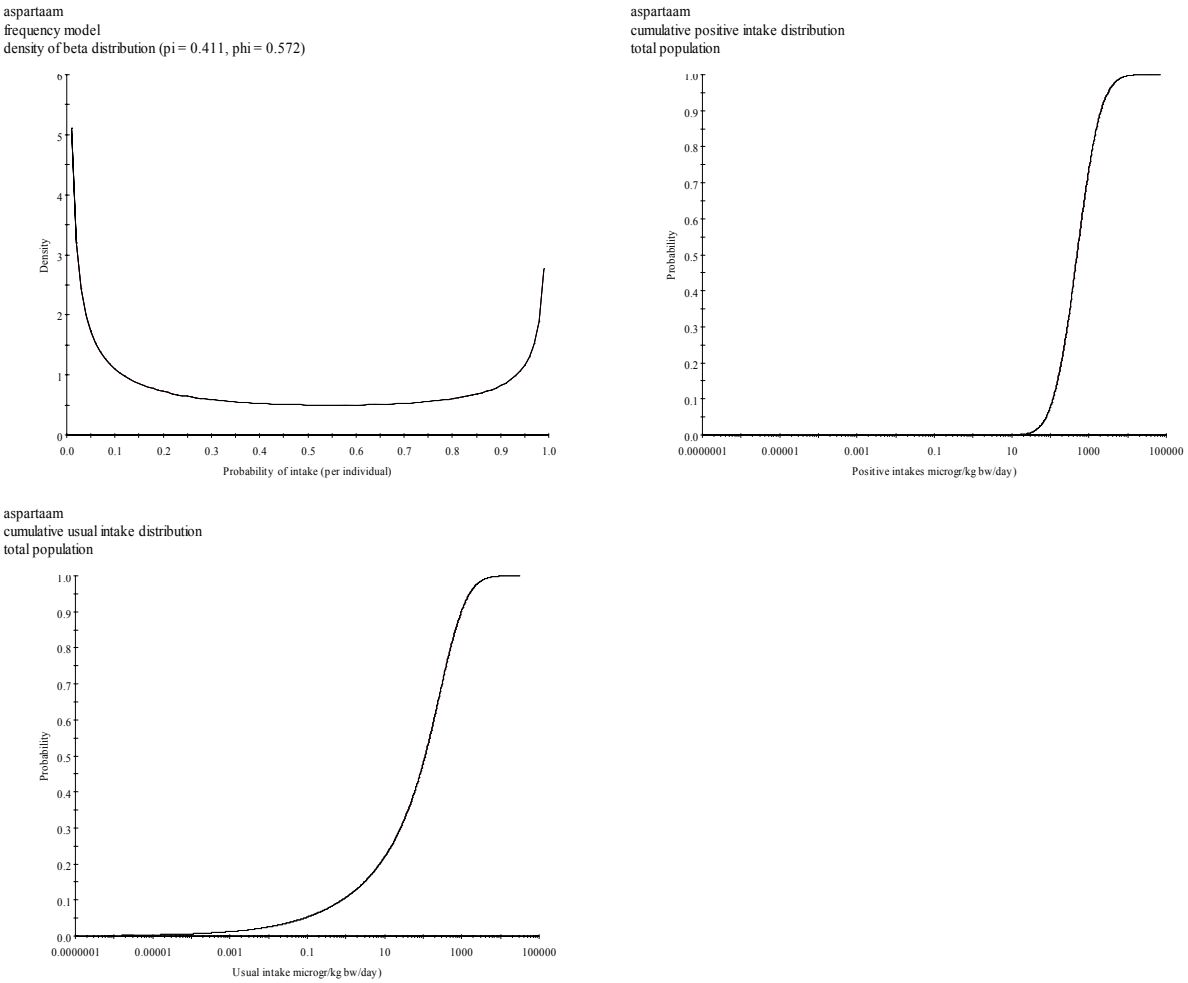


Figure 42: Density of the intake frequency distribution, cumulative distribution of the amount and usual intake

Find in Figure 43 a Normal QQ-plot to see the result of the transformation to normality and a plot which reveals any correlation between the frequency of consumption and the amount consumed. The left Box-Cox diagram shows the data of consumers having an intake on 1 out of 2 days, the second diagram shows consumers having intakes on 2 out of 2 days.

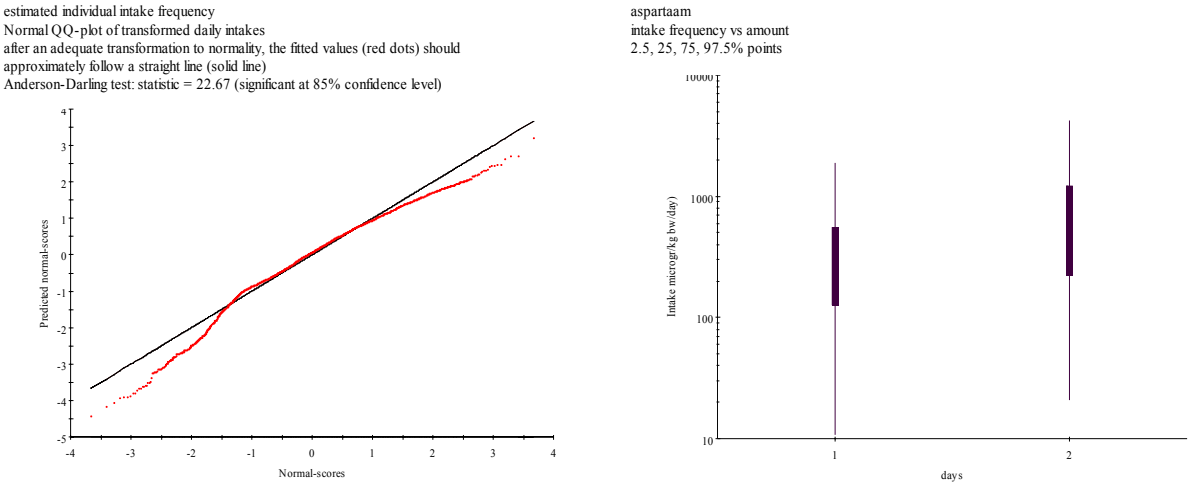


Figure 43: Normal QQ-plot and Box-Cox plots for intake frequency versus intake amounts

Choose 'Percentiles of usual intakes' to get percentiles of the usual intake distribution for the entire population are shown (see Table 18).

```
--------------------------------------------------------------------------------
Long term exposure (usual intake)
Based on the BetaBinomial frequency distribution and transformed (to
normality) distributed amounts

Compound:  aspartaam

--------------------------------------------------------------------------------


Usual intake percentiles of total population for percentiles (p_):

 sex       age       p50       p90       p95       p99      p99.9     p99.99
   -         -    112.30    958.86   1579.97   3795.26  10007.78  24961.30

See also technical information for more details
```

Table 18: Percentiles of the usual intake distribution

Choose 'Percentages of consumers lower than specified limits of usual intake' to get Table 19.

```
--------------------------------------------------------------------------------
Long term exposure (usual intake)
Based on the BetaBinomial frequency distribution and transformed (to
normality) distributed amounts

Compound:  aspartaam

--------------------------------------------------------------------------------


Usual intake percentages of total population lower than intake limit (q_):

           sex       age      q40.0     q500.0     q2000.0     q40000.0
             -         -      34.68      79.83      496.60        99.99
```

```
See also technical information for more details
```

Table 19: Percentages of Consumers lower than specified Limits of Usual Intake

4.3 Help on Charts

To view ComponentOne Charts, an ActiveX-aware browser is needed. Users can manipulate the chart they view by right-clicking the chart, which brings up the Property Editor (see Figure 44). Some more information is available under the link InfoCharts on the View Output screen. ComponentOne Chart graphs can be printed, or copied to the clipboard, for later inclusion in documents. Occasionally, after pasting ComponentOne Charts into Word the chart is not displayed (at all) and instead, an icon appears. To our experience, pasting charts from the clipboard encounters no difficulties when the Word document is opened first, then press the 'Copy to Clipboard'-button and paste the contents of the clipboard into the Word document.
Tabular output is available in separate ASCII output text files in the output directory. Alternatively, text can be copied and pasted from the 'view output' window into another document. In order to obtain a proper lay-out the function Paste Special from the Edit menu should be used, selecting "Unformatted text".
ComponentOne Chart is comprised of a 2D Control (ActiveX) for use in Windows applications. The control is stored in a so-called Cabinet-file, Olectra.CAB. A licence pack file Olectra.LPK is needed to register the control. To be able to view a ComponentOne Chart, the cabinet file and license pack have to be downloaded. Depending on the security level of your Internet Explorer, you may get the chart. You can change the security settings by doing the following:

- In Internet Explorer select {Tools | Internet Options} and choose the "Security" tab.
- Then select the Web zone "Internet" or "Local Internet" depending on whether you view the MCRA charts on the internet or local disk. Click the "Custom Level" button, and use the following settings (you may not see all of these settings):
    1. Automatic prompting for ActiveX Controls: Enable
    2. Binary and script behaviors: Enable
    3. Download signed ActiveX controls: Prompt
    4. Download unsigned ActiveX controls: Prompt
    5. Initialize and script ActiveX controls not marked as safe: Disable
    6. Run ActiveX controls and plug-ins: Enable
    7. Script ActiveX controls marked safe for scripting: Enable

To be sure those changes to the charts (after re-running the program) are displayed by the browser, you may need to do the following. Click:

1. Tools.
2. Internet Options.
3. General.
4. Settings.
5. Check: every visit to the page.
6. OK
7. OK

The setting for "Download unsigned ActiveX controls" is probably the most critical one. The first time you display an MCRA chart, you will get the Security Warning displayed here about installing OLECTRA.CAB. Click "Install" and you are done.

On some computers another version of the "Chart" ActiveX component might be installed, giving spurious error messages and/or no graphs when viewing this website. You then have to delete the file "olch2x7.ocx" from the "WINDOWS\system32" directory, and restart the website. This will copy the correct version of "olch2x7.ocx" to the "WINDOWS\system32" directory.

In Internet Explorer 6, after selecting an icon to display a chart in the View Output Menu, you may see an "active contents" security warning in the Internet Explorer Information Bar on top of your screen. Click on the warning, select Allow Blocked Contents and choose YES in the Security Warning displayed on the right.

4.3.1 To display the Property Editor

Click the right mouse button over any part of the chart and select properties of the pop-up menu. Select the tab that corresponds to the element of the chart that you want to edit. Click the 'OK' or 'Cancel'-button to close the Property Editor (see Figure 44).

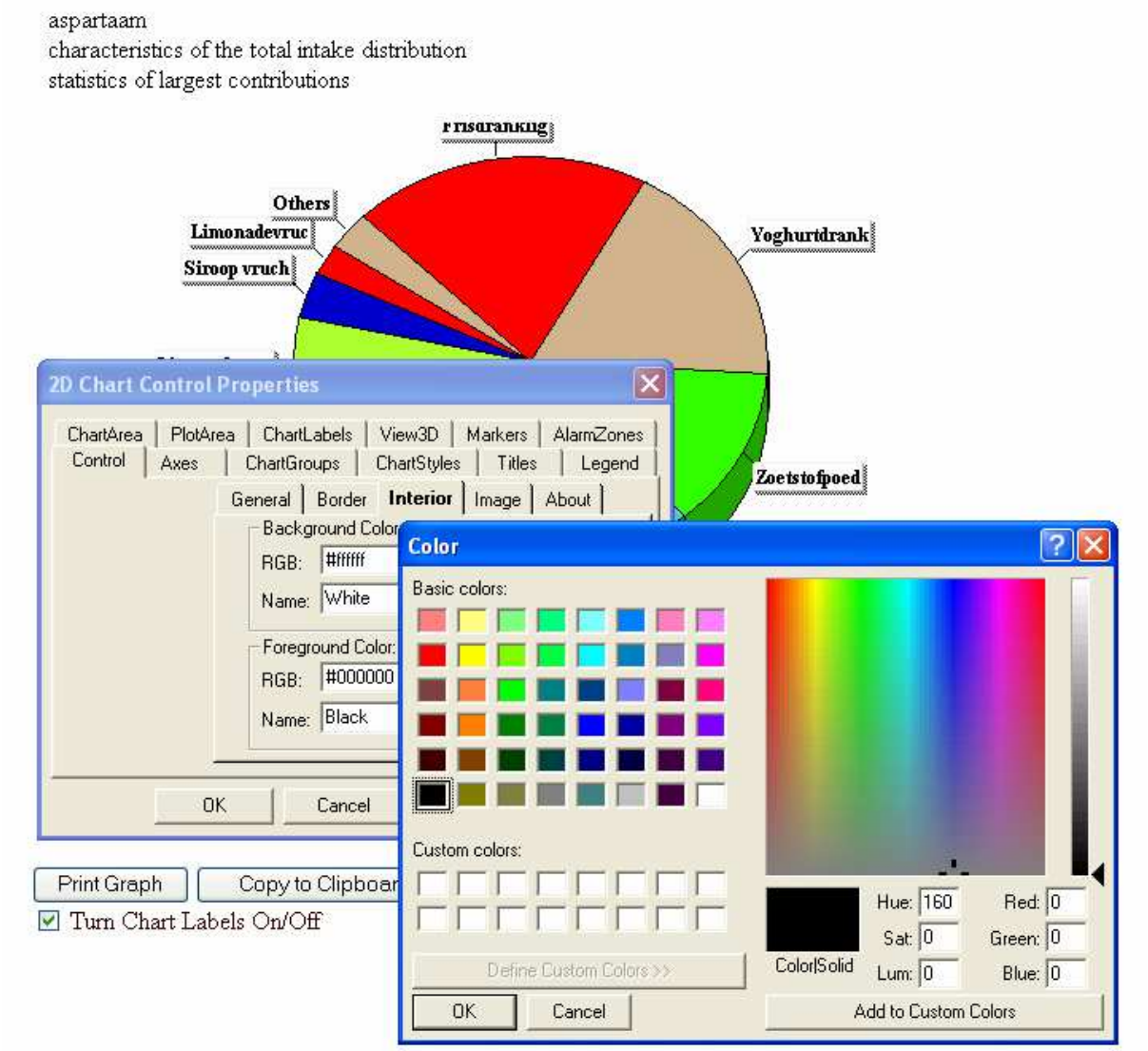


Figure 44: Property editor

Interacting with Chart

You can interact with the chart as it is running to examine data more closely or visually isolate a part of the chart. The interactions described here affect the chart displayed inside the ChartArea; other chart elements like the header are not affected. ComponentOne Chart provides users with 2 different mechanisms for zooming the chart: Graphics zoom and Axis zoom. Performing a Graphics zoom enlarges the selected area of a chart, while not necessarily showing the axes. Performing an Axis zoom changes the minimum and maximum data values to those selected, and redraws only that data

with axes. Scaling, moving, or graphics zooming the chart sets the PlotArea margin properties, so the chart will not automatically control margins anymore when other chart properties change.

To Scale the Chart:

1. Press CTRL, and hold down both mouse buttons (or middle button on 3-button mouse).
2. Move the mouse down to increase chart size, or move the mouse up to decrease chart size.

To Move the Chart:

1. Press SHIFT, and hold down both mouse buttons (or the middle button on 3-button mouse).
2. Move the mouse to change the positioning of the chart inside the ChartArea.

To Graphics Zoom an Area of the Chart:

1. Press CTRL, and hold down left mouse button.
2. Drag mouse to select zoom area and release the mouse button.

To Axis Zoom the Chart:

1. Press SHIFT, and hold down left mouse button.
2. Drag the mouse to select the zoom area and release the mouse button.

To Rotate the Chart (Bar/pie charts displaying 3D effect only):

1. Hold down both mouse buttons (or middle button on 3-button mouse).
2. Move mouse up or down to change the 3D inclination.
3. On bar charts, you can also move mouse left or right to change the 3D rotation angle.

To Reset to Automatic Scale and Position:

Press the “r” key to remove all scaling, moving, and zooming effects; chart regains control of PlotArea margins.

5 Model description

5.1 Basic model

Food consumption data may arise from different sources. Typically, national food consumption surveys or monitoring programs provide information on food intake in the general population. For example, from the Dutch Food Consumption Survey (1997) food consumption patterns (x_1 ,...,x_p), body weight (w) and age (a) are available for 6250 consumers on 2 consecutive days.

When concentrations are not measured on consumed foods, a composition database is necessary to convert the amounts of food as consumed (*e.g.* pizza) to amounts of foods (x_1 ,...,x_p) which are used in the model. Van Dooren *et al*. (1995) provide such a conversion for the Dutch situation.

Compound concentration data may be available from different sources. In some countries national monitoring databases exist, which are useful for the risk assessment of chemical compounds already in use. For example the Dutch KAP database (van Klaveren 1999) stores annually more than 200,000 records of measurements originating from food monitoring programs for meat, fish, dairy products, vegetables and fruit.

This chapter describes the stochastic (or Monte Carlo) models behind the MCRA program. These models assess **acute** (short-term) or **chronic** (long-term) risks due to the intake of chemical compounds from food by combining food consumption survey data and compound concentration data from *e.g.* monitoring programs.
The model for acute risk, as opposed to the model for chronic risk, allows for effects of food processing between monitoring and ingestion, it can model unit variability either from available data or using default assumptions, and it uses information on Limit of Reporting (LOR) and percent crop treated to check whether non-detects present a source of uncertainty.
The basic model for the intake of a special compound in an acute risk analysis is:

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

where y_{ij} is the intake by consumer i on day j (in microgram chemical compound per kg body weight), x_{ijk} is the consumption by consumer i on day j of food k (in g), c_{ijk} is the concentration of that chemical compound in food k eaten by consumer i on day j (in mg/kg, 'ppm'), and w_i is the body weight of consumer i (in kg). Finally, p is the number of foods accounted for in the model. **Note** that the definition of '**food**' is flexible: it may represent a Raw Agricultural Commodity (RAC), *e.g.* 'apple', but it may also specify subdivisions, *e.g.* 'apple, peeled' or 'apple, imported'.

In the **stochastic** model for the intake in an acute risk the quantities x_{ijk}, w_i and c_{ijk} are assumed to arise from probability distributions describing the variability for food consumption and weight, $p(x_1,...,x_p,w)$, and for compound concentrations in each food, $p_k(c)$. In principle, these probability distributions may be parametric (*e.g.* completely defined by the specification of some parameter values) or empirical (*e.g.* only implicitly defined by the availability of a representative sample). Given these probability distributions (or estimates thereof) MC-simulations can be used to generate an estimate of the probability distribution $p(y_{ij})$ to assess acute risks by intake of the compound (see 5.3).

The basic model for the intake in a chronic risk analysis is:

$$y_{ij} = \frac{\sum_{k=1}^{p} x_{ijk}\ c_k}{w_i}$$

where y_{ij} is the intake by consumer i on day j (in microgram chemical compound per kg body weight), x_{ijk} is the consumption by consumer i on day j of food k (in g), c_k is the average concentration of the chemical compound in food k (in mg/kg, 'ppm'), and w_i is the body weight of consumer i (in kg). Finally, p is the number of foods accounted for in the model.

Depending on the problem, MC-samples may be drawn from the complete data base, from a day- or age-restricted subset or from consumption-days only. In some cases there is insufficient information for specific subgroups in the population. For example, in a study on infants (age up to 12 months), a separately constructed food consumption database has been used (Boon *et al.* 2003).

When dietary components are consumed on a nearly daily basis, intake values y_{ij} may be used to estimate the probability distribution $p(y_{i.})$ for chronic risk assessment purposes (see 5.4).

5.2 How to deal with limited information on compound concentration data

In the probabilistic model, a distribution of food consumption data as well as a distribution of compound data are used. For both components of the model, a choice can be made between a non-parametric (see 5.2.1) or a parametric (see 5.3.1.1.2 and 5.3.1.1.3) approach. In a *parametric approach* the data are modelled with an appropriate distributional form (*e.g.* lognormal with parameters σ and μ). In a *non-parametric approach* the empirical distribution is used to sample from directly. Obviously the latter approach requires more data to obtain a satisfying representation of the full distribution. Therefore, parametric modelling becomes important in data-scarce situations (see 5.2.1).

Occasionally limited information emerge not as a consequence of the amount of data but how they are presented: data are reported using *e.g.* the mean and variance (see 5.2.2) or data are summarised as counts of observations falling into a series of classes (see 5.2.3). It is evident that a parametric approach is the only way out and that the parameters of the lognormal distribution should be inferred using the available data.

If for some foods there are far less concentration data than for others, it may be sensible to consider *pooling* procedures for means and or variances of the concentration distributions (see 5.3.1.1.3).
If the amount of data is limited, this may lead to a relatively large *sampling uncertainty*. Resampling methods may be used to assess the magnitude of this uncertainty (see 5.5).
Back to: Concentration model options

5.2.1 The choice between a parametric and non-parametric approach

How many compound concentration data are required for a sensible calculation of upper-tail percentiles in the intake distribution based on a non-parametric approach? The rule of thumb can be used that the chosen percentile should be contained directly in the data. For example, at least 20 measurements are needed to estimate the 95th percentile and at least 100 measurements to estimate the 99th percentile.

More generally, the number of measurements per food (n) should at least equal 1/(1-p%/100) to allow a rough empirical estimate of the pth percentile of the compound concentration distribution to be made. Of course, the risk assessment is only coarse with this minimum amount of data and more measurements per food are certainly worthwhile.

In situations where the number of measurements becomes a problem, an appropriate risk analysis should be based on further modelling. Essentially, the lack of data is compensated by *a priori* assumptions. Assuming a simple distributional form for the compound data, the number of measurements can be smaller in principle (at least 10, say). However, non-detect measurements provide no information about variability, and therefore we should now count the number of positive measurements. Figure 45 shows which approach could be best used depending on the total number of

measurements and the number of non-zero measurements. In principle, such a choice could be made separately for each food.

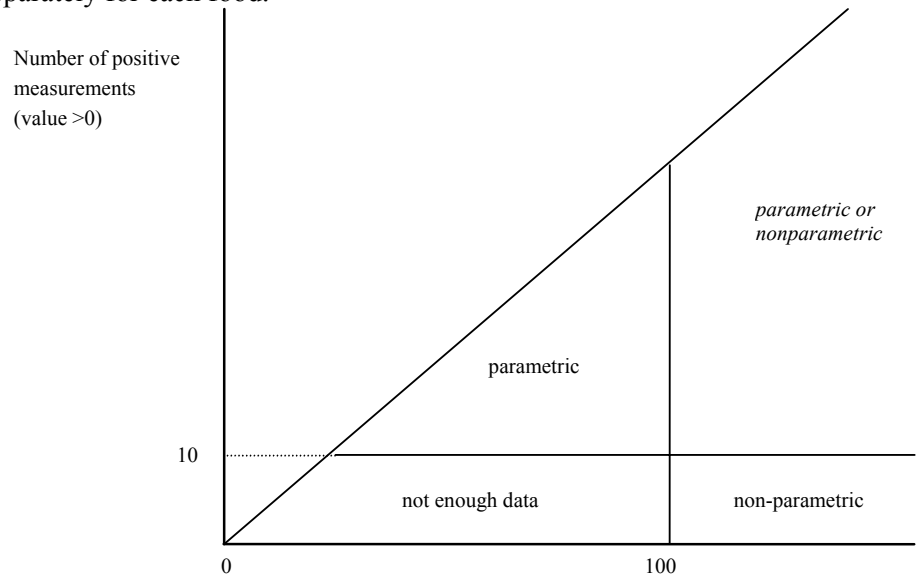


Figure 45: Use of non-parametric or parametric modelling for estimating the 99% percentile of the intake in relation to number of positive measurements and the total number of measurements.

5.2.2 Estimation based on histogram data

In EU reporting, compound data are sometimes reported in a tabulated (histogram) form: data are expressed as counts of observations falling into a series of groups. The observed counts are $n_1 \ldots n_c$, which fall into c classes with limits $c_1 \ldots c_c$. The number n_1 is the number of positive samples, which are nevertheless below the LOR (= c_1); n_2 is the number of positive samples that fall in between limits c_1 and c_2; n_c is the number of samples that fall in between limits c_{c-1} and c_c.

For histogram data, parameters μ and σ of the lognormal distribution can be obtained by fitting a normal distribution to a set of observations $n_1 \ldots n_c$. In an iterative way, expected counts for a standardised normal variable are calculated using the log-transformed group limits. Each round, parameters are updated until the process converges.

5.2.3 Estimation based on summary data

Occasionally, data are reported in a very condensed form. Summary statistics like the mean, quantiles and dispersion measures as the variance or the coefficient of variation are used to describe characteristics of the underlying concentration distributions. The reported statistics are calculated using all values (with concentrations below LOR sometimes replaced by ½*LOR), or using positive values (detects) only. In order to use the binomial-lognormal model, summary statistics based on all values must be corrected for the values replacing the concentrations below LOR. For the mean, the correction is straightforward, taking a zero or the midpoint-value (½*LOR). Likewise, the standard deviation or any measure of dispersion is corrected for the sum of squares due to all zero values and taking into account the corrected mean. The median is also corrected, but instead of correcting the value itself, a corrected quantile z_q is calculated corresponding to q, the lower fraction and z_q satisfying:

$$z_q = \Phi^{-1}\{q\} = \Phi^{-1}\{(\tfrac{1}{2}N - n_0)/(N - n_0)\}$$

with $\Phi(.)$, the cumulative probability function of the standard normal distribution, N, the total number of samples and n_0, the number of zeroes.

The maximum is the largest order statistic. Its expected value can be approximated by taking the appropriate population quantile, especially in large samples. Here, the problem is the other way around: the population quantile corresponding to the largest value given the sample size is to be estimated. For sufficiently large N an approximation to $E(q_{max})$ is provided by the value of z_q satisfying $\Phi(z_q) = N/(N+1)$. Blom (1958) and Harter (1961) made the following suggestions for smaller sample sizes:

$$z_q = \Phi^{-1}\{(N - \alpha)/(N - 2\alpha + 1)\}$$

with $\alpha = .315065 + .057974u - .009776u^2$ and $u = \log_{10}N$. Over a wide range of N α approximates the value 3/8. This empirical formula is a very accurate approximation to the exact value of $E(q_{max})$ and is used to estimate appropriate population quantiles for q_{max}. (David, 1970; Pearson and Hartley, 1972 ; Blom, 1958; Harter, 1961).

Three situations can be distinguished:

1) the reported information is insufficient to estimate both μ and σ, or
2) the reported statistics are sufficient to extract μ and σ, or
3) the information is redundant so various estimates for μ and σ are available.

Here, we first consider approaches for situation 2. Situation 1 requires additional information: a solution might be to use the information on comparable food-compound combinations to assess the necessary estimates. Situation 3, basically, is a pooling problem how to weigh and combine estimates that originate from different statistics.

5.2.3.1 Moments and other characteristics

A positive random variable X is said to be lognormally distributed with parameters μ and σ^2 if $Y = \ln X$ is normally distributed with mean μ and variance σ^2. The probability density function of X is:

$$f(x) = 1/(\sqrt{2\pi}\sigma x)\ \exp(-(\ln x - \mu)^2/2\sigma^2).$$

The corresponding normal distribution for Y is denoted by $N(\mu, \sigma^2)$.

Estimation of μ and σ using summary statistics is based on equations and characteristics derived from the moment generating function of the lognormal distribution. Required parameters are estimated by solving the formulas of the first two moments for μ and σ.
The following characteristics for variable X derived from the moment generating function are given:

mean:	$\exp(\mu + 1/2\sigma^2)$	(1)
variance:	$\exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1)$	(2)
mode:	$\exp(\mu - \sigma^2)$	(3)
quantile (q_q):	$\exp(\mu + z_q\sigma)$,	(4)
vc:	$\sqrt{(\exp(\sigma^2) - 1)}$	(5)

with vc the coefficient of variation, q a given lower fraction and z_q the corresponding standard normal deviate. The 50th quantile, the median, is a special case with $z_q = 0$. The geometric mean of X is equal to the median.

5.2.3.2 Estimation

Approach 1: estimation based on two quantiles, $q_{q1} \neq q_{q2}$.

Using (4) gives:

$\sigma = \log(q_{q1}/q_{q2}) / (z_{q1} - z_{q2})$. Substituting σ yields μ.

Approach 2: estimation based on a quantile and the mean.

Solving for σ using (1) and (4) gives:

$$\sigma^2 - 2z_q\sigma - 2\log(\text{mean}/q_q) = 0, \text{ with roots for } \sigma \text{ according to:}$$

$$z_q \pm \sqrt{(z_q^2 + 2\log(\text{mean}/q_q))} \qquad (6)$$

For moderate to small sample sizes the estimation of σ fails because the discriminant is negative, *i.c.* the argument of the square root function. Empirical simulations show that a negative discriminant happens more often for small sample sizes and for estimation based on extreme quantiles like the maximum. Figure 46 shows the empirical relation between the sample size and the fraction of failures for estimation methods involving the mean with respectively, the maximum and median. For the maximum, failures occur already at sample sizes n = 30 to 40, for the median n = 15 to 20. Negative discriminants occur when estimation is based on empirical (sampled) values instead of theoretical (calculated) values assuming a normal underlying distribution. The amount of failures for small sample sizes is in accordance with large sample theory. When the maximum is involved and estimation fails, an estimate of σ is assessed by equating the discriminant to zero. Empirical results show that this works out very well for sample size n > 4, although σ is slightly biased upwards being a conservative estimate. In case of the median no solution to this problem is available so the estimate of σ is set to a missing value.

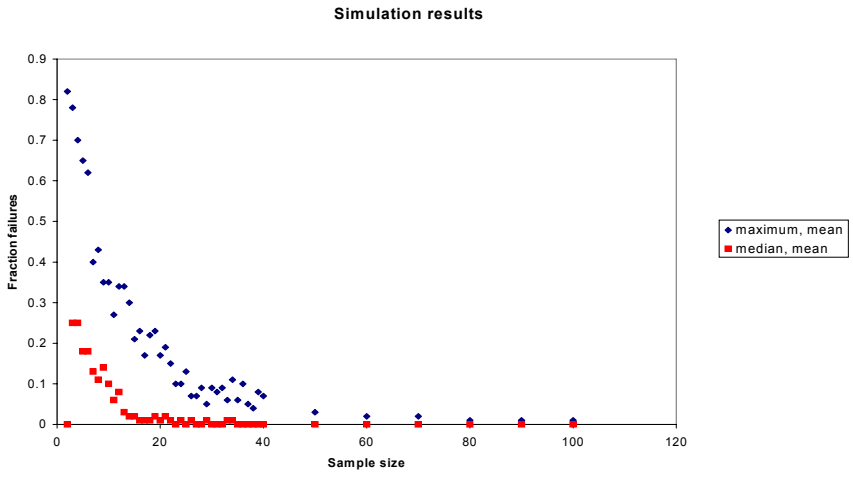


Figure 46: Simulated fraction of failures versus sample size for estimation of σ based on the mean and respectively the maximum and median

In general, for n large enough, say n > 40, σ has two roots. Usually, the mean is larger than the median. Then, σ is estimated with:

$$z_q + \sqrt{(z_q^2 + 2\log(\text{mean/median}))} \text{ with condition } \sigma > 2z_q..$$

In case of the mean and maximum σ is estimated with:

$$z_q - \sqrt{(z_q^2 + 2\log(\text{mean/max}))} \text{ with condition } \sigma < 2z_q..$$

Note that maximum is always greater than the mean. Here, the smallest root is taken as an estimate because empirical results show that the largest root yields unlikely high measures of dispersion and therefore should be rejected.

Approach 3: estimation based on mean and variance or coefficient of variation.

The coefficient of variation, vc = √(variance)/mean. Using (5), parameter σ is estimated with:

$$\sqrt{(\log(vc^2 + 1))}$$

and μ is estimated solving (1).

Approach 4: estimation based on a quantile and coefficient of variation.

For estimation of σ, see approach 3. Using (4), parameter μ is estimated with:

$$\log(\text{quantile}) - z_q\sigma$$

For the median, estimation of μ simplifies to:

$$\log(\text{median})$$

5.3 Acute risk assessment

5.3.1 Concentration model: modelling of compound concentrations in consumed food

5.3.1.1 Distributional assumptions

Compound concentrations in the various foods are independent and therefore can be modelled by univariate distributions.

5.3.1.1.1 Non-parametric modelling of concentrations (empirical)

In the empirical (non-parametric) approach, concentrations are sampled at random from the available data and combined with the consumption data to generate a new distribution of intake values. To assess the risk-intake, percentiles of the intake distribution are estimated.

Back to: Empirical

5.3.1.1.2 Parametric modelling of concentrations (binomial/lognormal no pooling)

In the parametric approach, compound concentrations per food are sampled from parametric distributions. A special feature of compound data is that the large majority of measured concentrations (often more than 80%) is recorded as zero (non-detects). These values may correspond to true zero concentrations (for example because the compound is never used in the specific food), or they may correspond to low concentrations which are below a pre-established reporting limit (LOR). In any case, the compound concentration distribution is very skew, with a large spike at zero and an extended tail to higher values. For statistical modelling a two-step procedure is chosen. First, the presence of a concentration ≥ LOR on food is modelled with a binomial distribution with a parameter p representing the probability of a reported concentration. Probability p depends on the chemical compound and the food and is estimated as the fraction of detects. Secondly, the non-zero compounds are modelled with the lognormal distribution. After consideration of several possibilities using the program *BestFit*, this distribution has been selected as being both theoretically sensible and practically useful. The parameters μ and σ are the mean and standard deviation of the log-transformed non-zero compound concentrations.

In the basic model (see 5.1)

$$c_{ijk} = I_{ijk} \cdot cpos_{ijk}$$

with I_{ijk} indicating whether a compound concentration is sampled (I_{ijk}=1) or not (I_{ijk}=0), and $cpos_{ijk}$ the compound concentration in the subpopulation of positive values. The probability of I_{ijk} being 1 or 0 depends on the number of detects found for food *k* and I_{ijk} is sampled separately for each consumer *i* on occasion *j*.
Back to: Binomial/lognormal (no pooling)

5.3.1.1.3 Parametric modelling of concentrations (binomial/lognormal with pooling)
When data are limited, it may be advantageous to apply the parametric approach for modelling of the positive concentrations. In MCRA the positive concentrations are modelled as lognormal with parameters μ and σ^2, representing mean and variance of the natural logarithm of the concentrations. However, estimation is often hampered because data on compounds in specific foods are sparse or even missing. In those cases, grouping of foods into food groups enlarges the number of measurements per group and may give sufficient data to base estimates upon. We must assume that concentration distributions are the same for the grouped foods. A second related question is the reliability of estimates, based on a few number of degrees of freedom. The following procedure is designed to cope with the above problems.

