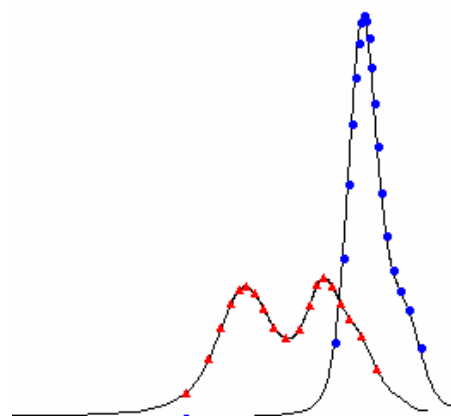
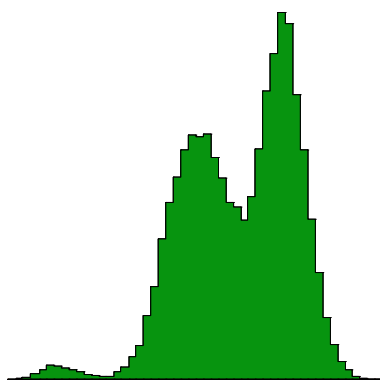
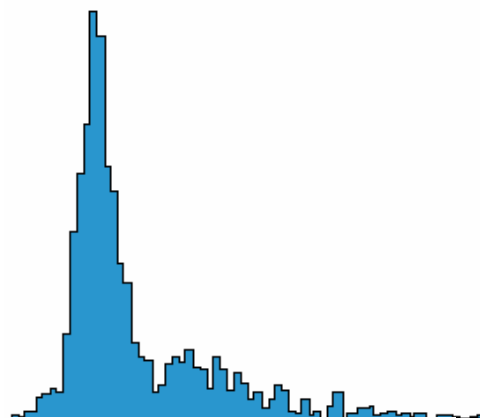
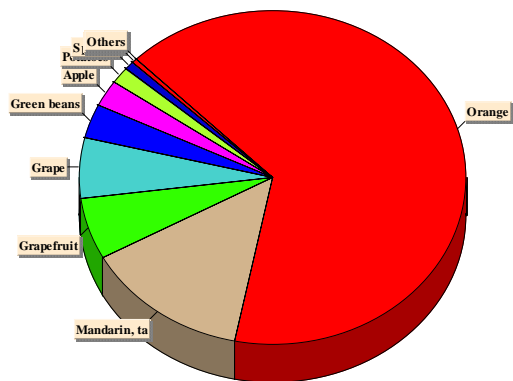


# MCRA, Release 4

a web-based program for Monte Carlo Risk Assessment

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

December 05

Biometris is the integration of the Centre for Biometry of Plant Research International and the Department of Mathematical and Statistical Methods of Wageningen University and is a part 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

<b>Contents</b>	<b>3</b>
<b>1 Foreword</b>	<b>7</b>
<b>2 Getting started with MCRA</b>	<b>8</b>
<b>2.1 Introduction</b>	<b>8</b>
<b>2.2 Registration</b>	<b>9</b>
<b>2.3 MCRA Main menu</b>	<b>9</b>
<b>2.4 Data needed</b>	<b>10</b>
<b>2.5 Manage input/output</b>	<b>11</b>
<b>2.6 Starting the MCRA program</b>	<b>12</b>
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 the products	17
<b>3 Specifying MCRA options and running a job</b>	<b>19</b>
<b>3.1 Overview</b>	<b>19</b>
<b>3.2 Acute risk analysis</b>	<b>20</b>
3.2.1 Concentration model options	22
3.2.1.1 <i>Empirical</i>	22
3.2.1.2 <i>Binomial/lognormal (no pooling)</i>	22
3.2.1.3 <i>Binomial/lognormal with pooling</i>	23
3.2.2 Concentration data options	23
3.2.2.1 <i>Replacement of non-detects</i>	23
3.2.2.2 <i>Modelling processing effects</i>	24
3.2.3 Unit variability model options	24
3.2.3.1 <i>Estimated parameters for unit variability</i>	25
3.2.3.2 <i>Beta distribution</i>	25
3.2.3.3 <i>Lognormal distribution</i>	26
3.2.3.4 <i>Bernoulli distribution</i>	26
3.2.4 Additional options concerning IESTI and consumption days	27
3.2.5 Intake model options	28
3.2.5.1 <i>Only empirical estimates</i>	28
3.2.5.2 <i>Empirical estimates and betabinomial/lognormal with age</i>	28
<b>3.3 Chronic risk analysis</b>	<b>29</b>
3.3.1 Concentration model options	30
3.3.2 Intake model options	30
3.3.2.1 <i>Betabinomial/lognormal</i>	30
3.3.2.2 <i>Betabinomial/lognormal (with age)</i>	31
3.3.2.3 <i>Discrete/semiparametric (Nusser)</i>	32
<b>3.4 Uncertainty analysis</b>	<b>33</b>
3.4.1 Uncertainty analysis options for acute risks	33
3.4.2 Uncertainty analysis options for chronic risks: betabinomial/lognormal and betabinomial/lognormal (with age)	34
3.4.3 Uncertainty analysis options for chronic risks: discrete/semiparametric (Nusser)	34

<b>3.5 Output</b>	<b>34</b>
3.5.1 Output options for acute risks: only empirical estimates	34
3.5.2 Output options for acute risks: empirical estimates + betabinomial/lognormal with age	35
3.5.3 Output options for chronic risks: betabinomial/lognormal	35
3.5.4 Output options for chronic risks: betabinomial/lognormal with age	35
3.5.5 Output options for chronic risks: binomial/lognormal + discrete/semiparametric (Nusser)	36
<b>3.6 Running an MCRA job</b>	<b>36</b>
<b>3.7 Checking the processing time</b>	<b>36</b>
<b>4 Output</b>	<b>40</b>
<b>4.1 Acute risk assessment: basic analysis</b>	<b>40</b>
<b>4.2 Chronic risk assessment: betabinomial/lognormal</b>	<b>48</b>
<b>4.3 HELP on CHARTs</b>	<b>54</b>
4.3.1 To display the Property Editor	55
<b>5 Model description</b>	<b>57</b>
<b>5.1 Basic model</b>	<b>57</b>
<b>5.2 How to deal with limited information</b>	<b>58</b>
5.2.1 The choice between a parametric and non-parametric approach	58
5.2.2 Estimation based on histogram data	59
5.2.3 Estimation based on summary data	59
5.2.3.1 Moments and other characteristics	60
5.2.3.2 Estimation	60
<b>5.3 Concentration model: modelling of compound concentrations in consumed food</b>	<b>62</b>
5.3.1 Distributional assumptions	62
5.3.1.1 Non-parametric modelling of concentrations (empirical)	62
5.3.1.2 Parametric modelling of concentrations (binomial/lognormal no pooling)	62
5.3.1.3 Parametric modelling of concentrations (binomial/lognormal with pooling)	63
5.3.2 Modelling of missing data and replacement of non-detects	63
5.3.3 Modelling of processing effects	64
5.3.3.1 No processing factor	64
5.3.3.2 Processing (fixed factors)	65
5.3.3.3 Processing (distribution based)	65
<b>5.4 Modelling of unit variability</b>	<b>66</b>
5.4.1 Introduction, variability in deterministic modelling	66
5.4.2 Approaches to unit variability in probabilistic modelling: specifying distributions	67
5.4.2.1 Beta model for unit variability	68
5.4.2.2 Lognormal model for unit variability	69
5.4.2.3 Bernoulli model for unit variability	70
5.4.2.4 Estimation of intake values using the concept of unit variability	70
5.4.3 Additional: Comparison of probabilistic with deterministic estimates of acute risk	70
<b>5.5 Intake model</b>	<b>72</b>
5.5.1 Only empirical estimates	72
5.5.2 Empirical estimates and betabinomial/lognormal with age	72
5.5.2.1 Monte Carlo data generation of intake	72
5.5.2.2 Modelling the intake frequency distribution	72
5.5.2.3 Modelling ln(intake) amounts	72
5.5.2.4 Estimating the acute risk variability of positive intake amounts	73
5.5.2.5 Estimating the acute intake distribution	73
<b>5.6 Chronic risk assessment</b>	<b>73</b>
5.6.1 Introduction	73
5.6.2 Modelling usual daily intake	73

5.6.3	Betabinomial/lognormal	74
5.6.3.1	<i>Modelling the intake frequency distribution</i>	74
5.6.3.2	<i>Modelling ln(intake) amounts</i>	74
5.6.3.3	<i>Estimating usual intake distributions</i>	74
5.6.4	Betabinomial/lognormal with age	74
5.6.4.1	<i>Modelling the intake frequency distribution</i>	75
5.6.4.2	<i>Modelling ln(intake) amounts</i>	75
5.6.4.3	<i>Estimating usual intake distributions</i>	75
5.6.5	Discrete/semiparametric (Nusser)	75
5.6.5.1	<i>Power or log transformation</i>	76
5.6.5.2	<i>Spline fit</i>	76
5.6.5.3	<i>Estimation of the parameters of the usual intake distribution</i>	76
5.6.5.4	<i>Back transformation and estimation of usual intake</i>	77
<b>5.7</b>	<b>Uncertainty of risk assessments: bootstrapping data sets</b>	<b>78</b>
<b>5.8</b>	<b>Binning</b>	<b>79</b>
<b>6</b>	<b>Selection of consumers, foods and/or products</b>	<b>80</b>
<b>6.1</b>	<b>Subset selection: consumers</b>	<b>81</b>
<b>6.2</b>	<b>Subset selection: consumption and concentration data</b>	<b>83</b>
<b>6.3</b>	<b>User-defined worst-case values</b>	<b>87</b>
<b>7</b>	<b>On-line editing of data</b>	<b>90</b>
<b>7.1</b>	<b>List or add concentrations</b>	<b>92</b>
<b>7.2</b>	<b>List or add variability factors</b>	<b>94</b>
<b>7.3</b>	<b>List or add processing factors</b>	<b>95</b>
<b>7.4</b>	<b>List or edit ARfD or ADI</b>	<b>96</b>
<b>8</b>	<b>References</b>	<b>98</b>
<b>9</b>	<b>APPENDIX A: Preparing the data</b>	<b>100</b>
<b>9.1</b>	<b>Overview</b>	<b>100</b>
<b>9.2</b>	<b>Products tables</b>	<b>101</b>
9.2.1	Products	101
9.2.2	Productcategory	102
9.2.3	Productgroup	102
9.2.4	Product_subgroup	103
<b>9.3</b>	<b>Food consumption tables</b>	<b>103</b>
9.3.1	Foodconsumption	103
9.3.2	Foodconversionmodel	104
9.3.3	Individual	104
9.3.4	Foodsurvey	105
9.3.5	Processing	105
9.3.6	Processingfactor	106
<b>9.4</b>	<b>Concentration of compounds tables</b>	<b>107</b>
9.4.1	Compounds	107
9.4.2	Country	107
9.4.3	Concentration	108
9.4.4	Summarydata	109
9.4.5	Histogramdata	110

9.4.6 VariabilityProd	111
9.4.7 VariabilityCompProd	111
9.4.8 VariabilityProcCompProd	112
9.4.9 Agriculturaluse	112
<b>10 APPENDIX B: Other output depending on specifications of options</b>	<b>114</b>
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	121
10.6 Acute risk assessment: betabinomial distribution with age	123
10.7 Acute risk assessment: binomial/lognormal with pooling	126
10.8 Chronic risk assessment: discrete/semiparametric (Nusser)	129
10.9 Chronic risk assessment: betabinomial/lognormal with age	135
10.10 Chronic risk assessment: betabinomial/lognormal with age and uncertainty	141
<b>11 APPENDIX C: Errors in displaying the page and scripting errors</b>	<b>144</b>

# 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 agricultural products. 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 but the use of multiple names is avoided as much as possible.

The methods for probabilistic modelling described here are implemented in the program MCRA (Monte Carlo Risk Assessment). MCRA is a computational tool for dietary risk assessment. MCRA can calculate intake distributions for both short-term (acute) and long-term (chronic) exposures. 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 exposure based on the average concentration and the empirical distribution of consumption between consumers and between different consumption 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 bootstrapping. 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 concentrations below analytical reporting limits 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 calculations.

The current MCRA system 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, written in the statistical package GenStat (2005), 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 GenStat 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 4. It covers the current release 4.0 (release 4 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 coordinates 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 the 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 food products. MCRA provides the following options:

- Acute probabilistic risk assessment: MCRA will calculate the exposure distribution (mg or microgram compound per kg body weight) from input data on consumption and compound concentrations in the food.
- In addition, age dependent percentiles and bootstrap confidence intervals may be estimated using simulated acute intakes as input data.
- Percentiles: the exposure distribution can be characterised by percentiles, *i.e.* compound concentration levels exceeded with only a small specified probability (for example the 99<sup>th</sup> percentile p99 is exceeded only in 1% of the cases).
- Uncertainty due to small samples: bootstrap sampling of Monte Carlo variation, of consumers and of compound concentrations to assess the uncertainty of the percentiles in the form of an approximate confidence interval.
- Calculation of point estimates (IESTI) and comparison with Monte Carlo results.
- Conversion of food consumption to the consumption of primary agricultural products, e.g. convert pizza consumption to consumption of wheat, tomato, cheese, etc.
- 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 the data on age, weight or sex of consumers, day of consumption, foods, primary products, and year, country and sampling type of concentration data.
- Insertion of worst case values for products without concentration measurement values.
- Calculate exposure distribution for consumption-days only
- Chronic risk assessment: MCRA calculates the usual intake distribution or chronic exposure when the total number of consumption days per consumer is 2 or more. In MCRA, basically, two methods are implemented: the first method is using the betabinomial distribution to model the intake frequency. The lognormal distribution is used to model logarithmically transformed intakes. Both distributions are integrated to estimate the usual intake distribution for the entire population. In an extended version, the usual intake is related to age and age-dependent percentiles and bootstrap confidence intervals are estimated. The second method follows an approach proposed by Nusser *et al.* (1996, 1997) and Dodd (1996). The chronic exposure distribution is characterised by percentiles and bootstrap confidence intervals on these percentiles.



## 2.2 Registration

To use MCRA, navigate your 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. You can find the latest developments and most recent information in the On Line Manual.

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

Address <http://mcra.rikilt.wur.nl/scripts/mcra.exe?page=newuser>

[Login](#)  
[Register](#)

### Apply to get access to the MCRA web application

Fill in this form and submit it to apply.  
Within 3 working days you will receive an e-mail message whether you will be granted access. All fields are required.

Username

Password

Confirm password

E-mail address

Full name

Organisation

Department

Postal address

**Figure 1: Registration form**

Registered users can login by clicking on [Login to MCRA for registered users](#) in the first line of the home page. By specifying their username and password the first screen of the MCRA Website is shown: **MCRA Main Menu**.

## 2.3 MCRA Main menu

In Figure 2 the possibilities within the Main menu are shown:

## MCRA: Monte Carlo Risk Assessment

Stochastic modelling of chemical intake from food

### Main menu

- Manage input/output
- MCRA 4.0 **NEW**
- MCRA 4.0 (field trial data and Dutch consumption data) **NEW**
- Older versions
- Registered user information
- Logout from MCRA website
- Help

On line Manual of Release 4, last update: 12/15/2005  
Manual of Release 4, Print version date: 12/15/2005

**Figure 2: MCRA Main Menu**

- Clicking 'MCRA 4.0' starts the preparations needed to do an MCRA analysis.
- Clicking 'Manage input/output' gives the possibility to upload (and download) your own data (see 2.4 ) or to download output from a former MCRA analysis.
- Clicking 'Older versions' starts the preparations needed to do an MCRA analysis with older versions of the program. However, older versions are not supported anymore.
- Clicking 'MCRA 4.0 (field trial data and Dutch consumption data)' also starts the preparations needed to do an MCRA analysis. However it gives the possibility to edit your own compound concentration data.
- Clicking 'Registered user information' enables you to view which information about you is stored in the user database.
- Clicking 'Help' gives explanation about the possibilities of the buttons in Main menu.
- Clicking 'Logout from MCRA website' is the official way 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

Which data do you need to have in order to run MCRA simulations? All data for MCRA are stored in Microsoft Access database tables according to fixed formats. So, if you want to use own data, you will need to prepare MS Access database files off-line, and then upload them to your personal user area on the MCRA website. The 'MCRA 4.0 (field trial data and Dutch consumption data)' button in the Main menu (see 2.3 ) offers some possibilities to edit your data on-line. However, on-line editing is restricted to compound concentration values, variability and processing factors, acute reference dose (ARfD) and average daily intake (ADI) (see Ch. 7 ).

In Chapter 9 a full description is given how data should be saved in MS Access database files. When using your own databases you must have exactly the same names for tables and fields (columns).

Basically input data for MCRA originate from two sources: food consumption surveys and monitoring programs on compound concentration data. Typically surveys on food consumption patterns provide information on food intake or food as eaten as such. Monitoring programs measure compound concentration levels on primary agricultural product. However MCRA estimates dietary risks for primary agricultural products only. Therefore a recipe database is necessary to convert amount of food as eaten to amount of primary product. In the MCRA system table 'foodconversionmodel' translates consumption data to amounts of primary agricultural products. Compound concentration data are

measured on primary agricultural products, so now the data are suited to be used in MCRA. An important role plays table 'products' in which all coding and labels on both foods as well as primary agricultural products are stored.

The MCRA system has a central database which offers a selection of example data. So it is possible to run example analyses without having data yourself. However, MCRA is primarily designed to work with user databases, or with a mix of user data and centrally available data. For example, you can provide your own data on compound concentrations and combine these with the consumption survey data in the central database to produce a risk assessment for this compound. BE CAREFUL when using tables originating from different databases: the codes of the primary agricultural products, foods of centrally supplied data, your own data and/or downloaded data should be consistent with each other.

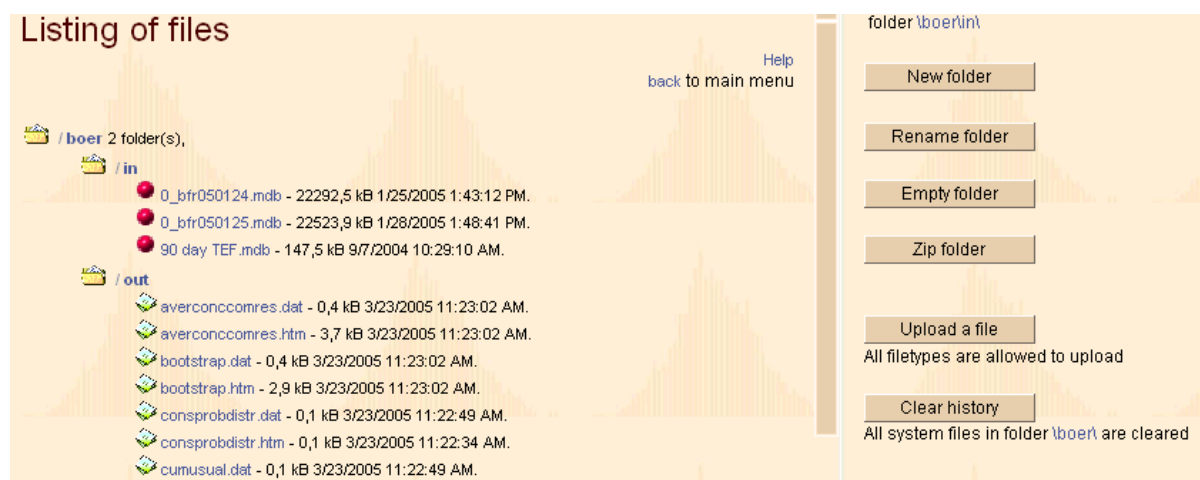
Consumption data are portions of food (consumed at different days) of consumers of whom the age, weight and sex must be known. Food data should be converted to primary agricultural products by means of a food conversion model. So, speaking about the food pizza, the percentages of the primarily agricultural products wheat, tomato, cheese etc. in a pizza must be known.

The second type of data, compound concentration data, are the amounts of compound found on monitoring samples of primarily agricultural products.

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' button in the Main menu brings you to a screen where you can upload your data files (see Figure 3). Each user has a personal data area, with two subdirectories named IN and OUT. The IN directory can be used to upload your own MS Access database files. Files can be uploaded directly or in zipped form. You can also zip, rename or delete files. MS Access database files and zip files can also be downloaded. Note: subdirectories IN and OUT should never be deleted.



**Figure 3: Manage input/output**

Output from MCRA will be written to your personal OUT directory. To download output for off-line viewing do the following. Go to the central menu (see Figure 6), click the View-output button and click DownloadOutput (see Figure 4). Output files are downloaded in a zipped format. The download includes a file 'viewoutput.htm' which gives you the same options to study your 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 your personal directory. Your own files on the 'in' or 'out' subdirectory are not cleared.

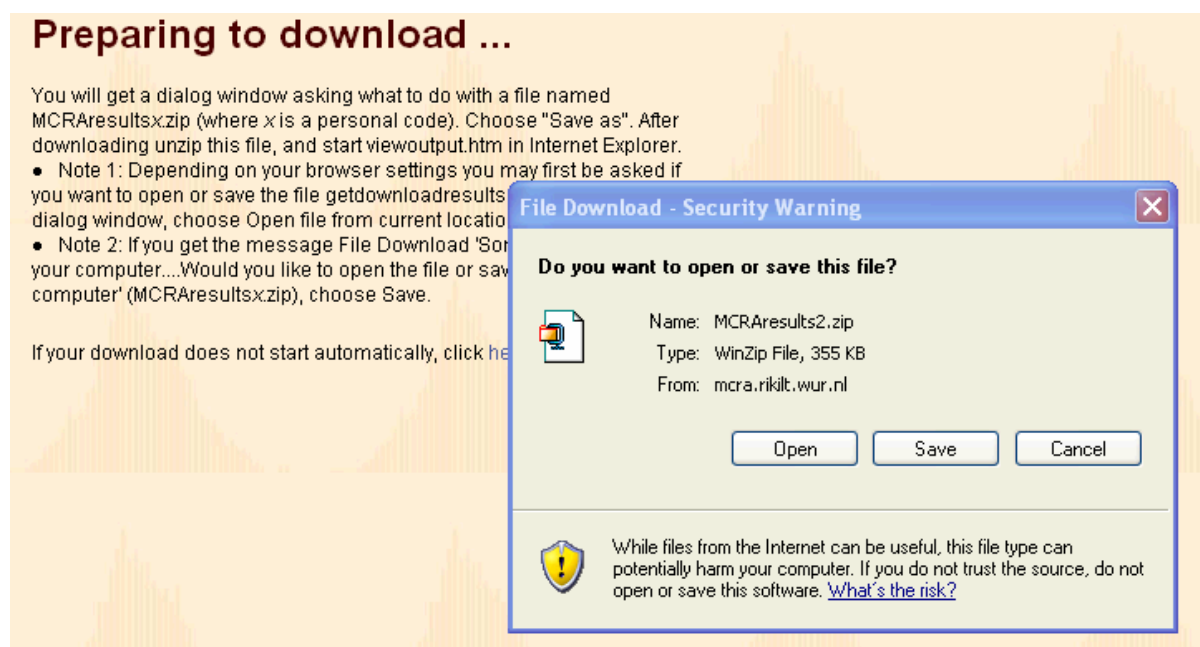


Figure 4: Download output for off-line viewing

## 2.6 Starting the MCRA program

If you want to run MCRA, click the 'MCRA 4.0' button in the Main menu (see Figure 2). Then depending on whether you are a new user or not one of the 2 screens of the MCRA central menu (see Figure 5 and Figure 6 ) is shown.

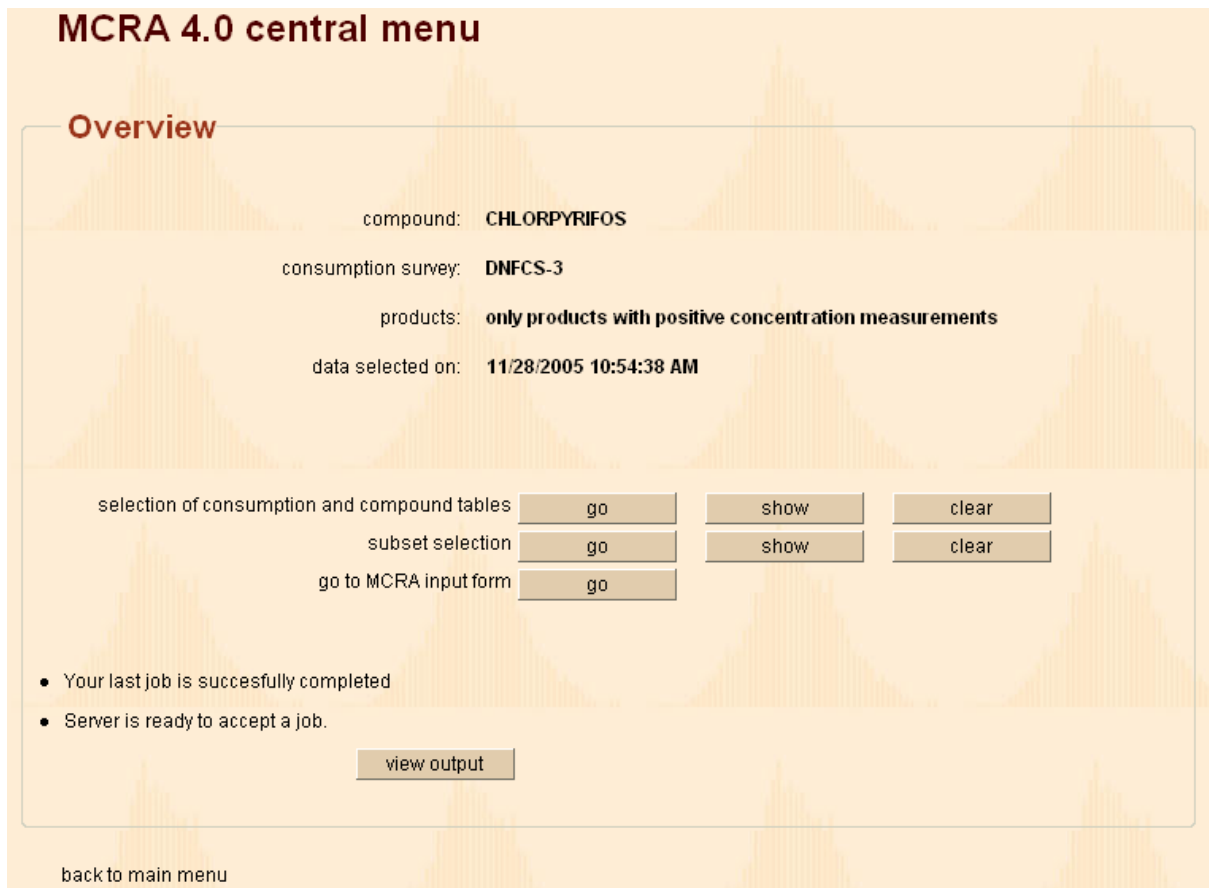
The MCRA system is essentially a cluster of activities build around a central menu. Each activity can be started from the central menu window and after finishing the activity the user can return to the MCRA central menu to do a new analysis.

New users are automatically brought to the MCRA central menu screen of Figure 5. Clicking the 'go' button starts the selection of food consumption and compound concentration tables:



**Figure 5: MCRA central menu, start selection of tables, survey and compound**

In all subsequent cases the MCRA central menu shows the latest selection of compound and consumption survey (see Figure 6).

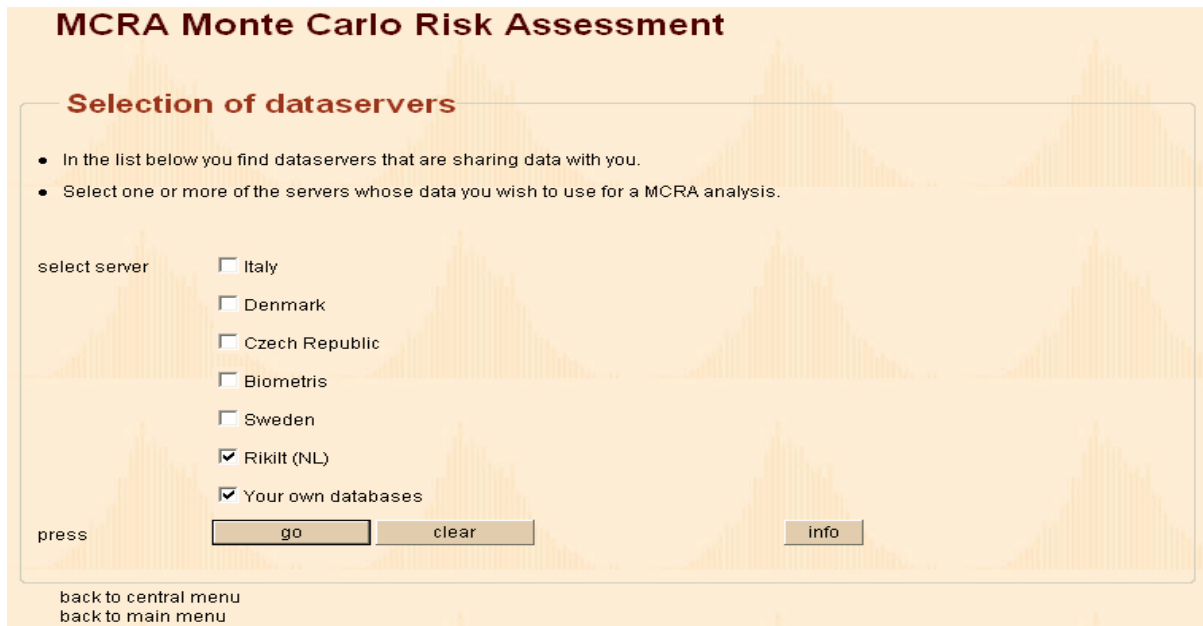


**Figure 6: MCRA central menu**

The first thing you have to do is selection of the food consumption and compound concentration tables. When you are a new user this is the only possibility in the MCRA central menu. (see Figure 5). When you already have made a selection, for example in an earlier session, you can go on with the same selections as before by clicking the go-to-MCRA-input-form ‘go’ button (see Figure 6). However, if you want to select new data you can click the selection-of-consumption-and-compound-tables ‘clear’ button and you get the MCRA central menu screen of Figure 5.

### 2.6.1 Selection of the tables

In MCRA food consumption data are related to compound concentration data. You can select tables from remote data servers with a central database or supply your own data by clicking the selection-of-consumption-and-compound-tables 'go' button in the screen of the MCRA central menu (see Figure 6). After clicking the 'go' or 'clear' button in the central menu a screen is entered (see Figure 7) displaying 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 7, data server 'Rikilt (NL)' and 'Your own databases' are checked. In Figure 8 all databases that are available to you (according to your user credentials) are shown. From here you can select databases for further use in the MCRA analysis



**Figure 7: Selection of data servers**

As mentioned before coding in different tables should be consistent. Therefore the safest way to select your data is to select them using one source. If not, convince yourself that data coming from different sources are consistent and suited for your purposes.

From each of these sources you have to know what kind of data they include. Clicking the 'info' button and in the (not shown) next screen clicking the buttons with country names gives the latest information. You can select whole databases or select tables from different databases making combinations from centrally supplied data or user data.