1. **Pooling variances within food groups.** For each food the variance σ^2 and mean μ is estimated. Then, foods are assigned to groups which are composed of related foods, *e.g.* a foodgroup containing sorts of cabbages or a group containing all kind of berries. Foods where agricultural use is allowed are remained separate from foods where agricultural use is not allowed. The homogeneity of variances in the different groups is assessed using Bartlett's test (Snedecor & Cochran, 1980). The test statistic determines whether variances within a group are to be pooled automatically ($p > 0.05$) or not ($p \leq 0.05$).
2. **Pooling means within food groups.** After pooling the variances, an overall test for differences of means within each group is performed, based on analysis of variance. Means within groups are pooled automatically if the probability $p > 0.05$.
3. **Using overall variance if there are < 10 degrees of freedom.** Estimates of variances based on less than 10 df are considered not very reliable. Therefore, variances based on < 10 df are compared to the overall variance (pooled over all foods except the tested food itself, *i.c.* corrected) and tested for equality. Variances are replaced by the overall variance (uncorrected) whenever the hypothesis of equality of variances is not rejected; if rejected, the original variances are maintained.

For a parametric risk assessment all variances and means must be present. This requirement implies that very often rearrangement of foods into (sub) groups preceeds the actual simulation of the intake distribution.

To summarise, actions are:
- calculate variances and means for each food
- classify foods into groups
- test homogeneity of variances and equality of means within groups of foods. Results are: not significant ($p > 0.05$) or significant ($p \leq 0.05$).
- take foods(-groups) with df < 10
- compare variance with overall variance (corrected). Replace variance with overall variance (uncorrected) for non-significant test results.

Back to: Binomial/lognormal with pooling

5.3.1.2 Modelling of missing data and replacement of non-detects
Missing data should be indicated by 9999 in the database tables. In principle such values are ignored in the analysis.
Most monitoring measurements of chemical compounds are non-detects, i.e. no quantitative measurement is reported. For this reason data are entered in the Concentration table by specifying the total number of measurements made together with the LOR. We use LOR to mean exactly what the

term says: measurements below LOR are not reported, whereas values equal to or higher than LOR are represented by numerical values in the database.
In the analytical and food risk fields analytical limits are often indicated as LOD (limit of detection) or LOQ (limit of quantification). Unfortunately, it is not always clear what is meant with these terms. In any case official recommendations are to always report any available numerical values even if they are below LOD or LOQ limits (IUPAC 1995).
For legal applications of compounds data may be available about the percentage of the crop which receives treatment. When a chemical compound can enter the food chain only via crop treatment, and when the percentage of crop treated is (approximately) known to be $100p_{crop\text{-}treated}$, then this knowledge may be used to infer that $100(1\text{-}p_{crop\text{-}treated})$% of the monitoring measurements should be real zeroes, contributing nothing to pesticide intake, whereas other non-detects in the monitoring data could have any value below the LOR. For $100(p_{non\text{-}detect} + p_{crop\text{-}treated} - 100)$% of the monitoring measurements, 0 and LOR represent best-case and worst-case estimates. A simple way (tier 1 approach) to consider the uncertainty associated with non-detects is to compare intake distributions for these best-case and worst-case situations.
Back to: Replacement of non-detects

5.3.1.3 Modelling of processing effects

Concentrations in the consumed food may be different from concentrations in the food as measured in monitoring programs (typically raw food) due to processing, such as peeling, washing, cooking etc.
In general, we assume the model:

$$cpos_{ijk} = f_k \cdot cr_{ijk}$$

where cr_{ijk} is the concentration in the raw food, and where f_k is a factor for a specific combination k of RAC and processing. Values will typically be between 0 and 1, although occasionally the processing factor may also be >1 (*e.g.* drying as applied for grapes and figs).
The user of the model will have to specify processing factors for each food k as defined in the food consumption data base. For this purpose, it is advised to maintain a data base of processing factors, indexed by chemical compound, RAC and processing type (*e.g.* washing, peeling or other processing). Before running the model, it may then be necessary to specify how the necessary processing factors are derived from the data base entries and/or other information. Example: if there are no processing factors known for captan in pears, it may be decided to use the corresponding factors for apples instead.
Often processign effects may be variable, and this may be entered in the Monte Carlo modelling by specifying two values for each processing factor:

1. $f_{k,nom}$: the nominal value, typically some sort of central value from an experimental study
2. $f_{k,upp}$: an upper 95% confidence limit, which typically will be set by an expert (even if statistical information on variability of the factor is available)

A typical data base entry might thus read:

RAC	processing	$f_{k,nom}$	$f_{k,upp}$
apple	washing	0.5	0.7

In the MC-modelling, processing factors can be used in either of three ways (for each food k to be chosen by the user):

5.3.1.3.1 No processing factor
Just take $f_k = 1$. This is in most (though not all) cases a worst-case assumption. No data on processing are needed and therefore this route is useful in a first tier approach.

5.3.1.3.2 Processing (fixed factors)
Use $f_k = f_{k,upp}$. Available information on specific processing effects is used, although still in a cautionary way (in accordance with the precautionary principle). Note that $f_{k,nom}$ values need not to be specified.

5.3.1.3.3 Processing (distribution based)

Sample f_k using a normal distribution. Log or logit transformed values of $f_{k,nom}$ and $f_{k,upp}$ are used to define the first two moments of the normal distribution. Two situations are distinguished depending on the type of transformation.

a) The logarithms of $f_{k,nom}$ and $f_{k,upp}$ are equated to the mean and the 95% one-sided upper confidence limit of a normal distribution. This normal distribution thus is specified by a mean $\ln(f_{k,nom})$ and a standard deviation $\{\ln(f_{k,upp}) - \ln(f_{k,nom})\}/1.645$. Values are drawn from this distribution in the MC-simulations. Processing factors f_k will be nonnegative. Note: $f_{k,upp}$ and $f_{k,nom}$ values equal to 0 are replaced by a low default value (0.01); this is useful computationally to avoid problems with logarithms.

b) The logits of $f_{k,nom}$ and $f_{k,upp}$ are equated to the mean and the 95% one-sided upper confidence limit of a normal distribution. This normal distribution thus is specified by a mean $\text{logit}(f_{k,nom})$ and a standard deviation $\{\text{logit}(f_{k,upp}) - \text{logit}(f_{k,nom})\}/1.645$. Values are drawn from this distribution in the MC-simulations. Processing factors f_k will be between 0 and 1. Note: $f_{k,upp}$ and $f_{k,nom}$ values equal to 0 and 1 are replaced by default values (0.01 and 0.99); this is useful computationally to avoid problems with logits.

The user should keep in mind that, in case of a lognormal distribution, $f_{k,nom}$ defines the median, while $f_{k,upp}$ quantifies skewness. The same holds for the logistic. Usually, a logarithm will be the standard transformation, but for very skew distributions (see Figure 47) occasionally values above 1 are sampled (upper row, 1rst, 3rd and 5th plot). A logit transformation should be considered instead.

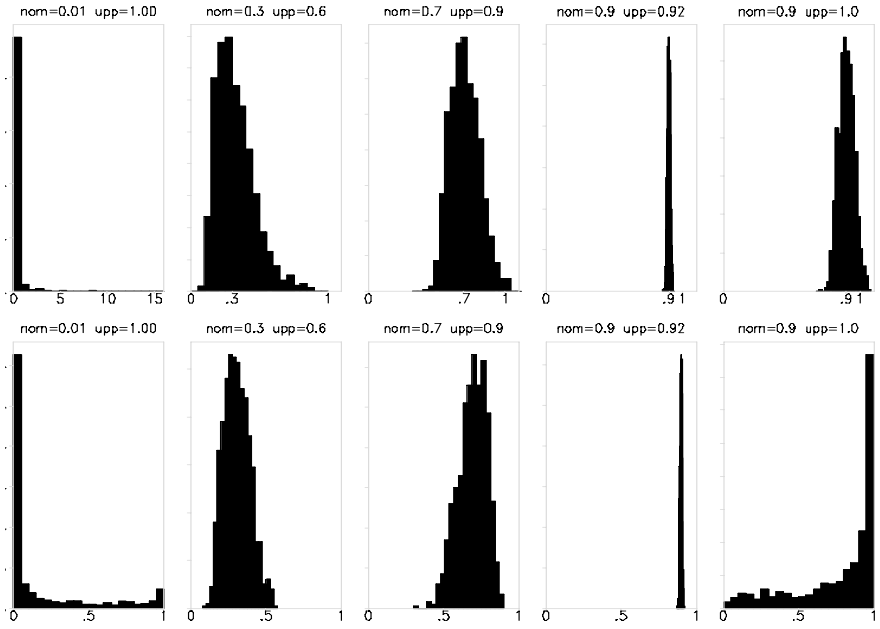


Figure 47: Lognormal (upper row) and logistic (lower row) distributions for various values of $f_{k,nom}$ (= nom) and $f_{k,upp}$ (= upp)

Back to: Modelling processing effects

5.3.2 Modelling of unit variability

5.3.2.1 Introduction, variability in deterministic modelling

Variability in compound concentrations between individual units is a relevant factor in the assessment of short-term dietary intake of chemical compounds. It is addressed separately because monitoring measurements cm_k are typically made on homogenised composite samples, both in controlled field trials and in food monitoring programs. Such a composite sample for food k is composed of nu_k units

with nominal unit weight wu_k each. The weight of a composite sample is therefore $wm_k = nu_k \times wu_k$. This weight is often larger than a consumer portion, *e.g.* a typical composite sample of 20 sweet peppers weighs 3.2 kg, whereas daily consumer portion weights in the Dutch Food Consumption Survey 1997 ranged from 0.08 g to 458 g.
How should monitoring data be used to estimate the raw food concentration levels cr_{ijk} in consumer portions? Although the mean level of cm_k may be a fair estimate of the mean level of cr_{ijk}, the variability of cm_k is not appropriate to estimate the variability of cr_{ijk}. In smaller portions more extreme values may occur more readily, and thus acute risks may be higher than would follow from a direct use of the composite sample data.
Therefore, the FAO/WHO Geneva Consultation recommended to include a *variability factor* (v) in the non-probabilistic calculation of an international estimate of short-term intake (*IESTI*) (FAO/WHO 1997). The *IESTI* has been adopted by the Joint Meeting of FAO and WHO experts on Pesticide Compounds in food in 1999, and was modified in 2000 to reflect that the supply for actual consumption on a given day is likely to be derived from a single lot (JMPR 1999, 2000). In both the original and the modified definition, the variability factor is used in a similar way. The basic idea is that the compound concentration for the first unit eaten is multiplied by v, whereas this factor is not applied for any remaining part of the daily consumption.

In the original presentation v was meant to reflect "*the ratio of a highest concentration in the individual product unit to the corresponding concentration seen in the composite sample*" (FAO/WHO 1997). It was not clearly stated what was meant with "*a highest concentration*". Should this be the maximum concentration found or should it be a high percentile, *e.g.* p95 or p97.5? In practical terms this did not matter too much, because little data were available. Therefore the FAO/WHO Consultation recommended to take *initial* values of v equal to "*the number of units in the composite sample as given in Codex sampling protocols*". This will provide a conservative estimate of the compound concentration in the first unit, based on the assumption that all of the compounds present in the composite sample are present in this single unit. If Codex sampling protocols are used, then the number of units per composite sample is 5 for large crops (unit weights > 250 g) and 10 for medium crops (unit weights 25-250 g). For small crops (< 25 g) a variability factor v = 1 was recommended. More recently, it has been proposed to replace the default value 10 with 7. For foods which are processed in large batches, *e.g.* juicing, marmalade/jam, sauce/puree, a variability factor v = 1 is proposed. To summarise:

unit weight, *wu*	FAO/WHO default variability factor, *v*
< 25 g	1
25 –250 g	7
> 250 g	5
juicing, marmalade/jam, sauce/puree	1

Table 20: Default variability factors for IESTI calculations

The Consultation specifically recommended to replace these default values with more realistic values obtained from studies on actually measured units. A working group of the International Conference on Pesticide Residues Variability and Acute Dietary Risk Assessment held in York in 1998 suggested to define v, for samples taken from controlled trials, as the 97.5th percentile of the unit levels divided by the sample mean (Harris *et al*. 2000), and this is used in the current version of MCRA as the defining relation.
Back to: Unit variability model options

5.3.2.2 Approaches to unit variability in probabilistic modelling: specifying distributions

How should variability between units be incorporated in probabilistic modelling of acute risks? In probabilistic modelling we generate consumption amounts and compound concentrations which will be multiplied, summed over foods and divided by body weight to estimate the intake. However, the compound concentration cm_k will usually be derived from a distribution based on measurements on composite samples. Assume that a batch of food contains N units (N large, for the statistics we assume

infinite). The monitoring measurement cm_k is made on a composite sample of nu_k units (for example, nu_k = 5). These units are assumed to be representative of the batch. Unit concentrations cr_{ijk} are to be simulated for one or more units from this batch that will be part of a consumption portion in the MC-simulation. Basically, there are three possibilities depending on the availability of data:

1. use actual measurement data on individual units;
2. use variability factors or other summary statistics based on measured individual units;
3. use conservative assumptions.

In MCRA only methods under categories 2 and 3 are implemented. The first approach has been pioneered in the context of a large UK survey on pesticides in fruit (Hamey 2000).

In MCRA the following three models, discussed below in more detail, are implemented:

1. **Beta model**, requires knowledge of the number of units in a composite sample, and of the variability between units (realistic or conservative estimates);
2. **Bernoulli model**, requires only knowledge of the number of units in a composite sample (results are always conservative);
3. **Lognormal model**, requires only knowledge of the variability between units (realistic or conservative estimates).

Preferably realistic estimates of unit variability are to be used, either expressed as coefficients of variation *cv* (standard deviation divided by mean) or as variability factors *v* (defined in MCRA as 97.5th percentile divided by mean). However, often such information is not directly available. In such cases it is customary to select high values for the variability factor, either based on collections of variability factors for other compounds/foods, or calculated as the theoretical maximum derived from the number of units in a composite sample.

How to translate the concept of conservatism to the probabilistic model? In a non-probabilistic model a higher value of *v* gives a higher *IESTI*, but in a stochastic model a higher variability means more spread around a central value. In general this means that higher values, but also lower values can be generated. In order to retain an overall conservatism it is therefore necessary to replace all simulated values below the monitoring level (cm_k) with cm_k itself.

It is common to use default conservative values, such as the FAO/WHO variability factors in Table 20. However, one should be aware that two entirely different interpretations are possible:

1. The default variability factor may be defined in the same way as a data-based variability factor (*v* = 97.5th percentile/mean). For example, it may be an expert opinion based on seeing many actual data sets from trials, that a certain value v can be used as a conservative value for other situations (see *e.g.* Table 1 in Harris *et al*. 2000). Then we might use the beta or the lognormal model, censoring these distributions at cm_k to guarantee conservative behaviour. For the beta model additional information on the number of units in a composite sample is needed.
2. Alternatively, one can revert to the original definition and interpret FAO/WHO variability factors as the number of units in the composite sample ($v = nu_k$). In this case, without other information, the only workable model is the Bernoulli model.

Back to: Estimated parameters for unit variability

5.3.2.2.1 Beta model for unit variability

With this model MCRA will generate values for individual unmeasured units of a measured composite sample. If cm_k is the concentration measured (or simulated) for the composite sample in monitoring for food *k*, then the concentration in any unit can be no larger than $c_{max} = nu_k * cm_k$, where nu_k is the number of units in the composite sample. Under the Beta model simulated unit values are drawn from a bounded distribution on the interval (0, c_{max}). The parameter for unit variability is specified as a coefficient of variation cv_k of the unit values in the composite sample, or as a variability factor.

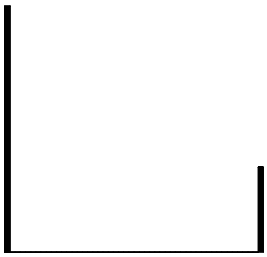
The standard beta distribution is defined on the interval (0, 1) and is usually characterised by two parameters *a* and *b,* with $a>0$, $b>0$ (see *e.g.* Mood *et al.* 1974). Alternatively, it can be parameterised by the mean $\mu=a/(a+b)$ and the variance $\sigma^2=ab(a+b+1)^{-1}(a+b)^{-2}$, or, as applied in MCRA, by the

mean μ and the squared coefficient of variation $cv^2=ba^{-1}(a+b+1)^{-1}$. Note that the coefficient of variation is the same for the unscaled and the scaled distributions.
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:

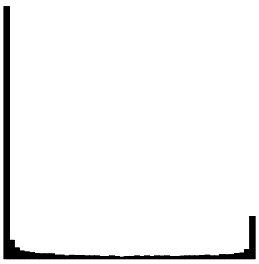
$$a = b(nu_k - 1)^{-1}$$
$$b = \frac{(nu_k - 1)(nu_k - 1 - cv_k^2)}{nu_k cv_k^2}$$

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.

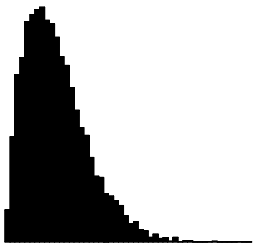
When the unit variability is specified by a variability factor $v_k = \frac{p97.5_k}{cm_k}$ instead of a coefficient of variation cv_k then MCRA applies a bisection algorithm to find a such that the cumulative probability $P[Beta(a,b)] = 0.975$ for $b = a(nu_k - 1)$.
Sampled values from the beta distribution are rescaled by multiplication with c_{max} to unit concentrations cr_{ijk} on the interval (0, c_{max}).
In the case that variability has been estimated by a conservative high value, all sampled values lower than cm_k are replaced by cm_k.
In Figure 48, for several values of the coefficient of variation and number of units the beta distribution is shown with estimated parameters a and b. When the parameter for unit variability is high (upper left plot) the ratio of the spikes on the extremes (3:1) represent the 75% probability at $cr_{ijk} = cm_k$ and 25% probability at $cr_{ijk} = c_{max}$. In the upper right plot, the parameter for unit variability is smaller and some unit values in between the two extremes are sampled. The ratio of the spikes is about 5:1, which is according to the number of units in the composite sample. In the lower left plot, variability is low and unit values are sampled around the monitoring compound. In the extreme case, when unit variability is close to zero the monitoring compound itself is sampled and a spike occurs (not shown). The lower right plot shows an intermediate situation, moderate to high variability.



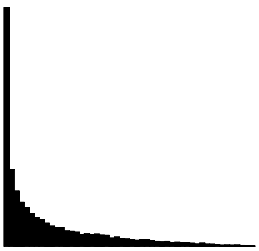
cv_k =1.732; nu_k=4; a=0.00005; b=0.00015



cv_k =1.20; nu_k=6; a=0.4; b=2



cv_k=0.62; nu_k=6; a=2; b=10



cv_k =1.46; nu_k=4; a=0.1; b=0.3

Figure 48: Standard Beta distribution for different values of the coefficient of variation cv_k and number of units nu_k in the composite sample. x axis from 0 to 1.

Back to: Beta distribution

5.3.2.2.2 Lognormal model for unit variability

With the Beta and Bernoulli models, MCRA simulates concentrations for units in the composite sample, such that the concentration of an individual unit can never be higher than the monitoring measurement multiplied by the number of units in the composite sample $c_{max} = nu_k * cm_k$.

With the Lognormal model for unit variability MCRA simulates concentrations for new units in the batch from which the composite sample was taken. Effectively the number of units in a batch is very large, so in this case there is no practical upper limit to the concentration that can be present.

The lognormal distribution is considered as an appropriate model for many empirical positive concentration distributions. With the Lognormal model MCRA assumes a lognormal distribution for unit compound concentrations. Let this distribution be 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 $\mu = \ln(cm_{ik})$.

Also for the Lognormal model MCRA allows two choices to specify the parameter for the unit variability. The parameter is specified as a coefficient of variation (cv_k) or as a variability factor (v_k).

The coefficient of variation cv is turned into the standard deviation σ on the log-transformed scale with:

$$\sigma = \sqrt{\ln(cv^2 + 1)}$$

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: $\sigma^2 - 2*1.96\sigma - 2\log(v) = 0$, with roots for σ according to:

$$\sigma = 1.96 \pm \sqrt{(1.96^2 + 2\log(v))}$$

The smallest positive root is taken as an estimate for σ (see also 5.2.3.2).

In the case that variability has been estimated by a conservative high value, all sampled values lower than cm_k are replaced by cm_k.

Back to: Lognormal distribution

5.3.2.2.3 Bernoulli model for unit variability

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. As a worst case approach we may take cv_k as large as possible. When cv_k is equal to the maximum possible value $\sqrt{nu_k - 1}$, the (unstandardised) Beta distribution simplifies to a Bernoulli distribution with probability $(nu_k - 1)/nu_k$ for the value 0 and probability $1/nu_k$ for the value $c_{max} = nu_k * cm_k$..

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 $cr_{ijk} = cm_k$ and 20% probability at $cr_{ijk} = c_{max}$.

Back to: Bernoulli distribution

5.3.2.2.4 Estimation of intake values using the concept of unit variability

- For each iteration i in the MC-simulation, obtain for each food k a simulated intake x_{ik} , and a simulated composite sample compound 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 κ_{ikl} from a Beta or Bernoulli distribution. Calculate concentration values as $c_{ikl} = \kappa_{ikl} * cm_{k,\,max} = \kappa_{ikl} * cm_k * nu_k$. Sum to obtain the simulated concentration in the consumed portion:

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

- For the Lognormal distribution: draw nux_{ik} simulated logconcentration values lc_{ikl} from a normal distribution with mean $\mu = \ln(cm_{ik})$ and standard deviation σ. Back transform and sum to obtain the simulated concentration in the consumed portion:

$$cr_{ik} = \sum_{l=1}^{nux_{ik}} w_{ikl} e^{lc_{ikl}} \Big/ x_{ik}$$

5.3.2.3 Additional: Comparison of probabilistic with deterministic estimates of acute risk

The IESTI (International Estimated Short-Term Intake) is a deterministic estimate of the short-term intake of a compound on the basis of the assumptions of high daily food consumption per consumer and highest compounds from supervised trials. The IESTI is expressed per kg body weight and has only been defined for single foods.

MCRA calculates IESTI for comparison with MC-percentiles.

Calculations of IESTI (according to FAO 2002) recognise four different case (1, 2a, 2b and 3). In cases 1 to 3 the following definitions are used:

LP:	Highest large portion reported, calculated as the 97.5th percentile of the distribution of consumed portions on days with positive consumption of the food (kg food/day)
HR:	Highest residue (=compound) in composite sample, mg/kg
bw:	Mean body weight, kg; in MCRA values may be input by the user, or weighted means are calculated over consumers with the number of days on which they consumed the food as weights
U:	Unit weight of the edible portion, kg.
v:	Variability factor – the factor applied to the composite compound to estimate the concentration in a high-compound unit
MR:	Median residue (=compound) in food, mg/kg

Although the FAO Manual refers to supervised trials only, MCRA calculates concentrations (HR or MR) from any compound concentration data set supplied (may also be monitoring data).

Concentrations (HR or MR) may be multiplied with a processing factor on beforehand, in MCRA this depends on the options chosen for processing.

Case 1:

The compound in a composite sample reflects the concentration in meal-sized portion of the food (unit weight is below 25 gr).

$$\text{IESTI} = \frac{LP * HR}{bw}$$

Case 2:

The meal sized-portion, such as a single fruit or vegetable unit might have a higher concentration than the composite (whole fruit or vegetable unit weight is above 250 gr). Case 2 is further divided into case 2a and 2b.

Case 2a:
Unit edible weight of raw food is less than large portion weight.

$$\text{IESTI} = \frac{U * HR * v + (LP - U) * HR}{bw}$$

The formula is based on the assumption that the first unit contains compounds at the HR*v level and the next one contains compounds at the HR level, which represents the compound in the composite from the same lot as the first one.

Case 2b:
Unit edible weight of raw food exceeds large portion weight.

$$\text{IESTI} = \frac{LP * HR * v}{bw}$$

The formula is based on the assumption that there is only one consumed unit and it contains compounds at the HR*v level.

Case 3:
For those processed foods where bulking or blending means that the median represents the likely highest concentration.

$$\text{IESTI} = \frac{LP * MR}{bw}$$

When an acute reference dose is available, the calculated IESTI values are also expressed as a percentage of the ARfD.

IESTI is a deterministic estimate to reflect the unit variability within a composite sample. In the probabilistic approach, unit variability is explicitly modelled and the result is an estimate of the intake distribution (per food). These two different approaches handle the same problem, but it is undefined to which MC-percentile the IESTI value should be compared. In MCRA the user is free to choose a percentage point for this comparison.

A point to note is that IESTI is calculated from positive consumptions on each separate food. To allow a fair comparison, the MC-percentiles are calculated in the same way. Note, however, that in a multi-food MC-analysis, even if one restricts the attention to consumption days only, the percentiles are typically based on consumption data which are partly zero (days with consumption zero for some but not all foods).
Back to: Additional options concerning IESTI and consumption days

5.3.3 Intake model

5.3.3.1 Only empirical estimates

This is an acute risk assessment using empirical compound data.

5.3.3.2 Empirical estimates and betabinomial/normal

An acute risk assessment may be followed by an analysis where the acute intake distribution is related to a covariable and/or cofactor. The simulated acute intake are input and the betabinomial distribution is used to estimate an intake frequency model describing the effects of explanatory variables. The positive intake values are transformed to normality and an intake amount model is estimated, again relating the amount value to a covariable and/or cofactor. MC-sampling is used to obtain a simulated acute intake distribution for each combination of levels of covariable and cofactor.

Back to: Empirical estimates and betabinomial/normal

5.3.3.2.1 Monte Carlo data generation of intake

Through MC-sampling, a large number of intakes is generated by combining randomly chosen consumption patterns of consumers *i* on day *j* with randomly chosen concentrations in the consumed foods. The replicates generated for consumer day *ij* are further indexed by *k* to represent differences due to concentration variability. We ignore the finiteness of the concentration data, that is, we ignore the identity of the chosen concentration values in the original concentration dataset.

Back to: Empirical estimates and betabinomial/normal

5.3.3.2.2 Intake frequency model

Let n_i and $npos_i$ be the total number of simulated intakes per consumer, and the number of simulated positive intakes, respectively. Then $npos_i$ is modelled as a function of *e.g.* age (and/or other consumer characteristics), using a betabinomial distribution with binomial totals n_i and overdispersion parameter ϕ (independent of age). The fitted binomial probabilities are $\hat{\pi}_x = f(x_i)$, where x_i is the age of consumer *i*, and the estimated overdispersion parameter is $\hat{\phi}$.

Back to: Empirical estimates and betabinomial/normal

5.3.3.2.3 Intake amount model

For the positive intakes, consider power of logarithmically transformed values y_{ijk}. (see 5.4.3.1 Average over replicates to obtain consumer day averages $y_{ij.}$. These values are modelled in a ML analysis with random terms consumer and consumer.day as a function of age (and/or other consumer characteristics), with the number of values per consumer day (n_{ij}) as weights w_{ij} to correct for differences in the precision at the consumer day stratum. The fitted values from the model are $\hat{\mu}_x = f(x_i)$, where x_i is the age of consumer *i*.

Back to: Empirical estimates and betabinomial/normal

5.3.3.2.4 Estimating the acute risk variability of positive intake amounts

Correct the full set of simulated positive intakes by $y'_{ijk} = y_{ijk} - \hat{\mu}_{x(i)}$. Estimate the variance $\sigma^2_{y'}$ of y'_{ijk}. We denote the estimated variance as $\hat{\sigma}^2_{y'}$. Now for each selected age *x* the transformed positive intake distribution is modelled as normal with mean $\hat{\mu}_x = f(x)$ and variance $\hat{\sigma}^2_{y'}$.

Back to: Empirical estimates and betabinomial/normal

5.3.3.2.5 Estimating the acute intake distribution

Acute intake distributions dependent on a covariable and/or cofactor are obtained by numerical integration. For each combination of levels of the covariable and cofactor, intake frequency values and transformed intake amounts are simulated and multiplied. This results in a number of distributions each one representing the acute intake distribution corresponding to a specific combination of levels of the covariable and cofactor. Covariable- and cofactor-dependent percentiles can be derived.

Back to: Empirical estimates and betabinomial/normal

5.4 Chronic risk assessment

In the MCRA program we have 2 models available to assess chronic risks:

1) the betabinomial/normal model (see 5.4.2 , 5.4.3 , 5.4.5),
2) the discrete/semi-parametric (ISUF) model without covariable and cofactor (see 5.4.6).

5.4.1 Introduction

In a chronic risk assessment the main interest goes to the fraction of consumers with a usual intake per day higher than an intake limit. Usual intake is defined here as the long-run average of daily intakes of a (chemical) compound by a consumer.
Usually, food consumption data are available for consumers on 2 (or more) consecutive days. We assume an equal number of days for each consumer. This is in confirmity with our method of data entry for consumption (see 9.4.1). As a consequence, days without consumptions do have zero intake.
Through the assumed independence of consumption data and compound concentration values (a most reasonable assumption) the daily intake of consumer i on day *j* can be calculated as the sum over foods of consumption amount per kg body weight times average concentration. See the basic model in 5.1. The average concentration of all available concentration measurements on a food is taken, with non-detect measurements entered as zero, $\frac{1}{2}LOR$ or LOR, or any other fraction of LOR as specified in the input options.
To calculate the usual intake in the first 2 available models two aspects will be taken into account: the number of days that a consumer eats that compound and if so, the amount of that compound. The number of days will be assumed to have a binomial distribution with a probability p different for each consumer. These probabilities will be assumed to come from a beta distribution. The eaten compound will be assumed to have a normal distribution mostly after transformation. In the next paragraphs the modelling will be further discussed.

5.4.2 Modelling the intake frequency distribution

Let *n* and *npos* be the total number of days per consumer (for all consumers equal) and the number of days with a positive intake, respectively. Then *npos* is modelled using a betabinomial distribution with binomial totals *n* and probabilities *p*. The probabilities, *p*, are assumed to follow a beta distribution:

$$f(p)= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}$$

With $B(\alpha,\beta)= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$, the probability that *npos* equals *x* can then be written as:

$$P(\underline{npos}=x) = \binom{n}{x} \frac{B(\alpha+x, n+\beta-x)}{B(\alpha,\beta)}, \qquad x=0, 1 \ldots n$$

This distribution is known as the betabinomial distribution.

The mean and variance of a beta distribution are:

$\alpha/(\alpha+\beta)$
and $\alpha\beta(\alpha+\beta+n)/[(\alpha+\beta)^2(\alpha+\beta+1)]$, respectively.

Re-parameterizing by $\pi=\alpha/(\alpha+\beta)$ and $\varphi=1/(\alpha+\beta+1)$ is a more stable and interpretable parameterization. It can be shown that the mean and variance of *npos* are equal to $n\pi$ and $n\pi(1-\pi)[1+(n-1)\varphi]$, respectively.

Note that the first part of the variance $n\pi(1-\pi)$ equals the binomial variance; the second part is the so-called overdispersion factor.

Fitting the betabinomial model with maximum likelihood gives estimates $\hat{\pi}$ and $\hat{\varphi}$ for the parameters π and φ. Back-transformation gives the following estimates for α and β:

$$\hat{\alpha} = \hat{\pi}(1-\hat{\varphi})/\hat{\varphi} \text{ and } \hat{\beta} = (1-\hat{\pi})(1-\hat{\varphi})/\hat{\varphi}$$

The distribution of the probability that a consumer eats a compound at a certain day is then:

Beta($\hat{\alpha}$, $\hat{\beta}$).

Back to: Chronic risk analysis

5.4.3 Modelling the positive intake amounts

5.4.3.1 Power or log transformation

First, to achieve a better normality, the positive daily intake amounts are transformed. The user can choose a logarithmic transformation $f(y) = \ln(y)$ (no parameters to be estimated) or a power transformation $f(y) = y^{\gamma}$ (one parameter to be estimated). In the latter case the optimal power is determined on the grid $\{10, 2, 1, \frac{1}{1.5}, \frac{1}{2}, \frac{1}{2.5}, \frac{1}{3}, \frac{1}{3.5}, ..., \frac{1}{100}\}$, with a further refinement grid search around the best fitting value. If a power $\frac{1}{100}$ gives the best fit in this grid search, then the logarithmic transformation is selected (Note that a logarithmic transform corresponds theoretically to $\gamma = 0$). The goodness of fit is determined by minimising the residual sum of squares: $(z(i) - \beta_1 y^{\gamma})^2$ of a regression of normal Blom scores on the power-transformed daily intakes. Normal Blom scores are (Tukey 1962):

$$z_{(i)} = \Phi^{-1}\left(\frac{i - \frac{3}{8}}{n + \frac{1}{4}}\right)$$

where i is the rank of the nth non-zero daily intake, n, the total number of non-zero intakes and $\Phi^{-1}(\cdot)$ is the inverse of the standard normal cumulative distribution function.

5.4.3.2 Model with between-consumer and within-consumer variance component

5.4.3.2.1 Log transformation

The transformed positive intake amounts are modelled in a ML analysis with random terms *consumer* and interaction *consumer.day* to estimate the between-consumer and within-consumer variance component:

$$\ln(\underline{y}_{ij}) = \mu + \underline{c}_i + \underline{u}_{ij}$$

where $\underline{c}_i$ and $\underline{u}_{ij}$ are the consumer effect and interaction effect respectively. These effects are assumed to be normally distributed $N(0, \sigma^2_{between})$ resp. $N(0, \sigma^2_{within})$.
If the positive intake amounts are logarithmically transformed it can be shown that the expectation and variance of the positive intake amount per random consumption day of a random consumer are:

$$E(\underline{y}_{ij}) = \exp(\mu + \tfrac{1}{2}\sigma^2_{within})$$
$$Var(\underline{y}_{ij}) = \sigma^2_{between}$$

5.4.3.2.2 Power transformation

The power transformed (with power q) positive intake amounts are modelled in a ML analysis with random terms *consumer* and interaction *consumer.day* to estimate the between-consumer and within-consumer variance component:

$$\underline{y}_{ij}^{q} = \mu + \underline{c}_i + \underline{u}_{ij}$$

where $\underline{c}_i$ and $\underline{u}_{ij}$ are the consumer effect and interaction effect respectively. These effects are assumed to be normally distributed N(0, $\sigma^2_{between}$) resp. N(0, σ^2_{within}).

5.4.4 Modelling usual daily intake

5.4.4.1 Analytical integration

For logarithmic transformed intake amounts, a analytical solution is available (not implemented in the MCRA program).
The usual intake is defined as the intake amount per random intake day (over both intake and non-intake days) of a random consumer. To obtain the usual intake the E($\underline{y}$) from 5.4.3.2 has to be multiplied by the probability π from 5.4.2 . If π was constant for all consumers the usual intake would have a lognormal distribution with mean $\ln(\pi) + \mu + \frac{1}{2}\sigma^2_{within}$ and variance $\sigma^2_{between}$. But because we have assumed in 5.4.2 that consumers have different p's coming from a beta distribution, the probability that a consumer has a usual intake lower than say an intake limit z equals:

$$P(py \le z) = \int_p (\underline{p} = p \wedge \underline{y} \le \frac{z}{p}) = \int_p (\underline{p} = p \wedge \ln \underline{y} \le (\ln(z) - \ln(p)))$$

$$\int_{p=0}^{1} \frac{\Gamma(\hat{\alpha}+\hat{\beta})}{\Gamma(\hat{\alpha})\Gamma(\hat{\beta})} p^{\hat{\alpha}-1}(1-p)^{\hat{\beta}-1} \Phi\left(\frac{\ln(z) - \ln(p) - \hat{\mu} - \frac{1}{2}\hat{\sigma}^2_{within}}{\hat{\sigma}_{between}}\right) dp$$

where Φ is the cumulative normal distribution.