## MCRA: Monte Carlo Risk Assessment

### Selection of databases

- In the list below you find servers and databases.
- Select one or more databases that you wish to use for a MCRA analysis.

select databases from list:

Rikilt (NL)	<input checked="" type="checkbox"/> dutch pesticides.mdb
	<input type="checkbox"/> mcradbtest.mdb
	<input type="checkbox"/> jiri.mdb
	<input type="checkbox"/> pesticides0203.mdb
Your own data	<input checked="" type="checkbox"/> mcradb.mdb
	<input type="checkbox"/> mcradbold.mdb
	<input type="checkbox"/> mcradiox.mdb
	<input type="checkbox"/> mcramyco2.mdb
	<input type="checkbox"/> mcra_3.1test.mdb
	<input type="checkbox"/> qq.mdb
	<input type="checkbox"/> solanin03.mdb
	<input type="checkbox"/> systemft.mdb
	<input checked="" type="checkbox"/> waldo.mdb
	<input checked="" type="checkbox"/> worstcase.mdb

press

back to central menu  
back to main menu

**Figure 8: Selection of databases**

Figure 9 displays the selected databases and shows which tables are available in each database. You may select a whole database at once (check 'All Tables') or make a combination of tables from different databases as done in Figure 9.

Note that while scrolling your mouse over the checkboxes the table name is displayed in a red, green or grey textbox. It is compulsory to select all tables for which a red textbox appears (Foodconsumption, Foodconversionmodel, Products, Individual, Compounds, and Country). It is also compulsory to make a choice between one of the tables containing compound concentration data (Concentration, Summarydata and Histogramdata). These tables have their table name displayed in a green textbox (see also 3.2.1 ). All tables with a grey textbox are optional (Productgroup, Productsubgroup, Agriculturaluse, Processing, Processingfactor, VariabilityProd, VariabilityCompProd and VariabilityProcCompProd). Make sure you select all compulsory tables otherwise the table selection menu keeps returning with a warning which tables are missing. Make also sure that in your selection of tables each type of table is represented only once. By checking check-columnnames MCRA performs a check whether all necessary columns are available in the selected tables and if not, produces a report of errors found.

Click the 'go' button to select the checked tables.

**MCRA: Monte Carlo Risk Assessment**

**Table selection for MCRA analysis**

Select data tables by checking the boxes below, then click the select-tables 'go' button  
Do not select more than one table with the same name.

Database:	Food consumption	Food consumption model	Individual	Compounds	Concentration	Summary data	Histogram data	Country	Products	Product group	Product sub group	Agricultural use	Processing	Processing factor	Variability Prod	Variability Comp Prod	Variability Proc Comp Prod	All Tables
dutch pesticides.mdb	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	-	-	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
mcradb.mdb	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	-	-	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
waldo.mdb	-	-	-	<input type="checkbox"/>	-	-	-	<input type="checkbox"/>	-	-	-	-	<input type="checkbox"/>	-	-	-	-	<input checked="" type="checkbox"/>
worstcase.mdb	-	-	-	<input type="checkbox"/>	<input type="checkbox"/>	-	-	<input type="checkbox"/>	<input type="checkbox"/>	-	-	-	<input type="checkbox"/>	-	-	-	-	<input checked="" type="checkbox"/>

check columnnames

select tables

clear all settings

[back to main menu](#)

**Figure 9: Table selection for MCRA analysis**

### 2.6.2 Selection of food consumption survey and of compound

After clicking the select-tables 'go' button the screen of Figure 10 is shown. For this example only 'All Tables' of the 'dutch pesticides.mdb' database is checked in Figure 9.

If there are consumption data from more than one survey in the selected food consumption table you are required to choose one survey from the scroll-down menu.

An MCRA analysis is performed on one compound separately. So, if there are more compounds in the selected Compounds table you are required to select one compound from the scroll-down menu (see Figure 10).



**MCRA: Monte Carlo Risk Assessment**

**Selected tables for MCRA**

**You selected the following CENTRAL files and tables:**

- dutch\_pesticides.mdb Foodconsumption
- dutch\_pesticides.mdb Foodconversionmodel
- dutch\_pesticides.mdb Products
- dutch\_pesticides.mdb Productgroup
- dutch\_pesticides.mdb Product\_subgroup
- dutch\_pesticides.mdb Individual
- dutch\_pesticides.mdb Compounds
- dutch\_pesticides.mdb Concentration
- dutch\_pesticides.mdb Country
- dutch\_pesticides.mdb Agriculturaluse
- dutch\_pesticides.mdb Processing
- dutch\_pesticides.mdb Processingfactor
- dutch\_pesticides.mdb VariabilityProd
- dutch\_pesticides.mdb VariabilityCompProd
- dutch\_pesticides.mdb VariabilityProcCompProd

The database contains more than 1 survey, make a selection:

The database contains 325 compounds, make a selection:

click after making your choice

[back to selection of tables for MCRA](#)  
[back to main menu](#)

**Figure 10: Selection of consumption survey and of compound**

### 2.6.3 Selection of the products

After selecting tables a check is made on the number of primary agricultural products involved. For consumed products (determined via the foodconversionmodel table), three situations may occur:

1. on some products only positive concentration values are measured,
2. on some products only non-detects are found,
3. on some products no concentration measurements are made (*i.e.* no positives nor non-detects)

Now, the user is confronted with three choices (see Figure 11):

1. continue with only the products for which positive concentration measurements are available
2. continue with products for which concentration data are available *i.e.* positive concentrations and/or non-detects
3. continue with all products *i.e.* also products without concentration measurements and let the program check the Agriculturaluse table (see 9.4.9 )

## MCRA: Monte Carlo Risk Assessment

### Table selection for MCRA analysis

The name of the food consumption survey is: **DNFCS-3**

In this survey the number of foods is: **557**

and the number of products: **118**

The selected compound is: **CHLORPYRIFOS**

Of the 118 consumed products:

for 25 products positive concentration values are available.

for 70 products measurements have been made but only nondetects were found. Include these products in the analysis if you want to replace them with the limit of reporting (LOR).

for 23 products no measurements have been made. Include these products in the analysis if you have information on concentrations for those products for which agricultural use is allowed.

selected consumption and compound tables

current subset selection

Continue with 25 products (with positive concentrations)

Continue with  $25 + 70 = 95$  products (measured)

Continue with  $25 + 70 + 23 = 118$  products (all)

press after making your choice

back to select another compound  
back to selection of tables for MCRA  
back to main menu

**Figure 11: Product selection for MCRA analysis**

After you have made your choice and clicking the 'go' button you enter the MCRA central menu (see Figure 6). Clicking the start MCRA 'go' button brings you to the MCRA options menu treated in Chapter 3 .

More information about selection of the consumer population, foods and/or products can be found in Chapter 6 .

## 3 Specifying MCRA options and running a job

### 3.1 Overview

After a selection is made of the required and additional tables, clicking the ‘go’ button of the central menu (see Figure 6) brings you to the MCRA input form, see Figure 12. In this chapter all options of the input form are discussed. Often, you encounter references to paragraphs in Chapter 5 for more theoretical background.

When the MCRA input form is entered for the first time, it looks like Figure 12 (default options shown).

The screenshot shows the MCRA program options input form. The form is titled "MCRA program options" and "Monte Carlo Risk Assessment". It is divided into several sections:

- Type of analysis:** Includes radio buttons for "risk type" (acute, chronic) and "uncertainty analysis" (yes, no). There is also an "optional" dropdown menu for "Choose default option set" and a "submit MCRA job" button.
- Acute risk model:** Includes dropdown menus for "concentration model" (empirical), "unit variability model" (no unit variability), and "intake model" (only empirical estimates). It also has text input fields for "number of Monte Carlo iterations" (100000), "number of computational chunks" (10), "random seed" (0), "number of bootstrap samples" (100), and "number of iterations per bootstrap sample" (10000).
- Concentration data:** Includes a "change options" button and the text "Processing and non-detects".
- Additional:** Includes a "change options" button and the text "IESTI and consumption days only".
- Uncertainty analysis:** Includes a "change options" button and the text "Bootstrap options".
- Output:** Includes a "change options" button and the text "Graphics and tables".

A "back to central menu" link is located at the bottom left of the form.

Figure 12: MCRA input form.

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

- In the top section, the type of analysis, acute or chronic, is chosen together with the option to perform an uncertainty analysis or not.
- In the left section, a number of options related to the choices made in the top section are specified. If an acute risk model is chosen, three more specifications have to be made concerning the concentration and unit variability model for the concentrations and the intake model which may be extended with a parametric function for the age effect. For a chronic risk assessment two models have to be specified: the concentration and intake model. Note that a unit variability model is not relevant.
- Finally, the right part of the screen displays a number of special option blocks depending on the choices made in the top and left section. E.g. for an acute risk model at least 3 option blocks are displayed: Concentration Data, Additional and Output. For a chronic risk model Concentration Data and Output are shown and, depending on the chosen intake model, an option block Intake model. In Figure 12 a fourth option block is shown: Uncertainty Analysis, because in the top section uncertainty analysis is set to yes. By clicking a change-options button the options of that block become visible (see Figure 13, default options shown).

After clicking any button or choosing any radio button or any item in a scroll-down menu, the screen is rebuilt, implementing your choices. Note that after changing a value in a text box you have to leave the field (replace the pointer) to implement the value.

Each description (like ‘risk type’, ‘uncertainty analysis’, ‘concentration model’ etc. in Figure 12) has its own mouse-over function revealing a screen-tip. By double-clicking the description name the on-line user manual is opened from where you can find further information.

You can save your settings by clicking the save-user-defaults button in the top section. With the scroll-down menu Choose-default-option-setting one of 4 settings can be chosen: Current, Previous, System and User. By clicking the ‘previous’ setting, all program settings from the last performed MCRA analysis are recalled and displayed in the input form. By clicking the ‘system’ setting default system settings are recalled and by clicking the ‘user’ setting the saved user defaults are implemented. When no MCRA analysis has been performed in the past, previous and user settings are equal to system settings.

**Concentration data**  
 close options Processing and non-detects  
 replace nondetects by (fraction of) LOR no replacement of nondetects  
 processing factors no processing  
 number of iterations per bootstrap sample 10000

**Additional**  
 close options IESTI and consumption days only  
 estimation of IESTI  yes  no  
 standard body weight is 60  
 compare IESTI with MonteCarlo-percentile  yes  no  
 MC percentage for comparison with IESTI 99  
 own variability factors  yes  no  
 select  all consumption days  consumption days only

**Uncertainty analysis**  
 close options Bootstrap options  
 bootstrap consumptions  yes  no  
 bootstrap concentrations  yes  no

**Output**  
 close options Graphics and tables  
 percentiles at percentages 50 90 95 99 99.9 99.99  
 percentage for summary/graph of right tail of exposure distribution 95  
 or exposure value instead of percentage \*  
 percentage for drill-down 97.5

**Figure 13: Right section after clicking the ‘change options’ buttons in the input form.**

After specifying all options the MCRA analysis is started by clicking the yellow ‘submit MCRA job’ button in the top section (see 3.6 )

### 3.2 Acute risk analysis

If in the MCRA input form (see Figure 12) acute risk is specified, the next specifications in the left section of the screen can be modified (see Figure 14, default options shown):

**Acute: model options**

concentration model	empirical
number of Monte Carlo iterations	100000
number of computational chunks	10
unit variability model	no unit variability
random seed	0
intake model	only empirical estimates
number of bootstrap samples	100
number of iterations per bootstrap sample	10000

**Figure 14: Left section of the MCRA input form if type of analysis is Acute**

As you can see there are scroll-down menus for the concentration model, the unit variability model and the intake model. In 3.2.1 , 3.2.3 and 3.2.5 the settings of these models are treated. But first some explanation about the other specifications is given.

### **Number of Monte Carlo iterations**

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 an iteration. So in Figure 14, the total number of Monte Carlo iterations is 100,000 divided over 10 chunks (10.000 MC iterations per chunk). Keep in mind that the number of products, certainly when the processing and unit variability model are applied, determine to a large extend what amount of iterations per chunk is feasible.

### **Number of computational chunks**

In general, the capacity of the internal memory restricts the size of the simulation. To overcome this problem, a Monte Carlo Risk Assessment can be performed in computational chunks. The results of each chunk are stored for later use. The chunk size (= iterations/chunks) is equal to the total number of Monte Carlo iterations divided by the number of chunks. Chunk size and number of chunks affect processing time. When the number of iterations within a chunk is too high, the performance of the computer is seriously degraded and swapping will occur. You can observe this after clicking ‘show progress’ (see Figure 31) by an extremely large processing time. Advised is to rerun the simulation with a higher number of chunks. The estimated CPU-time in the progress bar may be an indication to determine the optimal number of chunks. Best strategy is to keep the number of chunks as low as possible conditional upon chunk size. Factors influencing the chunk size are: the total number of products and/or the total number of combinations of products, processing types and unit variability.

The number of Monte Carlo iterations per chunk determines the maximum possible value for the number of iterations per bootstrap. When this value is higher than the number of iterations per chunk, the value is overruled by the program and reset to the maximum possible value.

### **Random seed**

The Monte Carlo simulation uses a pseudo-random number generator that is initialised by setting the seed. To get time-based values, set seed to zero and the generated sequence of random numbers is based on a default value which is printed in the program output. Using this value in a second run will result in identical simulation results provided that the model or number of iterations did not change.

### **Number of bootstrap samples**

If an uncertainty analysis is chosen, you have to specify the number of bootstrap samples. In Figure 14 the number is 100.

## Number of iterations per bootstrap sample

If an uncertainty analysis is chosen, you also have to specify the number of iterations per bootstrap sample. In Figure 14 the number of iterations per bootstrap sample equals the number of Monte Carlo iterations per chunk (is 10.000).

### 3.2.1 Concentration model options

Compound concentration data can be 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). The selection between these three types of concentration data is already made by the selection of tables in 2.6.1 .

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

Parametric modelling becomes important in data-scarce situations. The lognormal distribution with parameters  $\mu$  and  $\sigma$  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
<b>Option settings in concentration model</b>			
Non-parametric: empirical	x	-	-
Parametric: binomial/lognormal (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](#)

Possible settings in the input form for the concentration model are:

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

In the next paragraphs these settings are explained.

#### 3.2.1.1 Empirical

In the non-parametric approach we choose ‘empirical’: concentrations are sampled at random from the available data and combined with food consumption data to generate the intake distribution of exposure values.

[See also: Non-parametric modelling of concentrations \(empirical\)](#)

#### 3.2.1.2 Binomial/lognormal (no pooling)

In the parametric approach, compound concentrations per food product are sampled from parametric distributions based on full, histogram or summary data. Parameters  $\mu$  and  $\sigma$  of the lognormal distribution are estimated using the log-transformed non-zero compound concentrations (full data) or condensed data (summary or histogram data). You should choose this setting only when enough data are available to estimate  $\mu$ 's and  $\sigma$ 's for all products.

Estimation of the variance and/or mean may fail because sometimes compound measurements on specific food products are sparse or even missing. In case of missing parameters a warning message will be printed. MCRA should then be rerun with the setting ‘binomial/lognormal with pooling’.

A related question is the reliability of estimates based on a few degrees of freedom. To overcome these problems, basically, concentration data on other products are used to give sufficient data to base estimates upon. Products are classified into groups of similar products and missing or unreliable parameters are estimated using all concentration data in a group. This process of using concentration data on similar products to base estimates for  $\mu$  and  $\sigma$  upon is called pooling (see 3.2.1.3 )

[See also: Parametric modelling of concentrations \(binomial/lognormal no pooling\)](#)

### 3.2.1.3 Binomial/lognormal with pooling

Specifying pooling means that products are automatically assigned to groups of products and pooled. The identification of groups of products is based on the first 3 columns of the hierarchical product code (productcategory, productgroup, productsubgroup). Products for which the first three columns are equal are placed into the same group of products.