5.4.4.2 Numerical integration

If the positive amounts are transformed by a power transformation the power transformed values can not generally be written in terms of a probability distribution as in 5.4.4.1 : the distribution of the usual intakes has to be calculated numerically.
However, in the MCRA program for both power and logaritmic transformation, the usual intake distribution is obtained by numerical integration.
The distribution of the usual intakes can be obtained as follows:

1. Draw 1 value of a normal distribution with mean μ and variance $\sigma^2_{between}$
2. Calculate the inverse transformation of the value of Step 1.

   2a. For a logarithmic transformation: $x = \exp(\mu + \sigma_{between}\, ln\, y) + \frac{1}{2}\sigma^2_{within}$.

   2b. For a power transformation: $x = (\mu + \sigma_{between}\, e)^{\lambda} + \lambda(\lambda - 1)(\mu + \sigma_{between}\, e)^{\lambda - 2}\, \sigma^2_{within}/2$

   with q = $1/\lambda$, the power to approximately normality, e standard normal distributed N(0, 1) (Dodd *et al*. 2006, p1646).

3. Draw 1 value of the Beta distribution
4. Multiply the value of Step 2. with the value of Step 3.

The result is one random draw from the distribution of usual intakes.
Repeat Steps 1 till 4 a great number of times, say 50000.
Back to: Chronic risk analysis

5.4.5 Covariable and/or cofactor in the models

The intake frequency and transformed intake amount model can be extended to describe the effect of a covariable and/or cofactor. Then, usual intakes are dependent on explanatory variables.

5.4.5.1 Cofactor

5.4.5.1.1 Intake frequency model

The model is:

$$\text{logit}(\underline{\pi}) = \beta_{0l},\ l=1\ldots L$$

where L is the number of levels of the cofactor and $\hat{\pi}$ is the fitted probability of 5.4.2 .

5.4.5.1.2 Intake amount model

The model is:

$$\text{transf}(\underline{y}_{ij}) = \beta_{0l} + \underline{c}_i + \underline{u}_{ij},\ l=1\ldots L$$

with notation as in 5.4.3 and 5.4.5.1.1

5.4.5.1.3 Usual intake model
The usual intake is calculated per level of the cofactor as described in 5.4.4

5.4.5.2 Covariable

5.4.5.2.1 Intake frequency model

The model is:

$$\text{logit}(\underline{\pi}) = \beta_0 + \beta_1\, f(x_1;\, df),$$

where x_1 is the covariable, f is a spline or polynomial function and *df* the degrees of freedom. The degree of the function is determined by backward or forward selection.

5.4.5.2.2 Intake amount model

The model is:

$$\text{transf}(\underline{y}_{ij}) = \beta_0 + \beta_1\, f(x_1;\, df) + \underline{c}_i + \underline{u}_{ij},$$

with notation as in 5.4.3 and 5.4.5.2.1

5.4.5.2.3 Usual intake model
The usual intake is calculated for user specified values of the covariable.

5.4.5.3 Covariable and cofactor

5.4.5.3.1 Intake frequency model

The interaction model is:

$$\text{logit}(\underline{\pi}) = \beta_{0l} + \beta_{1l}\, f(x_1;\, df), \qquad l=1\ldots L$$

The main effect model is:

$$\text{logit}(\underline{\pi}) = \beta_{0l} + \beta_1\, f(x_1;\, df), \qquad l=1\ldots L$$

where x_1 is the covariable, f is a spline or polynomial function and df the degrees of freedom. The degree of the function is determined by backward or forward selection.

5.4.5.3.2 Intake amount model

The interaction model is:

$$\text{transf}(\underline{y_{ij}}) = \beta_{0l} + \beta_{1l}\, f(x_1; df) + \underline{c_i} + \underline{u_{ij}}, \qquad l=1\ldots L$$

The main effect model is:

$$\text{transf}(\underline{y_{ij}}) = \beta_{0l} + \beta_1\, f(x_1; df) + \underline{c_i} + \underline{u_{ij}}, \qquad l=1\ldots L$$

with notation as in 5.4.3 and 5.4.5.3.1

5.4.5.3.3 Usual intake model
The usual intake is calculated for all levels of the cofactor for a number of values of the covariable.
Back to: Chronic risk analysis

5.4.6 Discrete/semi-parametric (ISUF)

Nusser *et al.* (1996) describe how to assess chronic risks for data sets with positive intakes (a small fraction of zero intakes was allowed, but then replaced by a small positive value). The 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 intake from data sets with a substantial amount of zero intakes became feasible by modelling separately zero intake on part or all of the days via the estimation of intake probabilities as detailed in Nusser *et al.* (1997) and Dodd (1996). In MCRA, a discrete/semi-parametric model is implemented allowing for zero intake and heterogeneity of variance following the basic ideas of Nusser *et al.* (1996, 1997) and Dodd (1996).

Nusser *et al.* (1996, 1997) describe a procedure for the assessment of chronic risks using non-normal dietary intake data. Principally, their method consists of four steps:

1. transforming the daily intake data to approximate normality using a power function or log transformation
2. fitting a grafted polynomial function to the power or log transformed daily intakes. The polynomial provides some flexibility against power transformed components that are still deviating from normality,
3. estimating the parameters of the usual intake distribution in the transformed scale, and
1. estimating the percentiles of the distribution of usual intakes in the original scale.

Back to: Discrete/semi-parametric (ISUF)

5.4.6.1 Power or log transformation

To achieve a better normality, daily intakes are transformed. The user can choose a logarithmic transformation ($t = f(y) = \ln(y)$, no parameters to be estimated) or a power transformation ($t = f(y) = y^{\gamma}$, one parameter to be estimated). In the latter case the optimal power is determined on the grid $\left\{10, 2, 1, \frac{1}{1.5}, \frac{1}{2}, \frac{1}{2.5}, \frac{1}{3}, \frac{1}{3.5}, \ldots, \frac{1}{100}\right\}$, with a further refinement grid search around the best fitting value. If a power $\frac{1}{100}$ gives the best fit in this grid search, then the logarithmic transformation is selected (Note that a logarithmic transform corresponds theoretically to $\gamma = 0$). The goodness of fit is determined by minimising the residual sum of squares: $(z(i) - \beta_1 y^{\gamma})^2$ of a regression of normal Blom scores on the power-transformed daily intakes. Normal Blom scores are (Tukey 1962):

$$z_{(i)} = \Phi^{-1}\left(\frac{i - \frac{3}{8}}{n + \frac{1}{4}}\right)$$

where i is the rank of the nth non-zero daily intake, n, the total number of non-zero intakes and $\Phi^{-1}(\cdot)$ is the inverse of the standard normal cumulative distribution function.

Back to: Discrete/semi-parametric (ISUF)

5.4.6.2 Spline fit

To achieve a better normality, a second transformation (optional) is performed: a spline function $t = g(z)$ is fitted to the logarithmically or power transformed data t as a function of the normal Blom scores. The spline function is a grafted polynomial consisting of cubic polynomials between p = 3 joint points (knots) and linear functions in the two outer regions. The daily intakes are transformed by interpolating from t to $x = g^{-1}(t)$, using the fitted spline function.

After a successful transformation the daily intakes x will resemble Blom normal scores and their mean and total variance will therefore be approximately 0 and 1. The normality of the transformed values x is checked with the Anderson-Darling test. In the case of a spline transformation, if normality is rejected at the 85% confidence level, then the number of knots p is increased and the spline fit is repeated (until a maximum of 22 knots).

Back to: Discrete/semi-parametric (ISUF)

5.4.6.3 Estimation of the parameters of the usual intake distribution

Variance components for between and within-consumer information are fitted to the transformed non-zero daily intakes x using the model:

$$x_{ij} = x_i + u_{ij}$$

$$x_i \sim N(\mu, \sigma_B^2); \quad u_{ij} \sim N(0, s_i^2); \quad E(s_i^2) = \sigma_0^2; \quad \text{var}(s_i^2) = \sigma_A^2$$

In this model the total variance of the daily intakes is divided into a between-consumer component and a within-consumer component. The within-consumer variance component can be heterogeneous, that is, it can be different for different consumers. In the model the between-consumer variance σ_1^2 and the mean and the variance of the within-consumer variance component distribution (σ_B^2 and σ_A^2) are estimated using standard statistical methods (ANOVA). Further, a test statistic *MA4* is calculated to test whether the heterogeneity of variances is significant (see Dodd 1996 for details).

The estimate s_B^2 of the between-consumer variance is the basis for the estimation of the distribution of usual intake. The distribution of usual intakes on non-zero intake days in the x scale is represented by a set of 400 normal Blom scores (which themselves represent the standard normal distribution) multiplied by s_1: $x_i = s_B z_{(i)}$. The same calculation is applied to user-requested percentiles $z_p = \Phi^{-1}(p)$.

Back to: Discrete/semi-parametric (ISUF)

5.4.6.4 Back transformation and estimation of usual intake

The 400+ values x_i are back-transformed to the original scale. This is simple if no spline function has been estimated. If a spline function has been used, then it is a rather complicated procedure, because the spline function g was developed for daily intakes, not usual intakes. The following steps are made:

1. First the 400+ values x_i are expanded in a set of 9 * 400 values representing the distribution of daily intakes around each of the 400 points;
2. These 9 * 400+ values are back transformed using the functions g and f, and the sets of 9 values are then recombined (by weighted averaging) into 400 usual intake values y_i;

3. A spline function g_1 ,especially adapted for usual intakes, is now fitted to the 400 data pairs (x_i, t_i), where $t_i = f(y_i)$;
4. Finally the usual intakes on non-zero intake days are represented by the back-transform using this improved function: $y_i = f(g_1(x_i))$.

The user-requested percentiles y_p are the additional values (i > 400) in the 400+ set. The 400 y_i values define the cumulative distribution function by:

$$F(y_i) = \frac{i - \frac{3}{8}}{400 + \frac{1}{4}}.$$

The distribution is adapted in order to account for days with zero intake of consumers (defined here as consumers who have a positive probability of intake on any day, and therefore a non-zero usual intake). This is done by estimating the distribution of individual intake probabilities. This distribution is approximated via a number of classes (*e.g.* 21 or 51, can be selected by the user) arranged by the proportion of days on which there is a positive intake (p_m). Using a binomial distribution for each class, the fraction of consumers in each class (θ_m; $m = 0,...,M$) is estimated by optimising the fit of the predicted proportions of consumers with 0, 1, 2, ... intake days to the observed proportions. The number of parameters to be estimated is usually higher than the number of possible outcomes for a consumer (*e.g.* 3 when there are two days per consumer), and therefore a smooth approximation is made using a modified minimum chi-squared estimator. See Dodd (1996) for details. Only the fraction of non-consumers (θ_0) is estimated separately with no restriction to be similar to the other θ_m. It can be noted that the distribution of individual intake probabilities can be better estimated when the number of days per consumer in the consumption survey becomes higher. With only 2 days per consumer the procedure gives a rather artificial distribution, often with an estimated θ_0 of zero This step can be time-consuming. Therefore, the number of iterations in the estimation procedure can be limited by the user. In our experience it is not generally necessary to use 50,000 iterations as in Dodd (1996).

The estimated distribution of individual intake probabilities ($\hat{\theta}_0,...,\hat{\theta}_M$) is used to transform the distribution of usual intake on non-zero intake days (F_y) to the distribution of usual intake for consumers (F_C) and finally to the distribution of usual intake for the entire population (F_U). These transformations are based on the relation:

$$F_U(u) = \theta_0 + \sum_{m=1}^{M} \theta_m F_y(u/p_m)$$

which basically says that to obtain a certain level of usual intake u we should consider a different level (u/p_m) for the class of consumers which consume only on a fraction p_m of days. See Dodd (1996) for details of the computational procedure. Linear interpolation based on the 400 values of the F_y distribution is then used to compute representations of the cumulative distribution functions for consumers only and the entire population.

Back to: Discrete/semi-parametric (ISUF)

5.5 Uncertainty analysis: resampling data sets and resampling from distributions

In probabilistic risk assessment of dietary intake we use distributions which describe the *variability* in consumption within a given population of consumers and the *variability* of the occurrence and level of chemical compounds on the consumed foods. However, these calculations do not consider the amount of *uncertainty* that is due to the limited size of the underlying datasets. Typically, in a large number of simulations very many different combinations of consumption and compound concentrations are made. This leads to a smooth distribution of simulated intakes, and the impression of a very precise

estimation of intake percentiles or other quantities of interest. It is essential to realise that the accuracy of the inference depends on the accuracy of the basic data.

When doing an uncertainty analysis in MCRA a number of iterations is chosen, and in each iteration new inputs are resampled for a complete Monte Carlo analysis:

1. Datasets (concentration data, individual consumption data) are resampled from the original database (bootstrap methodology)
2. Parametric inputs, such as processing factors and their variabilities are resampled from parametric distributions.

5.5.1 Resampling datasets

A computer-based instrument to assess the reliability of outcomes is the *bootstrap* (Efron 1979, Efron & Tibshirani 1993). 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 mean, the standard deviation, the 95th 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.

In MC-modelling of acute risks two datasets are combined: consumption data and compound concentration data. It makes sense to apply resampling to both datasets separately, in order to characterise the uncertainty in the final intake. In MCRA the uncertainty algorithm (when selected) is applied to:

1. the multivariate consumption patterns and associated body weights: actually the data set of consumer identifiers is resampled, and all consumer information (consumption patterns for all consumption days, body weight, and age) is coupled to the selected consumer identifiers.
2. the univariate compound concentration data sets: these are resampled independently for all foods. In principle, the uncertainty algorithm is applied to the dataset consisting of both non-detects and positive values; in practice, for a dataset with n_0 non-detects and n_1 positive values, the number of positive values in a resampled set is obtained as a draw from a binomial distribution with parameter $n_1/(n_0+n_1)$ and binomial total n_0+n_1. Then, this number of values is selected randomly from the set of n_1 positive values.

In MCRA the resulting uncertainty distribution of percentiles of the intake distribution is summarised by specifying empirical 2.5th , 25th, 75th and 97.5th percentiles. The outer percentiles constitute a central 95% confidence interval for the variability percentiles. However, for this it is necessary that the number of resampled sets B is high enough. The number of resampled sets should be chosen depending on the confidence level wanted for the uncertainty interval. Typically 500-2000 resampled sets will be reasonable for a 95 % confidence interval (Efron & Tibshirani 1993, pp. 14-15, 275).

The same uncertainty algorithm can also be applied to deterministic estimates which are calculated from data sets. For example the maximum concentration found in a resampled set will be different, if the actual maximum value in the original dataset has *not* been selected. Also data-based estimates of large portion and average body weight will vary.

5.5.2 Resampling parametric distributions

In MCRA 6 this is implemented for processing factors and the variability of processing factors.

Processing effects are modelled either by a fixed processing factor, or by a lognormal or logistic-normal distribution (depending on the distribution type set in table Processingfactor).
In the former case (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 (procnom) and an estimate of p95 of the uncertainty distribution (procnomuncupp). The calculation is σ_{unc} ={logit(procnomuncupp) – logit(*procnom*)}/1.645 in case of disttype 1 (logistic-normal distribution) and σ_{unc} ={ln(procnomuncupp) – ln(*procnom*)}/1.645 in case of disttype 2 (lognormal distribution). Values lower than 0.01 or higher than 0.99 (disttype 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 the case of a processing factor distribution (describing the variability of processing factors) two uncertainties can be specified. First, the uncertainty about the central value μ can be specified as before using a parameter procnomuncupp. Secondly, the uncertainty about the variability standard deviation σ_{var} can be specified by the number of degrees of freedom *df* of a modified chi-square distribution which is used to generate new values of σ_{var}. Setting *df* very high means litte uncertainty, and σ_{var} will be almost equal in all iterations of the uncertainty analysis. Setting *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.

Back to: Uncertainty analysis

5.6 Binning

Binning is a method to summarise the simulated data (total intake, intake per food, consumption per food, concentration per food) in frequency intervals for further use in deriving the intake distributions. The alternative would be to store observations for subsequent use, but this would require for moderate simulations already a large amount of storage capacity and an excessive administration.
The mean value (*mean*) of the first, say 5%, of the total number of simulations is used to define the left limit of the central bin. For values above the mean, 1100 bins are used for storage. The upper limits of the upper bins are defined as 1% higher than the lower limit. So, for upper bin *i* the upper limit is calculated as *mean* x 1.01^{i}. For values below the mean also 1100 bins are defined with lower limits defined by *mean* x 1.01^{-i}. After the process of binning is completed, the quantile value of a specific percentile is determined by linear interpolation between the bin limits. These 2200 bins together provide efficient storage for numbers spanning more than 9 decades ($1.01^{2200}=3.2\text{x}10^{9}$), which should be amply sufficient for most practical problems.
To get accurate results, it is rather important that the mean value of the first 5% of the simulations represents, approximately, the true mean of the sampled distribution. Therefore, the total number of simulations should not be chosen too small. During the simulation, if a simulated intake is higher than the upper limit, this value is replaced by the new maximum.
When the mean value is missing, *e.g.* due to zero intakes, the program resorts to an average *mean* value, *e.g.* the average of the mean values of foods with nonzero intakes.

6 Selection of consumers and foods

In the next sections, the centrally supplied database 'validation.mdb' is used. All tables are selected and data for survey DNFCS-3 and compound CHLORPYRIFOS are retrieved from the database. Find in Figure 49 the outcome of the selection. In total, 1230 foods are eaten. Application of the conversion algorithm shows that the number of derived foods is 33 + 43 = 76. Note that the selected foods do not depend on the choice of the compound. The retrieved foods are entirely determined by consumption patterns, *e.g.* the choice of a survey.

The information in the lower part of the screen is related with the choice of the compound. Here, chlorpyrifos is selected and this compound is found on 33 derived foods. This means that there is at least one sample of each of the 33 foods with a positive concentration. On 43 derived foods only non-detects are found. This means that for each of the 43 foods all samples were negative, i.e. no chlorpyrifos found. No worstcase values were found in the database.

Check one of the two radio buttons to specify which foods must be used in the risk analysis. If non-detects are not replaced by the LOR (see input form), the first and second radio button in Figure 49 are equivalent. Note that the first button is computationally more efficient because less foods are involved.

Press the 'go'-button in Figure 49 to enter the central menu.

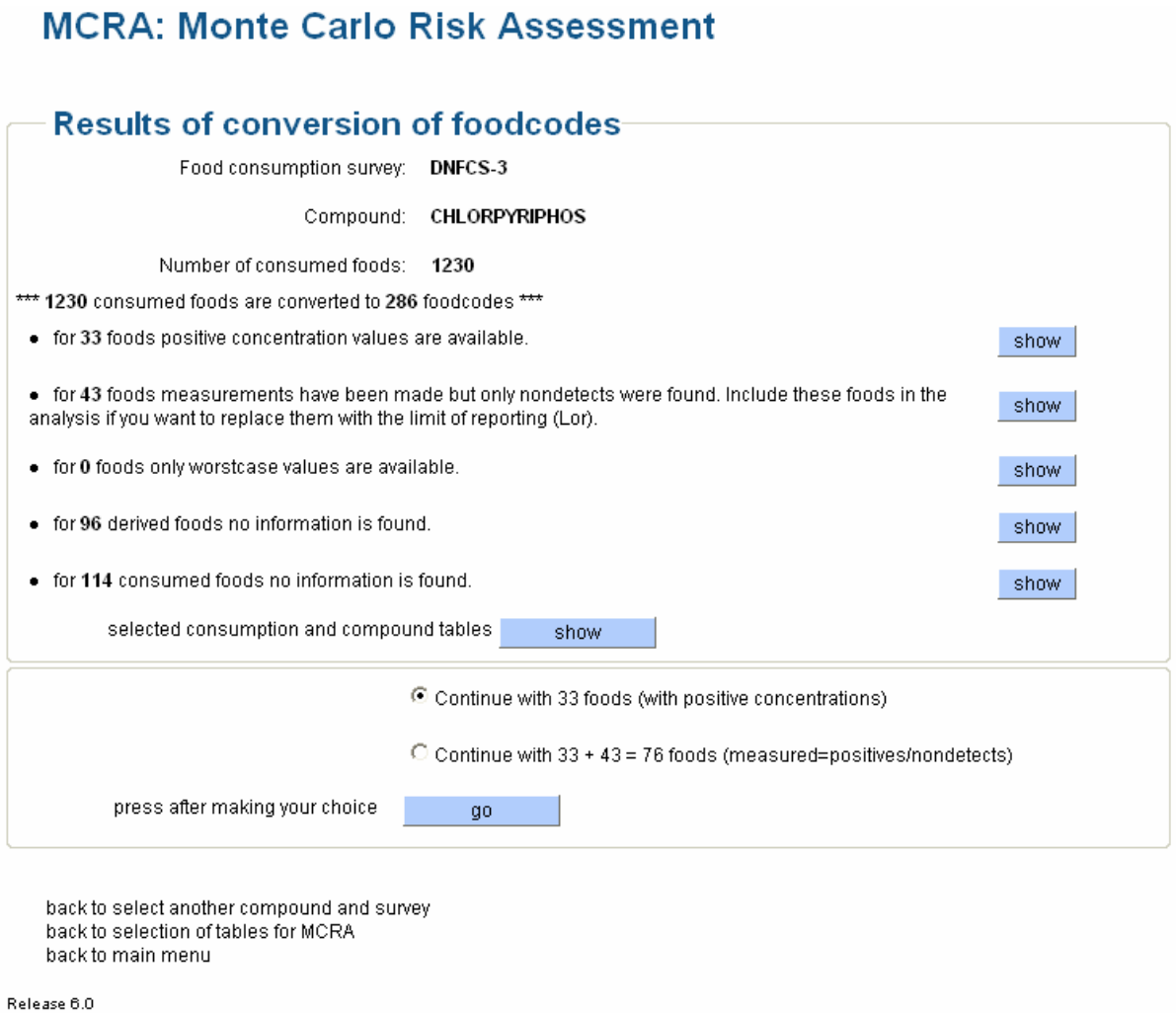


Figure 49: Selected foods for survey DNFCS-3 and compound CHLORPYRIFOS

With the subset-selection 'go'-button in the MCRA central menu (see Figure 50) subset selections are made on the consumer population or on the foods. Through the use of scroll-down menus the user 1)

selects ages, weights and sexes (and covariable and cofactor if chosen) to define the population of consumers; 2) selects day(s) of consumption, foods, year, country and sampling type of the compound concentration data. In Figure 50 the subset-selection 'show'-button shows all selected levels for those variables on which selections are active. The corresponding 'clear'-button removes all selections already made.

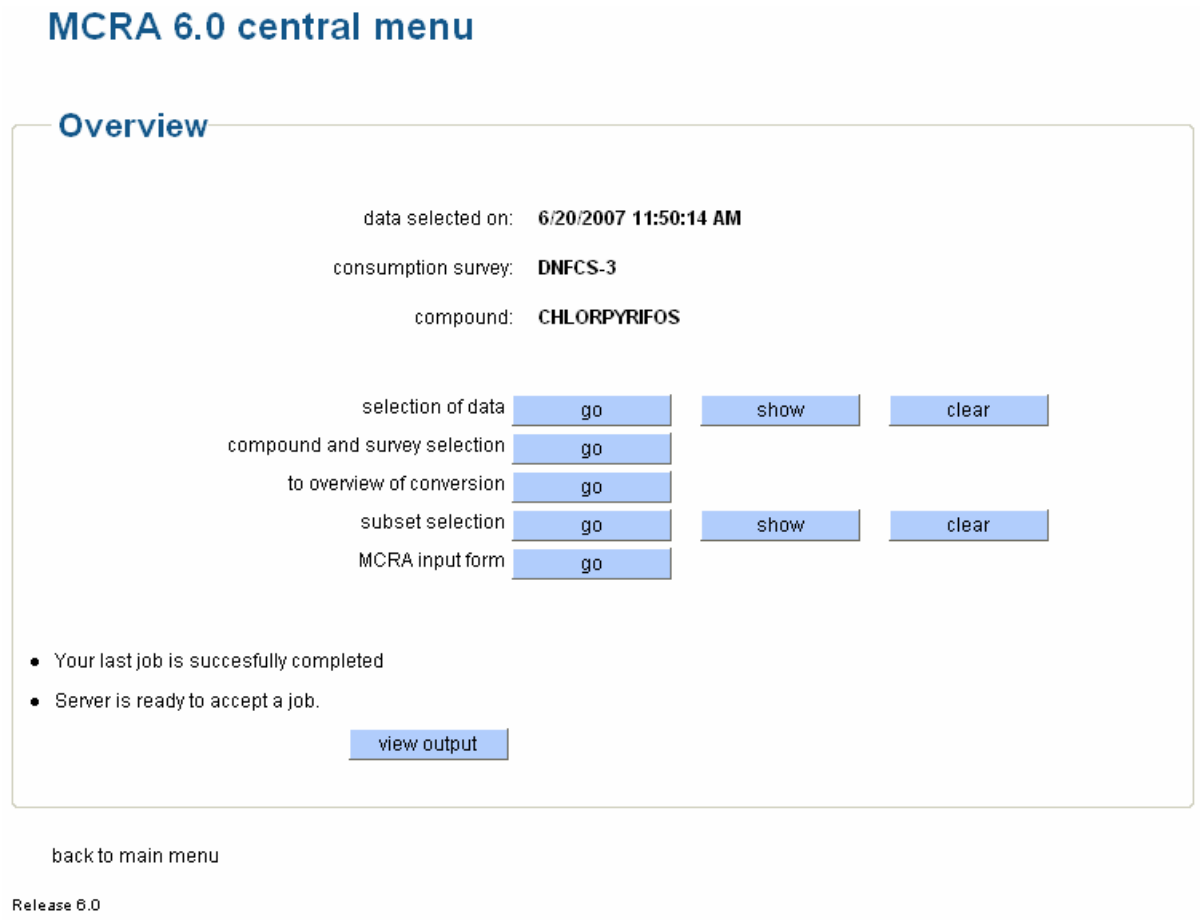


Figure 50: MCRA central menu, subset selection

If you want to do the same type of analysis for one of the other options, click the to-overview-of-conversion 'go'-button in the centralmenu.

6.1 Subset selection: consumers

After clicking the subset-selection 'go'-button in the MCRA central menu (see Figure 50) the first subset selection screen is shown (see Figure 51). Here, the consumer population is defined using characteristics on consumers (age, weight, sex, length and vegetarian). The first time, the characteristics of the total population are shown, so for age the minimum and maximum values are 1 and 97 year (both current selection as database). Note that combining levels of variables occasionally results in empty subsets *e.g.* the combination of ages within the range 1 to 8 and weights in the range 70 kg to 150 kg is an empty subset. So avoid making combinations of variables that yield empty subsets. The number of consumers currently selected is always mentioned in the screen, *e.g.* 6250 (see Figure 51).

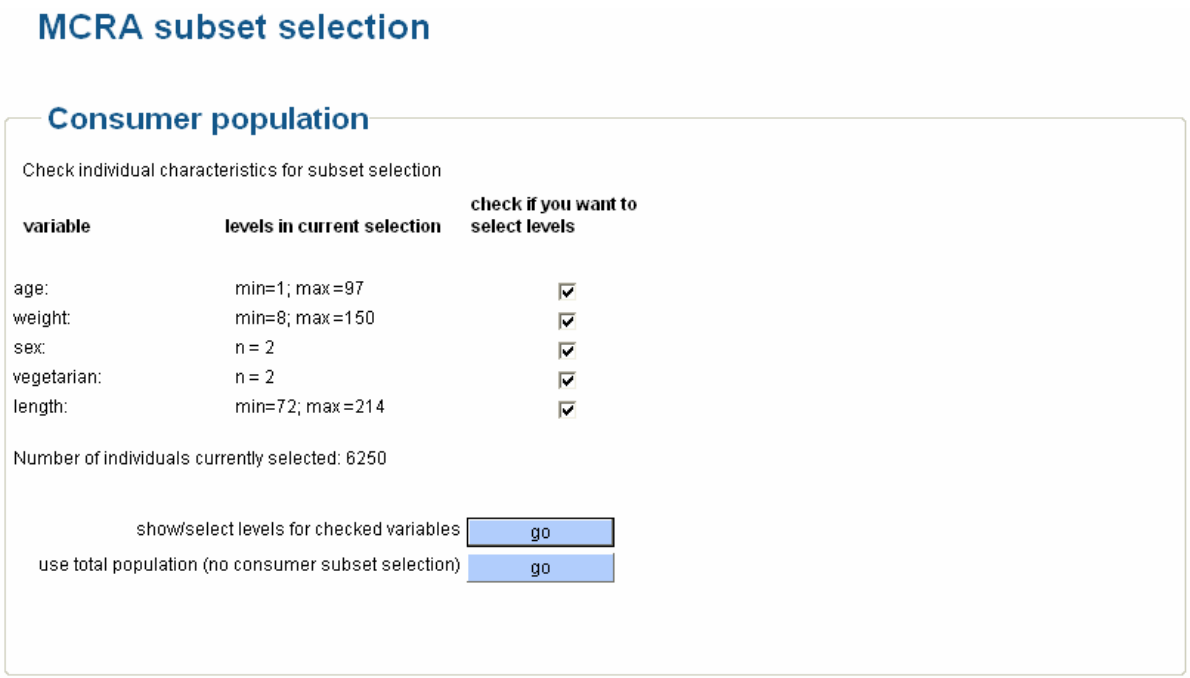


Figure 51: MCRA subset selection: consumer population

To make a selection of consumers, check one or more of the checkboxes and click the upper 'go'-button. The screen in Figure 52 appears after checking all checkboxes. Choose values for age, weight (the unit depends on the survey you are currently using, see table FoodSurvey in 9.4.3) or length and levels of sex or vegetarian. Change some levels and implement the changes by clicking the subset-selection-based-on-chosen-levels 'go'-button. The system automatically adjusts the levels of all variables.

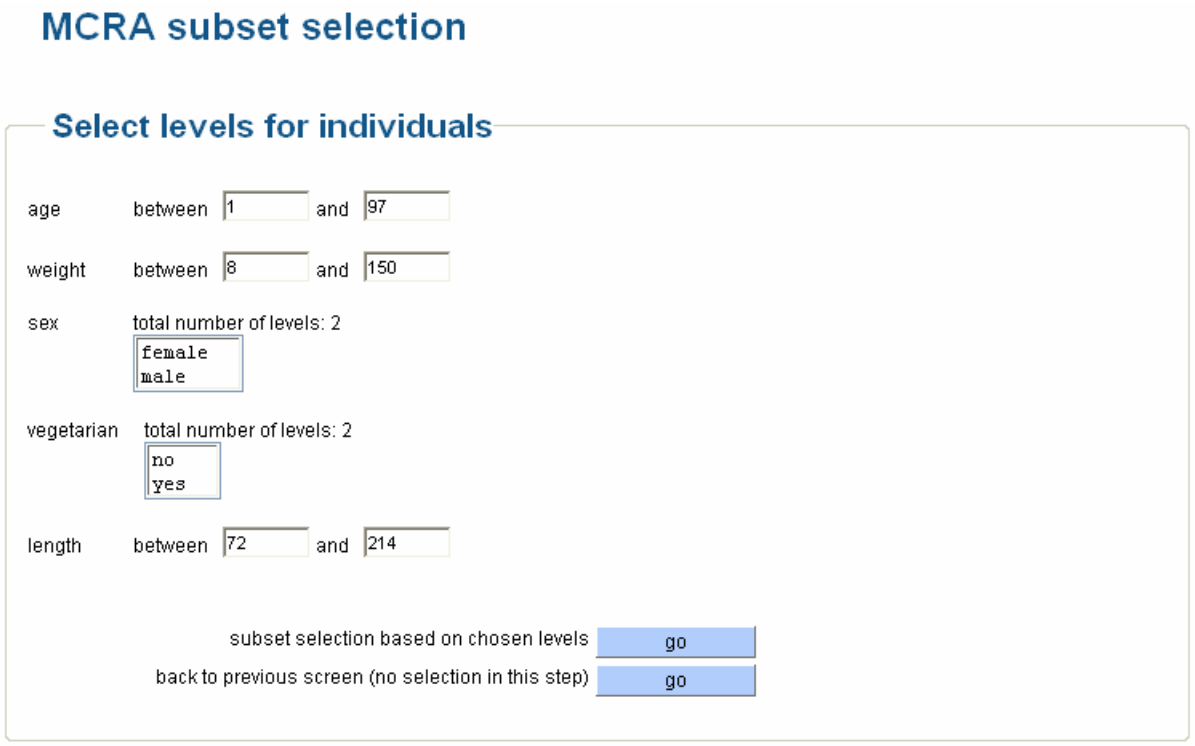


Figure 52: MCRA subset selection: select levels for consumers

For example, select a population between 1 and 6 years. The minimum and maximum weight is 8 and 30 kg, respectively. The selected number of children is 530 (screens are not shown). If you want to include previously excluded levels in a further selection step, just click the radio button labeled 'database' in the 'select from:'-column and the original levels in the database are retrieved again. Note that age and weight are related, so results may be unexpected when combinations are selected which are inconsistent. So, in the children example, if you want to return to your original population, check for both age and weight the radio button 'database', check both checkboxes and press the 'go'-button again. Then, your original population with 6250 consumers is selected again.

6.2 Subset selection: consumption and concentration data

Select the total population. The next screen shows you the current selected levels of each variable (see Figure 53). Also the levels of each variable stored in the database are shown. In our example, the number of derived foods with positive concentration values is 33 and these foods are derived from 306 consumed foods.