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

**1. Test homogeneity of variances within the groups of products**

if variances are homogeneous,  
pool variances

**test homogeneity of means within the groups of products**

if means are homogeneous,  
pool means.

**2. Test homogeneity of variances of products 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: Parametric modelling of concentrations \(binomial/lognormal with pooling\)](#)

## 3.2.2 Concentration data options

For the concentration data, you will always find an option block in the right section of the MCRA input form. Here you can specify how to model missing data and non-detects and how to model processing effects. The possible options in this block are shown in Figure 15 (default options shown).

The screenshot shows a form titled "Concentration data" with a "close options" button. Below the title, under the heading "Processing and non-detects", there are three settings:

- replace nondetects by (fraction of) LOR: no replacement of nondetects
- multiplication constant for LOR: 0.5
- processing factors: no processing

**Figure 15: Option block Concentration data**

### 3.2.2.1 Replacement of non-detects

Possible 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 can enter the food chain only via crop treatment, and when the percentage of crop treated is (approximately) known, then this knowledge may be 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) can be replaced with 0 or the LOR multiplied by a multiplication constant. If percent crop treated data are available (table AgriculturalUse see 9.4.9), then replacement by LOR can be 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

Possible settings of modelling of processing effects 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 raw product. 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 are not exactly known or information is of limited quality. These uncertainties may be entered into the model by specifying two values:  $f_{k,nom}$ , the nominal value, typically some sort of mean; and,  $f_{k,upp}$ , an upper 95% confidence limit. Distribution based processing factors require both values whereas for fixed factors only  $f_{k,upp}$  need to be specified. 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 'Processingfactor' ( $f_{k,nom} = \text{proc\_nom}$ ,  $f_{k,upp} = \text{proc\_upp}$ ) and processing codes and labels from table 'Processing'. 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  $\mu$  and  $\sigma$ . 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 Processing. Choose  $\text{disttype} = 1$  for a logistic-normal distribution or  $\text{disttype} = 2$  for a log-normal distribution. To process simultaneously some products 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 products. 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

Possible settings of unit variability are 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 raw product, 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 can be selected:

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).

[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)	<ul style="list-style-type: none"> <li>• no censoring at <math>cm_k</math></li> <li>• no upper limit to the unit concentration</li> </ul>	<ul style="list-style-type: none"> <li>• no censoring at <math>cm_k</math></li> <li>• unit values never higher than <math>nu_k \cdot cm_k</math></li> </ul>
Estimates of unit variability are conservative (C)	<ul style="list-style-type: none"> <li>• unit values will be left-censored at <math>cm_k</math></li> <li>• no upper limit to the unit concentration</li> </ul>	<ul style="list-style-type: none"> <li>• unit values will be left-censored at <math>cm_k</math></li> <li>• unit values never higher than <math>nu_k \cdot cm_k</math></li> </ul>

**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

When choosing this setting an option block Unit variability: Beta distribution appears in the right section of the MCRA input form (see Figure 16, default options shown).

**Unit variability: Beta distribution**

close options **Beta**

unit weight  $\leq 25$

$25 >$  unit weight  $\leq 250$

unit weight  $> 250$

parameter is  variability factor  coefficient of variation

estimates are  realistic estimates  conservative estimates

unit variability is compound/processing dependent  no  yes

unit variability is compound dependent  no  yes

**Figure 16: Option block Unit variability: Beta distribution**

The parameter for unit variability is specified as a variability factor  $\nu$  or as a coefficient of variation  $cv$  of the unit values in the composite sample. Variability factors  $\nu$  (97.5<sup>th</sup> 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  $\nu$  are recommended: for large crops ( $wu_k > 250g$ ) value  $\nu = 5$ ; for medium crops ( $wu_k$  25- 250g)  $\nu = 7$ ; and for small crops ( $wu_k \leq 25g$ )  $\nu = 1$  (FAO/WHO, 1997). For products which are processed in large batches, e.g. juicing, marmalade/jam, sauce/puree,  $\nu = 1$ . The latter information is specified in field bulking/blending of table Processing (see 9.3.5 ). If the parameter for variability is missing, zero variability is assumed, and the unit concentrations are equal to the sampled composite sample concentrations.

A choice to be made is 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 16, but can be modified by the user.

[See also: Beta model for unit variability](#)

### 3.2.3.3 Lognormal distribution

When choosing this setting an option block Unit variability: Lognormal distribution appears in the right section of the MCRA input form (see Figure 17, default options shown).

**Unit variability: Lognormal distribution**

close options      **Lognormal**

parameter is     variability factor     coefficient of variation

estimates are     realistic estimates     conservative estimates

unit variability is compound/processing dependent     no     yes

unit variability is compound dependent     no     yes

**Figure 17: Option block: Unit variability: Lognormal distribution**

In Figure 17 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 coefficient of variation  $cv$  or as a variability factor  $\nu$ .

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 97.5<sup>th</sup> percentile of the concentration in the individual measurements divided by the corresponding mean concentration seen in the composite sample. Finally a choice is made whether a realistic or a conservative approach is modelled. In the conservative approach, unit concentrations of the composite sample are left-censored at the value of the monitoring compound. When a realistic approach is defined, 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

When choosing this setting an option block, Unit variability: Bernoulli distribution, appears in the right section of the MCRA input form (see Figure 18, default options shown).

**Figure 18: 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  $\nu$  are recommended: for large crops ( $wu_k > 250g$ ) value  $\nu = 5$ ; for medium crops ( $wu_k 25- 250g$ )  $\nu = 7$ ; and for small crops ( $wu_k \leq 25g$ )  $\nu = 1$  (FAO/WHO, 1997). For products which are processed in large batches, e.g. juicing, marmalade/jam, sauce/puree,  $\nu = 1$ . The latter information is specified in field bulking/blending of table Processing (see 9.3.5 ). The number of units within a consumption is calculated and for each unit a Bernoulli distribution is used to sample the monitoring compound itself with probability  $(\nu-1)/\nu$  or a multiple  $\nu$  of it with probability  $1/\nu$  (see Figure 18).

[See also: Bernoulli model for unit variability](#)

### 3.2.4 Additional options concerning IESTI and consumption days

For the acute risk model there is always an option block of Additional options in the right section of the MCRA input form that handles IESTI and consumption days (see Figure 19, default options shown).

**Figure 19: 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 microgr/kg body weight/day and estimated per product.

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 product) of the Monte Carlo simulation. In the output (not shown) two kinds of estimates of the Monte Carlo 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 product 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 Monte Carlo Risk Assessment.

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

Possible settings of the intake model are:

- **only empirical estimates** [default]
- **empirical estimates and betabinomial/lognormal with age**

#### 3.2.5.1 Only empirical estimates

This is a straightforward acute risk analysis.

#### 3.2.5.2 Empirical estimates and betabinomial/lognormal with age

When choosing this setting an option block Intake model appears in the right section of the MCRA input form (see Figure 20, default options shown).

**Intake model**

close options

**Betabinomial/lognormal with age**

age effect for intake frequency model  yes  no

function to model effect  spline  polynomial

testing method  backward  forward

minimum degrees of freedom 0

maximum degrees of freedom 4

testing at level 0.01

age effect for intake amount model  yes  no

function to model effect  spline  polynomial

testing method  backward  forward

minimum degrees of freedom 0

maximum degrees of freedom 4

testing at level 0.01

**Figure 20: Option block Intake model in the right section of the MCRA input form screen**

Note that estimation of an age effect for an acute risk assessment is additional to the standard analysis. The simulated intake values of an acute analysis are used as input and: 1) an intake frequency function is estimated with a betabinomial model using a spline or polynomial function to model age effects; 2) non-zero intakes are log transformed and the transformed intake amounts are analysed using REML, again, using a spline or polynomial function to model age effects, and; 3) the information of both analyses is combined to estimate an age dependent acute intake distribution and to derive the requested age dependent percentiles.

[See also: Empirical estimates and betabinomial/lognormal with age](#)

### Intake frequency

The intake frequency function models the probability of a consumer 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). So the probability of having an intake may be related to the age of the consumer. The betabinomial distribution is very suited to sample consumer probabilities of having an intake using age as explanatory variable.

In Figure 20, first, we have to decide on modelling the age effect or not. The next choice is between using 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 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. We also have to decide on the method of testing: backward means that testing starts with a spline or polynomial of the highest degree. In each elimination round the number of degrees of freedom was decreased one at a time, and was stopped when the resulting decrease in fit was 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 Figure 20).

[See also: Modelling the intake frequency distribution](#)

### Intake amount distribution

The non-zero intakes are logarithmically transformed and the  $\ln(\text{intakes})$  are analysed with REML using a spline or polynomial function to model age effects. The analysis provides us with age dependent mean intakes and the standard deviation of the  $\ln(\text{intake})$  distribution. In Figure 20, second part, we decide on modelling the age effect or not; the use of a spline or polynomial function; backward or forward selection; the degrees of freedom; and the testing level. For the  $\ln(\text{intake})$  amount model choosing the spline function implies that the degree is automatically selected by the algorithm. For a polynomial function, the user has to decide on the degrees of freedom and method.

[See also: Modelling  \$\ln\(\text{intake}\)\$  amounts](#)

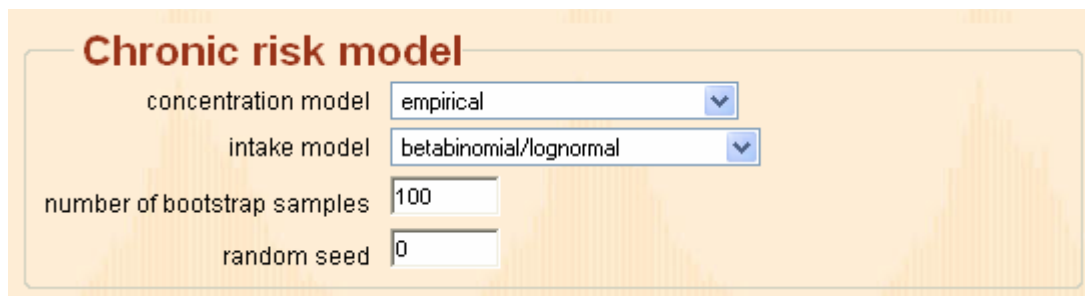
### Estimation of age dependent percentiles

The intake frequency is used to sample age dependent intake probabilities. Together with the age dependent mean  $\ln(\text{intakes})$  and the estimated standard deviation of the second analysis, we now have all what is necessary to estimate an acute intake distribution with a parametric age effect. The number of simulated age dependent intakes is set equal to the number of iterations supplied in the MCRA input form. After simulating the age dependent intake distribution, the requested percentiles are estimated.

[See also: Estimating the acute intake distribution](#)

## 3.3 Chronic risk analysis

Specify in the MCRA input form (see Figure 12) chronic risk and uncertainty, then the specifications in the left section of the screen can be modified (see Figure 21, default options shown):



**Chronic risk model**

concentration model	<input type="text" value="empirical"/>
intake model	<input type="text" value="betabinomial/lognormal"/>
number of bootstrap samples	<input type="text" value="100"/>
random seed	<input type="text" value="0"/>

**Figure 21: Left section of the MCRA input form if type of analysis is Chronic**

In dietary risk assessment, chronic exposure 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 compounds (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, that option consumption-days only is not relevant for chronic risks and that chronic risk assessment is only performed when the total number of days per consumer is 2 or more.

[See also: Chronic risk assessment](#)

### **Random seed**

This option is only visible when uncertainty is specified. In that case the settings are the same as for the acute risk model. See 3.2

### **Number of bootstrap samples**

This option is only visible when uncertainty is specified. In that case the settings are the same as for the acute risk model. See 3.2

### **3.3.1 Concentration model options**

The same settings as for the acute risk model are available:

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

and also the options of the concentration data are the same (see 3.2.2 )

### **3.3.2 Intake model options**

The possible settings here are:

- **betabinomial/lognormal** [default]
- **betabinomial/lognormal (with age)**
- **discrete/semiparametric (Nusser)**

In the next subparagraphs these settings are discussed.

#### ***3.3.2.1 Betabinomial/lognormal***

Input for a chronic risk assessment is the intake calculated as the consumption on each day of each consumer multiplied by the average value of the compounds divided by body weight (see 3.3 ). Applying the betabinomial/lognormal model, 1) an intake frequency function is estimated using the betabinomial model; 2) non-zero intakes are log transformed, the transformed intake amounts are analysed using REML and parameters of the usual intake distribution are estimated; 3) the information of both analyses is combined to estimate a usual intake distribution which is used to derive the requested usual intake percentiles.

[See also: Betabinomial/lognormal](#)

### **Intake frequency**

The intake frequency function models the probability of a consumer 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. The betabinomial distribution is very suited to sample consumer probabilities of having an intake.

[See also: Modelling the intake frequency distribution](#)

## Intake amount distribution

The non-zero intakes are logarithmically transformed and the  $\ln(\text{intakes})$  are analysed with REML estimating the parameters of the non-zero usual intake distribution. The total variance of the non-zero daily intakes is divided into a between-consumer component and a within-consumer component. The variance of the between-consumer variance component is the basis for the estimation of the the usual intake distribution.

[See also: Modelling  \$\ln\(\text{intake}\)\$  amounts](#)

## Estimation of usual intake percentiles

The intake frequency is used to sample intake probabilities. Together with the non-zero usual intake distribution a usual intake distribution is estimated for the entire population, e.g. both consumers as non-consumers. This is the distribution that is used for the estimation of the usual intake percentiles.

[See also: Estimating usual intake distributions](#)

### 3.3.2.2 Betabinomial/lognormal (with age)

When choosing this setting an option block Intake model appears in the right section of the MCRA input form (see Figure 22, default options shown).

**Intake model**

close options

**Betabinomial/lognormal with age**

age effect for intake frequency model  yes  no

function to model effect  spline  polynomial

testing method  backward  forward

minimum degrees of freedom 0

maximum degrees of freedom 4

testing at level 0.01

age effect for intake amount model  yes  no

function to model effect  spline  polynomial

testing method  backward  forward

minimum degrees of freedom 0

maximum degrees of freedom 4

testing at level 0.01

**Figure 22: Option block Intake model if intake model is betabinomial/lognormal (with age)**

Input for a chronic risk assessment is the intake calculated as the consumption on each day of each consumer multiplied by the average value of the compounds divided by body weight (see 3.3 ). Applying the betabinomial/lognormal with age model 1) an intake frequency function is estimated with a betabinomial model using a spline or polynomial function to model age effects; 2) non-zero intakes are log transformed and the transformed intake amounts are analysed using REML, again using a spline or polynomial function to model age effects, and; 3) the information of both analyses is combined to estimate an age dependent usual intake distribution and to derive the requested age dependent usual intake percentiles.

[See also: Betabinomial/lognormal with age](#)

## Intake frequency

The intake frequency function models the probability of a consumer 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). So the probability of having an intake may be

related to the age of the consumer. The betabinomial distribution is very suited to sample consumer probabilities of having an intake using age as explanatory variable.

In Figure 22, first, we have to decide on modelling the age effect or not. The next choice is between using 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 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. We also have to decide on the method of testing: backward means that testing starts with a spline or polynomial of the highest degree. In each elimination round the number of degrees of freedom was decreased one at a time, and was stopped when the resulting decrease in fit was 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 Figure 22).

[See also: Modelling the intake frequency distribution](#)

### Intake amount distribution

The non-zero intakes are logarithmically transformed and the  $\ln(\text{intakes})$  are analysed with REML using a spline or polynomial function to model age effects. The analysis provides us with age dependent  $\ln(\text{intakes})$  and estimates of the parameters of the non-zero usual intake distribution. The total variance of the non-zero  $\ln(\text{intakes})$  is divided into a between-consumer component and a within-consumer component. The variance of the between-consumer variance component is the basis for the estimation of the distribution of the usual intake. In Figure 22, second part, we decide on modelling the age effect or not; the use of a spline or polynomial function; backward or forward selection; the degrees of freedom; and the testing level. For the intake amount model, choosing the spline function implies that the degree is automatically selected by the algorithm. For a polynomial function, the user has to decide on the degrees of freedom and method.

[See also: Modelling  \$\ln\(\text{intake}\)\$  amounts](#)

### Estimation of age dependent percentiles

The intake frequency is used to sample age dependent intake probabilities. Together with the age dependent mean  $\ln(\text{intakes})$  and the estimated between-consumer variance component of the REML analysis, we can estimate a usual intake distribution with a parametric age effect. The requested usual intake percentiles are estimated using a simulated age dependent usual intake distribution.

[See also: Estimating usual intake distributions](#)

#### 3.3.2.3 Discrete/semiparametric (Nusser)

When choosing this setting an option block Intake model appears in the right section of the MCRA input form (see Figure 23, default options shown).

**Intake model**

close options

**Discrete/semi-parametric (Nusser)**

transformation  power  logarithmic

spline fit  yes  no

number of iterations to estimate frequency distribution

number of bins for discretisation

**Figure 23: Option block Intake model if intake model setting is discrete/semiparametric (Nusser)**

Input for a chronic risk assessment is the intake calculated as the consumption on each day of each consumer multiplied by the average value of the compounds divided by body weight (see 3.3 ). Before estimating the chronic percentiles of the distribution applying the discrete/semi-parametric



(Nusser) model non-normal exposure data are transformed to approximate normality following an approach proposed by Nusser *et al.* (1996, 1997) and Dodd (1996).

[See also: Discrete/semiparametric \(Nusser\)](#)

### **Transformation: power or logarithmic**

Transformations are specified using a power transformation or a logarithmic transformation. Usually, a power transformation is satisfactory.

[See also: Power or log transformation](#)

### **Spline fit or not**

After transforming the intakes, 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 intake frequency distribution**

The algorithm to estimate the intake frequency distribution may be modified by specifying the number of iterations. The specified number in Figure 23 is 5 so the number of iterations is  $5 \times 1000 = 5000$ .

[See also: Back transformation and estimation of usual intake](#)

### **Number of bins for discretisation**

The resolution of the probability grid for the intake frequency distribution is equal to 20. So, the probability mass is discretised at a grid ranging from 0 to 1 and a steplength equal to 0.05.

[See also: Back transformation and estimation of usual intake](#)

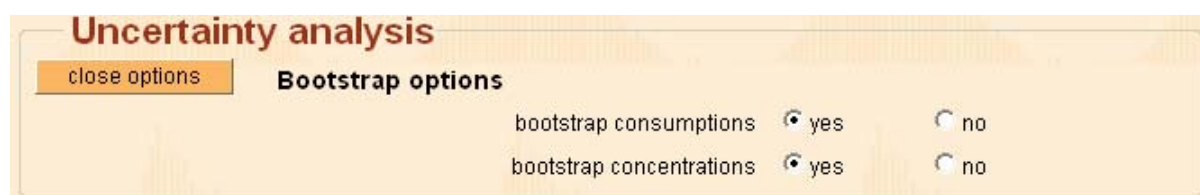
## **3.4 Uncertainty analysis**

Choose in the top section of the MCRA input form to perform an uncertainty analysis or not.

[See also: Uncertainty of risk assessments: bootstrapping data sets](#)

### **3.4.1 Uncertainty analysis options for acute risks**

When choosing uncertainty analysis together with an acute risk model an option block Uncertainty analysis appears in the right section of the MCRA input form (see Figure 24, default options shown).



Uncertainty analysis	
close options	<b>Bootstrap options</b>
	bootstrap consumptions <input checked="" type="radio"/> yes <input type="radio"/> no
	bootstrap concentrations <input checked="" type="radio"/> yes <input type="radio"/> no

**Figure 24: Option block Uncertainty analysis if an acute risk model is chosen**

The uncertainty of output statistics (e.g. mean or percentiles of the exposure distribution) is assessed by bootstrapping. The bootstrap can be applied on the level of fresh Monte Carlo samples, on the level of consumers and on the level of the compounds. To examine the uncertainty due to fresh Monte Carlo samples in each analysis only, specify 'no' for both bootstrap consumption data and compound data. Applying a bootstrap for consumption and compound data means that from each level with replacement a bootstrap sample is drawn and the corresponding intake distribution is calculated. Then, each bootstrap sample 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 bootstrap samples and the number of iterations per bootstrap sample is specified in the left section of the MCRA input form (see Figure 14). There, 100 bootstrap samples are specified and

each bootstrap sample consists of 10,000 values. The number of values within a bootstrap sample restricts which percentiles are displayed. Here, the highest possible percentage for which uncertainty information can be calculated is the 99.99<sup>th</sup> percentile, for a bootstrap-sample containing 1000 values this is the 99.9<sup>th</sup> percentile.

### 3.4.2 Uncertainty analysis options for chronic risks: betabinomial/lognormal and betabinomial/lognormal (with age)

There is no difference with the uncertainty analysis for acute risks described in 3.4.1 .

### 3.4.3 Uncertainty analysis options for chronic risks: discrete/semiparametric (Nusser)

When choosing uncertainty analysis together with chronic risk model: discrete/semiparametric (Nusser) an option block Uncertainty analysis appears in the right section of the MCRA input form (see Figure 25, default options shown).

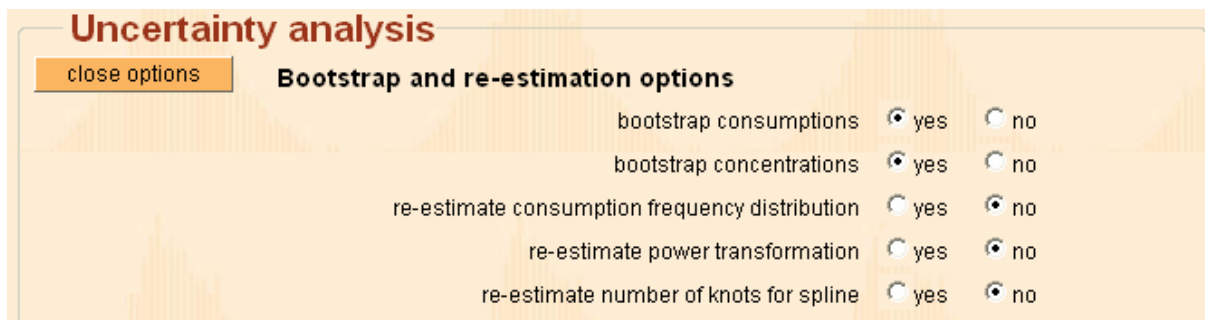


Figure 25: Option block Uncertainty analysis if a chronic risk model is chosen

The uncertainty is assessed using a bootstrap for consumption and compound data.

Option ‘re-estimate consumption frequency distribution’ is only relevant when consumptions are bootstrapped. 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 approximate 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

Output options can be specified in the output block. In Figure 26 (default options shown), percentages are specified, each percentage separated by a space.

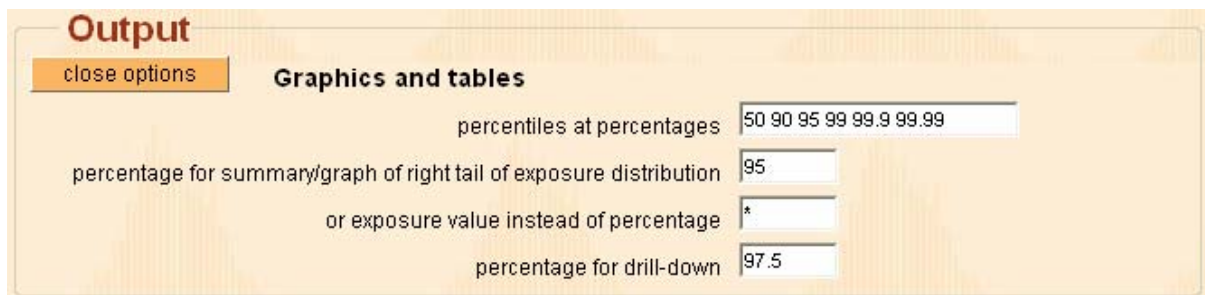


Figure 26: Option block Output if an acute risk model is chosen with only empirical estimates

The next option is needed for summarising the contribution of products to the right tail of the exposure distribution and to display a graph of the upper tail. A percentage may be specified, but specifying an exposure value instead overrules the percentage. The percentage for drill-down option is used to select the nine consumers lying 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 contents found on each consumption. To find the upper nine consumers, e.g. the nine consumers with the highest intake, specify percentage 100. Note that the information on the nine selected consumers is only approximately, giving a rough indication about the kind of consumers and their intake. The MCRA program performs an extra loop to gather the necessary information about consumers. This extra loop is only and only used to display information about the nine selected consumers.

### 3.5.2 Output options for acute risks: empirical estimates + betabinomial/lognormal with age

Output options can be specified in the output block (see Figure 27, default options shown). See also 3.5.1 for explanation on the options. In the lower part of the screen, options related to age are shown. The minimum and maximum age are retrieved from the database, but can be overruled by specifying own values. The steplength is automatically determined but can be overruled. Default, approximately 20 steps are taken and the calculated steplength is rounded to the nearest integer. To get predictions for specific ages, specify your extra ages space separated. Extra ages may lay outside the specified range determined by the minimum and maximum age.

**Output**  
close options

**Graphics and tables**

percentages at percentiles 0.01 0.02 0.04 0.06 0.08

percentiles at percentages 50 90 95 99 99.9 99.99

percentage for summary/graph of right tail of exposure distribution 99.5

or exposure value instead of percentage \*

percentage for drill-down 97.5

show results from minimum age 9999

with steps of 5

to maximum age 9999

extra ages \*

**Figure 27: Option block Output if an acute risk model is chosen with empirical estimates + betabinomial/lognormal with age**

### 3.5.3 Output options for chronic risks: betabinomial/lognormal

Output options can be specified in the output block (see Figure 28, default options shown). See also 3.5.1 for explanation on options.

**Output**  
close options

**Graphics and tables**

percentages at percentiles 0.01 0.02 0.04 0.06 0.08

percentiles at percentages 50 90 95 99 99.9 99.99

**Figure 28: Option block Output if a chronic risk model is chosen with betabinomial/lognormal**

### 3.5.4 Output options for chronic risks: betabinomial/lognormal with age

Output options can be specified in the output block (see Figure 29, default options shown). See also 3.5.1 and 3.5.2 for explanation on options.

**Output**  
close options

**Graphics and tables**

percentages at percentiles 0.01 0.02 0.04 0.06 0.08

percentiles at percentages 50 90 95 99 99.9 99.99

show results from minimum age 9999

with steps of 5

to maximum age 9999

extra ages \*

**Figure 29: Option block Output if a chronic risk model is chosen with betabinomial/lognormal with age**

### 3.5.5 Output options for chronic risks: binomial/lognormal + discrete/semiparametric (Nusser)

Output options can be specified in the output block (see Figure 30, default options shown). See also 3.5.1 for explanation on the options.

**Output**  
close options

**Graphics and tables**

percentiles at percentages 50 90 95 99 99.9 99.99

**Figure 30: Option block Output if a chronic risk model is chosen with binomial/lognormal + discrete/semiparametric (Nusser)**

## 3.6 Running an MCRA job

After setting the options in the MCRA input form, you have to click the ‘MCRA submit job’ button in the top section. If you wish to be notified when the MCRA analysis is completed check the radio button to *yes*. After submitting the form you enter the MCRA central menu screen (see Figure 31).

## 3.7 Checking the processing time

After submitting the form, all model specifications are passed to the server and the analysis is initiated. First, all data are exported to <user>data.mdb, a system file which is located on the user directory. Export of data takes only place when data are changed e.g. after selection of new tables or after subset selection. In all other cases, e.g. different model options, exportation of data is already performed and this step is cancelled.

The user may follow the progress of the analysis by clicking the ‘show progress’ button in the central menu (see Figure 31). Here also information is given about the estimated CPU time Performing an exposure assessment may take considerable time if the data files are large and/or if the number of iterations is high. The ‘update window’ button is used to update information coming from the server and, when the job is completed to renew the screen. Now a button ‘view output’ appears (see Figure 33). By clicking the ‘kill job’ button a job is terminated with a fatal error status.

**MCRA 4.0 central menu**

**Overview**

compound: **CHLORPYRIFOS**

consumption survey: **DNFCS-3**

products: **only products with positive concentration measurements**

data selected on: **12/13/2005 2:38:02 PM**

selection of consumption and compound tables

subset selection

go to MCRA input form

- Your job is currently running on workstation 1. By pressing  the elapsed and total expected processing time is displayed.
- Server is busy.
- 

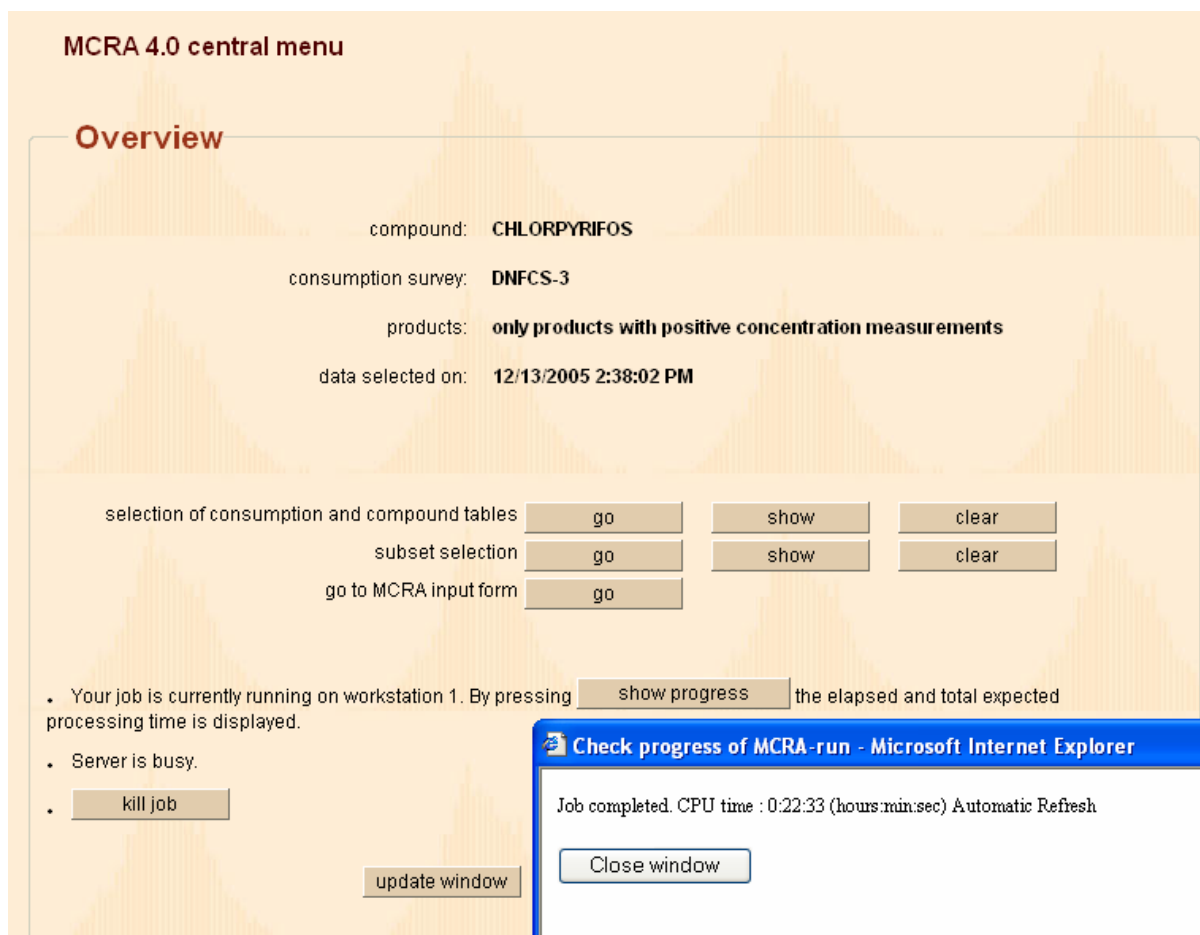
**Check progress of MCRA-run - Microsoft Internet Explorer**

Computational chunk (uncertainty) 2/100

Estimated cpu time (h:m:s) elapsed / total : 0:0:29 / 0:25:33

Automatic Refresh

**Figure 31: Central menu, show progress**



**Figure 32: Central menu, job completed, close window**

After clicking 'update window' button you get the screen of Figure 33



**Figure 33: Central menu, job completed, view output**

Clicking the 'view output' button gives Figure 35 in 4.1 .