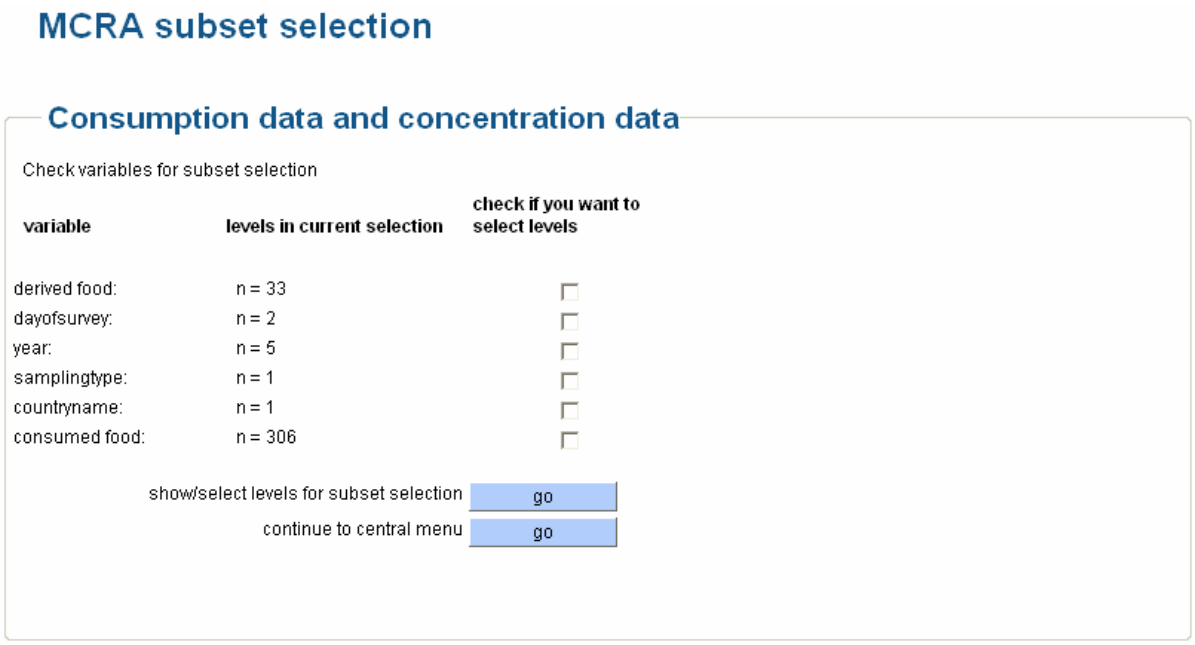


Figure 53: MCRA subset selection: consumption data and concentrations

The subset selection screen for consumption and concentration data is used to select levels for a number of variables. Variables on which subset selection can be made are:

- derived foods: to restrict the analysis to specific derived foods;
- (consumption) day: to restrict typical consumption survey data to specific days (*e.g.* only the first);
- year: to restrict the compound concentration data to specific years;
- samplingtype: to include only compound concentration data from a specific samplingtype (*i.e.* monitoring);
- country: to include only compound concentration data from foods originating from specific countries.
- consumed foods: to restrict the analysis to specific consumed foods;

In general, do not select combinations of levels which may result in empty subsets, *e.g.* grapefruits are not grown in the Netherlands. So, combination of these levels for foodname and country results in empty subsets. In this example, it is quite clear why things go wrong. Often it is less clear what you

might expect from a combination of levels, especially when ‘current selection’ is combined with ‘database’ in order to retrieve the original levels for a variable, so be aware of empty subsets.
The example is continued by checking the checkboxes for derived and consumed foods.Click the upper ‘go’button. A new screen appears with scroll-down menus for derived and consumed foods(see Figure 54) Choose the three BAMI GORENG foods in the first scroll-down menu and click the upper ‘go’-button.

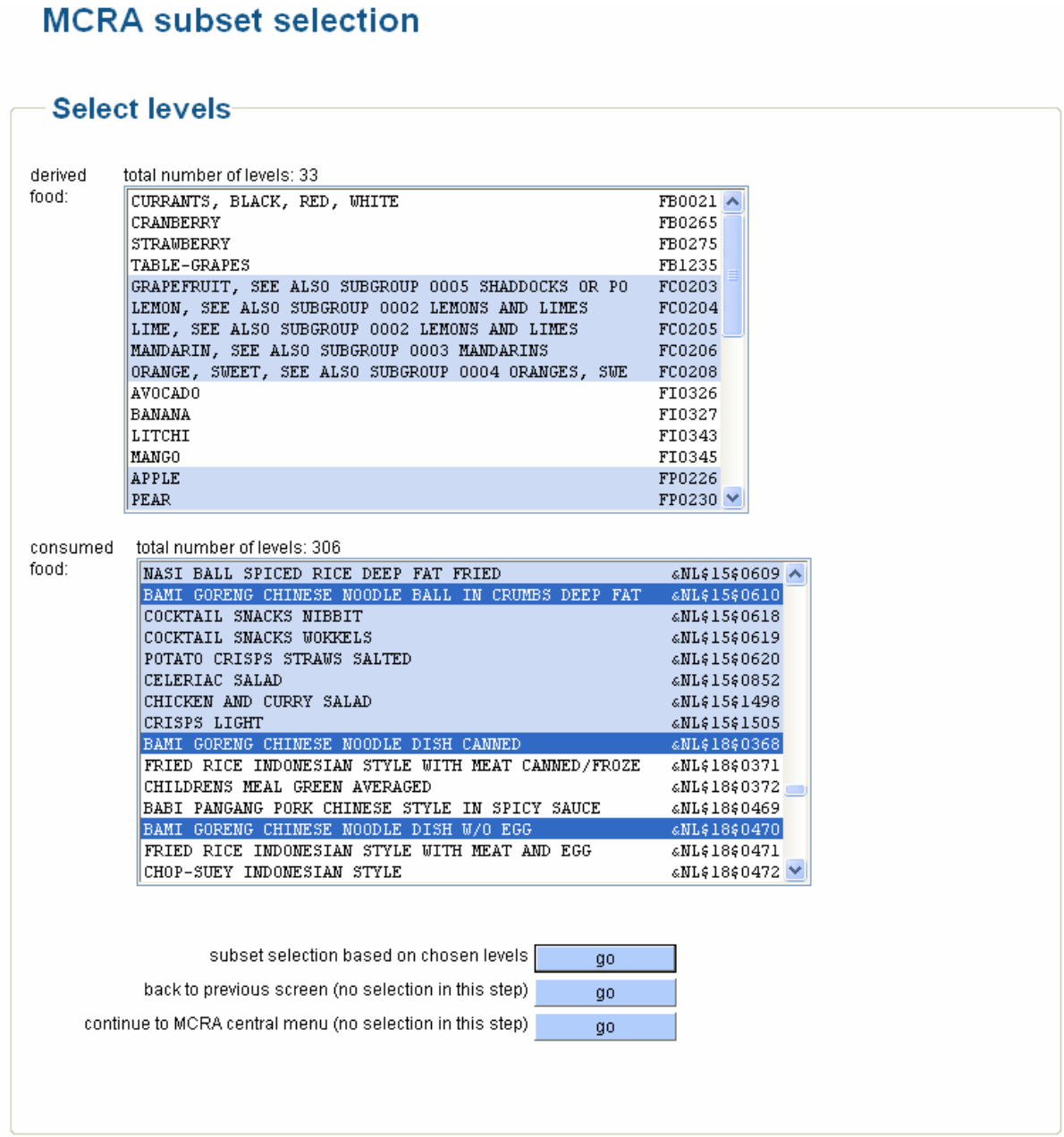


Figure 54: MCRA subset selection: select levels

The system automatically adjusts the levels of all variables: all foods that are not an ingredient of a BAMI GORENG food are removed as you can check by pressing the ‘show’-button in the central menu. Here, the steps above are repeated (check foodname checkbox and press upper button). In Figure 55, the selected foods are shown: LEMON, CELERY, ONION, SWEET PEPPER and CARROT are ingredients of BAMI GORENG foods.

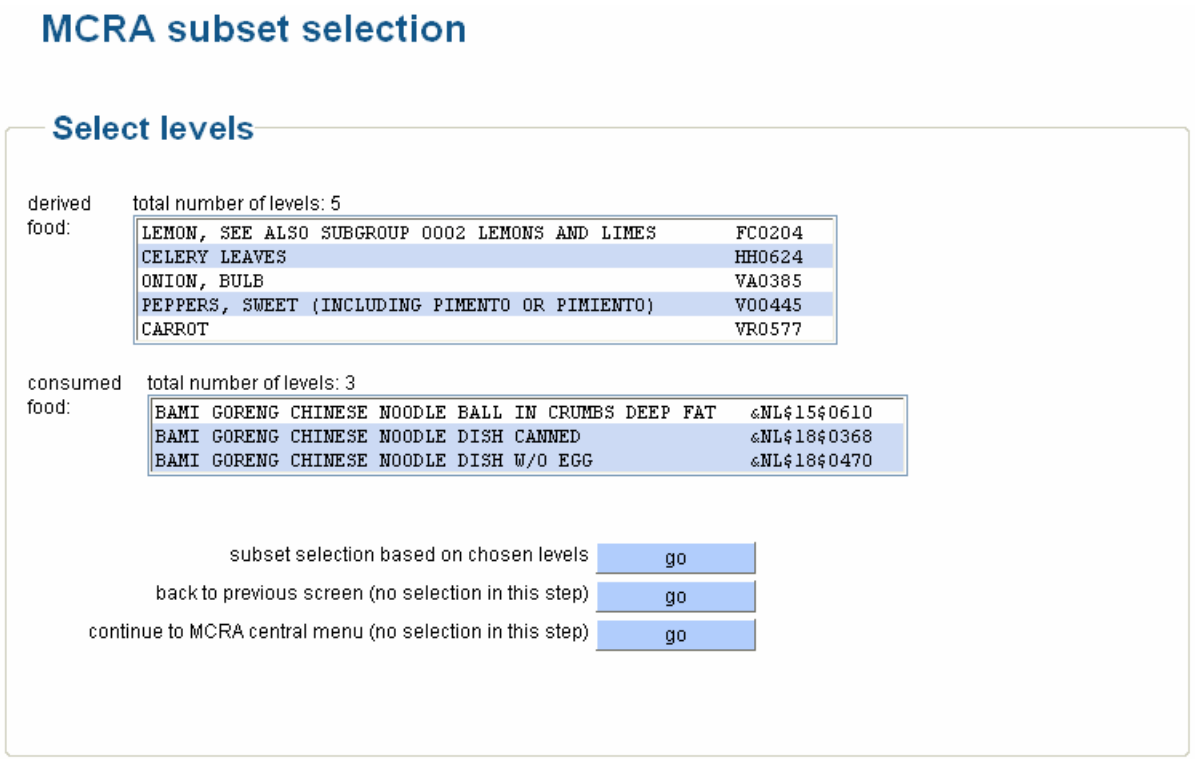


Figure 55: MCRA subset selection: select levels

7 On-line editing of data

Instead of selecting data from a database stored in the central or user directory as discussed in 2.6.1 , it is also possible to edit data on-line. Choose 'MCRA 6.0 (field trial data and Dutch consumption data)' in the MCRA main menu (see Figure 2). Now you can add and edit your own concentration data on-line. You are able to combine your own concentration data with centrally supplied food consumption data; food codes for the concentration data are automatically generated from the food consumption database.

From the central menu (see Figure 6), click the start-selection-of-consumption-and-compound-tables 'go'-button. Then the available databases are shown (see Figure 56).

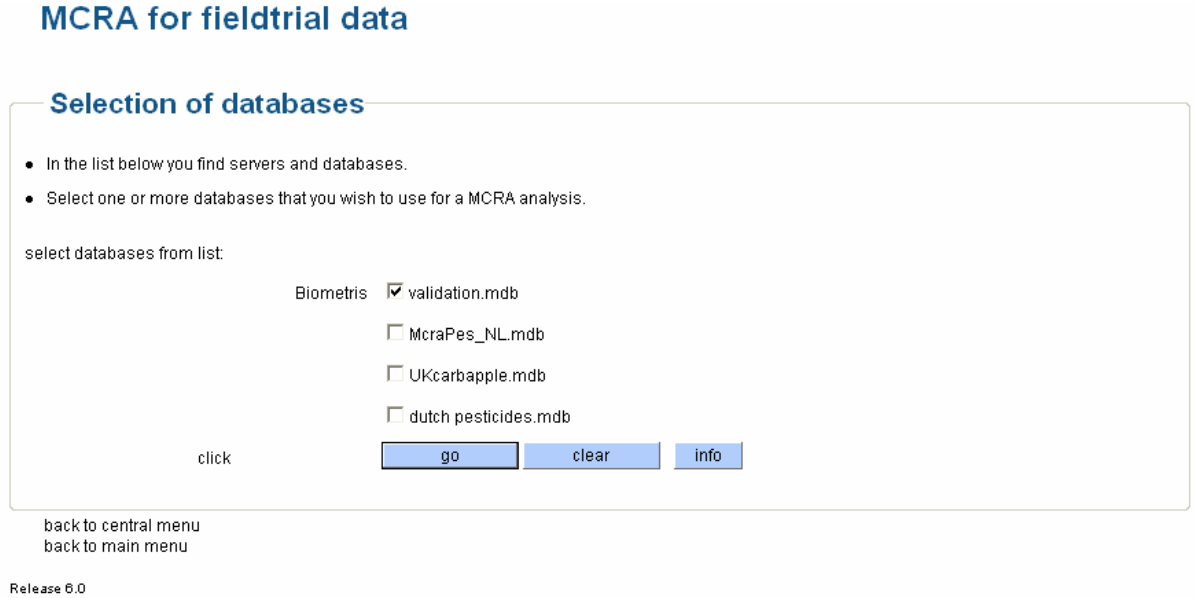


Figure 56: Database selection menu, MCRA for field trial data

Check one or more databases and click the ‘go’-button. The table selection menu appears (see Figure 57). Its layout is changed and only tables related to consumption *e.g.* Foodconsumption, Foods and Individual are shown. It is compulsory to select these three tables (see also 2.6.1). Tables related to concentration data are not being displayed. Check ‘All Tables’ and the menus for on-line adding and editing data are reached.

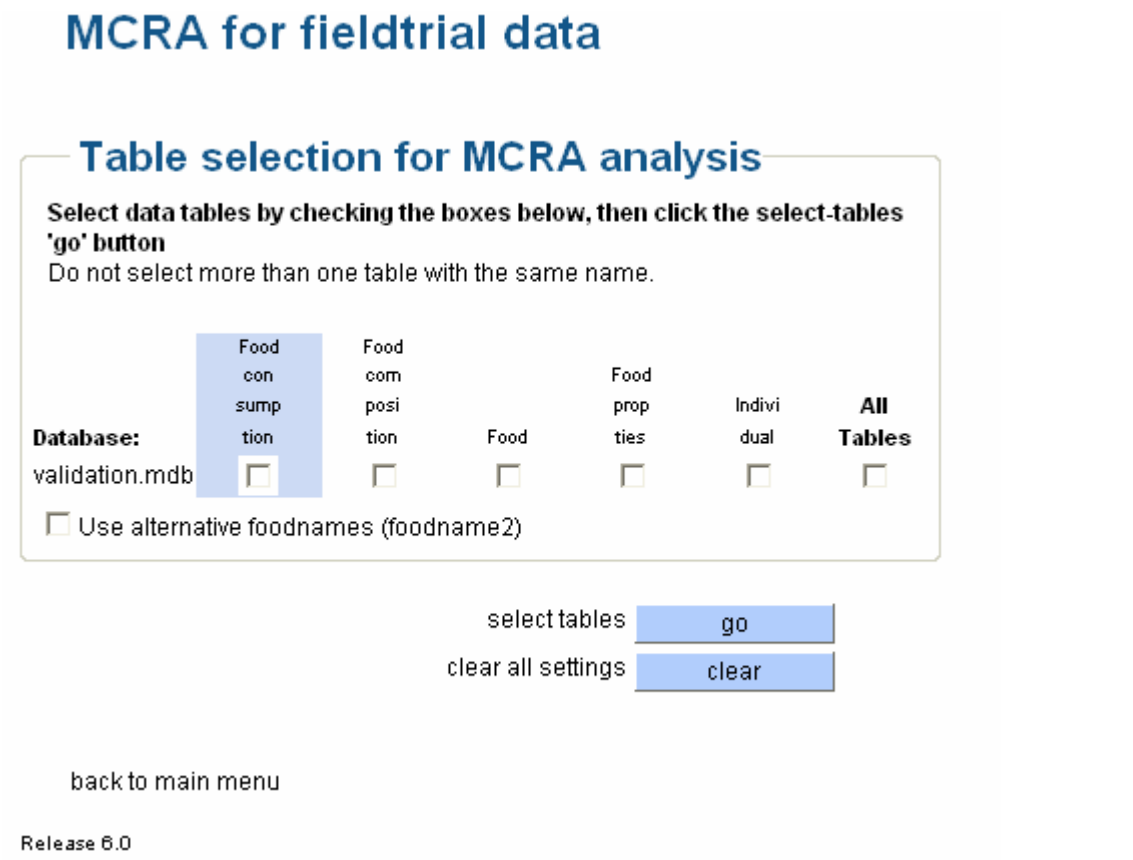


Figure 57: Table selection menu, MCRA for field trial data

The first time, create a new database (see Figure 58). In all subsequent cases, databases created on-line are shown in the scroll-down menu. Note that for a proper functioning of the editing menus only buttons and links in the window screen should be used. Do not use the 'back'-button in the toolbar during editing: this will mess up the order of the foods and a proper functioning of the menu is not guaranteed anymore.

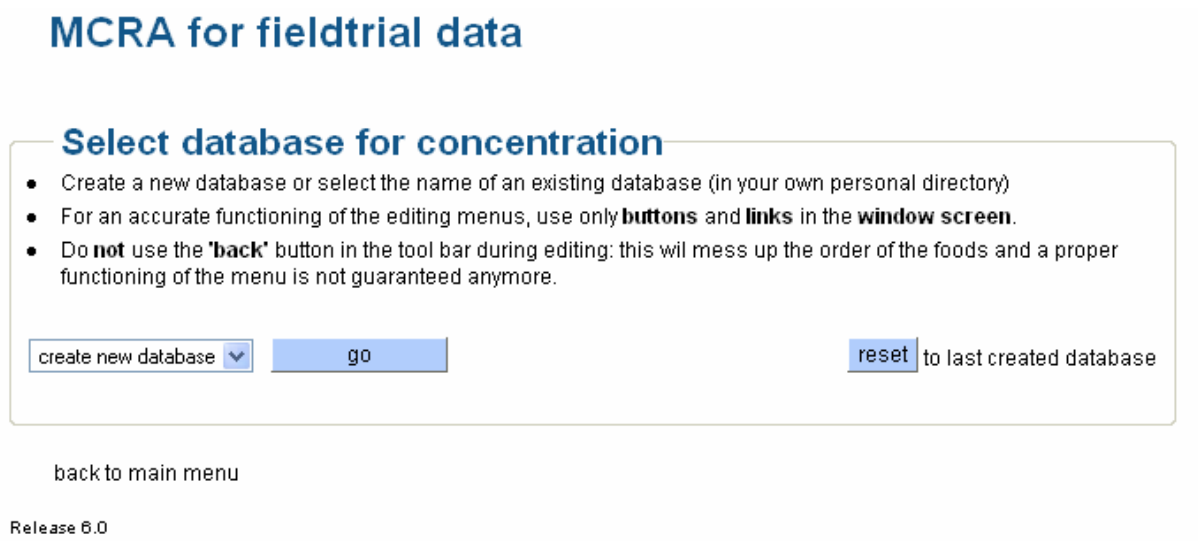


Figure 58: Select database for concentration, create new database

Specify the name of the database (see Figure 59) and press the 'go'-button. Note you should only use letters and digits, no other characters are allowed.

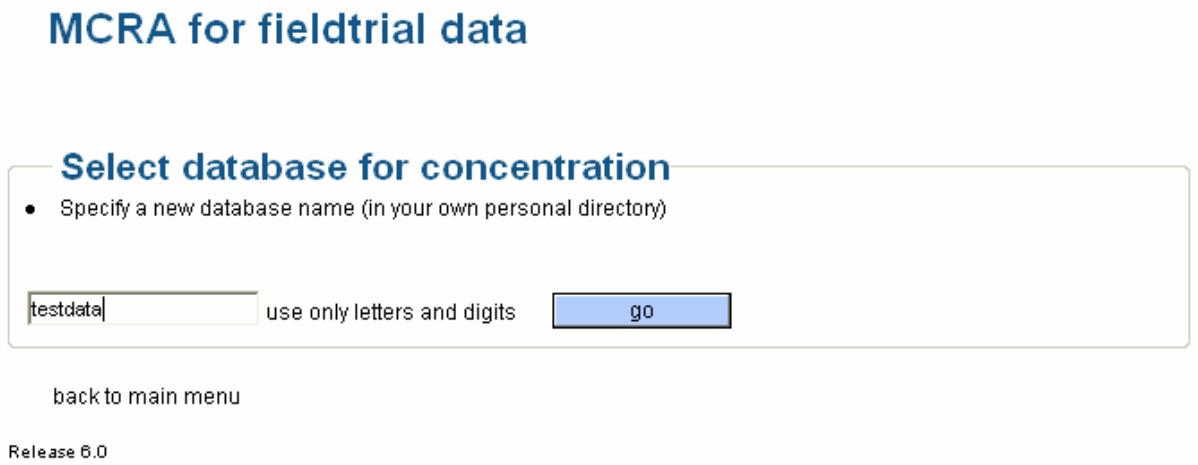


Figure 59: Select database for concentration, specify a new database name

If you have entered compound concentrations on earlier occasions, Figure 60 appears. If not the 'list or add variability factors', 'list or add processing factors' and 'list or edit ARfD or ADI'-buttons are not available.
Click the 'list-or-add-concentrations'-button and a screen displaying an add-records link and a message '*** Concentration database is empty, add records ***' is shown (only if you did not enter concentration values before).

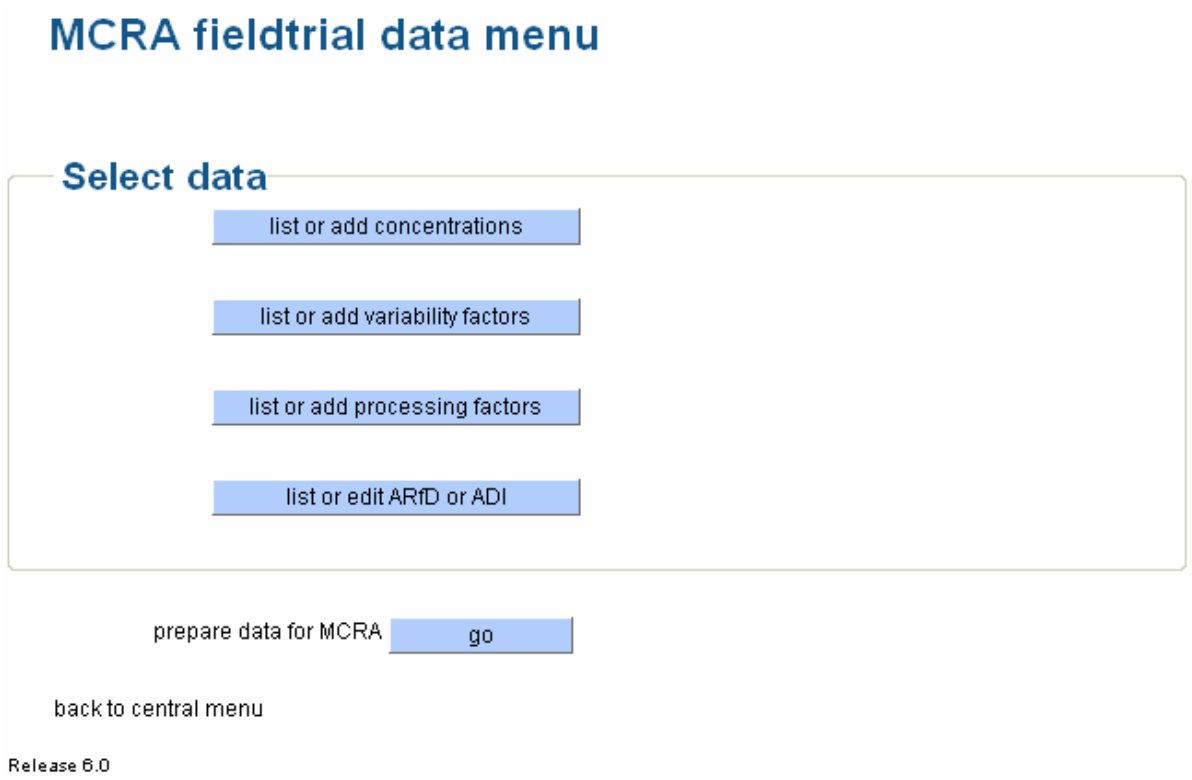


Figure 60: Select data

7.1 List or add concentrations

Click the 'list or add concentrations'-button to add compound concentration values. Specify the name of a compound (see Figure 61) if you are entering values for the first time.

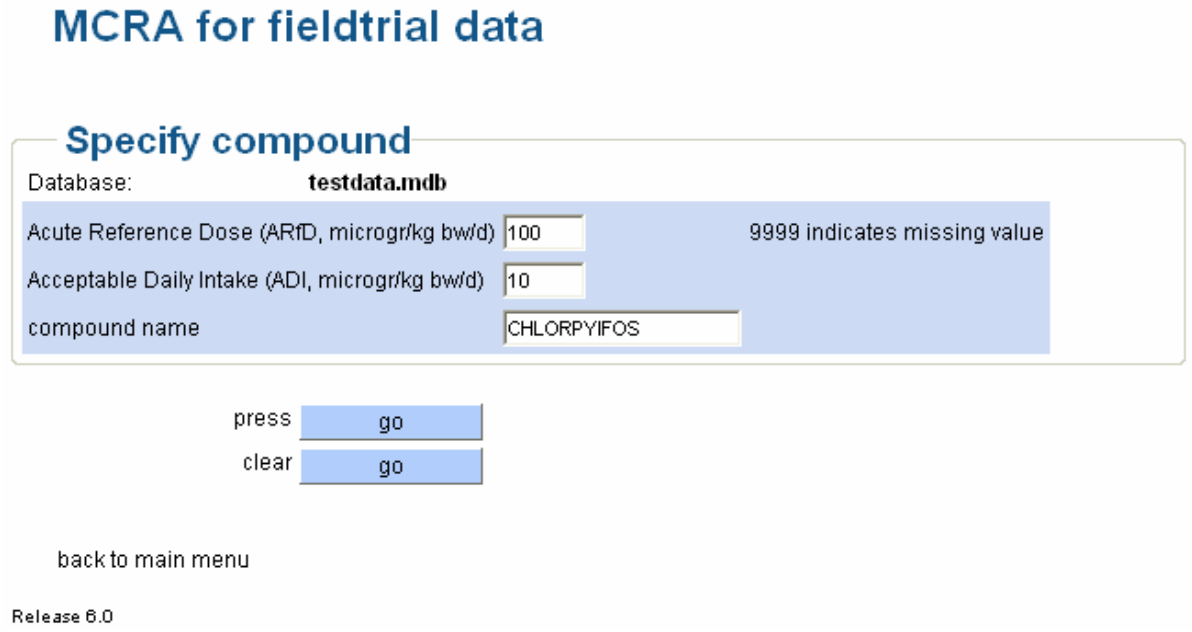


Figure 61: Specify compound name

The menu for specifying a compound name is skipped when the database already contains compound concentration values. Select a food in the scroll-down menu, add compound concentration values (see Figure 62) and click the enter-food 'go'-button.

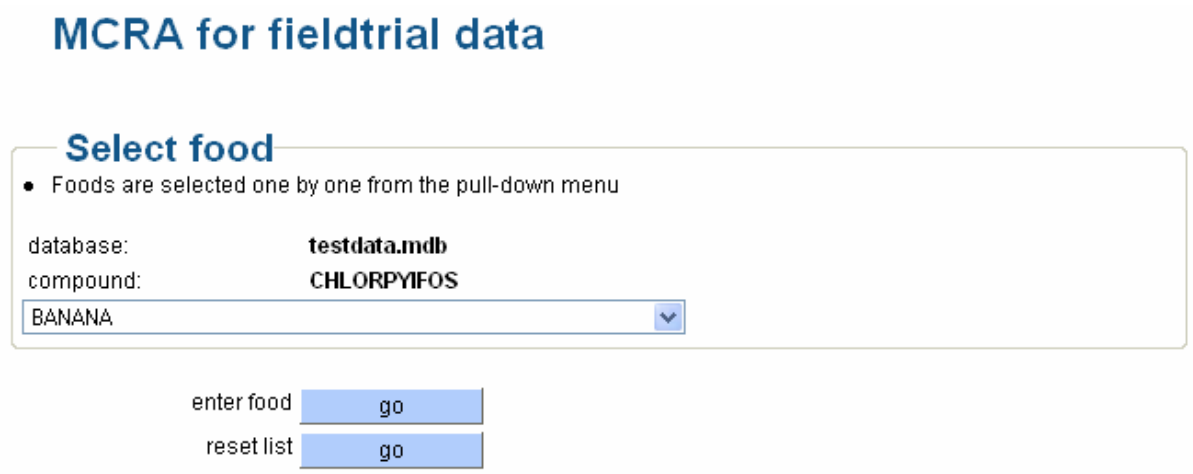


Figure 62: Select food

In Figure 63, for food Banana and compound Chlorpyrifos, the number of samples and the concentration value are entered. After saving this record, a message '*** one record added ***' appears and another concentration value may be entered. Alternatively, click the select-new-food 'go'-button for entering values for a new food.

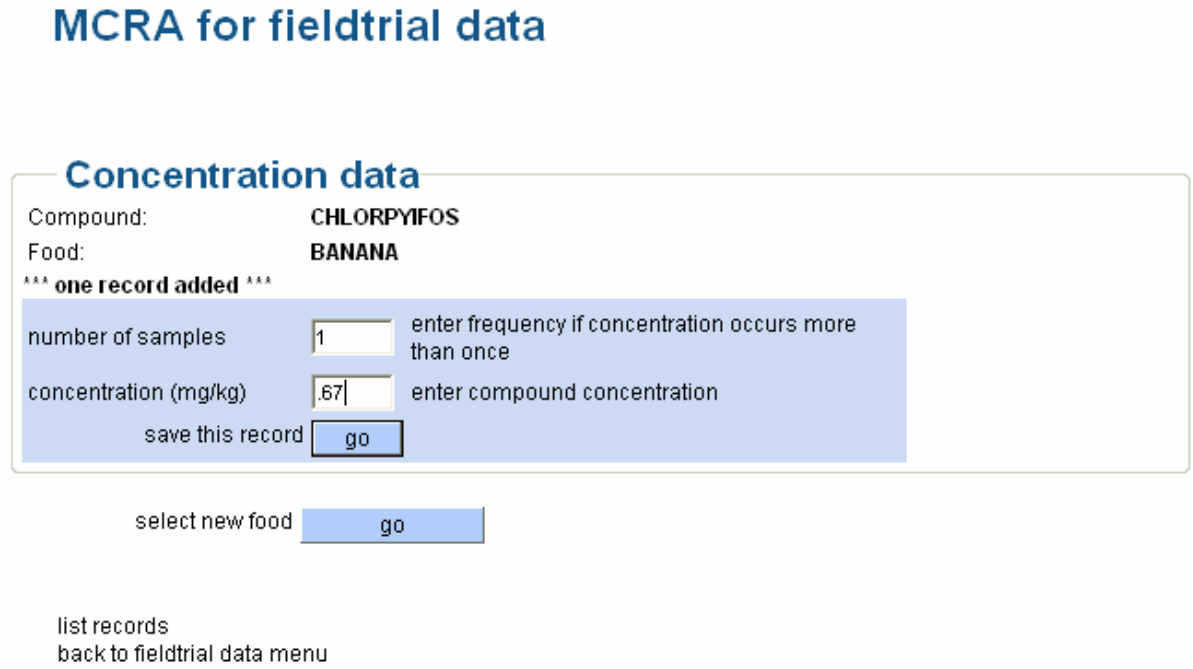


Figure 63: Enter concentration data

Click the list-records link (see Figure 63) for an overview of the compound concentration values in the database (see Figure 64).

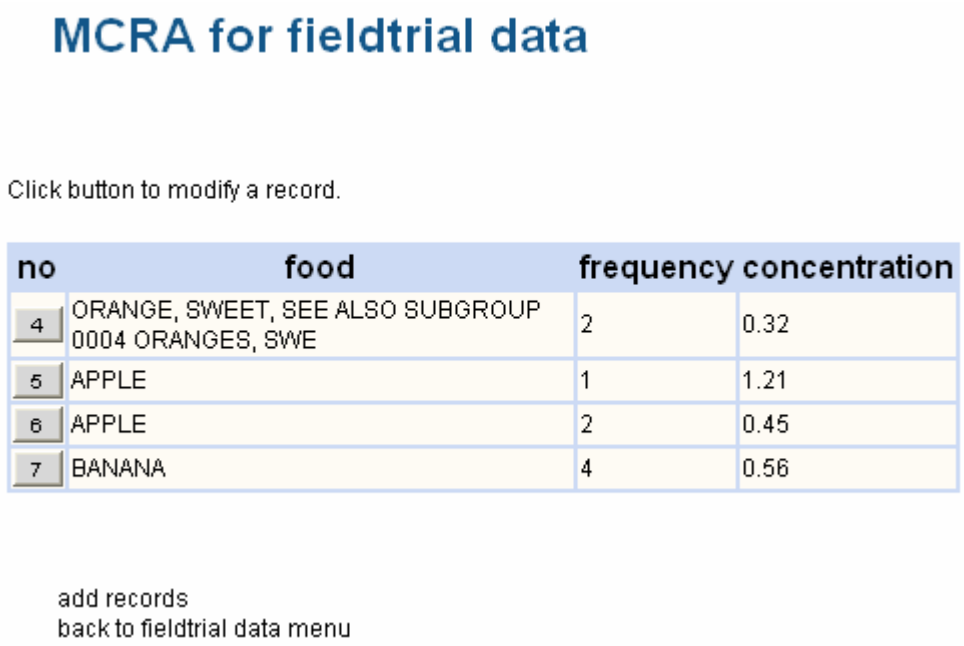


Figure 64: List concentration values

Click a button displaying the row number to retrieve a record. The value for the number of samples or the compound concentration value may be changed. Save or delete the record or choose a new food (see Figure 65).