## 4 Output

The results of the analysis can be viewed by clicking the ‘view output’ button in Figure 33. (It is always possible to view your output of your last job even if you left your MCRA program). This brings you to a screen where both tabular and graphical results can be viewed. Depending on the type of analysis there are different icons and different items in the display (see Figure 35 and Figure 40).

In 4.1 the different forms of the output are explained with an example of a basic acute risk analysis. In 4.2 an example is given of a chronic risk analysis. Other examples you can find in Ch. 10 .

If you want to download the results for off-line viewing see 2.5.

### 4.1 Acute risk assessment: basic analysis

In this example output is shown for an acute risk assessment for organo phosphate pesticide chlorpyrifos which can have neurotoxic effects. This example uses the dutch pesticides database from Rikilt (NL). It is a complete database with food survey tables, compound concentration tables and product tables. After selection ‘All tables’ the choice of the survey DNFCS-3 and the compound CHLORPYRIFOS a further selection is made of 12 products (see 6.2 ): GREEN BEANS, SPINACH, POTATOES, CARROT, SWEET PEPPER, APPLE, PEACH, GRAPE, LEMON, ORANGE, MANDARIN TANGERINES and GRAPEFRUIT.

Finally, after these selections the MCRA input form is reached (see Figure 34).

Figure 34: MCRA input form for basic acute risk analysis

Table 3 lists the main options of this screen for this basic analysis:

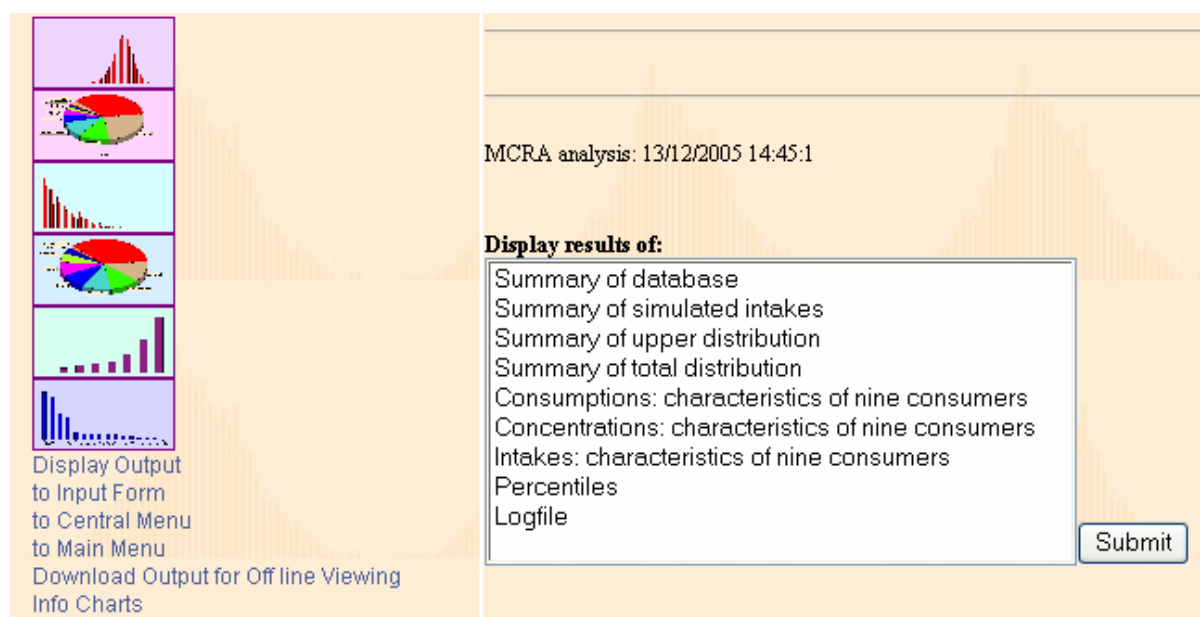
Input form	
risk type	Acute
uncertainty analysis	No
concentration model	empirical
Number of Monte Carlo iterations	200000



Number of computational chunks	10
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	99

**Table 3: Input form options: basic acute risk analysis**

After clicking the ‘submit MCRA job’ in Figure 34 the analysis is started. When the analysis is finished you can view the results using the ‘view output’ button in the central menu. This gives the screen of Figure 35:



**Figure 35: View output window for basic acute risk analysis**

After choosing ‘Logfile’ and clicking the ‘Submit’ button in Figure 35 the logfile is shown. In Table 4 you find the main characteristics concerning this analysis taken from the logfile.

<b>Log-file</b>	
Survey name	DNFCS-3
Compound code	CHLORPYRIFOS
Number of products	12
Acute reference dose (ARfD)	100
Acceptable daily intake (ADI)	10
Number of detects	364
Number of non-detects	2408
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	12191

**Table 4: Information taken from the log-file**

After choosing 'Summary of database' and clicking the 'Submit' button in Figure 35 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 product you will find the code and label.

In the upper part the average consumption over all consumers and all days (MeanCons) together with the average consumption on consumption days only (MeanConsDay) is given. In addition, the number of consumption days (NconsDays) and total number of days (Ndays) are displayed also expressed as percentage consumption days (%PosCons). So, there are 12500 days (2 days for 6250 consumers). Green Beans is consumed on 1266 days (10.1%). The average consumption of Green Beans on these days is  $11.8 \times 12500/1266 = 116.5$  g.

```

Compound:  CHLORPYRIFOS
*****
SUMMARY DATABASE
Code       : product code
Product    : product label
MeanCons   : 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
%PosCons   : percentage consumption days
MeanConc   : mean concentration all samples (after processing)
MeanPosConc : mean concentration positives (after processing)
NPosConc   : number of positive concentrations
NConc      : total number of concentration measurements
%PosConc   : percentage positive concentrations
*****

```

FOOD CONSUMPTION DATA						
Code	Product	MeanCons (g)	MeanConsDays (g)	NConsDays	NDays	%PosCons (%)
17111.10	GREEN BEANS	11.80	116.5	1266	12500	10.1
18115.20	SPINACH	8.08	223.0	453	12500	3.6
18401.10	POTATOES	138.52	173.1	10005	12500	80.0
18402.20	CARROT	12.12	36.4	4165	12500	33.3
18506.10	SWEET PEPPER	3.33	16.5	2533	12500	20.3
19201.10	APPLE	61.50	99.4	7737	12500	61.9
19304.10	PEACH	2.08	7.3	3538	12500	28.3
19401.10	GRAPE	13.23	34.2	4836	12500	38.7
19501.10	LEMON	1.54	4.2	4589	12500	36.7
19502.10	ORANGE	57.06	99.6	7163	12500	57.3
19503.10	MANDARIN TA	8.83	41.2	2677	12500	21.4
19507.10	GRAPEFRUIT	4.06	28.5	1785	12500	14.3

COMPOUND CONCENTRATION DATA						
Code	Product	MeanConc (mg/kg)	MeanPosConc (mg/kg)	NPosConc	NConc	%PosConc (%)
17111.10	GREEN BEANS	0.0120	0.4367	6	218	2.8
18115.20	SPINACH	0.0041	0.6400	1	158	0.6
18401.10	POTATOES	0.0005	0.0600	1	119	0.8
18402.20	CARROT	0.0014	0.1100	2	161	1.2
18506.10	SWEET PEPPER	0.0023	0.2000	3	256	1.2
19201.10	APPLE	0.0014	0.0431	13	397	3.3
19304.10	PEACH	0.0065	0.1100	6	101	5.9
19401.10	GRAPE	0.0176	0.1782	68	689	9.9
19501.10	LEMON	0.0105	0.0825	8	63	12.7
19502.10	ORANGE	0.0475	0.1161	139	340	40.9
19503.10	MANDARIN TA	0.0583	0.1305	84	188	44.7
19507.10	GRAPEFRUIT	0.0685	0.1703	33	82	40.2

**Table 5: Summary of the database, consumptions and compounds**

The lower part of Table 5 displays information concerning the compound concentrations. The mean concentrations of all samples (MeanConc) and of positives only (MeanPosConc) are calculated. The

number of positives (NPosConc) and the total number of concentration measurements (NConc) are given as well as the percentage of positive concentrations (%PosConc). In this example for Green Beans 6 positive concentrations are found out of 218 samples (2.8%). The mean concentration of the positive samples is  $0.012 \times 218/6 = 0.4367$  mg/kg.

After choosing 'Summary of simulated intakes' and clicking the 'Submit' button in Figure 35 essentially you will find 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 6 200,000) times from the days in the food consumption database (with replacement) irrespective of consumer. For each consumed product per day a concentration is sampled from the distribution of concentrations of that product. Multiplying consumptions and concentrations gives the intake per product. Summing up the intakes over the products and dividing the sum by the bodyweight of the consumer of the sampled day gives the total intake of the compound (here CHLORPYRIFOS) expressed in microgram per kg bodyweight per day. For both the consumption and concentration part you will find one additional statistic, DeltaC and DeltaR respectively. This statistic displays the difference expressed as percentage between the mean (MeanCons resp. MeanConc) of the simulated data and the mean of the data as found in your database. The better the simulation resembles the database the lower the percentages indicating the deviation between simulated data and input data.

```

Compound:  CHLORPYRIFOS
*****
SUMMARY SIMULATION
Code       : product code
Product    : product label
MeanCons   : 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
N          : total number of iterations
%PosCons   : percentage consumption days
MeanConc   : mean concentration in simulations with positive
            amount consumed (after processing)
DeltaR     : difference (%) compared to average concentration
            in database
NPosConc   : number of positive concentrations in simulations
            with positive amount consumed.
NConc      : total number of concentration measurements
            (detects and non-detects) in simulations with
            positive amount consumed.
%PosConc   : percentage positive concentrations
ProcFact   : mean processing factor
*****

```

FOOD CONSUMPTION DATA							
Code	Product	MeanCons (g)	DeltaC (%)	MeanConsDays (g)	NConsDay	N	%PosCons (%)
17111.10	GREEN BEANS	11.71	( -0.7)	116.16	20165	200000	10.1
18115.20	SPINACH	8.01	( -0.9)	224.46	7133	200000	3.6
18401.10	POTATOES	138.68	( 0.1)	173.34	160008	200000	80.0
18402.20	CARROT	12.13	( 0.1)	36.25	66949	200000	33.5
18506.10	SWEET PEPPER	3.34	( 0.2)	16.38	40777	200000	20.4
19201.10	APPLE	61.23	( -0.4)	98.83	123918	200000	62.0
19304.10	PEACH	2.08	( 0.4)	7.30	57131	200000	28.6
19401.10	GRAPE	13.11	( -0.8)	33.88	77407	200000	38.7
19501.10	LEMON	1.52	( -1.7)	4.14	73296	200000	36.6
19502.10	ORANGE	56.76	( -0.5)	98.86	114833	200000	57.4
19503.10	MANDARIN TA	8.81	( -0.2)	40.99	42988	200000	21.5
19507.10	GRAPEFRUIT	4.03	( -0.8)	28.01	28786	200000	14.4

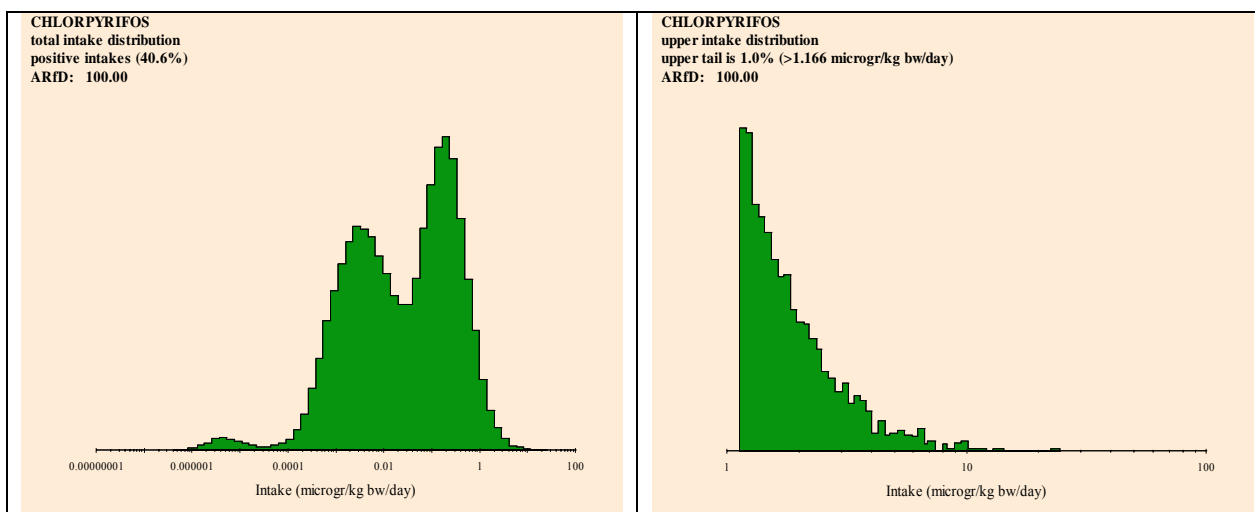
  

COMPOUND CONCENTRATION DATA							
Code	Product	MeanConc (mg/kg)	DeltaR (%)	NPosconc	NConc	%PosConc (%)	ProcFact
17111.10	GREEN BEANS	0.0117	( -2.6)	555	20165	2.8	1.00
18115.20	SPINACH	0.0037	( -9.2)	41	7133	0.6	1.00
18401.10	POTATOES	0.0005	( -2.3)	1314	160008	0.8	1.00
18402.20	CARROT	0.0014	( -0.1)	820	66949	1.2	1.00

18506.10	SWEET PEPPER	0.0022	( -6.0)	460	40777	1.1	1.00
19201.10	APPLE	0.0014	( 1.4)	4106	123918	3.3	1.00
19304.10	PEACH	0.0064	( -2.1)	3333	57131	5.8	1.00
19401.10	GRAPE	0.0173	( -1.7)	7618	77407	9.8	1.00
19501.10	LEMON	0.0107	( 2.5)	9417	73296	12.8	1.00
19502.10	ORANGE	0.0475	( 0.2)	47062	114833	41.0	1.00
19503.10	MANDARIN TA	0.0580	( -0.6)	19073	42988	44.4	1.00
19507.10	GRAPEFRUIT	0.0691	( 0.8)	11594	28786	40.3	1.00

**Table 6: Summary of simulated intakes**

In the upper left part of Figure 35 five icons are displayed. Clicking the upper icon gives a histogram of the simulated total intakes (see the left plot in Figure 36). Clicking the icon, two icons lower, shows you the upper tail of this intake distribution (see the right plot of Figure 36). Here, the specified percentage is 1%, corresponding with an intake of 1.166 microgr/kg\_bw/day. The acute reference dose (ARfD) for Chlorpyrifos is equal to 100 microgr/kg bw/day.