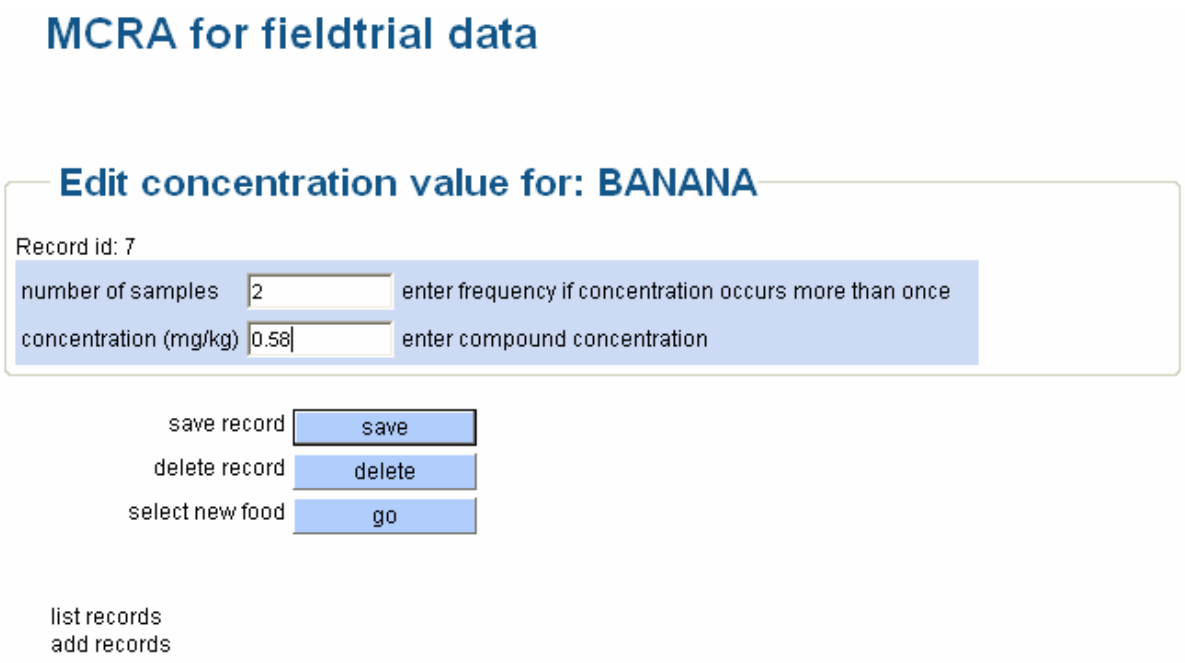


Figure 65: Edit concentration value

7.2 List or add variability factors

After supplying all compound concentrations values, click the back-to-fieldtrial-data-menu link (see Figure 63), and go back to the select data menu (see Figure 60). Now all foods are known and adding of variability factors, processing factors and ARfD/ADI values may be started. After entering all values, it still remains possible to supply new concentrations values for old and/or new foods. The

process of entering variability factors and/or processing factors for new foods may be repeated as well.
Click the 'list or add variability factors'-button. If the database does not contain any variability factors, a message '*** Variability database contains no records yet ***' is displayed. Click the add-records link and enter the menu for adding variability factors (see Figure 66). The variability factor and the number of units of the composite sample should be specified. For each food in the concentration table, values should be supplied. Only foods for which no values are supplied are displayed and the process of supplying values is repeated until all necessary values are given. Then and only then, the message '*** Variability factors are specified for all foods present in the concentration table ***' or '*** For all foods, variability factors are already specified ***' is displayed. Records may be changed afterwards by clicking one of the buttons in the listing menu.

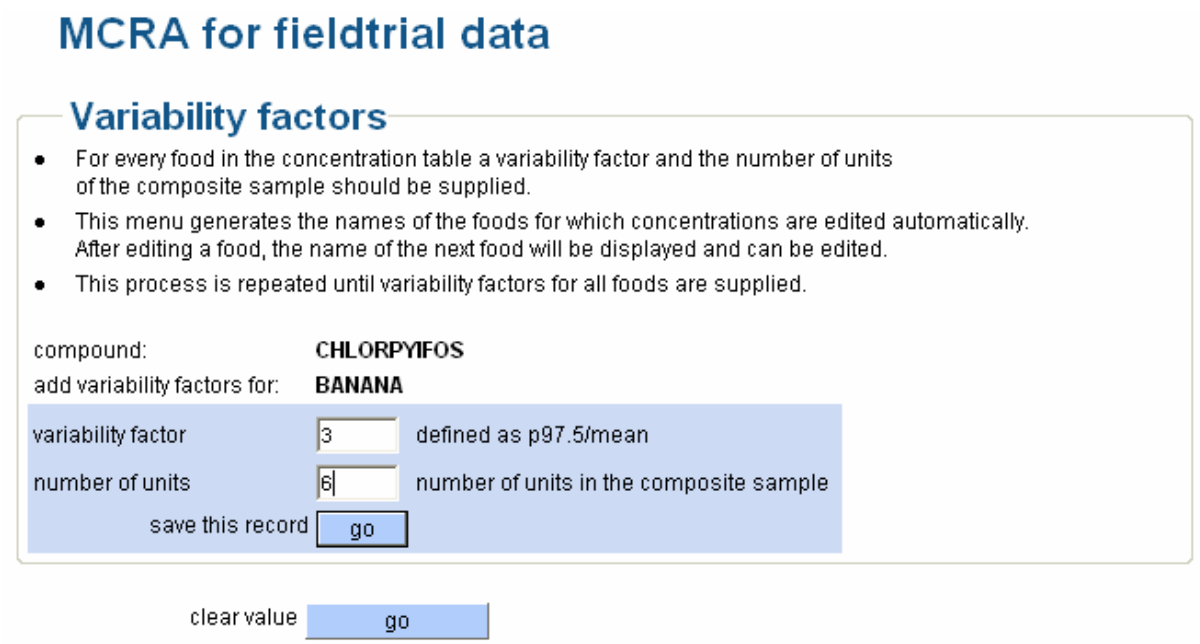


Figure 66: Variability factors

7.3 List or add processing factors

In Figure 67 the menu for entering processing factors is displayed. If you did not enter any values before, a message '*** Processing factor database contains no records yet ***' is displayed. Click the add-records link. Select from the scroll-down menu a processing type, enter a processing factor and save the record. Repeat this step until the message 'All factors for Banana are specified' appears and click the select-next-food 'go'-button.

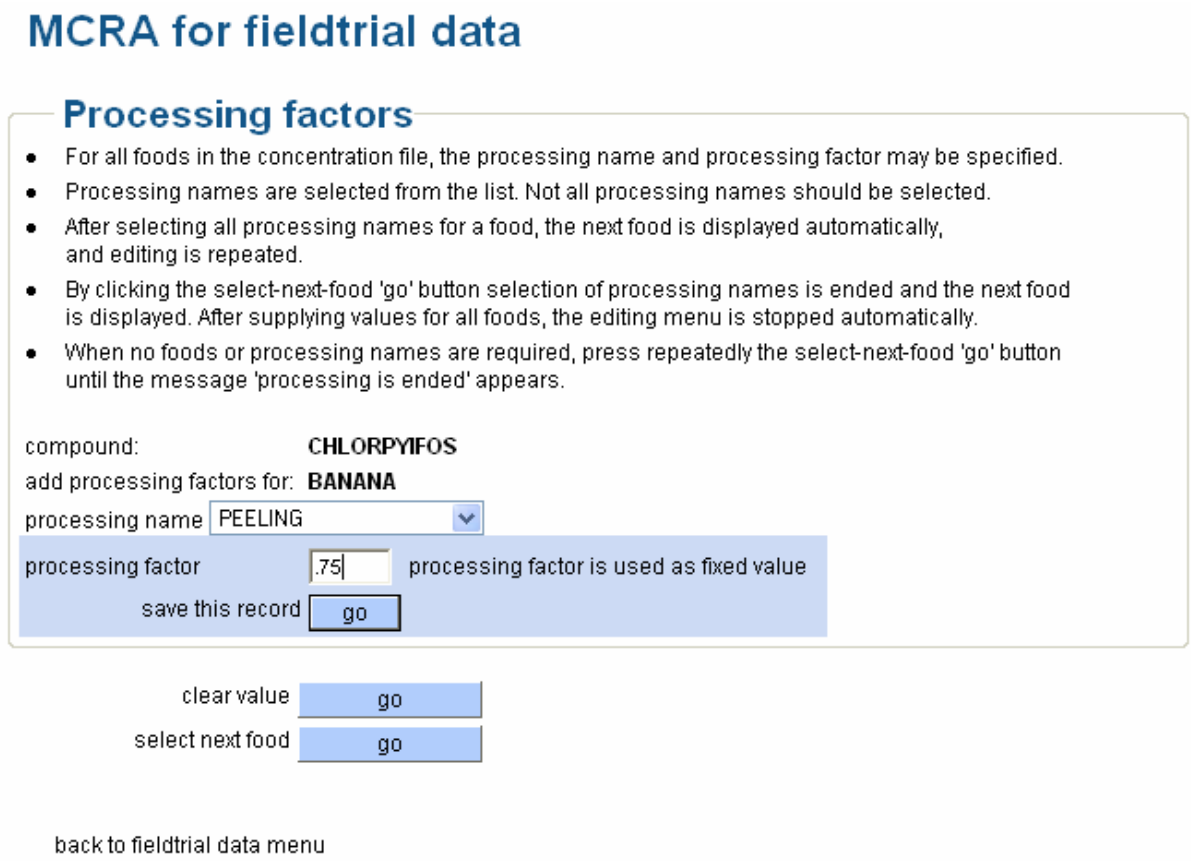


Figure 67: Processing factors

To finish on-line editing for a food, click the 'select-next-food'-button and continue entering values for the next food. It is not necessary to enter values for all combinations of foods and processing types: missing processing factors are by default replaced by value 1 (worst case scenario).
To enter processing factors for only a few number of foods and processing type combinations, enter your values and click repeatedly the 'select-next-food'-button until the message '*** Processing factors for foods are specified ***' is displayed. If you want to enter values at a later stage, just click the 'list or add processing factors'-button in the select data menu. A list of processing factors found in the database is given and after clicking the add-records link, new processing values may be entered. Note that the scroll-down menus are dependent on the food and contain only those processing types for which no values are supplied. Records may be changed afterwards by clicking one of the buttons the listing menu.

7.4 List or edit ARfD or ADI

Click the 'list or edit ARfD or ADI'-button in Figure 60 to enter values for the acute reference dose (ARfD) or average daily intake (ADI). This brings you to Figure 68.

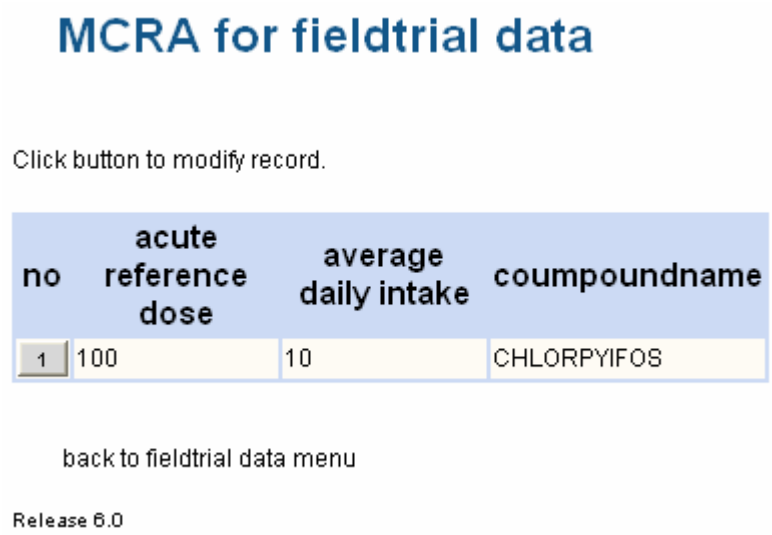


Figure 68: ARfD and ADI

8 References

- Apex Software Corporation (1999). ComponentOne WebChart 7.0.10, Charting tool for browser-independent Web server applications.
- @Risk (1996). Advanced risk analysis for spreadsheets, Windows version. Pallisade Corporation, Newfield, NY, USA.
- Bestfit.(1997). Probability distribution fitting for Windows. Pallisade Corporation, Newfield, NY, USA.
- Blom, G (1958). Statistical estimates and transformed beta-variables. Wiley, New York.
- Boon, PE, van der Voet, H & van Klaveren, JD (2003). Validation of a probabilistic model of dietary exposure to selected pesticides in Dutch infants, *Food Additives and Contaminants*, 20, Suppl. 1: S36-S49.
- Crossley, SJ (2000). Joint FAO/WHO Geneva consultation – acute dietary intake methodology. *Food Additives and Contaminants*, 17: 557-562.
- David, HA (1970). Order statistics. John Wiley & Sons, New York.
- de Boer, WJ, van der Voet, H, PE Boon, G van Donkersgoed & JD van Klaveren (2004). MCRA, a web-based program for Monte Carlo Risk Assessment, Release 3, User Manual. Report March 2004. Biometris and RIKILT, Wageningen University and Research Centre, Wageningen.
- de Boer, WJ & van der Voet, H (2000). Dietary risk assessment concerning acute exposure to residues and contaminants using summary data. Note WDB-2000-01, Centre for Biometry Wageningen, Wageningen.
- Dodd, KW (1996). A technical guide to C-SIDE. Technical Report 96-TR 32, Department of Statistics and Center for Agricultural and Rural Development, Iowa State University, Ames, Iowa. Available at http://www.card.iastate.edu/publications/DBS/PDFFiles/96tr32.pdf
- Dodd, K.W. Guenther, P.M., Freedman, L.S., Subar, A.F., Kipnis, V., Midthune, D., Tooze, J.A. and Krebs-Smith, S.M. (2006). Statistical methods for estimating usual intake of nutrients and foods: a review of the theory. *Journal of the American Dietetic Association*, 106: 1640-1650.
- Efron, B (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, **7**: 1-26.
- Efron, B & Tibshirani, RJ (1993). An introduction to the bootstrap. *Chapman & Hall*, New York.
- FAO/WHO (1997). Food consumption and exposure assessment of chemicals. Report of an FAO/WHO Consultation, Geneva, Switzerland. 10-14 February 1997.
- FAO (2002). Submission and evaluation of pesticide residues data for the estimation of maximum residue levels in food and feed. FAO, Rome.
- GenStat (2005). GenStat for Windows. Release 8.1, eigth edition, VSN International Ltd., Oxford.
- Hamey, PY (2000). A practical application of probabilistic modelling in assessment of dietary exposure of fruit consumers to pesticide residues. *Food Additives and Contaminants*, 17: 601-610.
- Harris, C *et al.* (2000). Summary report of the International Conference on pesticide residues variability and acute dietary risk assessment. *Food Additives and Contaminants*, 17: 481-485.
- Harter, HL. Expected values of normal order statistics. Biometrika 48: 151-165.
- IUPAC (1995). Nomenclature in evaluation of analytical methods including detection and quantification capabilities (IUPAC Recommendations 1995). *Pure and Applied Chemistry* 67: 1699-1723.
- JMPR (1999, 2000). Reports of the joint FAO/WHO meetings of experts on Pesticide residues in food.
- Kistemaker C, Bouman M and Hulshof KFAM (1998). De consumptie van afzonderlijke producten door Nederlandse bevolkingsgroepen - Voedselconsumptiepeiling 1997-1998. Zeist, TNO-Voeding (Report No: 98.812).
- Nusser SM, Carriquiry AL, Dodd KW & Fuller WA (1996). A semi-parametric transformation approach to estimating usual daily intake distributions. *Journal of the American Statistical Association*, 91: 1440-1449.

- Nusser SM, Fuller WA, and Guenther PM (1997). Estimating usual dietary intake distributions: adjusting for measurement error and nonnormality in 24-hour food intake data. In: Lyberg L, Biemer P, Collins M, DeLeeuw E, Dippo C, Schwartz N, and Trewin D (editors), *Survey Measurement and Process Quality*, Wiley, New York. p. 689-709.
- Pearson, ES and Hartley, HO. Biometrika tables for statisticians (1977). Vol II.
- Shimizu, K, and Crow, EL (eds). (1988). Lognormal distributions: theory and applications. Marcel Dekker, INC. New York.
- Snedecor, GW & Cochran, WG (1980). Statistical Methods (7th edition). Iowa State University Click, Ames, Iowa.
- van der Voet, H, de Boer, WJ & Keizer, LCP (1999). Statistical instruments for dietary risk assessment concerning acute exposure to residues and contaminants. Report August 1999, Centre for Biometry Wageningen, Wageningen.
- van der Voet, H, de Boer, WJ & Boon, P (2001). Modelling exposure to pesticides. Note HVT-2001-03, Centre for Biometry Wageningen, Wageningen.
- van der Voet, H. and Slob, W. (2007). Integration of probabilistic exposure assessment and probabilistic hazard characterization. *Risk Analysis*, 27: 351-371.
- van Dooren, MMH , Boeijen, I, van Klaveren, JD and van Donkersgoed G (1995). Conversie van consumeerbare voedingsmiddelen naar primaire agrarische produkten. RIKILT-report. Wageningen, RIKILT-DLO (Report No: 95.17).
- van Klaveren, JD (1999). Quality programme for agricultural products. Results residue monitoring in the Netherlands. RIKILT Institute of Food Safety, Wageningen.

9 APPENDIX A: Preparing the data

9.1 Overview

MCRA requires that all information needed for a risk assessment is stored in MS Access databases. Data are organised into tables which are divided into 3 groups with information on

- food consumption data;
- compound concentration data, and
- linking and additional data.

See Table 21 for an overview. To run MCRA, tables called 'required' should always be supplied. Selection of 'additional' tables depends on subset requirements (see Ch. 5) and model specifications (see Ch. 3).

Required tables (minimal configuration)	
Food	food codes and labels (see 9.3.1)
Food consumption	**Description**
FoodConsumption	consumption of foods (see 9.4.1)
Individual	consumer characteristics (see 9.4.2)
Concentration of compounds	**Description**
Compound	compound codes, labels, agricultural and toxicological limits (see 9.5.1)
Country	country codes, labels (see 9.5.2)
One and only one of:	
ConcentrationValues	compound concentration data (full data) (see 9.5.3)
ConcentrationSummaryStatistics	compound concentration data (as summary statistics) (see 9.5.4)
ConcentrationDiscreteValues	compound concentration data (table of frequency counts) (see 9.5.5)
Additional tables (for querying or specific options)	
FoodProperties	food codes and labels, and food specific information(see 9.3.2)
FoodComposition	food codes and labels, compositions (see 9.3.3)
FoodMarketShare	food codes and labels, marketshares (see 9.3.4)
Food consumption	**Description**
FoodSurvey	Name of survey (see 9.4.3)
ProcessingType	processing codes and labels (see 9.4.4)
Processing	processing factors (see 9.4.5)
Concentration of compounds	**Description**
VariabilityProd	unit variability factors (see 9.5.6)
VariabilityCompProd	unit variability factors, compound-specific (see 9.5.7)
VariabilityProcCompProd	unit variability factors, processing- and compound-specific (see 9.5.8)
AgriculturalUse	information on the agricultural use of compounds (*e.g.* use allowed, percent crop treated) (see 9.5.9)
ConcentrationWorstcaseValues	information on worstcase values (*e.g.* compound and food specific worstcase values) (see 9.5.10)

Table 21: Overview of tables

Tables are organised into columns (fields) and rows (records). In the next paragraphs, the format of tables is described.

General remarks:

- Table and column names should be exactly as indicated in the sections below and each table should contain **all** fields, except for tables **ConcentrationSummaryStatistics** (see 9.5.4) and **ConcentrationDiscreteValues** (see 9.5.5).
- Missing values are indicated with code 9999, unless stated otherwise. In general, an empty cell is also interpreted as a missing value. Occasionally, the use of empty cells leads to errors in retrieving data. Therefore, it is advised to use the code 9999 to indicate missing values

In sections 9.3 , 9.4 and 9.5 the format is explained: the table name is given followed by field names and a description with in parentheses the datatype. Each section ends with some notes and an example.

9.2 Harmonised CODEX codes

In the MCRA program we use harmonised CODEX codes in the interest of Pan-European risk assessment. This coding offers flexibility to enter food consumption data and chemical concentration data at any desired level of food coding (*e.g.* food as eaten, ingredient, raw agricultural commodity, processed food, brand level, etc.).

9.2.1 Definitions

A food code is a string consisting of symbols:

- letters (case-unsensitive, so x and X are the same letter),
- digits, and/or
- special symbols, such as ~!#$^&*()+-=[]{};':",./<>?`

Some special symbols are reserved for special use (see below), and can not be used freely in own codes:

- &
- $
- -
- *

Some symbols are not allowed at all, because this would interfere with the way the strings are analysed:

- %
- _
- @

The first symbol should be:

- a letter (indicates a CODEX code or a code derived from a CODEX code), or
- & followed by a 2-letter country code (indicates a national food code)

CODEX codes start with two letters and four digits, and should comply with the CODEX Classification of foods and animal feeds. The code XX9999 (usually followed by a subtype code) can be used for all foods which cannot be placed in the Codex classification system.

Any code (CODEX code or national code) can be followed by:

- $ plus a subtype code, and/or
- - plus a processing code
- *- plus a processing code. Here the asterisk (*) serves as a wildcard for the preceding code: the processing information is valid for all codes that start with the code preceding the *.

Subtype codes and processing codes can have any format. Multiple levels of subtype code are allowed (*e.g.* &NL00$123$456). Only one level of processing code is allowed (*e.g.* FP0226-2). Subtype codes should precede processing codes (*e.g.* &NL00$123$456-2).

Within EU-Safefoods we will harmonise subtype codes and processing codes as far as they apply to CODEX codes. For this purpose lists will be maintained at RIKILT, and any new subtype code or processing code is checked against this list, and when found appropriate added to this list.

The table Food has to contain descriptive names for all food codes that occur in the tables **FoodConsumption** and **Concentration.....** and that the user wants to include in the analysis. Names will be in English, but an additional column with alternative names can be used.

9.2.2 Conversion of codes

We distinguish 3 types of food code conversion, which may be provided in three different tables:

1. Food processing (table **Processing**)
   Processing factors will be applied to concentration data.
2. Food composition (table **FoodComposition**)
   Composition percentages will be used to transform the consumed amounts.
3. Subtypes/ Market share (table **FoodMarketShare**)
   Market share percentages will be used for as probabilities of selecting concentration data for each of the subtypes.

9.2.3 Conversion rules

- For each *code* in the **FoodConsumption** data set try to find the most appropriate concentration information by searching information in a specific order according to the steps in the scheme below.
- If a code is converted into one or more other codes, then for any such other code re-start the search scheme before continuing to the next step or substep in the scheme with the old code.
- The search is ended if concentration data have been found for *code* or as many as possible derived codes. When a code is converted to multiple new codes (composition, subtypes), then the search continues for all these new codes.
- If no link can be made to concentration data, then the consumption of this *code* is considered irrelevant for the current intake assessment.

In 9.2.4 a scheme is given to link the food consumption and compound concentration data.

9.2.4 Scheme to link food consumption and compound concentration data

Find in Figure 69, an schematic outline of the search for food codes. After a successful search, the code is found in one of the concentration values tables. If the code is not found, searching starts in one of the link tables. If the code is found, the search starts again in one of the concentration values tables and the search is repeated. If a code is not found at all, the search for a code is unsuccessful.

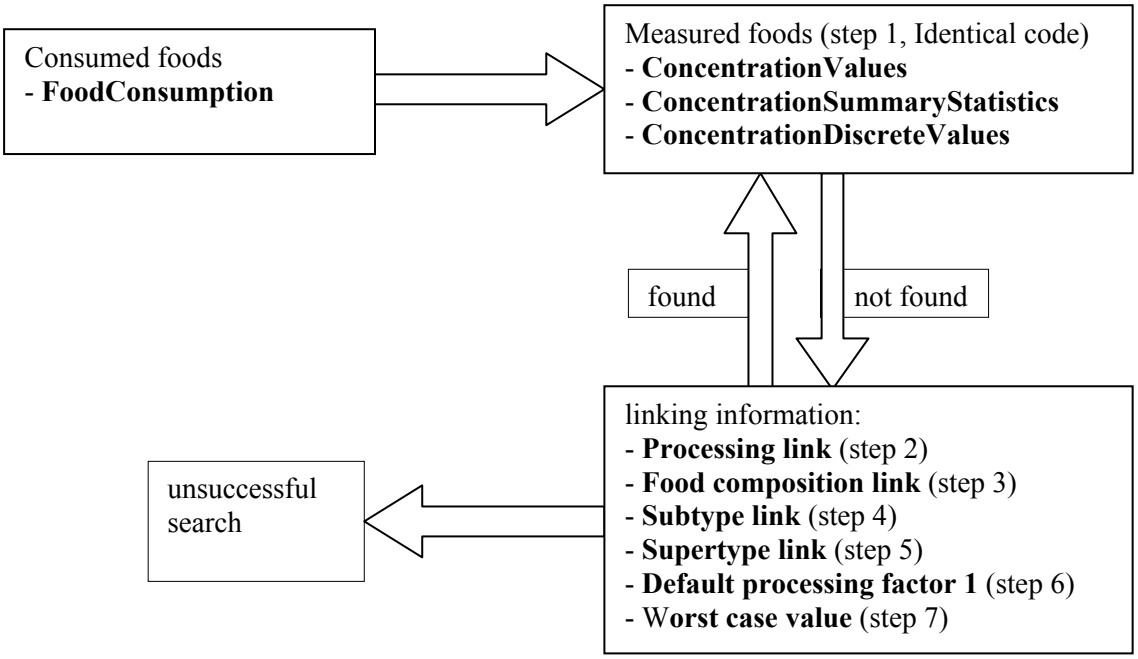


Figure 69: Link scheme to find CODEX codes

Step 1. (Identical code)
Try to find *code* in the field **foodmeasured** of the **ConcentrationValues** (9.5.3)**, ConcentrationSummaryStatistics** (9.5.4 or **ConcentrationDiscreteValues** (9.5.5) table. If found, the search is ended successfully. (Note that one and only one of these tables should be present. These were formerly referred to as full data, summary data and histogram data, respectively)

Step 2. (Processing link)

a. Try to find *code* in the field **foodprocessed** of the **Processing** table (9.4.5), and convert to the code specified in the field **foodunprocessed**.
b. (wildcard match) Try to find a wildcard match code in the field **foodprocessed** of the **Processing** table. Wildcard match codes consist of an initial string (*startcode*, may be empty), an asterisk (*), and possibly a processing part (-*processingtype*). * may be any string *endcode* (not containing a -) such that *code* equals *startcodeendcode* or *startcodeendcode-processingtype*.
    a. If *code* contains a processing part (-*processingtype*), then the wildcard match code should also end with -*processingtype.* Convert to the code specified in the field **foodunprocessed**, where *endcode* is substituted for any * in the new code.
    b. If *code* contains no processing part, then the wildcard match code should also contain no processing part. Convert to the code specified in the field **foodunprocessed**, where *endcode* is substituted for any * in the new code.

Step 3. (Food composition link)

a. Try to find *code* in the field **food** of the **Foodcomposition** table (9.3.3), and convert to one or more ingredient codes found in the field **ingredient**
b. If *code* contains a processing part (*maincode-processingtype*), then try to find *maincode* in the field **food** of the **Foodcomposition** table. Convert to one or more ingredient codes and add -*processingtype* to the new codes.

Step 4. (Subtype link)
Starting from *code*, try to find subtype codes equal to *code$** in the field **foodtype** of the **FoodMarketShare** table (9.3.4 , where the strings represented by the wildcard are not allowed to contain a $ themselves (meaning that we look only for codes one level down in the type hierarchy).
Check that for the selected codes the market share percentages in the field **marketshare%** sum to 100%.

If this is not OK, then the result depends on the user option "Allow conversion to subtypes not summing to 100% and rescale".

1. If this option is allowed, then the marketshare percentages are rescaled to a sum of 100 %.
2. If this option is not allowed, then the search in step 4 is considered unsuccessful.

Step 5. (Supertype link)
This step will only be taken if the user has explicitly allowed this option.

a. If *code* contains a subtype part but no processing part (*maincode$subcode*), then convert to *maincode*.
b. If *code* contains a subtype part and a processing part (*maincode$subcode-processingtype*), then convert to *maincode-processingtype*.

Step 6. (Default processing factor 1)
If *code* contains a processing part (*-processingtype*), then remove it.

Step 7. (Worst case value)
This step will only be taken if the user has explicitly allowed this option.
Try to find *code* in the field **food** of the **ConcentrationWorstcaseValues** table (9.5.10). If found, the search is ended successfully.

9.2.5 Example of use

The **FoodConsumption** table (9.4.1) may contain &NL070251 (Apple pie Dutch):

individual	dayofsurvey	foodconsumed	amountconsumed	foodsurvey
1012	1	&NL070251	150	DNFCS

If measurements are available for FP0226$Elstar (Apple Elstar), FP0226$JonaGold (Apple JonaGold) and GC0654 (Wheat), then we need a conversion.

The **FoodComposition** table (9.3.3) may then specify the composition data that apple pie contains peeled apple and wheat:

food	ingredient	proportion%
&NL070251	FP0226-2	58.09
&NL070251	GC0654	14.52

The **Processing** table (9.4.5) may contain a processing factor for peeling of apples:

compound	foodprocessed	foodunprocessed	proctype	procnom	procupp
120701	FP0226-2	FP0226	2	0.3	0.5

where the field **proctype** specifies explicitly the type of processing (2 = peeling in this case), and fields **procnom** and **procupp** are processing factor nominal and upper values.
The **FoodMarketShare** table (9.3.4) may contain market shares for subtypes of apple:

foodtype	marketshare%
FP0226$Elstar	30

FP0226$JonaGold	70

Finally, the **ConcentrationValues** table (9.5.3) should then contain measured concentrations for FP0226$Elstar (Apple Elstar), FP0226$JonaGold (Apple JonaGold) and GC0654 (Wheat):

compound	foodmeasured	year	month	samplingtype	country	numberofsamples	value
120701	FP0226$Elstar	2006	11	M	NL	1	0.34
120701	FP0226$Elstar	2006	11	M	NL	6	-0.01
120701	FP0226$JonaGold	2006	11	M	NL	1	0.20
120701	FP0226$JonaGold	2006	11	M	NL	1	0.05
120701	GC0654	2006	11	M	NL	6	-0.01

The **Food** table (9.3.1) should at least contain the following entries:

food	foodname
&NL070251	Apple pie Dutch
FP0226$Elstar	Apple Elstar
FP0226$JonaGold	Apple JonaGold
GC0654	Wheat

Note that entries for intermediate stages such as FP0226-2 or FP0226 are not obligatory.

9.3 Food tables

9.3.1 Food (compulsory)

field name	description
food (text)	food code
foodname (text)	food label
foodname2 (text, optional)	alternative food label, *e.g.* national language

- Foodname2 is used for alternative foodnames.

Example:

Food : Table

food	foodname	foodname2
&NL0002	Potatoe	aardappel
&NL0003	Pie	taart
&NL0004	Yoghurtdrink	yoghurtdrank
FP0226	Apple	appel

9.3.2 FoodProperties (optional, (for unit variability compulsory))

field name	description
food (text)	food code
foodname (text)	food label
unitweight (number)	nominal weight of a unit (gr)
edibleportion (number)	edible portion (corrected large portion weight, gr)
largeportion (number)	weight of a large portion (gr)

- For unknown nominal unit weight use value 0.
- Missing values for edibleportion and largeportion: 9999.

Example:

FoodProperties : Table

food	foodname	unitweight	edibleportion	largeportion
DT1114	TEA, GREEN, BLACK (BLACK, FERMEN	20	9999	13
FB0020	BLUEBERRIES	25	190	200
FB0021	CURRANTS, BLACK, RED, WHITE	10	160	167
FB0264	BLACKBERRIES	12	125	125

9.3.3 FoodComposition (optional)

field name	description
food (text) ingredient (text) proportion% (number)	food code ingredients of the food proportion of each ingredient in the food (in percentages)

- Specifies the composition of foods and corresponding proportions.

Example:

FoodComposition : Table

food	ingredient	proportion%
&NL1001	AC2341	40
&NL1001	AB0101	30
&NL1001	CF1211	5
&NL1001	FP0226	25
&NL10201	FP0230	80
&NL10301	FI0353-4	10

9.3.4 FoodMarketShare (optional)

field name	description
foodtype (text) marketshare% (number)	subtype of food market share of each subtype (in percentages)

- Specifies food marketshares of subtypes.

Example:

FoodMarketshare : Table

foodtype	marketshare%
&TE1$1$1	10
&TE1$1$2	30
&A1$1	33
&A1$2	67
&A1$1$3	10
&A1$1$4	90

9.4 Food consumption tables

9.4.1 FoodConsumption (compulsory)

field name	description
individual (number) dayofsurvey (number)	consumer identification number day (sequential number in food consumption survey)

foodconsumed (text) amountconsumed (number) foodsurvey (text)	food code consumed portion of food (g) name of survey

- Contains data on consumed foods. Days without consumptions are not recorded. The number of available days per consumer is inferred from this table and is assumed to be the same for each consumer in the survey.
- No missing values allowed.

Example:

FoodConsumption : Table

individual	dayofsurvey	foodconsumed	amountconsumed	foodsurvey
1	1	&NL00011	200	DNFCS-3
1	2	&NL0001	300	DNFCS-3
1	2	&NL0002	150	DNFCS-3
1	2	&NL0101$3	35	DNFCS-3
2	1	&NL0002	150	DNFCS-3
2	2	&NL0004	212	DNFCS-3

9.4.2 Individual(compulsory)

field name	description
individual (number) foodsurvey (text) age (number) weight (number) sex (text)	consumer identification number name of survey age (*e.g.* in years, months or days) body weight (*e.g.* in kg or g) gender

- Specify in table FoodSurvey (see 9.4.3) the unit for age and weight.
- No missing values allowed.

Example:

Individual : Table

individual	age	weight	sex	foodsurvey
1	25	70	male	DNFCS-3
2	40	75	female	DNFCS-3
3	12	21	female	DNFCS-3
4	67	68	female	DNFCS-3

9.4.3 FoodSurvey (optional)

field name	description
foodsurvey (text) year (number) country (text) agein (text) weightin (text)	name of survey year of survey country of survey unit of age unit of weight

- Defines characteristics of the survey.
- No missing values allowed.

Example:

FoodSurvey : Table

foodsurvey	year	country	agein	weightin
DNFCS-3	2006	NL	y	kg
EU-BABY	2005	NL	m	gr

9.4.4 ProcessingType (optional, (for processing compulsory))

field name	description
proctype (number)	code of processing type
procname (text)	description of processing type
disttype (number)	indicator (1/2): • simulated processing factors are restricted to the interval (0,1) using a logistic-normal distribution (1), • or simulated processing factors are restricted to positive values using a log-normal distribution (2)
bulkingblending (number)	indicator (0/1): for types of processing applied on large batches, *e.g.* juicing, sauce/puree (obligatory), • 0 = no bulking/blending ; • 1 = bulking/blending

- Information on bulking and blending is only relevant for modelling of processing effects in combination with unit variability and IESTI calculations, but should always be present in the table even when these effects are not explored.
- No missing values allowed.

Example:

ProcessingType : Table

proctype	procname	disttype	bulkingblending
1	RAW	1	0
2	PEELING	1	0
3	COOKING IN WATER	1	0
4	BAKING OF BREAD	1	0
5	CANNED/CONSERVED	1	0
6	BREWING	1	0
7	DRYING	2	0
8	FRYING/BAKING IN FAT	1	0
9	JUICING	1	1
10	MILLING	2	0
11	MARMELADE/JAM	1	1
12	OIL EXTRACTION	2	0
13	SAUCE/PUREE	1	1
14	CLEANING	1	0
15	WASHING/CLEANING	1	0
16	WINE MAKING	1	0
99	UNKNOWN	1	0

9.4.5 Processing (optional)

field name	Description

compound (text)	compound code
foodprocessed (text)	food code processed
foodunprocessed (text)	food code unprocessed
proctype (number)	code of processing type
procnom (number)	nominal value (best estimate of 50th percentile) of processing factor (defines median processing factor)
procupp (number, optional)	upper value (estimate of 95th percentile or "worst case" estimate) of processing factor due to variability (from procnom and procupp a standard deviation for variability of the processing factor is derived)
procnomuncupp (number, optional)	upper 95th percentile of procnom due to uncertainty (from procnom and procnomuncupp a standard deviation for uncertainty of procnom is derived)
procuppuncdf (number, optional)	degrees of freedom of a chi-square distribution describing the uncertainty of the standard deviation for variability of the processing factor
procuppuncupp[1] (number, optional)	upper 95th percentile of procupp due to uncertainty (from procnom, procupp, procnomuncupp and procuppuncupp the degrees of freedom of a chi-square distribution describing the uncertainty of the standard deviation for variability is derived)

- This table is only relevant when the input option for processing is set to fixed or distribution.
- When the input option is set to fixed then in addition to the information in the first four columns only procnom or procupp needs to be specified. If both are specified the highest value will be used (worst case argument). For use in an uncertainty analysis also procnomuncupp may be specified.
- When the input option is set to distribution then in addition to the information in the first four columns procnom and procupp have to be specified describing the variability of processing factors. For use in an uncertainty analysis also procnomuncupp and/or procuppuncdf (or procuppuncupp[1]) may be specified.
- procupp should be higher than procnom.
- procnomuncupp should be higher than procnom.
- procuppuncupp should be higher than procupp.
- procuppuncdf should be positive, with values close to zero defining maximum uncertainty, and high values defining minimal uncertainty.
- Values lower than 0.01 are reset to 0.01; for processing types with disttype 1 (logistic) values higher than 0.99 are reset to 0.99.
- Procuppuncdf and procuppuncupp are alternative ways to specify uncertainty for the variability of processing factors[1]. The variability of processing factors is described by a standard deviation (at a logistic or logarithmic scale), and its uncertainty is described by setting the degrees of freedom (procuppuncdf) of a modified chi-squared distribution (see van der Voet and Slob, 2007 for an example). Alternatively, an upper uncertainty percentile on the upper variability percentile (procuppuncupp) can be specified, from which the appropriate number of degrees of freedom is derived by simulation[1].
- If a value for procuppuncdf is specified, procuppuncupp will be ignored.

Example:

Processing : Table

compound	foodprocessed	foodunprocessed	proctype	procnom	procupp	procnomuncupp	procuppuncdf
011004001	FP0226-9	FP0226	9	0.9	0.9	0.5	4
011004001	FP0226-13	FP0226	13	0.8	0.8	0.5	4
011004001	VO0448-15	VO0448	15	1.2	1.2	0.5	5
011004001	VO0448-9	VO0448	9	0.5	0.5	0.3	20
011004001	VO0448-5	VO0448	5	0.5	0.5	0.3	1

[1] This option is not implemented in the first release of MCRA 6.0 and will be implemented later.

9.5 Compound concentration tables

9.5.1 Compound (compulsory)

field name	Description
compound (text)	compound code
compoundname (text)	compound label (name of compound)
arfd (number)	ARfD (acute reference dose), in microgr/kg bw/day
adi (number)	ADI (acceptable daily intake), in microgr/kg bw/day
unit (number, optional)	-6 (default) or -9, see below

- Missing values for ARfD and ADI: 9999.
- Column unit contains a coding to determine the unit as used for compound concentration data and dietary intake. Coding is as follows:

	-6	-9
concentration:	mg/kg	microgram/kg
intake:	microgram/kg bw/day	nanogram/kg bw/day

- If column unit doesn't exist code -6 is assumed

Example:

Compound : Table

compound	arfd	adi	compoundname	unit
C01001001	0.01	0.001	ACEPHATE	-6
C01003001	0.05	0.01	DISULFOTON	-9
C01001002	0.01	9999	IPRODIONE	-6

9.5.2 Country (compulsory)

field name	Description
country (text)	code for country
countryname (text)	name of the country, label

- No missing values allowed

Example:

Country : Table

country	countryname
99	Unknown
AA	Aruba
NL	The Netherlands

9.5.3 ConcentrationValues (optional)

field name	description
compound (text)	compound code
foodmeasured (text)	food code
year (number)	sampling year
month (number)	number of month
samplingtype (text)	type of sampling (monitoring)
country (text)	country of sample
numberofsamples (number)	count of the number of times the specified concentration or limit of reporting (LOR) occurs
value (number)	concentration (mg/kg) or LOR (see below)

- The limit of reporting is specified in column value using a minus (-) sign to make the distinction between a measured concentrations, *e.g.* –0.02 (see example first row).
- Concentration values are stored in column value and the number of times each value occurs in column numberofsamples, *e.g.* 0.21 and 1, respectively.
- Missing LORs are reported as –9999. The MCRA program replaces missing LORs with 1) the maximum LOR found in the database, 2) if all LORs are missing, the lowest concentration found in the database. A warning is generated when 1) and 2) are not possible.
- No missing values allowed for the other columns.

Example:

ConcentrationValues : Table

compound	country	value	numberofsamples	foodmeasured	year	month	samplingtype
C01001001	NL	0.053	2	FP0226	2005	5	m
C01001001	NL	0.013	4	VX1201	2006	3	m
C01001002	NL	-0.02	10	VS2300	2006	3	m
C01001002	NL	-0.003	15	FP0226	2006	12	m
C01001001	NL	-0.05	25	FP0226	2006	1	m

9.5.4 ConcentrationSummaryStatistics (optional)

field name	description
compound (text)	compound code
foodmeasured (text)	food code
country (text)	code for country
limitofreporting (number)	limit of reporting (mg/kg)
numberofsamples (number)	size of sample (detects and non-detects)
numberofpositives (number)	number of positive concentration values (detects)
the mean: mean *or* meanall (number)	statistic for the mean
the median: med *or* medall (number)	statistic for the median
max (number)	statistic for the maximum
the variance: var *or* varall (number)	statistic for the variance
the percentile: perc *or* percall (number)	statistic for the percentile
percentile (number)	specifies the percentage of the statistics **perc** and **percall**

- Field names mean, meanall, med, medall, max, var, varall, perc, percall and percentile are optional and their order is free. Not all statistics need to be present in the table. See also last bullet.
- Statistics ending on '**all**' refer to statistics based on all samples including non-detects (concentrations below LORs), while statistics without suffix 'all' relate to statistics based on nonzero samples (non-detects) only.
- The use of equivalent statistics, like **mean** and **meanall**, for one food in the same row is not allowed.
- Be aware that statistics should be consistent *e.g.*: **med** is always smaller than **mean;** the calculated mean (nonzero samples only) that is derived from statistic **meanall** should be smaller than **max**; specifying **medall** implies that more than half the number of samples are detects (numberofpositives); specifying **percall** implies that the number of detects (numberofpositives) is greater than the percentage specified in column **percentile**.
- Missing LORs are reported as –9999. The MCRA program replaces missing LORs with 1) the maximum LOR found in the database, 2) if all LORs are missing. A warning is generated when 1) and 2) are not possible.

- Missing statistics are reported as 9999. Columns containing only missing values are not allowed and should be deleted.

Example:

ConcentrationSummaryStatistics : Table

compound	foodmeasured	country	limitofreporting	numberofsamples	numberofpositives	mean	median	max	var	perc	percentile
C012001001	FP0226	NL	0.001	200	100	2.5	1.8	9.5	3.38	6.3	95
C012001001	FB1235	NL	0.001	333	200	3.6	2.9	21	7	8.1	95

9.5.5 ConcentrationDiscreteValues (optional)

field name	description
compound (text) foodmeasured (text) country (text) limitofreporting (number) numberofsamples (number) c%01 (number) c%02 (number) c%05 (number) c%1 (number) c%2 (number) c%5 (number) c1 (number) c2 (number) *c<xxx>* (number) *c<xxx>* (number) *c<xxx>* (number) cE10 (number)	compound code food code code for country limit of reporting (mg/kg) size of sample (detects and non-detects) • number of samples with a concentration between the value extracted from the field name of the previous column (exception: for the first column a value 0 is taken) and the value extracted from the field name in the current column (mg/kg). • classes (*i.e.* columnnames) are free to choose so *c<xxx>*… may be replaced with any appropriate concentration *e.g.* c5, c10 etc.

- Field names representing the number of frequency counts are constructed as follows:
  c = indicates class limit,
  % = represents the decimal point (if necessary),
  xx = is the value of the class limit.
  Thus: field name c%02 specifies class limit 0.02, field name c2 specifies class limit 2, field name cE10 specifies class limit $1*10^{10}$.
- The number of non-detects measurements is given as the difference between the numberofsamples and the sum of frequency counts, *e.g.* see example first record 377 – 1 = 376,.
- Missing LORs are reported as –9999. The MCRA program replaces missing LORs with 1) the maximum LOR found in the database, 2) if all LORs are missing….. A warning is generated when 1) and 2) are not possible.
- For columns numberofsamples, c%02…cE10 no missing values is allowed: classes without frequency counts are reported as 0. When no data are available for a food, delete the entire row.

Example:

ConcentrationDiscreteValues : Table

compound	foodmeasured	country	limitofreporting	numberofsamples	c%01	c%1	c1	c10	c100	cE10
C012001001	FP0226	NL	0.001	200	0	3	15	85	12	1
C012001001	FB1235	NL	0.001	333	0	2	17	180	3	0

9.5.6 VariabilityProd (optional)

field name	description
food (text) varfac (number) coefvar (number)	food code variability factor coefficient of variation

nounitcomp (number)	number of units in the composite sample

- This table is used for specifying real empirical estimates of unit variability (*e.g.* from special studies) for the lognormal and the beta distribution and the number of units in a composite sample.
- Estimates for unit variability are independent of the compound.
- Missing values: 9999
- When the parameter for unit variability is a coefficient of variation and the number of units equals 1, unit variability is ignored for this food.

Example:

VariabilityProd : Table

food	varfac	coefvar	nounitcomp
DF5263	0.41	1.2	12
FB0264	0.6	2.3	15
FB0268	1.2	2.1	20
FB0272	0.95	0.98	25

9.5.7 VariabilityCompProd (optional)

field name	description
compound (text)	compound code
food (text)	food code
varfac (number)	variability factor
coefvar (number)	coefficient of variation
nounitcomp (number)	number of units in the composite sample

- This table is used for specifying real empirical estimates of unit variability (*e.g.* from special studies) for the lognormal and the beta distribution that are dependent on the compound. Values for unit variability in table VariabilityProd are replaced by the new ones.

Example:

VariabilityCompProd : Table

compound	food	varfac	coefvar	nounitcomp
C01001001	DF5263	0.3	1.1	12
C01001002	FB0264	0.76	2.3	15
C01001003	FB0268	1.4	2.4	20
C01001004	FB0272	1.1	1.4	25

9.5.8 VariabilityProcCompProd (optional)

field name	description
compound (text)	compound code
food (text)	food code
proctype (number)	processing type code
varfac (number)	variability factor
coefvar (number)	coefficient of variation
nounitcomp (number)	number of units in the composite sample

- This table is used for specifying real empirical estimates of unit variability (*e.g.* from special studies) for the lognormal and the beta distribution that are dependent on the combination of processing type and compound. Values for unit variability in table VariabilityProd and VariabilityCompProd are replaced by the new ones. This can be used for example to reset the variability factor to 1 for grape juice and raisins (dried grapes).

Example:

VariabilityProcCompProd : Table

compound	food	proctype	varfac	coefvar	nounitcomp
C01001001	DF5263	2	0.3	1.1	12
C01001002	FB0264	3	0.76	2.3	15
C01001003	FB0268	4	1.4	2.4	20
C01001004	FB0272	3	1.1	1.4	25

9.5.9 AgriculturalUse (optional)

field name	description
compound (text)	compound code
food (text)	food code
country (text)	code for country
year (number)	year
useallowed (number)	indicator (0/1) whether use of the compound for the food is allowed (1) or not (0)
perccroptreated (number)	maximum percentage of the food that is treated with the compound

- For combinations of compound and foods that are not listed in table AgriculturalUse MCRA will assume that use is not allowed.

Example:

AgriculturalUse : Table

compound	food	country	year	useallowed	perccroptreated
C012007001	FB1235	NL	2006	1	34
C012001002	AB1235	NL	2006	1	80
C012001001	SX1235	NL	2002	1	55

9.5.10 ConcentrationWorstcaseValues (optional)

field name	description
compound (text)	compound code
food (text)	food code
country (text)	code for country
year (number)	year
worstcasevalue (number)	worstcase value

- When information on detects and non-detects is missing, worstcase values may be used.

Example:

ConcentrationWorstcaseValues : Table

compound	food	country	worstcasevalue	year
C01001001	AC0102	NL	0.02	2006
C01002001	CF1211	BE	0.05	2006
C01002003	AS1211	EU	0.05	2006

10 APPENDIX B: Example output

Appendix B shows more output examples. For the acute risk assessment the example of 4.1 is used, for the chronic risk assessment the example of 4.2 .

10.1 Acute risk assessment: processing fixed factors

In this example output is shown for an acute risk assessment for Chlorpyrifos. In paragraphs 10.2 unto 10.7 the same data as in 4.1 are used. Set the 'processing factors' option is to processing (fixed factors). Table 22 lists the main options:

Input form	
Compound	CHLORPYRIFOS
risk type	acute
uncertainty analysis	no
concentration model	empirical
number of Monte Carlo simulations	100000
unit variability model	no unit variability
random seed	0
intake model	only empirical estimates
concentration data:	
processing factors	**processing (fixed factors)**
additional	system defaults
output	system defaults

Table 22: Input form options: processing

In Table 23, find the main characteristics concerning this analysis taken from the 'Additional output' file. Note that the number of foods and processing combinations is 11 whereas the number of foods is 9.

'Additional output' file	
Number of foods	9
Acute reference dose (ARfD)	100
Average daily intake (ADI)	10
Number of detects	665
Number of non-detects	3267
Number of foods and processing type combinations	11
No of consumers	6250
Population characteristics,	
minimum age	1
maximum age	97
minimum weight	8
maximum weight	150
sex	female, male
Total no of consumption days	12132

Table 23: Information in 'Additional output' file

For a summary of the data, see Table 5. In Table 24 you find a summary of the simulated intakes. Compared to Table 6, this table contains a second section with information on processed foods only. Find an additional column (ProcFact) with, for fixed factors, the value of the fixed processing factor and for distribution based factors, the mean of the sampled processing factor values.

```
--------------------------------------------------------------------------------
Summary of simulations of consumptions and compound concentrations with respect to:
```

Compound: CHLORPYRIPHOS

```
Code         : food code
Food         : food label
-------------------------------------------------------------------------
MeanConsum   : average consumption, all consumers, all days
DeltaC       : difference (%) compared to average consumption
               in database
MeanConsDays : average consumption, consumption days only
NConsDays    : number of consumption days in the data set
NDays        : total number of days
%ConsDays    : percentage consumption days
-------------------------------------------------------------------------
MeanConcen   : mean concentration in simulations with positive amount consumed
               (after processing)
DeltaR       : difference (%) compared to average concentration
               in database
NSamplPos    : number of positive concentrations in simulations with positive
               amount consumed
NSamples     : total number of concentration measurements
               (detects and non-detects) in simulations with positive
               amount consumed.
%SamplPos    : percentage positive concentrations
ProcFact     : mean processing factor
```

Food consumption data

Code	Food	MeanConsum (g)	DeltaC (%)	MeanConsDays (g)	NConsDays	NDays	%ConsDays (%)
FB1235	TABLE-GRAPES	12.99	-1.8	33.51	38776	100000	38.8
FC0203	GRAPEFRUIT,	4.05	-0.4	28.29	14308	100000	14.3
FC0204	LEMON, SEE A	1.56	1.1	4.21	37039	100000	37.0
FC0206	MANDARIN, SE	8.80	-0.4	41.09	21404	100000	21.4
FC0208	ORANGE, SWEE	56.81	-0.4	99.05	57352	100000	57.4
FP0226	APPLE	61.51	0.0	99.22	61994	100000	62.0
FS0247	PEACH	2.04	-1.9	7.17	28395	100000	28.4
VO0445	PEPPERS, SWE	3.32	-0.5	16.37	20270	100000	20.3
VR0589	POTATO	138.21	-0.2	172.99	79893	100000	79.9

Compound concentration data

Code	Food	MeanConc mg/kg	DeltaR (%)	NSamplPos	NSamples	%SamplPos (%)	ProcFact
FB1235	TABLE-GRAPES	0.0161	0.9	5191	38776	13.4	1.00
FC0203	GRAPEFRUIT,	0.0692	-0.1	6249	14308	43.7	1.00
FC0204	LEMON, SEE A	0.0126	-1.0	8208	37039	22.2	1.00
FC0206	MANDARIN, SE	0.0959	-0.6	16003	21404	73.0	0.90
FC0208	ORANGE, SWEE	0.0521	0.0	27678	57352	48.3	1.00
FP0226	APPLE	0.0030	-1.1	3423	61994	5.5	1.00
FS0247	PEACH	0.0075	-1.4	3756	28395	13.2	1.00
VO0445	PEPPERS, SWE	0.0009	-15.1	189	20270	0.9	1.00
VR0589	POTATO	0.0002	-1.6	290	79893	0.4	1.00

Food consumption data

Food & Processing		MeanConsum (g)	DeltaC (%)	MeanConsDays (g)	NConsDays	NDays	%ConsDays (%)	ProcFact
MANDARIN, SE	canned/conserv	1.92	-0.2	70.86	2711	100000	2.7	0.40
MANDARIN, SE	juicing	1.50	3.5	77.03	1949	100000	1.9	0.70
MANDARIN, SE	unknown	5.37	-1.4	31.13	17259	100000	7.3	1.00

Table 24: Summary of simulation, including processed foods

For processing, all tables as found in 4.1 are extended with additional information on the processed food. Note that the percentiles of the acute intake distribution after applying 'processing factors', see Table 25, are slightly higher than without processing (see Table 9)

```
-------------------------------------------------------------------------
Random sampling is based on seed :        0
Number of simulations (consumers):   100000  out of   6250
CHLORPYRIPHOS microgr/kg bw/day)  consumption:    49418 out of 100000

Compound:  CHLORPYRIPHOS

-------------------------------------------------------------------------
     Percentiles, maximum and average intake
-------------------------------------------------------------------------

   Percentage     Percentiles of CHLORPYRIPHOS  microgr/kg bw/day)

        50.00        0.00000
        90.00        0.22119
        95.00        0.42000
        99.00        1.18770
        99.90        3.45401
        99.99        6.71302
         mean        0.08012
      maximum       14.39752
```

Table 25: Percentiles for the acute intake distribution applying processing

10.2 Acute risk assessment: unit variability, Beta distribution

In this example, output is shown for an acute risk assessment and unit variability for organo phosphate pesticide Chlorpyrifos. Table 26 lists the main options:

Input form	
risk type	acute
uncertainty analysis	no
concentration model	empirical
Number of Monte Carlo simulations	100000
unit variability model	**beta distribution**
random seed	0
intake model	only empirical estimates
concentration data	system defaults
Additional	system defaults
Output	system defaults

Table 26: Input form options: unit variability

Set, see left section of the Input form (see Figure 16), unit variable to Beta. Then, in the right section a 'Unit variability' block appears as shown in Figure 70:

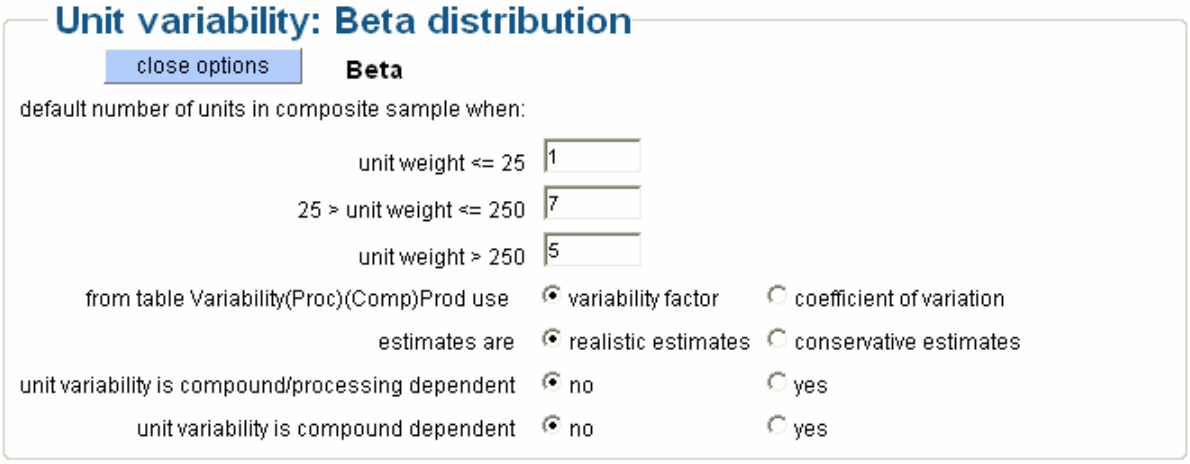


Figure 70: 'Unit variability' block if Beta distribution is chosen

In Table 27 you find the main characteristics concerning this analysis taken from the 'Additional output' file.

'Additional output' file

```
Model 1 : Beta distribution of unit concentrations.
Maximum possible unit concentration is NU times comp. sample concentration
(NU = number of units in composite sample).
Default values can be overruled by specifying NU values.
                          unit weight <= 25:  1
                  25 < unit weight <= 250:  7
                          unit weight > 250:  5
Estimated variability factors are p97.5/mean.
```

Table 27: Information in 'Additional output' file

In Table 28, the percentiles of the intake distribution are shown. Note that the percentiles are much higher (p99.99 = 13.95) than without unit variability (p99.99 = 7.358, see Table 9) and after applying processing (p99.99 = 6.71, see Table 25).

```
-------------------------------------------------------------------------
Random sampling is based on seed :        0
Number of simulations (consumers):   100000  out of   6250
CHLORPYRIPHOS (microgr/kg bw/day)  consumption:    49404 out of 100000

Compound:  CHLORPYRIPHOS

-------------------------------------------------------------------------
     Percentiles, maximum and average intake
-------------------------------------------------------------------------

   Percentage     Percentiles of CHLORPYRIPHOS  (microgr/kg bw/day)

        50.00        0.00000
        90.00        0.10089
        95.00        0.38606
        99.00        1.76710
        99.90        5.87387
        99.99       13.95083
         mean        0.08390
      maximum       28.09446
```

Table 28: Percentiles for the acute intake distribution after applying unit variability

10.3 Acute risk assessment: IESTI

In this example, output is shown for an acute risk assessment and IESTI for organo phosphate pesticide Chlorpyrifos. Table 29 lists the main options:

Input form	
risk type	acute
uncertainty analysis	no
Concentration model	empirical
number of Monte Carlo simulations	100000
unit variability model	no unit variability
random seed	0
intake model	only empirical estimates
Concentration data	system defaults
additional:	
estimation of IESTI	**yes**
standard body weight is	**60**
compare IESTI with Monte Carlo-percentile	**yes**
MC percentage for comparison with IESTI	**99**
use own variability factors	**No**
consumption days only	**All days**
Output	system defaults

Table 29: Input form options: IESTI

In Table 30 IESTI estimates are displayed together with the estimate expressed as percentage of the ARfD (%ARfD). Also a comparison is made with the MC-percentile per food for positive consumption days only (ConsPos) as well as all consumption days (AllDays). For Orange the IESTI estimate is 2.550 microgram/kg bw/day which is much higher than the MC-percentiles for positive consumption days only (0.99 microgram/kg bw/day) and slightly higher than the MC-percentiles for all consumption days (1.89 microgram/kg bw/day).

IESTI

Compound: CHLORPYRIPHOS

IESTI estimates (microgram/kg bw/day)
%ofARfD: estimates expressed as % of Acute Reference Dose

ConsPos: percentiles for positive intake days only (per food)
AllDays: percentiles for all days (including days without intake, per food)
LP: Large portion consumption (g/day)
UW: Unit weight (g/day)
HR: High residue = Largest value from database

IESTI compared with Monte Carlo percentiles for positive intake days only and All Days

Food	LP	UW	HR	IESTI	%ofARfD	ConsPos p99	AllDays p99
TABLE-GRAPES	340	500	1.200	34.000	34.000	0.08	1.18
GRAPEFRUIT,	301	340	0.530	13.294	13.294	0.08	1.08
LEMON, SEE A	30	108	0.150	0.075	0.075	0.01	0.07
MANDARIN, SE	210	133	0.690	2.415	2.415	0.42	1.02
ORANGE, SWEE	340	160	0.450	2.550	2.550	0.99	1.89
APPLE	316	112	0.610	3.213	3.213	0.08	1.37
PEACH	230	110	0.270	1.035	1.035	0.00	0.14
PEPPERS, SWE	115	160	0.360	0.690	0.690	0.00	0.22
POTATO	420	216	0.060	0.420	0.420	0.00	0.61

Table 30: IESTI estimates

10.4 Acute risk assessment: uncertainty

In this example, output is shown for an acute risk assessment and an uncertainty analysis for organo phosphate pesticide Chlorpyrifos. Table 31 lists the main options:

Input form	
risk type	acute
uncertainty analysis	**yes**
concentration model	empirical
number of Monte Carlo simulations	100000
unit variability model	no unit variability
random seed	0
intake model	only empirical estimates
number of resampled sets	**100**
number of simulations per resampled set	**20000**
concentration data:	system default
Additional	system defaults
Output	system defaults

Table 31: Input form options: uncertainty

In Table 32, the percentiles of the intake distribution for the specified percentages are displayed together with the 2.5, 25, 75, 97.5% points of the percentile uncertainty distribution. In this example, the 95% uncertainty interval for the p99.99 (7.90 microgram/kg bw/day) is (4.81, 9.78).

```
--------------------------------------------------------------------------------
Random sampling is based on seed :        0
Number of simulations (consumers):   100000  out of   6250
CHLORPYRIPHOS (microgr/kg bw/day)  consumption:    49339 out of 100000

Compound:  CHLORPYRIPHOS

--------------------------------------------------------------------------------
     Percentiles, maximum and average intake
--------------------------------------------------------------------------------

   Percentage     Percentiles of CHLORPYRIPHOS  (microgr/kg bw/day)

        50.00        0.00000
        90.00        0.23686
        95.00        0.44734
        99.00        1.21234
        99.90        3.28142
        99.99        7.89895
         mean        0.08358
      maximum       15.42857



                Uncertainty of percentiles distribution
                     2.5%          25%          75%        97.5%
       50.00      0.00000      0.00000      0.00000      0.00020
       90.00      0.21474      0.22829      0.24797      0.26665
       95.00      0.40499      0.43125      0.46066      0.48991
       99.00      1.08528      1.16305      1.27898      1.39271
       99.90      2.70767      3.11917      3.60726      4.17651
       99.99      4.80543      5.76208      7.53429      9.77615
        Mean      0.07465      0.08065      0.08769      0.09282
     Maximum      6.03799      8.14262     11.48968     18.31118
```

Table 32: Percentiles for the intake distribution and uncertainty

The information of Table 32 is plotted in Figure 71. The 2.5 and 97.5% points are displayed by the endpoints of the small line segments, whereas the thick bar indicates the distance between the 25%- and 75%-points of the percentile uncertainty distribution. As seen, percentiles for high percentages have a large uncertainty interval.

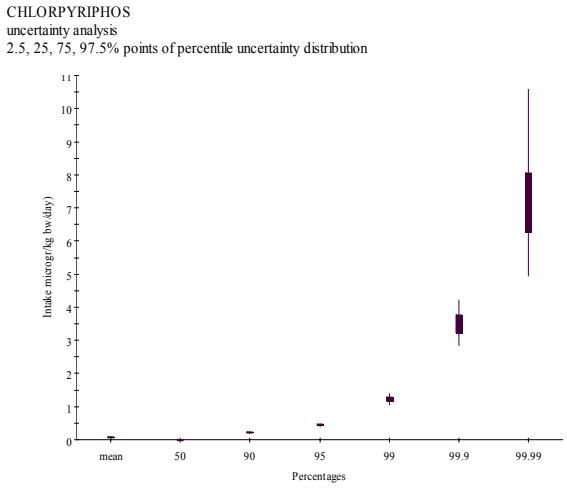


Figure 71: Uncertainty analysis

10.5 Acute risk assessment: diagnostics

In this example, output is shown for an acute risk assessment and an uncertainty analysis focusing on the diagnostics for organo phosphate pesticide Chlorpyrifos. For input options and information in 'Additional output' file, see Table 31 .

Diagnostics are only available after running an acute risk analysis in combination with uncertainty analysis. The provided diagnostic tools focus on the stability of the percentiles, or, re-phrasing, quantify 1) the amount of MC-variability and, 2) the amount of variability due to resampling consumption and compound data. By quantifying both quantities, we are able to assess the influence both sources of variability have on the estimated value of the percentiles.
The diagnostics are displayed in a number of graphs (as many as the number of requested percentiles see input screen). For each percentile a graph is available which can be used to draw inference about the optimal number of MC-simulations, the number of resampled sets and the number of simulations per resampled set. Recall that we run the analysis with 100.000 MC-simulations and 100 resampled sets with 20.000 simulations each.
To make inference, we divide the total number of MC-simulations in 2 samples of 50.000 simulations 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 partition, we estimate the percentiles of the available samples and then the variance of the percentiles. So, in partition $n = 1$, the estimate of the variance is based on 2 percentiles derived from samples of size 50.000; in partition $n = 2$ the estimate of the variance is based on 4 percentiles derived from samples of size 25.000, …, etc. The estimated variances of each partition are plotted against the number of MC-simulations per sample of each partition. We expect the variance to decrease as function of sample size, so for larger sample sizes MC-variability decreases. Therefore, through the observed variances a monotone decreasing spline function is fitted. For each variance the 90% confidence limits are calculated.
The uncertainty analysis provides an estimate of the variance of the percentiles derived from the 100 resampled sets of sample size 20.000.
Now the fitted spline function is interpolated to estimate the amount of MC-variability at 20.000 simulations and to calculate the contribution of MC-variability to the total resampling variability at

20.000 simulations. The fitted spline function can be used to estimate the MC-contribution to the total resampling variability for any arbitrary number of MC-simulations.

In the left plot of Figure 72 diagnostics are displayed for percentage point 50%. We can skip this plot because the percentile is 0.0000 microgram/kg bw/day. For percentage point 90% in the right plot we have an estimate of the percentile (0.13733 microgram/kg bw/day) and we are able to make inferences about the stability of the estimate.
On the x-axis the number of MC-simulations is displayed and on the y-axis the variance of the MC-percentile. The estimated variances in each partition are shown as black dots, the fitted monotone decreasing spline function as a black line. For each variance, the 90% confidence interval is indicated by a vertical line segment. The red dotted line indicates the interpolated variance of the bootstrap percentiles as a function of the number of simulations in each bootstrap sample. Note the horizontal black line with the open red boxes at 20.000 simulations, which is the estimate of the variance of the resampled percentiles. The black dot at 200.000 simulations is the extrapolated value for the MC-variability using the spline fit. At 20.000 simulations the MC-variability contributes 0.7% to the total resampling variability. For a theoretical resample of size 200.000 the MC-contribution would be 0.0%.

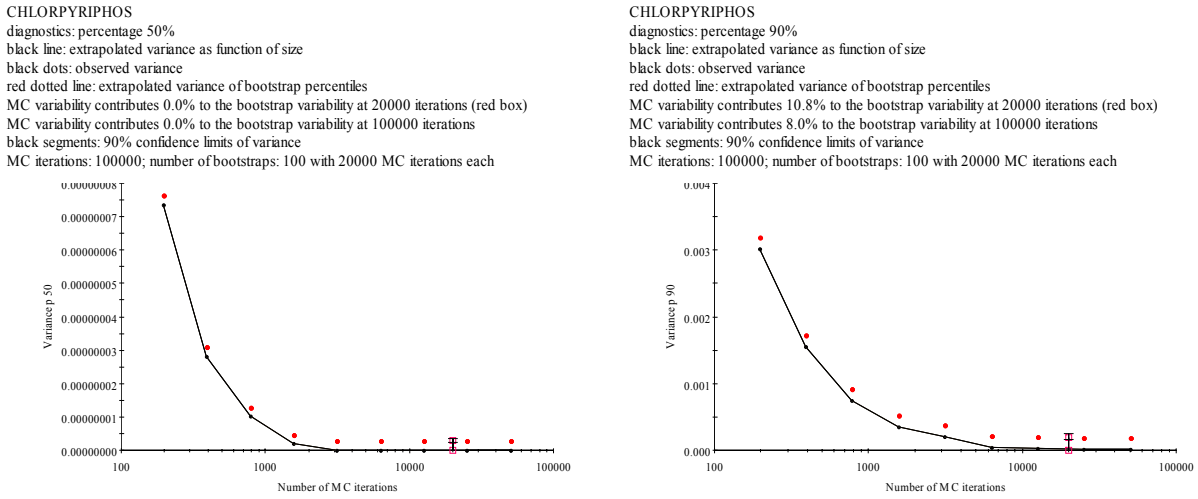


Figure 72: Diagnostic graphs for percentage 50% and 90%

In Figure 73, the diagnostics for percentage point 95% and 99% are displayed. The contribution of the MC-variability for p95 and p99 is 0.8% and 3.4%, respectively, indicating that these higher percentages have stable percentiles.

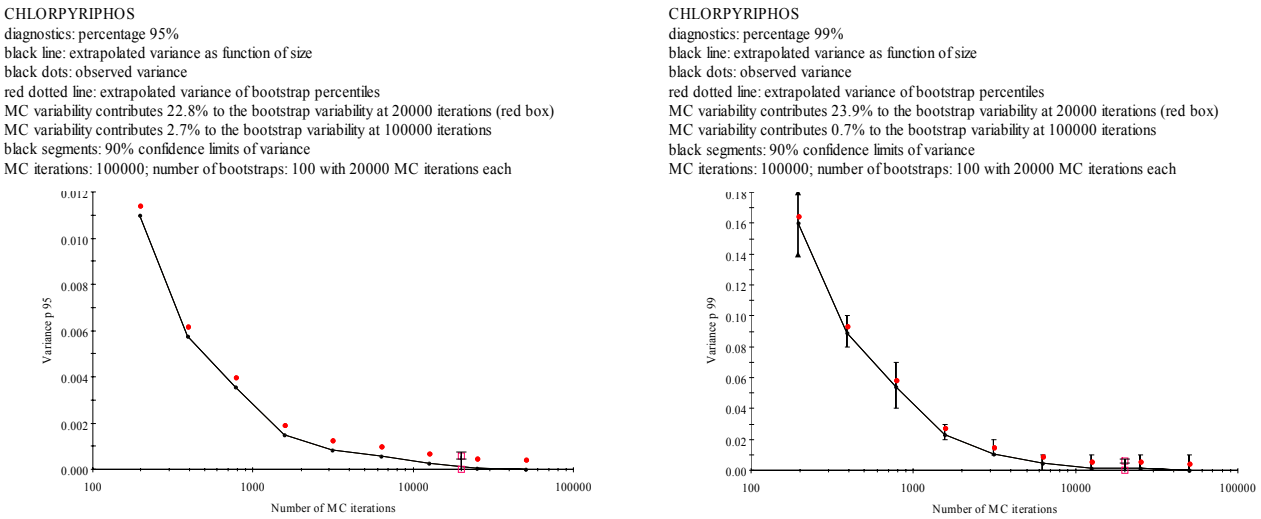


Figure 73: Diagnostic graphs for percentage 95% and 99%

In Figure 74 the diagnostics for percentage point 99.9% and 99.99% are displayed. The contribution of the MC-variability for p99.9 and p99.99 is 40.0% and 33.2%, respectively, indicating that the last percentiles are unstable. The confidence interval at 50.000 MC-simulations is displayed as an arrow for graphical reasons and indicates that a cut off is used. The real confidence interval is much higher. Extrapolation to 200.000 simulations shows that MC-variability contributes 0.3% to the total resampling variability (p99.99 right plot). Note that in Figure 74 not all estimated variances are displayed. As the number of samples in a partition increases, sample size decreases. This restricts the number of available percentiles (maximum possible percentage: (100 –100/size)).