**Figure 36: Total intake distribution and upper tail (1%)**

If the intake distribution exceeds the ARfD, that part of the intake distribution is displayed in red. This is obvious not the case in Figure 36.

Choosing 'Summary of total distribution' and clicking the 'Submit' button in Figure 35 gives the summary displayed in Table 7. There you find the contributions expressed as percentages (RelContr) of each product to the total intake distribution. A pie chart of the products with the 9 largest contributions can be displayed by clicking the second icon from the top in Figure 35. This pie chart is shown in the left plot of Figure 37. 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 product are given in Table 7. The last column (%Zeros) shows the number of zero intakes per product. If %Zeros is greater than 97.5% the p97.5% is lower than the mean!

Let's take Orange as an example. It contributes 65.3% to the total intake distribution and its average concentration is 0.047 microgr/kg bw/day. The p2.5%, median, p97.5% and %Zeros of the total intake distribution of Orange are: 0.000, 0.000, 0.512 microgr/kg bw/day and 76.6% respectively.

```

Compound:  CHLORPYRIFOS
*****
Contribution (%) per product to the total distribution
RelContr:  relative contribution (%) per product
p2.5%    :  2.5% perc. of intake distr. per product (microgr/kg bw/day)
Mean     :  mean of intake distr.per product (microgr/kg bw/day)
Median   :  median of intake distr.per product (microgr/kg bw/day)
p97.5%   :  97.5% perc. of intake distr. per product (microgr/kg bw/day)
%Zeros   :  percentage zeros per product

```

Product	RelContr %	Mean (microgr/kg bw/day)	Median (microgr/kg bw/day)	p2.5% (microgr/kg bw/day)	p97.5% (microgr/kg bw/day)	%Zeros %
GREEN BEANS	3.3	0.002	0.000	0.000	0.000	99.7
SPINACH	0.6	0.000	0.000	0.000	0.000	100.0
POTATOES	1.7	0.001	0.000	0.000	0.000	99.3
CARROT	0.4	0.000	0.000	0.000	0.000	99.6
SWEET PEPPER	0.2	0.000	0.000	0.000	0.000	99.8
APPLE	2.4	0.002	0.000	0.000	0.000	98.0
PEACH	0.4	0.000	0.000	0.000	0.000	98.3
GRAPE	5.8	0.004	0.000	0.000	0.006	96.2
LEMON	0.4	0.000	0.000	0.000	0.001	95.3
ORANGE	65.3	0.047	0.000	0.000	0.512	76.6
MANDARIN TA	13.6	0.010	0.000	0.000	0.116	90.3
GRAPEFRUIT	5.9	0.004	0.000	0.000	0.007	94.2

**Table 7: Contribution to the total intake distribution**

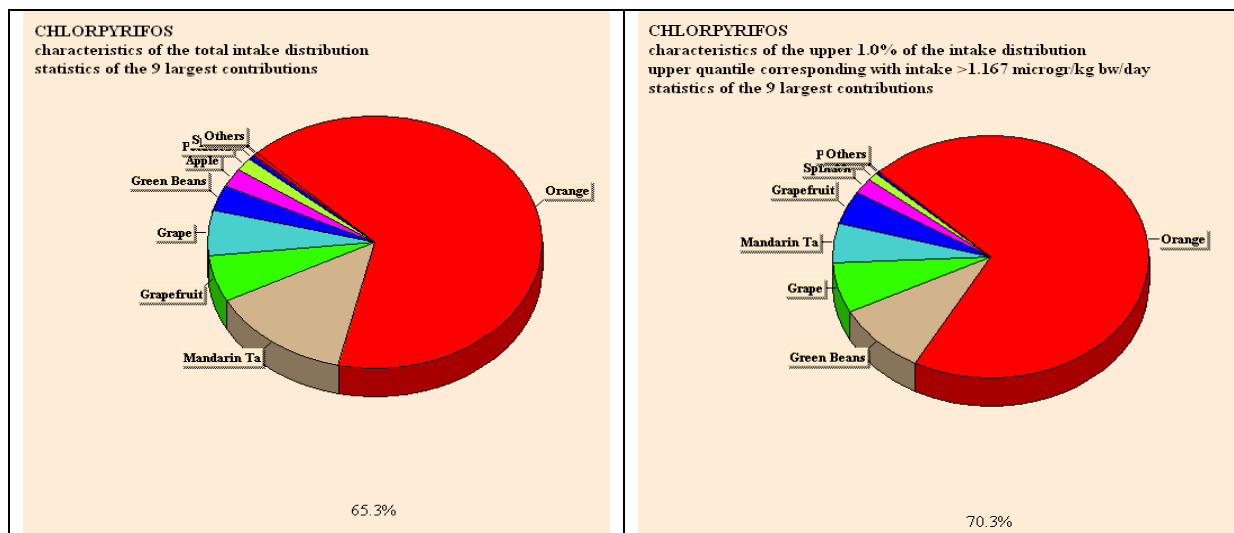
Choosing ‘Summary of upper distribution’ and clicking the ‘Submit’ button in Figure 35 gives the summary displayed in Table 8. There you find the contributions expressed as percentages (RelContr) of each product to the upper tail according to the specified percentage (here 1%). A pie chart of the products with the 9 largest contributions to the upper 1% of the intake distribution can be displayed by clicking the fourth icon from the top in Figure 35. This pie chart is shown in the right plot of Figure 37. 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 product are given in Table 8. This table shows the same statistics for that part of the intakes per product that correspond with the upper 1% of the intake distribution. So this table represents a drilldown of the upper 1% of the intake distribution and shows statistics for the intakes per products that constitutes the upper 1%. See also Figure 38, where the mean intake for each product in the upper tail is graphically displayed. The last column (%Zeros) shows the number of zero intakes per product.

Let’s take Orange again as an example. It contributes 70.3% to the upper 1% of the intakes and its average concentration is 1.525 microgr/kg\_bw/day. The p2.5%, median, p97.5% and %Zeros of the upper 5% of the intakes are: 0.000, 1.404, 4.512 microgr/kg bw/day and 18.7% respectively. Note that Mandarin Ta has the second largest contribution to the total intake distribution (see Table 7) whereas Green Beans has the second largest to the upper 1% of the intake distribution.

Product	RelContr %	Mean (microgr/kg bw/day)	Median (microgr/kg bw/day)	p2.5% (microgr/kg bw/day)	p97.5% (microgr/kg bw/day)	%Zeros %
GREEN BEANS	9.5	0.207	0.000	0.000	3.490	94.2
SPINACH	2.1	0.046	0.000	0.000	0.000	98.5
POTATOES	0.3	0.006	0.000	0.000	0.000	98.9
CARROT	0.1	0.002	0.000	0.000	0.000	99.4
SWEET PEPPER	0.0	0.000	0.000	0.000	0.000	99.8
APPLE	1.1	0.023	0.000	0.000	0.065	95.9
PEACH	0.1	0.002	0.000	0.000	0.000	97.2
GRAPE	6.8	0.148	0.000	0.000	1.917	88.8
LEMON	0.0	0.000	0.000	0.000	0.002	94.6
ORANGE	70.3	1.525	1.404	0.000	4.512	18.7
MANDARIN TA	5.2	0.112	0.000	0.000	1.532	82.7
GRAPEFRUIT	4.5	0.097	0.000	0.000	1.550	88.2

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

In Figure 37 you find a graphical display of the figures in Table 7 and Table 8. These are obtained by clicking the two icons in Figure 35 with a pie chart.



**Figure 37: Relative contribution of products to the total intake distribution and upper tail (1%)**

By scrolling your mouse over the piechart, you find the contribution of each product at the bottom of the chart. Here, the relative contribution for Orange is displayed.

Choosing 'Percentiles' and clicking the 'Submit' button in Figure 35 gives the summary displayed in 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 34) are shown. In this example, 0.01% of the population has an intake higher than 7.28073 microgr/kg bw/day. Also the mean and maximum are given.

```

*****
Random sampling is based on seed :          0
Number of iterations (consumers):  200000 out of 6250
CHLORPYRIFOS (microgr/kg bw/day) consumption: 200000 out of 200000
*****

Compound:  CHLORPYRIFOS

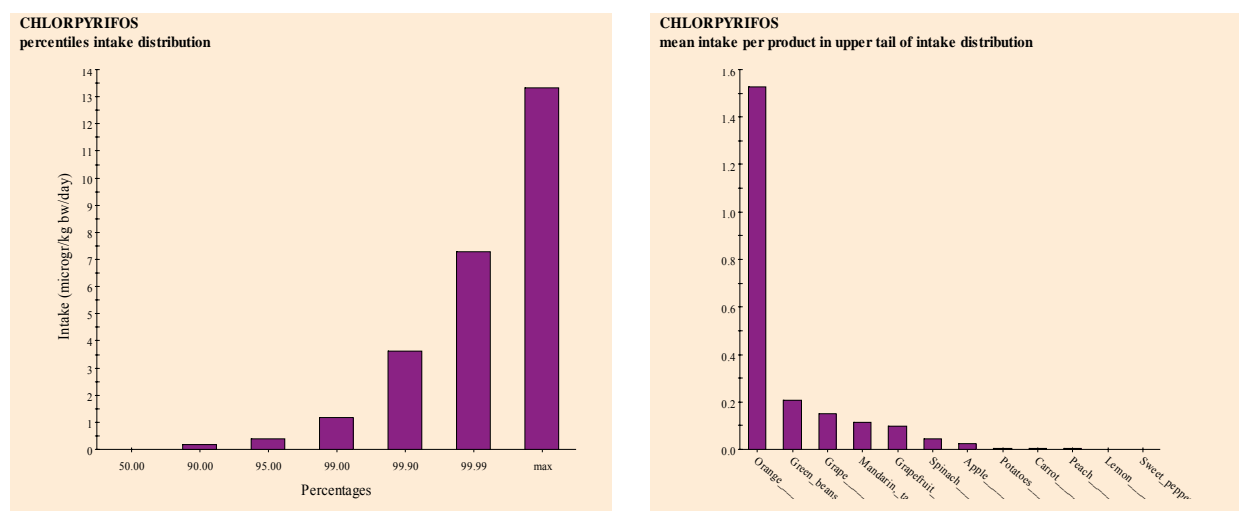
*****
          Percentiles, maximum and average intake
*****

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

50.00           0.00000
90.00           0.19172
95.00           0.38571
99.00           1.16667
99.90           3.61038
99.99           7.28073
mean            0.07234
maximum         13.34071
  
```

**Table 9: Percentiles for the acute intake distribution**

In the left plot of Figure 38 you find a graphical display of the figures in Table 9. This is obtained by clicking the corresponding icon in Figure 35. The right plot of Figure 38 is obtained by clicking the lowest icon in Figure 35 and shows the mean intake per product in the upper tail (the third column of Table 8).



**Figure 38: Percentiles and mean intake per product of the acute intake distribution**

Choosing respectively ‘Consumption: characteristics of nine consumers’, ‘Concentration: characteristics of nine consumers’ or ‘Intake: characteristics of nine consumers’ and clicking the ‘Submit’ button in Figure 35 gives the summaries displayed in Table 10, Table 11 and Table 12. In these tables the results of a drill down of consumers according to a specified percentage (here 97.5%) are displayed. For the nine consumers around the specified percentage the sampled consumption, compound concentration and intake are displayed (as well as the consumer number, weight and age for the intake in Table 12). So for consumer number 363571 (86 kg, 33 years), the total intake is 0.664 microgr/kg bw/day. This consumer consumed 2 products (Potatoes 429.0 gr, Grapefruit 150.3 gr). Only on Grapefruit a compound was found (0.380 mg/kg), so the total intake for this consumer is  $150.3 \times 0.380/86 = 0.664$  microgr/kg bw/day.

Compound: CHLORPYRIFOS									
Drill down: consumption (g/day) per product of the 9 consumers around the specified percentage (97.50%) for the intake distribution									
product	-4	-3	-2	-1	0	+1	+2	+3	+4
GREEN BEANS	.	4.7	.	.	.	.	.	.	.
SPINACH	.	.	.	.	.	416.2	.	.	.
POTATOES	429.0	6.8	30.00	152.15	150.0	150.0	343.2	191.0	191.0
CARROT	.	7.4	.	.	.	.	9.1	.	.
SWEET PEPPER	.	2.7	.	.	.	.	18.0	.	.
APPLE	.	.	296.39	9.68	.	80.0	540.0	120.0	120.0
PEACH	.	.	2.75	.	115.0	.	.	.	.
GRAPE	.	.	1.50	150.00	54.1	.	348.4	.	.
LEMON	.	.	0.48	2.34	0.7	.	4.1	.	.
ORANGE	.	650.0	0.70	120.00	200.0	2.2	470.0	240.0	240.0
MANDARIN TA	.	.	55.00	110.00	.	242.2	.	30.0	30.0
GRAPEFRUIT	150.3	.	0.70	.	.	.	.	.	.

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

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

	-4	-3	-2	-1	0	+1	+2	+3	+4
product									
GREEN BEANS	.	0.000	.	.	.	.	.	.	.
SPINACH	.	.	.	.	.	0.000	.	.	.
POTATOES	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
CARROT	.	0.000	.	.	.	.	0.000	.	.
SWEET PEPPER	.	0.160	.	.	.	.	0.000	.	.
APPLE	.	.	0.000	0.000	.	0.000	0.000	0.000	0.000
PEACH	.	.	0.000	.	0.000	.	.	.	.
GRAPE	.	.	0.000	0.000	0.400	.	0.000	.	.
LEMON	.	.	0.000	0.000	0.040	.	0.000	.	.
ORANGE	.	0.080	0.000	0.130	0.190	0.280	0.110	0.100	0.110
MANDARIN TA	.	.	0.130	0.250	.	0.230	.	0.150	0.070
GRAPEFRUIT	0.380	.	0.210	.	.	.	.	.	.

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

Compound: CHLORPYRIFOS									
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
Consumer	363571	374841	183954	345662	322991	172661	354161	372933	372933
Body Weight	86	79	11	65	90	85	78	43	43
Age	33	29	1	53	66	30	53	11	11
Total	0.664	0.664	0.663	0.663	0.663	0.663	0.663	0.663	0.663
Product									
GREEN BEANS	.	0.000	.	.	.	.	.	.	.
SPINACH	.	.	.	.	.	0.000	.	.	.
POTATOES	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
CARROT	.	0.000	.	.	.	.	0.000	.	.
SWEET PEPPER	.	0.005	.	.	.	.	0.000	.	.
APPLE	.	.	0.000	0.000	.	0.000	0.000	0.000	0.000
PEACH	.	.	0.000	.	0.000	.	.	.	.
GRAPE	.	.	0.000	0.000	0.240	.	0.000	.	.
LEMON	.	.	0.000	0.000	0.000	.	0.000	.	.
ORANGE	.	0.658	0.000	0.240	0.422	0.007	0.663	0.558	0.614
MANDARIN TA	.	.	0.650	0.423	.	0.656	.	0.105	0.049
GRAPEFRUIT	0.664	.	0.013	.	.	.	.	.	.