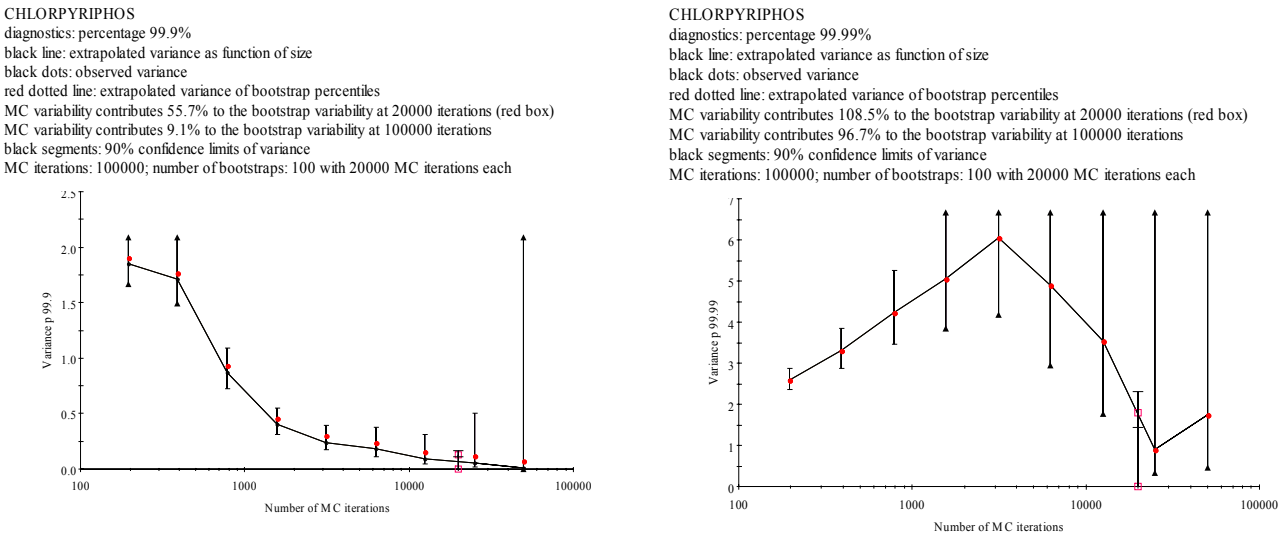


Figure 74: Diagnostic graphs for percentage 99.9% and 99.99%

10.6 Acute risk assessment: betabinomial distribution

In this example, output is shown for an acute risk assessment followed by a betabinomial/normal model to estimate age effects for organo phosphate pesticide Chlorpyrifos. The estimation of an age effect for an acute risk assessment is additional to the standard acute risk assessment (see 4.1). Table 33 lists the main options:

Input form

risk type	acute
uncertainty analysis	no
concentration model	empirical
number of Monte Carlo simulations	100000
unit variability model	no unit variability
random seed	0
intake model	**empirical estimates + betabinomial/normal**
concentration data	system defaults
additional	system defaults
intake model:	
age effect for intake frequency model	yes
function to model effect	polynomial
testing method	backward
minimum degrees of freedom	0
maximum degrees of freedom	4
testing at level	0.01
age effect for intake amount model	yes
transformation	logarithmic
function to model effect	polynomial
testing method	backward
minimum degrees of freedom	0
maximum degrees of freedom	4
testing at level	0.01
Output	system defaults

Table 33: Input form options: acute risk and betabinomial/normal

In Table 34 you find the main characteristics concerning this analysis taken from the 'Additional output' file. The intake frequency function is estimated using a betabinomial model and a polynomial function with 4 degrees of freedom to model age effects. Backward selection is applied meaning that model selection is started with a spline of the highest degree. The model for the logarithmic transformed intake amounts, ln(intake), is based on ML and a polynomial function to model age effects.

'Additional output' file

```
EXPOSURE SECTION
Acute risk assessment
 BetaBinomial/Normal model
  Intake frequency model is based on BetaBinomial model
    No effect of cofactor included
    Include effect of covariable (age)
    Function of covariable          : polynomial
       Minimum degrees of freedom   : 0
       Maximum degrees of freedom   : 4
       DF selection                 : backward
       Testing at level             : 0.01
  Model for intake amounts is based on ML
    No effect of cofactor included
    Include effect of covariable (age)
    Function of covariable          : polynomial
       Minimum degrees of freedom   : 0
       Maximum degrees of freedom   : 4
       DF selection                 : backward
       Testing at level             : 0.01

Intake Frequency: BetaBinomial model with covariable age (no cofactor)
================================================================================
=======

* backward selection of degrees of freedom for polynomial model with Prob = 0.01
```

Df	Ncycle	Phi	_2Loglik	ResDf	Chi	ChiDf	ChiProb
4	4	0.3077	34456.284	6244	2.61	1	0.106
3	4	0.3078	34458.890	6245	34.61	1	0.000
2	4	0.3092	34493.500	6246	-	-	-

- Degrees of freedom of polynomial according to backward deviance testing is 3

Estimates of parameters (autoscaled covariable)

Parameter	estimate	s.e.	t(*)
Constant	0.092192	0.014994	6.15
age Lin	-0.147148	0.014977	-9.82
age Quad	0.277016	0.013934	19.88
age Cub	-0.077561	0.013678	-5.67

Overdispersion phi	0.3078	0.00436	70.58
_2Loglikelihood	34458.89		
Degrees of freedom	6245		

Transformed Intake Amounts: ML model with covariable age (no cofactor)

Maximum likelihood (not REML)

* backward selection of degrees of freedom for polynomial model with Prob = 0.01

Df	Between	Within	_2Loglik	ResDf	Chi	ChiDf	ChiProb
4	2.1632	13.9313	58034.880	9775	3.38	1	0.066
3	2.1657	13.9312	58038.257	9776	2.96	1	0.085
2	2.1675	13.9323	58041.216	9777	32.59	1	0.000
1	2.1952	13.9216	58073.802	9778	-	-	-

* Degrees of freedom of polynomial according to backward deviance testing is 2

Estimates of parameters (autoscaled covariable)

Parameter	estimate	s.e.	t(*)
Constant	-4.057710	0.025295	-160.41
age Lin	-0.065550	0.025105	-2.61
age Quad	0.152223	0.024517	6.21

Between person variance	2.1675
Within person variance	13.9323
_2LogLikelihood	58041.22
Degrees of freedom	9777

Variance of logarithmic transformed intake distribution is: 4.19

***** Acute intake percentiles/percentages are calculated from:

sex	age	Intake probability	Transformed intake amount
-	1	0.7401	-3.6527
-	5	0.6722	-3.7430
.	.	.	.
-	93	0.5514	-3.1919
-	97	0.5451	-3.0515

Acute intake percentiles of population:

sex	age	p50	p90	p95	p99	p99.9	p99.99
-	1	0.02	0.26	0.55	2.30	10.47	48.33
-	5	0.01	0.21	0.46	1.96	9.37	30.29
.	.	.	.	.	.	.	.

-	93	0.02	0.29	0.62	2.67	12.17	41.01
-	97	0.02	0.34	0.73	3.10	16.40	60.88

Acute intake percentages of population lower than intake limit:

sex	age	q0.01	q0.02	q0.04	q0.06	q0.08
-	1	39.17	52.30	65.15	72.01	76.57
-	5	43.34	56.46	68.75	75.28	79.41
.	.	.	.	.	.	.
-	93	38.50	50.98	63.42	70.21	74.80
-	97	36.30	48.46	61.19	68.10	72.76

Table 34: Information in 'Additional output' file

As seen in Table 34, the age effect for the intake frequency is modelled with a polynomial with 3 degrees of freedom. Overdispersion parameter phi is equal to 0.3078, representing between consumer variation.
The ln(intake) amounts are modelled using ML, the fitted polynomial has 2 degrees of freedom. The variance of the ln(intake) distribution is 4.19. Find also the intake probabilities and intake amounts on the transformed scale for several ages (not all shown).
In the left plot of Figure 75 the intake frequency is shown. For 50 age classes, the mean intake frequencies are displayed (black dots). The age effect is represented by the curved line. The red dotted line indicates the 95% confidence interval for the fitted age effect. The blue lines are the 2.5 and 97.5% percentiles of the fitted betabinomial distribution and indicate that the probability of having an intake varies between 0 and 1 with a 0.95 probability.
The right plot displays the ln(intake) distribution. The age effect is small. The black dots represent the mean ln(intake) per age class. To get some idea of the variation in the data, the standard deviation of the distribution is also shown by a vertical line segment with green boxes at the end.

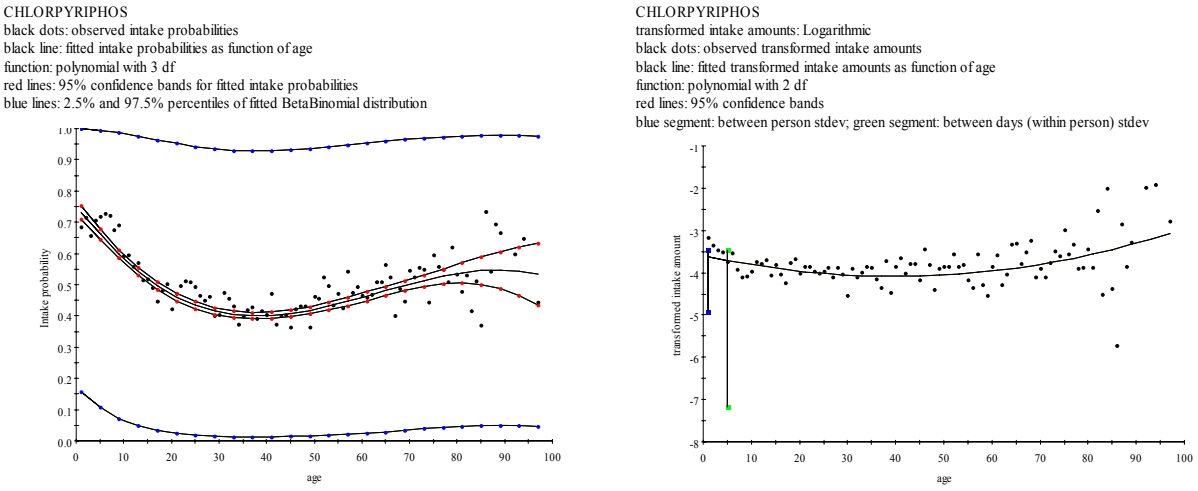


Figure 75: Age dependent intake frequency and ln(intake) distribution

In Figure 76, age dependent percentiles (derived from percentages) and percentages (derived from percentiles) are shown. The same information is found as tabular output (not shown).

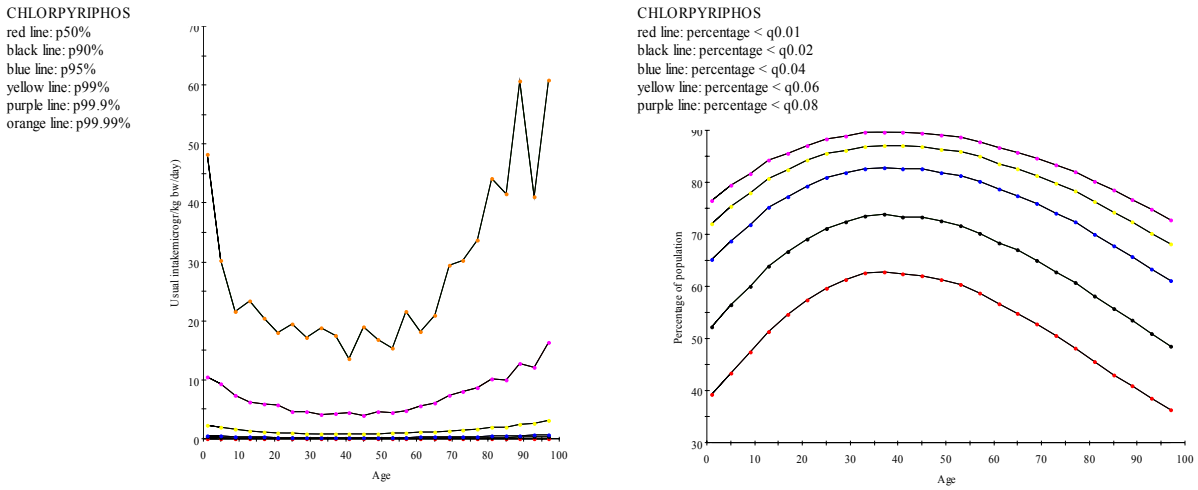


Figure 76: Age dependent percentiles and percentages

Find in Table 35 a short summary of an acute risk assessment using age as explanatory variable.

```
------------------------------------------------------------------------------
Short term exposure (acute intake)

Compound:  CHLORPYRIPHOS

------------------------------------------------------------------------------

Transformation to normality
Logarithmic transformation
------------------------------------------------------------------------------
Anderson-Darling test for Normality
Test statistic                          : 521.76
p-value                                 : 0

------------------------------------------------------------------------------
Variability within and between individuals
Variance components (transformed scale/unit variance)
             within individuals         : 0.86
            between individuals         : 0.14
                          ratio         : 6.16

Variance of the transformed distribution : 4.19
------------------------------------------------------------------------------
```

Table 35: Technical information

10.7 Acute risk assessment: binomial/lognormal with pooling

Example 1

In this example, output is shown for an acute risk assessment for organo phosphate pesticide Chlorpyrifos. Table 36 lists the main options:

Input form	
risk type	acute
uncertainty analysis	no
concentration model	**binomial/lognormal (with pooling)**
number of Monte Carlo simulations	100000

unit variability model	no unit variability
random seed	0
intake model	only empirical estimates
concentration data	system defaults
Additional	system defaults
Output	system defaults

Table 36: Input form options: binomial with pooling

For information in the 'Additional output' file, tables and figures, see 4.1 .
When specifying the binomial/lognormal distribution with pooling, a parametric form of modelling is used to simulate data and estimate the intake percentiles. For each food, the positives samples are taken to estimate the variance and mean on the lognormal scale. If pooling is requested, food groups need to be formed. Each food is characterised by a hierarchical food code. For CODEX codes, the first 2 characters in combination with factor Allowed define a group used in pooling. Factor Allowed indicates whether a chemical substance is allowed on a food or not. If the code is a subtype (*e.g.* X$y, the supertype is taken to form foodgroups (*e.g.* X).
Table 37 illustrates the pooling procedure. There are 6 food groups (ProdGr): 5 groups with a single food and one group with 4 foods: group FC contains foods Lemon, Orange, Mandarin and Grapefruit. Within a foodgroup variances and means of foods are pooled.
In the example, the original mean (Mean) and sigma (StdDev) are displayed together with the number of observations (nos). In columns Mu and Sigma, you find the result of pooling: parameters μ and σ of the lognormal distribution; the pooled number of degrees of freedom is in column Df.
In group FC, standard deviations are 0.94, 0.26, 0.65 and 0.66. A test of homogeneity of variances is not significant so the pooled standard deviation becomes 0.68. A test of homogeneity of means is significant so means are not pooled. For Potato no standard deviation is estimated because only one positive sample is available. In a second step, standard deviations of foods with less than 10 degrees of freedom are replaced by the overall variance if the test of homogeneity is not significant. For these foods, the pooled or original standard deviation is replaced by the value 0.75. Note, that for Peach (StdDev = 0.33, n = 4) the standard deviation is replaced by the overall standard deviation, but for Lemon (Sigma = 0.68, StdDev = 0.26) no replacement of sigma occurs: the test on homogeneity of variance is not significant due to the high degrees of freedom of the pooled estimate (Df = 151). AD is the value of the Anderson-Darling test statistic for Normality and ER indicates if the statistic is significant or not.

```
--------------------------------------------------------------------------------
Compound: CHLORPYRIFOS
--------------------------------------------------------------------------------
Summary of calculations and input of a PARAMETRIC SIMULATION
>0        : number of detects
nos       : total number of measurements, detects and non-detects
Frpos     : fraction of detects
Mu        : parameter mu of the lognormal distribution
Mean      : original means per food before pooling
Sigma     : parameter sigma of the lognormal distribution
StdDev    : original st.dev. per food before pooling
AD        : Anderson-Darling statistic
ER        : significance level of AD-statistic:
            ns = hypothesis of Normality not rejected
             s = hypothesis of Normality rejected
Df        : degrees of freedom of pooled sigma
Group     : combination of foodgroup and allowed
ProdGr    : foodgroup
Allwd     : code if compound is allowed on food (1) or not (0)
Food      : foodlabels
--------------------------------------------------------------------------------
  >0  nos  Frpos     Mu   Mean Sigma StdDev    AD   ER   Df Group ProdGr Allwd  Food
  54  501   0.11  -2.35  -2.35  0.96   0.96  1.12    s   53     1      FB    1  TABLE-GRAPES
  13   41   0.32  -2.02  -2.02  0.68   0.94  0.84    s  151     2      FC    1  GRAPEFRUIT,
   4   39   0.10  -2.84  -2.84  0.68   0.26  0.20   ns  151     2      FC    1  LEMON, SEE A
  44  105   0.42  -2.13  -2.13  0.68   0.65  0.19   ns  151     2      FC    1  MANDARIN, SE
  94  221   0.43  -2.38  -2.38  0.68   0.66  0.34   ns  151     2      FC    1  ORANGE, SWEE
  10  241   0.04  -3.20  -3.20  0.27   0.27  0.61    s    9     3      FP    1  APPLE
```

4	62	0.06	-3.26	-3.26	0.75	0.33	0.16	ns	217	4	FS	1	PEACH
2	111	0.02	-1.77	-1.77	0.75	1.06	-0.81	ns	217	5	VO	1	PEPPERS, SWE
1	40	0.03	-2.81	-2.81	0.75	.	.	.	217	6	VR	1	POTATO

Table 37: Pooling information

In Table 38 the percentiles for a parametric model are displayed. Note that the p99.99 changed from 6.69 to 7.06 microgram/kg bw/day.

```
Random sampling is based on seed :         0
Number of simulations (consumers):   100000  out of   6250
------------------------------------------------------------------------------
CHLORPYRIFOS (microgr/kg bw/day)  consumption:   100000 out of 100000


 Compound:  CHLORPYRIFOS
------------------------------------------------------------------------------
     Percentiles, maximum and average intake

   Percentage     Percentiles of CHLORPYRIFOS  (microgram/kg bw/day)
        50.00        0.00000
        90.00        0.20379
        95.00        0.40437
        99.00        1.15487
        99.90        3.26754
        99.99        7.05740
         mean        0.07323
     maximum       12.97748
```

Table 38: Percentiles for the acute intake distribution using the binomial/lognormal with pooling

Example 2

The variances and means of foods are pooled within a food group. Food groups are formed based on the first two characters in combination with factor Allowed (see example, factor Levels (= food code) and Allowed). Factor Allowed indicates whether a chemical compound is allowed on a product or not. Foodgroup 10101 contains foods Bean and Sperzieboon. Foodgroup 10201 is split into two subgroups (according to factor Allowed): one group with Chicory, Endive, Cabbage lettuce and Curly lettuce and a group with Roodlof and Spinach. On the last two foods the use of a chemical compound is not allowed.

In the example, the original mean and sigma are displayed together with the number of observations. The last three columns show the parameters μ and σ of the lognormal distribution together with the degrees of freedom after pooling. For Bean and Sperzieboon, a pooled μ (= -1.67) and σ (= 1.31*)* are used. Chicory, Endive, Cabbage lettuce and Curly lettuce only sigma is pooled (σ = 1.47), the original means are maintained. For Roodlof (1 observation) the overall sigma (= 1.36) is used to estimated the variance.

Labels	Levels	Allowed	Original Sigma	Mean	Nobs	Automatically pooling Sigma	mean	Nobs
BEAN	10101	1	1.60	-1.17	8	1.31	-1.67	12
SPERZIEBOON	10101	1	0.75	-2.33	6	1.31	-1.67	12
CHICORY	10201	1	1.38	-2.69	4	1.47	-2.69	382
ROODLOF	10201	0	*	-2.3	1	1.36	-2.3	729
ENDIVE	10201	1	1.52	-0.91	92	1.47	-0.91	382
CABBAGE LETTUCE	10201	1	1.46	-1.44	286	1.47	-1.44	382
CURLY LETTUCE	10201	1	1.08	-2.14	4	1.47	-2.14	382
SPINACH	10201	0	1.18	-0.57	10	1.36	-0.57	729
BRUSSELS SPROUT	10301	0	1.14	-2.7	2	1.36	-2.7	729
CHINESE CABBAGE	10301	1	1.62	-2.32	21	1.62	-2.32	20
OXHEART/CONICAL	10301	0	*	-2.3	1	1.36	-2.3	729
ONION (SMALL)	10301	1	0.08	-1.66	2	0.08	-1.66	1
FENNEL	10301	1	0.16	-2.38	3	0.16	-2.38	2
POTATO	10401	0	0.62	0.19	2	0.59	0.19	50
WINTER CARROT	10401	0	0.62	-2.55	14	0.59	-2.55	50
CARROT	10401	0	0.54	-2.71	36	0.59	-2.71	50
RADISH	10401	1	1.52	-2.91	6	1.36	-2.91	729
CELERIAC	10401	0	1.31	-2.07	2	0.59	-2.07	50
GRAPE	10501	0	1.14	-1.06	25	1.14	-1.06	24
STRAWBERRY	10501	1	1.14	-1.57	169	1.14	-1.57	168
RASPBERRY	10501	1	1.73	-1.04	9	1.36	-1.04	729
BLACKBERRY	10501	1	1.15	-0.89	17	1.15	-0.89	16
BLUE BERRY	10501	1	1.83	-1.24	3	1.36	-1.24	729
CURRANT	10501	1	1.87	-0.62	30	1.87	-0.62	29
OTHER FRUIT, NUT	10601	0	*	-1.51	1	1.36	-1.51	729

10.8 Chronic risk assessment: discrete/semi-parametric (ISUF)

In this example output is shown for a chronic risk assessment for aspartaam. In 4.2 the same data are analyzed with another model. The summaries of the data before modeling can there be found. Table 39 lists the main options:

Input form	
Risk type	chronic
uncertainty analysis	no
concentration model	empirical
intake model	discrete/semi-parametric (ISUF)
concentration data	system defaults
intake model:	
transformation	power
spline fit	yes
number of simulations to estimate frequency distribution	10
number of bins for discretisation	20
Output	system defaults

Table 39: Input form options: discrete/semi-parametric (Nusser)

In Table 40 you find the main characteristics concerning this analysis taken from the 'Additional output' file.

'Additional output' file	
Number of products	1
Acute reference dose (ARfD)	*
Acceptable daily intake (ADI)	5
Number of detects	113
Number of non-detects	0
No of consumers	6250
Population characteristics,	
minimum age	1

```
      maximum age                    97
      minimum weight                 8
      maximum weight                 150
      Sex                            female, male
Total no of consumption days         5136
Replace all non-detects
Multiplication factor for LOR        0.5
INTAKE SECTION
Chronic Risk Assessment
 Discrete/semi-parametric (Nusser)
    Power transformation
    followed by spline transformation
     No. of iterations to estimate theta: 5000
     No. of binomial proportions (M): 20
```

Table 40: Information in 'Additional output' file

Find in Figure 77 the untransformed intake distribution (left plot). The right plot shows the distribution after a power transformation (with exponent 0.148).

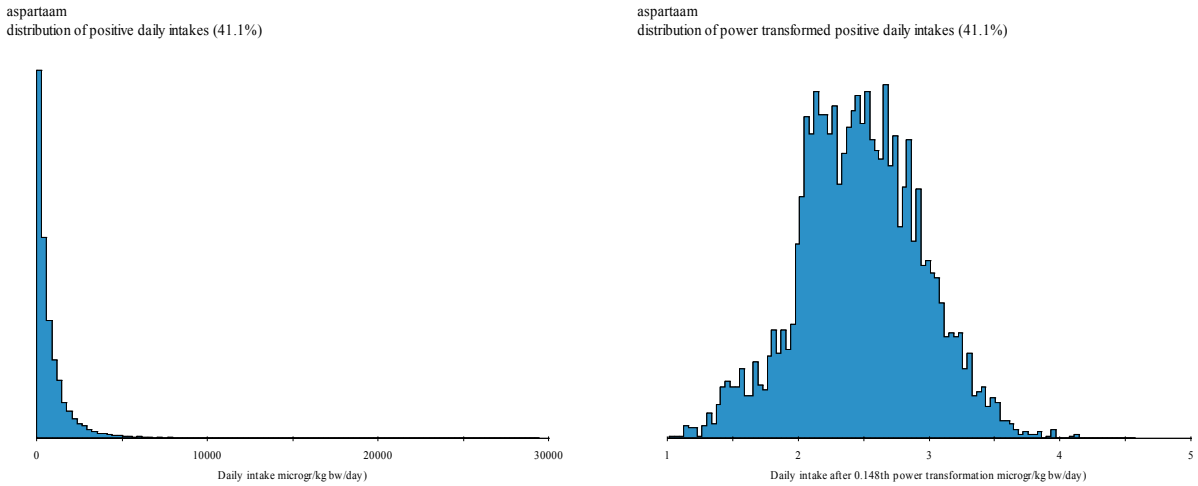


Figure 77: Untransformed and power transformed intake distribution

The Nusser method additionally fits a spline function to the power transformed values as a function of the normal Blom scores, see left plot Figure 78. The result of a power and spline transformation of the positive intakes is shown in the right plot.

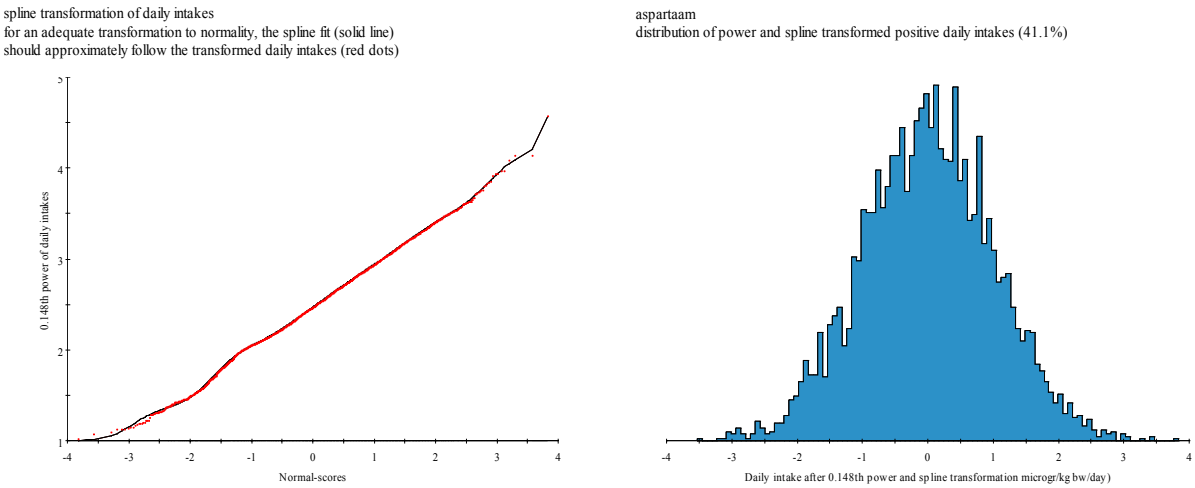


Figure 78: Transformation plot and power and spline transformed intakes

The fit seems adequate, which is confirmed by the QQ plot shown in Figure 79. The Anderson-Darling test on the transformed non-zero daily intakes is not significant.

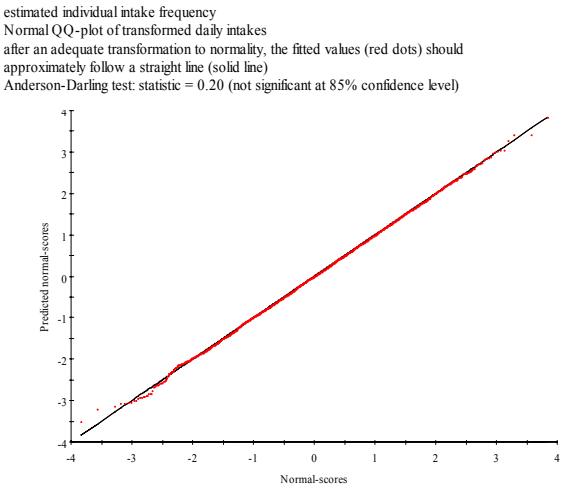


Figure 79: Normal probability plot

In Table 41 you find details on the spline functions (11 knots) and the Anderson-Darling test result (0.20), which shows no significant deviation from normality for the transformed values (p-value = 0.90). In the same output section you also find a test whether there is evidence for heterogeneity of variance between consumers, MA4=5.8447, p-value=0.0000.
Based on the daily intake the variance components are estimated, reported as within consumers is 0.30 and between consumers is 0.70. Therefore in this case the day-to-day variation in aspartaam intake is much lower than the variation between individual consumers.

```
------------------------------------------------------------------------------
Long term exposure (usual intake)

Compound:  aspartaam

------------------------------------------------------------------------------

Transformation to normality
```

```
Power transformation                        : 0.148
Number of knots spline function             : 11
-------------------------------------------------------------------------
Anderson-Darling test for Normality
Test statistic                              : 0.20
p-value                                     : 0.9

-------------------------------------------------------------------------
Heterogeneity of variance between individuals

Test statistic (MA4)                        : 5.8221   (3 for homogeneous
variances)
p-value                                     : 0.0000
-------------------------------------------------------------------------
Variability within and between individuals
Variance components (transformed scale/unit variance)
             within individuals             : 0.30
            between individuals             : 0.70
                          ratio             : 0.43

Intake frequency

Estimated fraction non-consumers (theta_0): 0.2818
```

Table 41: Technical information on usual intake

The next step is to estimate the intake frequency distribution. The estimated fraction of non-consumers theta_0 (consumers with zero usual intake) is estimated to be 0.281. In output file theta.txt (which you get after downloading results) you find the full results, see Table 42. So the fraction of consumers that always eats foods with aspartaam (p=1.00) is estimated to be 0.0601.

p	theta	
0.00	0.28183499	(= theta0)
0.05	0.03651911	
0.10	0.03631187	
0.15	0.03429980	
0.20	0.03298592	
0.25	0.02907445	
0.30	0.02736419	
0.35	0.02595573	
0.40	0.02525151	
0.45	0.02505030	
0.50	0.02535211	
0.55	0.02625755	
0.60	0.02766600	
0.65	0.02987928	
0.70	0.03279678	
0.75	0.03712274	
0.80	0.04195171	
0.85	0.04708249	
0.90	0.05191147	
0.95	0.05694165	
1.00	0.06006036	

Table 42: Estimates of the intake frequency distribution, theta

Graphical results of the values from Table 42 are shown in Figure 80.

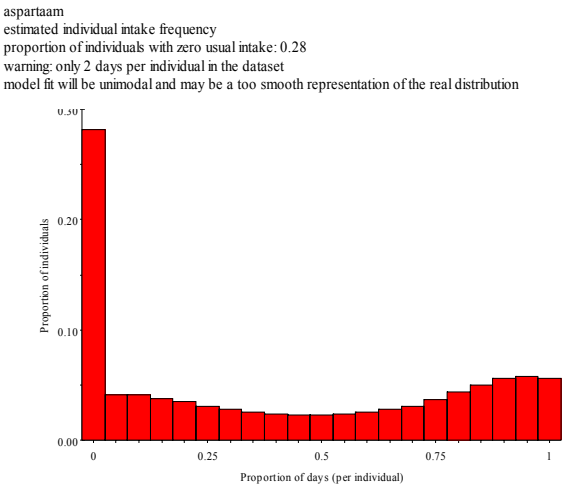


Figure 80: Intake frequency distribution

Finally three cumulative distributions are estimated: for intake days only, for intakers only, and for the total population, see the left plot of Figure 81. For for intake days only and for intakers only the distributions are also shown as probability densities in the right plot.

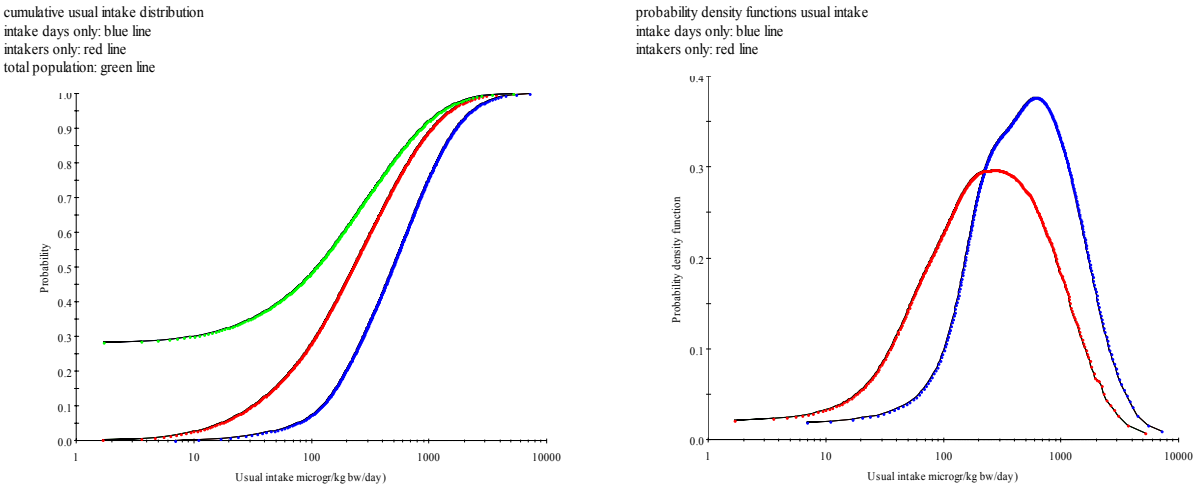


Figure 81: Cumulative usual intake and density functions

```
--------------------------------------------------------------------------------
Long term exposure (usual intake)

Compound:  aspartaam

--------------------------------------------------------------------------------

          Percentage                 Total
                                  population
               50.00               109.6090
               90.00               847.0583
               95.00              1269.0260
```

99.00	2447.1316
99.90	4078.5759
99.99	5166.2093

Table 43: Percentiles for usual intake distribution: entire population

In Table 44 you find information on selected percentiles for the other types of distributions.

Short term exposure

Percentage	Intake days only	Intakers only
50.00	487.8314	224.2399
90.00	1674.2954	1042.5254
95.00	2280.0644	1489.2372
99.00	3946.3718	2699.3570
99.90	7087.3211	4350.4755
99.99	10282.0186	5166.2093

Table 44: Technical information on usual intake: percentiles for usual intake distribution, positives only and consumers only

10.9 Chronic risk assessment: betabinomial/normal with covariable age and uncertainty

In this example output is shown for a chronic risk assessment for aspartaam modeling the effect of age on the usual intake. Table 45 lists the main options:

Input form	
risk type	chronic
uncertainty analysis	yes
concentration model	empirical
Number of MC simulations	10000
intake model	betabinomial/normal
concentration data	system defaults
INTAKE FREQUENCY MODEL:	
age effect	yes
function to model age effect	polynomial
minimum degrees of freedom	0
maximum degrees of freedom	4
testing method	backward
testing at level	0.01
INTAKE AMOUNT MODEL:	yes
transformation	logarithm
age effect	yes
function to model effect	polynomial
minimum degrees of freedom	0
maximum degrees of freedom	4
testing method	backward
testing at level	0.01
Output	system defaults

Table 45: Input form options: betabinomial/normal with covariable age and uncertainty

In Table 46 you find the main characteristics concerning this analysis taken from the 'Additional output' file.
The intake frequency function is estimated with a betabinomial model using a polynomial function to model age effects. Backward selection is applied meaning that model selection is started with a

polynomial of degree 4. The logarithmically transformed intake amounts are modeled using ML and a polynomial to describe age effects.

'Additional output' file

```
Chronic risk assessment
 BetaBinomial/Normal model
  Intake frequency model is based on BetaBinomial model
    No effect of cofactor included
    Include effect of covariable (age)
    Function of covariable          : polynomial
       Minimum degrees of freedom   : 0
       Maximum degrees of freedom   : 4
       DF selection                 : backward
       Testing at level             : 0.01
  Model for intake amounts is based on ML
    No effect of cofactor included
    Transformation                  : Logarithmic
    Include effect of covariable (age)
    Function of covariable          : polynomial
       Minimum degrees of freedom   : 0
       Maximum degrees of freedom   : 4
       DF selection                 : backward
       Testing at level             : 0.01


Intake Frequency: BetaBinomial model with covariable age (no cofactor)
================================================================================

* backward selection of degrees of freedom for polynomial model with Prob =
0.01

   Df  Ncycle    Phi       _2Loglik       ResDf       Chi    ChiDf     ChiProb
    4       4  .5106      11949.208        6244     27.41        1       0.000
    3       6  .5121      11976.616        6245         -        -           -

* Degrees of freedom of polynomial according to backward deviance testing is 4

Estimates of parameters (autoscaled covariable)
-----------------------------------------------

                     Parameter       estimate      s.e.      t(*)
                      Constant      -0.758153  0.053550    -14.16
                     age Lin        -0.970010  0.082574    -11.75
                     age Quad       -0.021940  0.097746     -0.22
                     age Cub        -0.305835  0.089282     -3.43
                     age Quart      -0.304492  0.060303     -5.05


            Overdispersion phi         0.5106   0.01171     43.61
               _2Loglikelihood        11949.21
            Degrees of freedom            6244



Transformed Intake Amounts: ML model with covariable age (no cofactor)
================================================================================

Maximum likelihood (not REML)
* backward selection of degrees of freedom for polynomial model with Prob = .01
```

Df Between	Within	_2Loglik	ResDf	Chi	ChiDf	ChiProb
4 1.1825	0.5379	16133.771	5130	0.55	1	0.457
3 1.1827	0.5379	16134.325	5131	19.00	1	0.000
2 1.1927	0.5375	16153.323	5132	-	-	-

* Degrees of freedom of polynomial according to backward deviance testing is 3

Estimates of parameters (autoscaled covariable)

Parameter	estimate	s.e.	t(*)
Constant	5.939241	0.021963	270.42
age Lin	-0.185661	0.022033	-8.43
age Quad	0.161915	0.020330	7.96
age Cub	-0.086084	0.019717	-4.37

Between person variance	1.1827
Within person variance	0.5379
_2LogLikelihood	16134.33
Degrees of freedom	5131

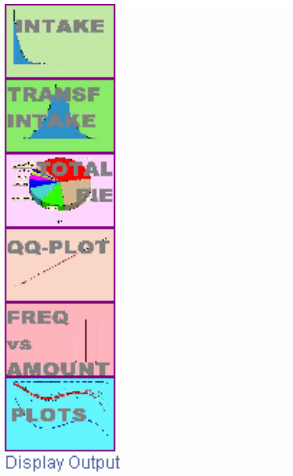
***** Usual intake percentiles/percentages are calculated from:

sex	age	Intake probability	Transformed intake amount
-	1	0.6996	6.6524
-	5	0.7055	6.3967
-	9	0.6881	6.1879
-	13	0.6524	6.0219
-	17	0.6025	5.8946
-	21	0.5435	5.8017
-	25	0.4813	5.7391
-	29	0.4218	5.7027
-	33	0.3697	5.6882
-	37	0.3273	5.6916
-	41	0.2952	5.7085
-	45	0.2727	5.7350
-	49	0.2586	5.7668
-	53	0.2510	5.7998
-	57	0.2481	5.8297
-	61	0.2473	5.8525
-	65	0.2457	5.8640
-	69	0.2397	5.8599
-	73	0.2255	5.8362
-	77	0.1998	5.7887
-	81	0.1616	5.7132
-	85	0.1147	5.6056
-	89	0.0680	5.4616
-	93	0.0320	5.2772
-	97	0.0114	5.0481

Table 46: Information in 'Additional output' file

As seen in Table 46, an age effect is found for the intake frequency. A polynomial with 4 degrees of freedom is fitted and overdispersion parameter phi is equal to 0.5106. Parameter phi represents between consumer variation. So each consumer has its own probability of having an intake. This probability is sampled from a betabinomial distribution with age dependent probabilities and dispersion factor phi. For the intake amount model a polynomial model with 3 degrees of freedom is fitted. Find in the output, the estimated intake probabilities and transformed intake amounts dependent on age

Click icon 'Plots' to get the next screen (see Figure 82).



Display plots of:

Frequency model
Probability of intake vs. age

Transformed intake amount model
Transformed intake amount vs. age

Percentiles of usual intake vs. age
p50% p90% p95% p99% p99.9% p99.99% percentiles of usual intake vs. age

Confidence intervals for percentiles of usual intake vs. age
Confidence intervals: 50% percentile of usual intake vs. age
Confidence intervals: 90% percentile of usual intake vs. age
Confidence intervals: 95% percentile of usual intake vs. age
Confidence intervals: 99% percentile of usual intake vs. age
Confidence intervals: 99.9% percentile of usual intake vs. age
Confidence intervals: 99.99% percentile of usual intake vs. age

Percentage of population < usual intake limits vs. age
Percentage of population < q0.01 q0.02 q0.04 q0.06 q0.08 of usual intake vs. age

Confidence intervals for percentage of population < usual intake limits vs. age
Confidence intervals: percentage of population < 0.01 of usual intake vs. age
Confidence intervals: percentage of population < 0.02 of usual intake vs. age
Confidence intervals: percentage of population < 0.04 of usual intake vs. age
Confidence intervals: percentage of population < 0.06 of usual intake vs. age
Confidence intervals: percentage of population < 0.08 of usual intake vs. age

Figure 82: Diversity of possible plots

The first two plots are shown in Figure 83.

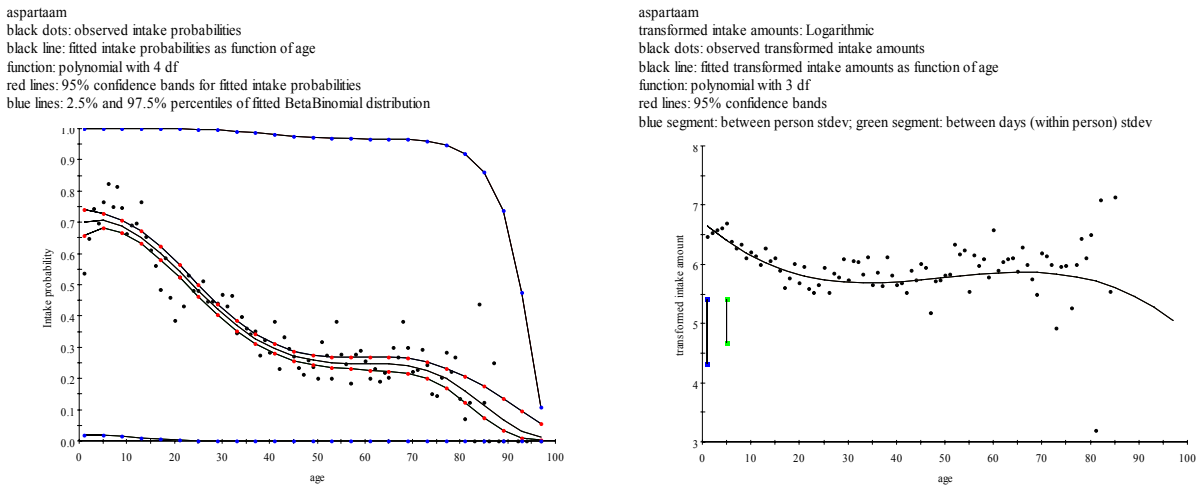


Figure 83: Age dependent intake frequency (left) and intake amount (right)

In the left plot of Figure 83 the intake frequency is shown. For 50 age classes, the mean intake frequencies are displayed (black dots). The fitted age effect is plotted through the dots as a black line. As seen, the probability of having an intake decreases with age. The red dotted line indicates the 95% confidence interval for the fitted age effect. Note the wider intervals at the edges of the plot showing that for very old and young consumers less information is available. The blue lines are the 2.5 and 97.5% percentiles of the fitted betabinomial distribution: the sampled intake probability of consumers according to the betabinomial are within these lines and may vary from about 0.04 till about 1 for young people and 0 till about 0.1 for old people.

The right plot of Figure 83 displays the ln(intake) distribution. The age effect is plotted as a black line through the observed mean ln(intake) amounts per age class. The components of variance for the between consumers and within consumers variation are 1.18 and 0.54, respectively (on the logarithmically transformed scale). The standard deviation of the between and within consumer variation are indicated by the vertical blue and green line segments.
From these models the usual intake is calculated. Click 'Display output' to get summaries of the analysis.
In Table 47, age dependent percentiles of the usual intake are shown together with uncertainty limits based on resampling.

Usual intake percentiles of total population for percentiles (p_):

sex	age	p50	p90	p95	p99	p99.9	p99.99
-	1	646.68	3046.04	4649.36	10163.81	24096.67	51706.86
-	5	505.59	2382.62	3670.83	8001.49	18413.50	33718.23
-	9	394.50	1908.54	2907.32	6327.66	14923.00	24587.06
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
-	93	0.00	15.30	77.61	421.97	1523.45	3870.07
-	97	0.00	0.03	3.75	139.71	686.38	1992.09

Uncertainty of percentiles distribution based on 100 resampled sets

p	sex	age						
est:	-	1	646.68	3046.04	4649.36	10163.81	24096.67	51706.86
2.5%	-	1	453.14	2109.81	3193.76	6905.88	15574.14	31639.97
25%	-	1	555.70	2558.83	3828.65	8304.43	19041.04	40613.89
75%	-	1	631.42	2936.67	4479.25	9865.74	23408.42	52559.19
97.5%	-	1	692.20	3115.41	4843.41	11113.77	28112.13	68942.51
est:	-	5	505.59	2382.62	3670.83	8001.49	18413.50	33718.23
2.5%	-	5	386.87	1851.10	2804.74	5971.28	13399.11	26905.62
25%	-	5	449.80	2129.26	3237.93	6949.38	15491.97	35066.43
75%	-	5	499.12	2431.88	3767.18	8330.75	19167.11	47158.12
97.5%	-	5	527.81	2612.19	4072.87	9485.99	23902.51	59468.72
est:	-	9	394.50	1908.54	2907.32	6327.66	14923.00	24587.06
2.5%	-	9	318.16	1569.25	2363.32	5067.06	11047.47	18723.74
25%	-	9	364.98	1757.08	2668.94	5768.92	12871.17	22382.79
75%	-	9	400.94	2019.04	3122.20	6957.87	16240.96	28205.90
97.5%	-	9	420.63	2142.14	3373.49	7867.73	19130.75	33821.71
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
est:	-	93	0.00	15.30	77.61	421.97	1523.45	3870.07
2.5%	-	93	4.24	273.58	492.58	1294.96	3931.73	8931.78
25%	-	93	25.69	769.90	1367.70	3633.35	9695.44	19557.33
75%	-	93	62.67	1112.89	1959.62	5092.72	13736.59	30358.94
97.5%	-	93	127.01	1699.63	2976.02	7730.14	20340.48	46220.44
est:	-	97	0.00	0.03	3.75	139.71	686.38	1992.09
2.5%	-	97	3.27	260.96	463.10	1238.96	3397.27	7698.74
25%	-	97	36.28	928.28	1652.30	4332.58	11612.30	22339.64
75%	-	97	102.59	1461.53	2534.35	6527.55	17633.98	37805.35
97.5%	-	97	216.36	2316.63	4036.64	10322.23	27333.32	57025.29

Table 47: Percentiles of usual intakes and uncertainty intervals

In Table 48 age dependent percentages of consumers lower than specified limits of usual intake are shown also with uncertainty limits based on resampling.

Usual intake percentages of total population lower than intake limit (q_):

sex	age	q0.01	q0.02	q0.04	q0.06	q0.08
-	1	0.03	0.03	0.05	0.06	0.08
-	5	0.02	0.03	0.05	0.05	0.08

	-	9	0.02	0.04	0.07	0.09	0.10
	.	.	.	.	.	.	.
	.	.	.	.	.	.	.
	-	93	71.78	73.32	74.91	75.85	76.56
	-	97	89.10	89.77	90.41	90.81	91.09

Uncertainty distribution of intake percentages of total population lower than intake limit (q_):
based on 100 resampled sets

q	sex	age					
est:	-	1	0.03	0.03	0.05	0.06	0.08
2.5%	-	1	0.00	0.00	0.01	0.01	0.02
25%	-	1	0.01	0.01	0.02	0.02	0.03
75%	-	1	0.01	0.02	0.03	0.04	0.05
97.5%	-	1	0.03	0.03	0.05	0.06	0.08
est:	-	5	0.02	0.03	0.05	0.05	0.08
2.5%	-	5	0.01	0.01	0.02	0.03	0.04
25%	-	5	0.01	0.02	0.03	0.05	0.06
75%	-	5	0.02	0.04	0.06	0.07	0.09
97.5%	-	5	0.04	0.06	0.09	0.12	0.14
est:	-	9	0.02	0.04	0.07	0.09	0.10
2.5%	-	9	0.01	0.03	0.05	0.07	0.08
25%	-	9	0.03	0.04	0.07	0.10	0.12
75%	-	9	0.05	0.08	0.11	0.15	0.18
97.5%	-	9	0.07	0.10	0.16	0.20	0.24
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
est:	-	93	71.78	73.32	74.91	75.85	76.56
2.5%	-	93	2.21	2.80	3.49	4.00	4.40
25%	-	93	5.31	6.31	7.51	8.35	9.02
75%	-	93	10.60	12.09	13.83	15.04	15.98
97.5%	-	93	21.01	23.26	25.74	27.37	28.47
est:	-	97	89.10	89.77	90.41	90.81	91.09
2.5%	-	97	1.04	1.37	1.81	2.11	2.36
25%	-	97	3.67	4.46	5.41	6.03	6.54
75%	-	97	9.44	10.87	12.47	13.51	14.32
97.5%	-	97	22.44	24.77	27.31	28.90	30.09

Table 48: Percentages of consumers lower than specified limits and uncertainty intervals

Figure 84 shows the age dependent percentiles and percentage for the usual intake distribution.

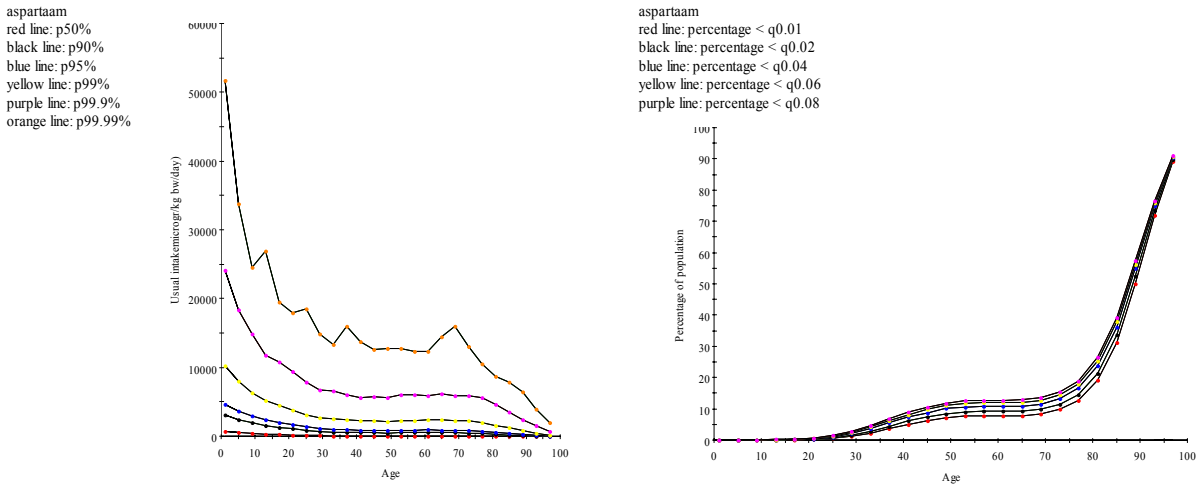


Figure 84: Age dependent percentiles derived from percentages and percentages derived from percentiles

In Table 49, a report of technical information on the analysis is shown.

```
------------------------------------------------------------------------------
Long term exposure (usual intake)

Compound:  aspartaam

------------------------------------------------------------------------------

Transformation to normality
Logarithmic transformation
------------------------------------------------------------------------------
Anderson-Darling test for Normality
Test statistic                          : 22.67
p-value                                 : 0

------------------------------------------------------------------------------
Variability within and between individuals
Variance components (transformed scale/unit variance)
             within individuals         : 0.31
            between individuals         : 0.69
                          ratio         : 0.45

------------------------------------------------------------------------------
```

Table 49: Technical information on usual intake

10.10 Chronic risk assessment: betabinomial/normal with covariable age and cofactor sex

In this example output is shown for a chronic risk assessment for aspartaam modeling the effect of age and sex. Table 50 lists the main options:

Input form	
risk type	chronic
uncertainty analysis	yes
Random seed	0
concentration model	empirical
number of MC simulations	100000
intake model	betabinomial/normal with age and sex
concentration data	system defaults
cofactor	sex
covariable	age
interaction	no
INTAKE FREQUENCY MODEL	
sex effect	yes
age effect	yes
function to model effect	polynomial
testing method	backward
minimum degrees of freedom	0
maximum degrees of freedom	4
testing at level	0.01

INTAKE AMOUNT MODEL	
transformation	logarithmic
sex effect	yes
age effect	yes
function to model effect	polynomial
testing method	backward
minimum degrees of freedom	0
maximum degrees of freedom	4
testing at level	0.01
Output	system defaults

Table 50: Input form options: betabinomial/normal with covariable age and cofactor sex

In Table 51 you find the main characteristics concerning this analysis taken from the 'Additional output' file. The intake frequency function is estimated with a betabinomial model using a polynomial function with 4 degrees of freedom to model the age effect. Backward selection is applied meaning that model selection is started with a spline of the highest degree. To model the transformed intake amounts a polynomial with 3 degrees of freedom is used. There is a clear sex effect for the intake frequency model, whereas the sex effect for the intake amount model has a P-value of 6.6%.

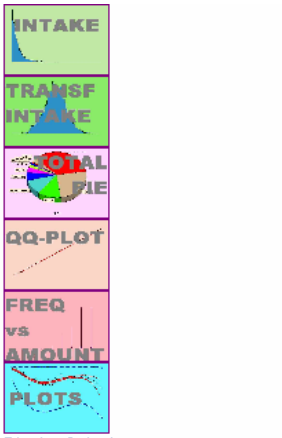
'Additional output' file

```
Intake Frequency: BetaBinomial model with cofactor sex and covariable age (no
interaction)
================================================================================

* backward selection of degrees of freedom for polynomial model with Prob =
0.01

   Df  Ncycle      Phi     _2Loglik   ResDf      Chi       ChiDf     ChiProb
    4       4   0.5092   11927.137    6243       27.96        1       0.000
    3       6   0.5108   11955.100    6244           -        -           -

* Degrees of freedom of polynomial according to backward deviance testing is
4

Estimates of parameters (autoscaled covariable)
-----------------------------------------------

                     Parameter       estimate      s.e.      t(*)
                      Constant      -0.664664  0.056905    -11.68
                     age Lin        -0.986971  0.082685    -11.94
                     age Quad       -0.032632  0.097706     -0.33
                     age Cub        -0.315594  0.089219     -3.54
                     age Quart      -0.306695  0.060141     -5.10
                      sex male      -0.224418  0.047878     -4.69

Parameters for factors are differences compared with the reference level:
         Factor   Reference level
            sex            female


            Overdispersion phi         0.5092   0.01173     43.40
               _2Loglikelihood       11927.14
            Degrees of freedom           6243



Transformed Intake Amounts: ML model with cofactor sex and covariable age (no
interaction)
```

```
================================================================================

Maximum likelihood (not REML)
* backward selection of degrees of freedom for polynomial model with Prob =
0.01

   Df Between      Within   _2Loglik    ResDf       Chi     ChiDf     ChiProb
    4  1.1811      0.5378   6130.454     5129      0.49         1       0.485
    3  1.1813      0.5378   6130.943     5130     19.41         1       0.000
    2  1.1915      0.5374   6150.348     5131         -         -           -

* Degrees of freedom of polynomial according to backward deviance testing is
3

Estimates of parameters (autoscaled covariable)
-----------------------------------------------
                     Parameter       estimate      s.e.      t(*)
                      Constant       5.975393  0.029456    202.86
                     age Lin        -0.188744  0.022086     -8.55
                     age Quad        0.162650  0.020324      8.00
                     age Cub        -0.086991  0.019714     -4.41
                      sex male      -0.081538  0.044323     -1.84

Parameters for factors are differences compared with the reference level:
         Factor   Reference level
            sex            female

       Between person variance         1.1813
        Within person variance         0.5378
               _2LogLikelihood       16130.94
           Degrees of freedom           5130


***** Usual intake percentiles/percentages are calculated from:
---------------------------------------------------------------
            sex       age   Intake probability   Transformed intake amount
         female         1               0.7237                       6.6959
           male         1               0.6767                       6.6144
         female         5               0.7288                       6.4379
           male         5               0.6823                       6.3564
         female         9               0.7117                       6.2272
           male         9               0.6636                       6.1457
              .         .                    .                            .
              .         .                    .                            .
         female        93               0.0326                       5.2954
           male        93               0.0262                       5.2139
         female        97               0.0113                       5.0625
           male        97               0.0090                       4.9810
```

Table 51: Information in 'Additional output' file

Click on icon 'Plots' to display the next screen (see Figure 85).



Display plots of:

Frequency model
Probability of intake vs. age (level female of sex)
Probability of intake vs. age (level male of sex)
Probability of intake vs. age (all levels of sex)

Transformed intake amount model
Transformed intake amount vs. age (level female of sex)
Transformed intake amount vs. age (level male of sex)
Transformed intake amount vs. age (all levels of sex)

Percentiles of usual intake vs. age
p50% percentile of usual intake vs. age (all levels of sex)
p90% percentile of usual intake vs. age (all levels of sex)
p95% percentile of usual intake vs. age (all levels of sex)
p99% percentile of usual intake vs. age (all levels of sex)
p99.9% percentile of usual intake vs. age (all levels of sex)
p99.99% percentile of usual intake vs. age (all levels of sex)
p50% p90% p95% p99% p99.9% p99.99% percentiles of usual intake vs. age (level female of sex)
p50% p90% p95% p99% p99.9% p99.99% percentiles of usual intake vs. age (level male of sex)

Percentage of population < usual intake limits vs. age
Percentage of population < q0.01 of usual intake vs. age (all levels of sex)
Percentage of population < q0.02 of usual intake vs. age (all levels of sex)
Percentage of population < q0.04 of usual intake vs. age (all levels of sex)
Percentage of population < q0.06 of usual intake vs. age (all levels of sex)
Percentage of population < q0.08 of usual intake vs. age (all levels of sex)
Percentage of population < q0.01 q0.02 q0.04 q0.06 q0.08 of usual intake vs. age (level female of sex)
Percentage of population < q0.01 q0.02 q0.04 q0.06 q0.08 of usual intake vs. age (level male of sex)

Figure 85: Diversity of possible plots if covariable and cofactor present and no uncertainty

Find in the left plot of Figure 86 for the intake frequency distribution the polynomial functions of age for both levels of sex. In the right plot, find for the transformed intake amounts the polynomial functions of age for both levels of sex.

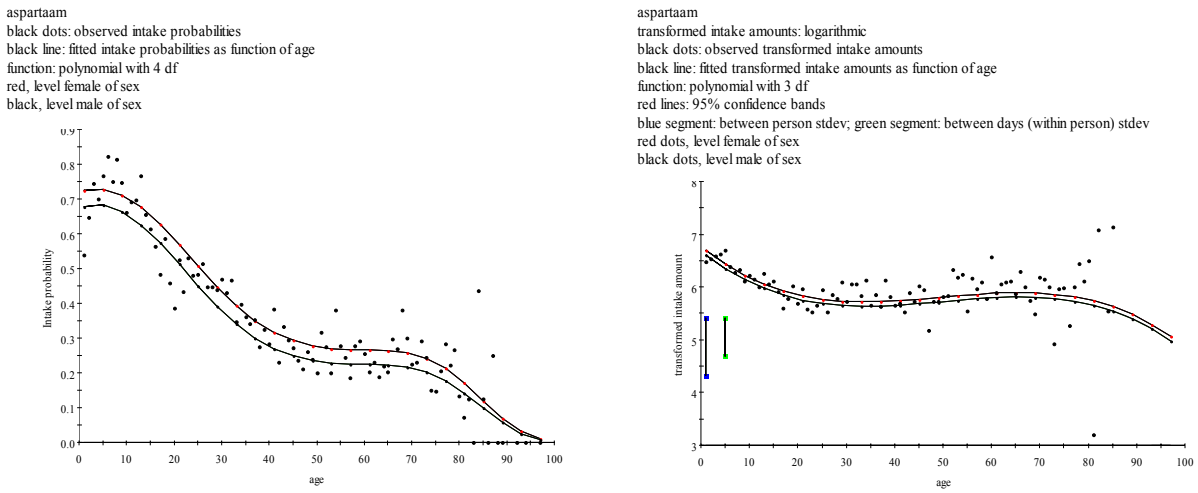


Figure 86: Age dependent percentiles derived from percentages

The intake frequency model and intake amount model are used to derive the usual intake. Click on 'Display output' in Figure 85 to get the result of the analysis. In Table 52 the age dependent percentiles of the usual intake are shown for each level of sex.

```
Usual intake percentiles of population:

        sex        age        p50        p90        p95        p99      p99.9     p99.99
```

female	1	703.16	3242.11	4941.88	10929.49	25370.44	50474.40
male	1	593.79	2870.60	4420.77	9810.18	22715.09	40497.57
female	5	547.95	2550.08	3866.46	8336.34	19202.33	33808.96
male	5	458.52	2220.18	3403.86	7426.94	17353.50	39425.23
female	9	429.53	2038.84	3072.94	6727.03	15705.71	33215.09
male	9	365.71	1768.00	2705.13	5887.64	13970.10	27149.02
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
female	93	0.00	15.31	81.25	429.84	1531.09	3698.93
male	93	0.00	6.72	52.85	334.93	1253.93	3634.92
female	97	0.00	0.03	4.11	139.18	718.62	2044.27
male	97	0.00	0.00	1.30	107.30	664.78	1928.50

Table 52: Percentiles of usual intakes

In Table 53 age dependent percentages of consumers lower than specified limits of usual intake are shown for each level of sex.

Usual intake percentages of population lower than intake limit:

sex	age	q0.01	q0.02	q0.04	q0.06	q0.08
female	1	0.02	0.02	0.03	0.05	0.05
male	1	0.03	0.05	0.07	0.09	0.11
female	5	0.01	0.02	0.02	0.04	0.05
male	5	0.03	0.04	0.07	0.09	0.11
female	9	0.02	0.03	0.05	0.06	0.08
male	9	0.04	0.06	0.09	0.11	0.14
.	.	.	.	.	.	.
.	.	.	.	.	.	.
female	93	71.61	73.14	74.79	75.71	76.40
male	93	76.69	78.01	79.35	80.11	80.68
female	97	88.94	89.60	90.30	90.72	91.01
male	97	91.01	91.59	92.18	92.49	92.72

Table 53: Percentages of consumers lower than specified limits

Figure 87 shows the age dependent percentiles for the usual intake distribution for both levels of sex.

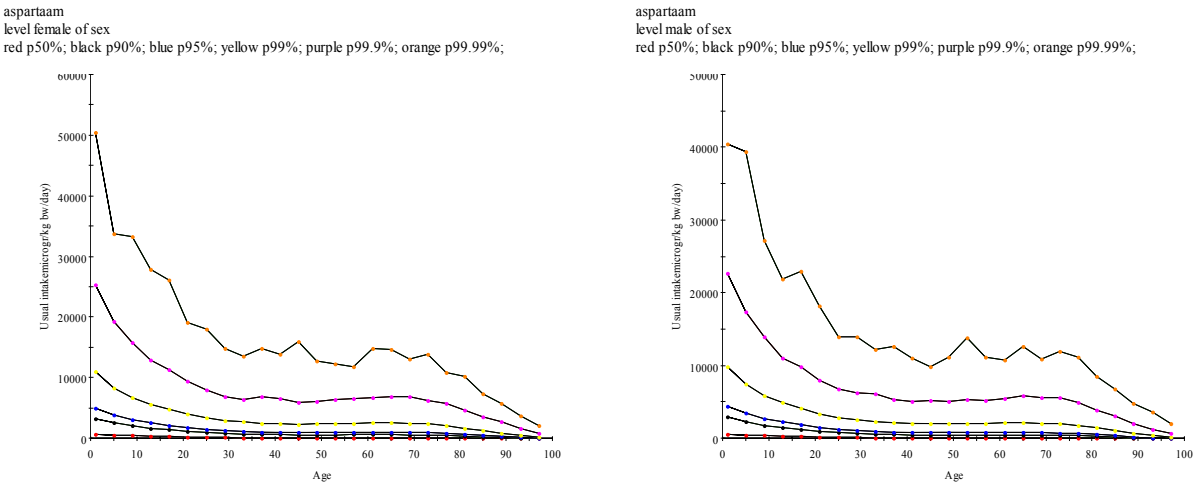


Figure 87: Age dependent percentiles for females (left) and males (right)

Figure 88 shows the age dependent percentages of consumers lower than the specified limits for both levels of sex.

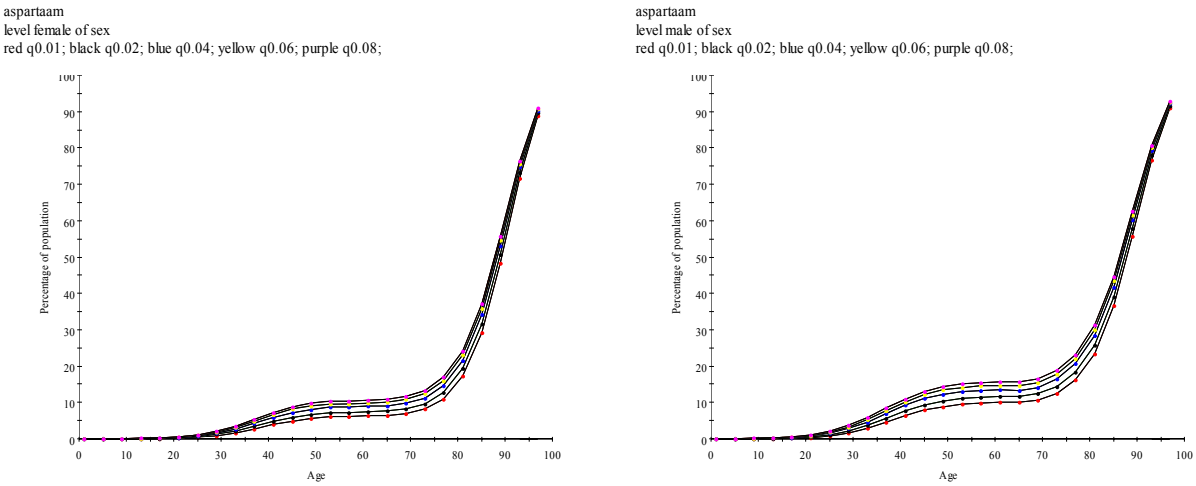
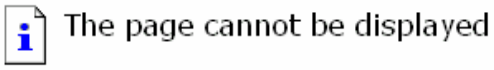


Figure 88: Age dependent percentages female consumers (left) and percentages male consumers (right) lower than the specified limits

11 APPENDIX C: Errors in displaying the page and scripting errors

Occasionally, ASP-scripts crash due to inconsistencies found in tables, as a consequence of wrong column names (see 9.1), and combinations of levels during subset selection that lead to empty subsets.

In general, the internet explorer error message is:



or an ASP debug-scripting-error is displayed.

When this occurs:

- try to reach the MCRA main menu (see Figure 2),
- go to manage input/output,
- move your mouse to directory 'IN' or 'OUT' or any other directory and left click,
- click the **'**Clear history'-button (see Figure 4).

If you cannot reach the MCRA main menu:

- close the internet explorer,
- login to the website again,
- go to manage input/output in the MCRA main menu (see Figure 2),
- move your mouse to directory 'IN' or 'OUT' or any other directory and left click,
- clear your history first by clicking the **'**Clear history'-button (see Figure 4).

Links
for correct link insert cursor in link and press Shift+F9:
Acute risk type