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

## 4.2 Chronic risk assessment: betabinomial/lognormal

In this example, output is shown for a chronic risk assessment for antibiotic lasalocide. The example concerns the intake via eggs. This database is not centrally supplied. Find in Figure 39 the MCRA input form:

**Figure 39: MCRA input form for chronic risk analysis betabinomial/lognormal**

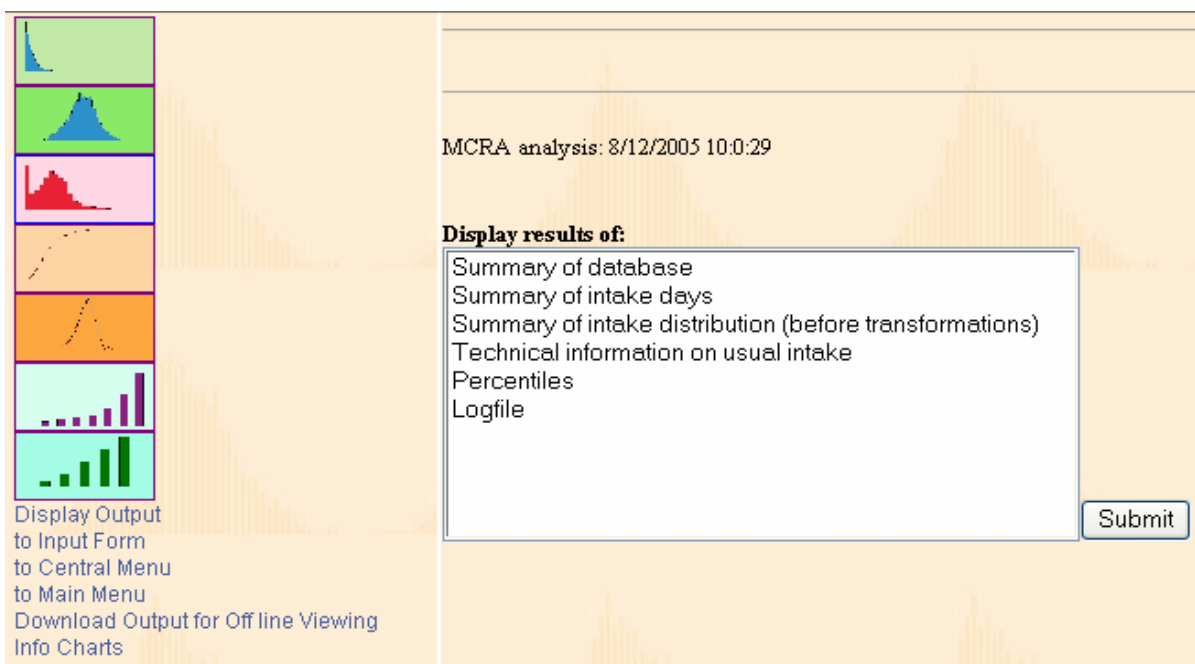


Table 13 lists the main options:

<b>Input form</b>	
risk type	chronic
uncertainty analysis	no
concentration model	empirical
intake model	betabinomial/lognormal
concentration data	system defaults
output	system defaults

**Table 13: Input form options: betabinomial/lognormal**

After clicking the ‘submit MCRA job’ the analysis is started. When the analysis is finished you can view the results using the ‘view output’ button in the central menu. This gives the screen of Figure 40:



**Figure 40: View output window for chronic risk analysis betabinomial/lognormal**

In Table 14 you find the main characteristics concerning this analysis taken from the logfile. The intake frequency function is estimated with a betabinomial model with smoothing parameter  $\phi$ . The model for the logarithmically transformed intake amounts is based on REML.

<b>Log-file</b>	
Number of products	1
Acute reference dose (ARfD)	*
Acceptable daily intake (ADI)	5
Number of detects	31
Number of non-detects	219
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	8630
Replace all non-detects	

Multiplicationfactor for LOR	0.5		
Intake Frequency: Betabinomial model			
=====			
Estimates of parameters			
-----			
Parameter	estimate	s.e.	t(*)
Mean probability	0.6904	0.004481	154.07
Overdispersion	0.1811	0.01295	13.99
Intake (amounts): REML model			
=====			
Estimates of parameters			
-----			
Parameter	estimate	s.e.	t(*)
Mean ln(intake)	-2.578	0.02017	-127.77
Between consumer variance	0.4590	0.05614	8.17
Within consumer variance	2.723	0.06470	42.09

**Table 14: Information in log-file**

The Dutch National Food Consumption Survey is used, which has 2 days for 6250 consumers, so 12500 days in total. Lasalocide measurements in eggs were made in 250 cases, 31 of which had a positive value. In this analysis the other 219 non-detects were replaced by  $\frac{1}{2}LOR = \frac{1}{2}0.05 = 0.025$  mg/kg.

Choosing 'Summary of database' and clicking the 'Submit' button in Figure 40 gives the summary displayed in Table 15.

```

Compound: LASALOCIDE
*****
SUMMARY DATABASE
Code      : product code
Product   : product label
MeanCons  : 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
%PosCons  : percentage consumption days
MeanConc  : mean concentration all samples (after processing)
MeanPosConc : mean concentration positives (after processing)
NPosConc  : number of positive concentrations
NConc     : total number of concentration measurements
%PosConc  : percentage positive concentrations
*****

FOOD CONSUMPTION DATA

Code      Product      MeanCons  MeanConsDays  NConsDays  NDays  %PosCons
          (g)          (g)
11311.00 WHOLE EGG C    18.00      26.1          8630       12500   69.0

COMPOUND CONCENTRATION DATA

Code      Product      MeanConc  MeanPosConc  NPosConc  NConc  %PosConc
          (mg/kg)  (mg/kg)
11311.00 WHOLE EGG C    0.4825     3.8910       31         250     12.4

```

**Table 15: Summary of the database, consumptions and compounds**

So on average everyone consumes 18 g of egg per day. This is an average of 26.1 g on 8630 consumption days and 0 g on 12500 - 8630 = 3870 non-consumption days. The average of the 31 positive lasalocide concentrations and 219 times 0 mg/kg is 0.4825 mg/kg.

Choosing 'Summary of intake days' and clicking the 'Submit' button in Figure 40 gives the summary displayed in Table 16.

Compound: LASALOCIDE			
LONG TERM EXPOSURE (USUAL INTAKE)			
Summary of intake days in the data			
proportion and number of consumers with positive intake on:			
	0 days	0.1346	841
	1 days	0.3501	2188
	2 days	0.5154	3221
	All intakes	Positive intakes	%
Number of observations	12500	8630	69.0
Number of consumers	6250	5409	86.5

**Table 16: Summary of intake days**

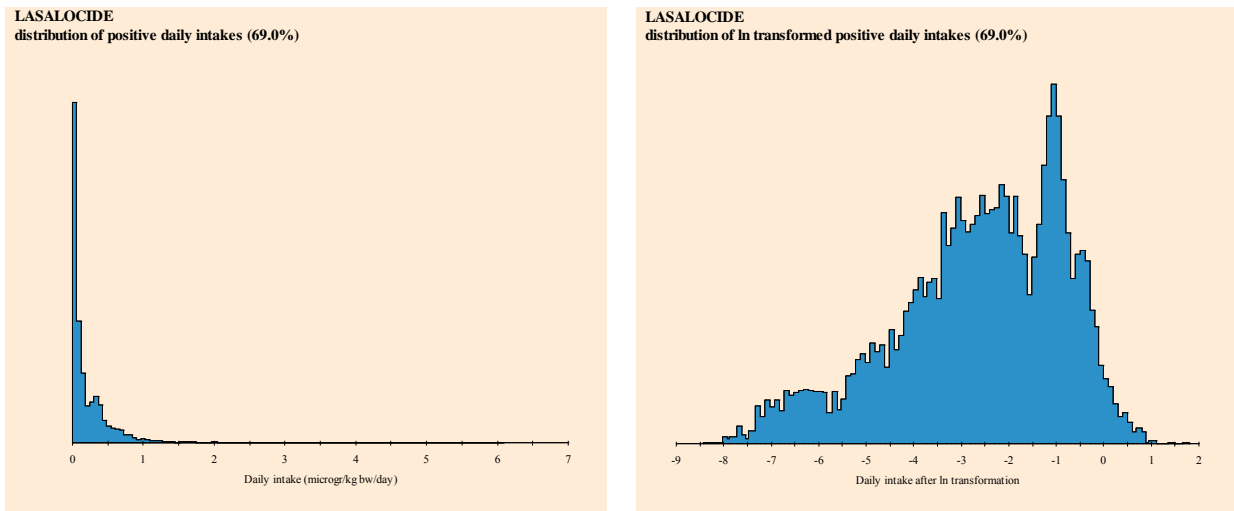
Choosing 'Summary of intake distribution' and clicking the 'Submit' button in Figure 40 gives the summary displayed in Table 17.

Compound: LASALOCIDE	
LONG TERM EXPOSURE (USUAL INTAKE)	
Summary of all intake data (zeros included)	
Number of observations =	12500
Mean =	0.154
Median =	0.031
Minimum =	0.000
Maximum =	6.368
Lower quartile =	0.000
Upper quartile =	0.181
Summary of positive intake data (without zeros)	
Number of observations =	8630
Mean =	0.223
Median =	0.095
Minimum =	0.000
Maximum =	6.368
Lower quartile =	0.026
Upper quartile =	0.322

**Table 17: Summary statistics of the calculated daily intakes**

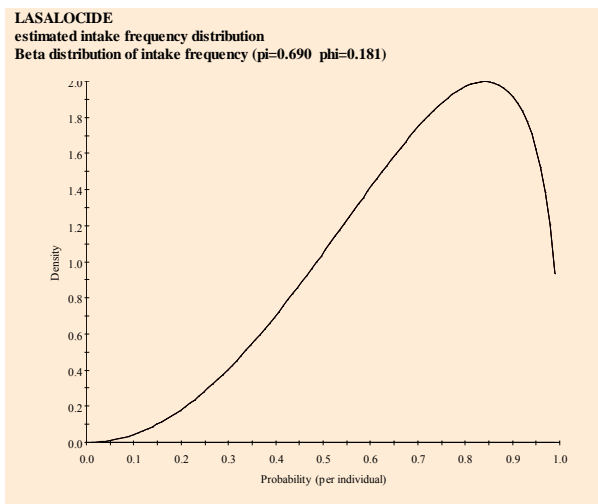
In the upper left part of Figure 40 seven icons are displayed. Clicking the upper icon gives the empirical distribution of the non-zero daily intake values, calculated as daily consumption values times average concentration (0.4825 in this case) divided by body weight (see the left plot in Figure 41). Note the spike near 0 (note that true zero intakes are already excluded here), and the bimodal character of the rest of the distribution (possibly related to the consumption of one or two whole eggs?). This distribution is definitely non-normal. A logarithmic transformation is requested for the non-zero daily intakes.

Clicking the second icon from above shows you the intake distribution after a logarithmic transformation (see the right plot of Figure 41). 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 85.



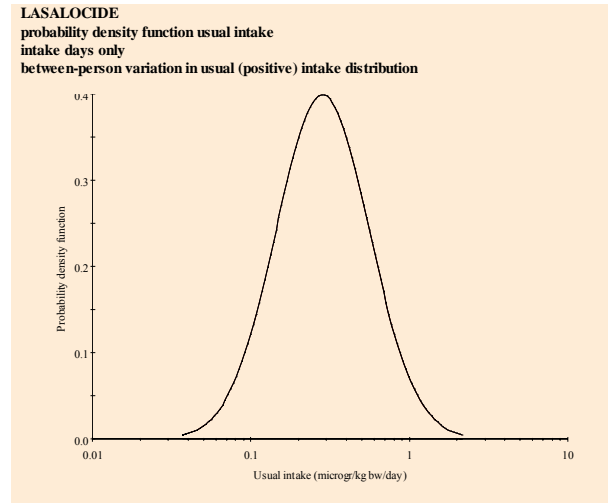
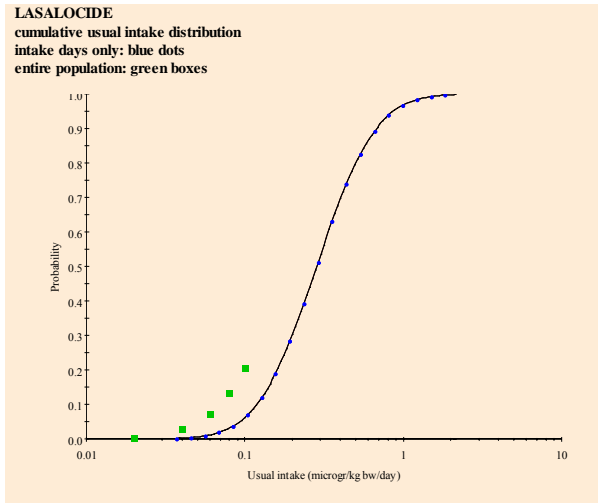
**Figure 41: Untransformed and ln transformed intake distribution**

Clicking the third icon from above shows you the intake frequency distribution (see Figure 42). The mean probability for a consumer to have an intake is equal to 0.69 but dispersion parameter  $\phi$  (= 0.181) of the betabinomial distribution provides that each consumer has its own probability to have an intake.



**Figure 42: Intake frequency**

Clicking the fourth and the fifth icon from above shows you respectively the left and the right plot of Figure 43. The left plot is the cumulative usual intake distribution (black line). The green boxes indicate the percentiles of the usual intake distribution for the entire population as requested in the input form of Figure 39. The right plot is the density of the usual intake distribution for positives only.



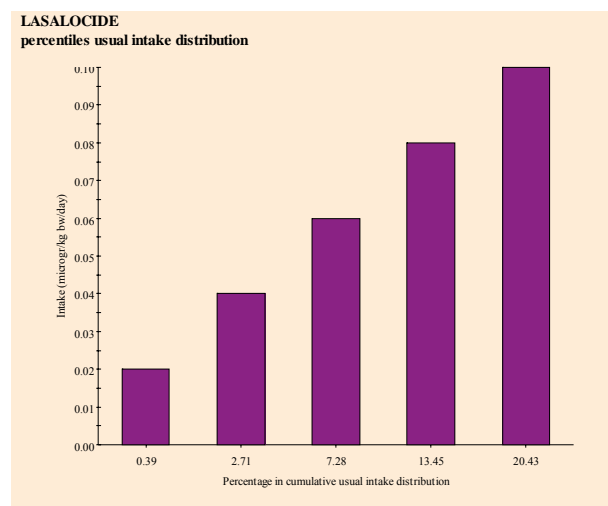
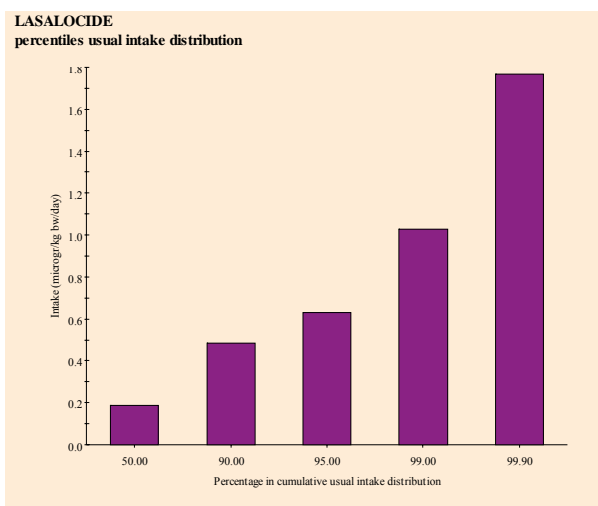
**Figure 43: Cumulative usual intake and density function**

Choosing 'Percentiles' and clicking the 'Submit' button in Figure 40 gives percentiles of the usual intake distribution for the entire population (see Table 18).

Compound: LASALOCIDE	
LONG TERM EXPOSURE (USUAL INTAKE)	
Percentage	Entire population
50.00	0.19760
90.00	0.50817
95.00	0.66043
99.00	1.07493
99.90	1.84598

**Table 18: Percentiles of the usual intake distribution**

Clicking the sixth and the seventh icon from above shows you respectively the left and the right plot of Figure 44. In the left plot the percentiles of Table 18 are graphically displayed whereas in the right plot percentages derived from percentiles are shown.



**Figure 44: Percentiles and percentages**

### 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 45). Some more information is available under the link [InfoCharts](#) on the View Output screen. ComponentOne Chart graphs can be printed, or copied to the Windows 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 that 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 45).

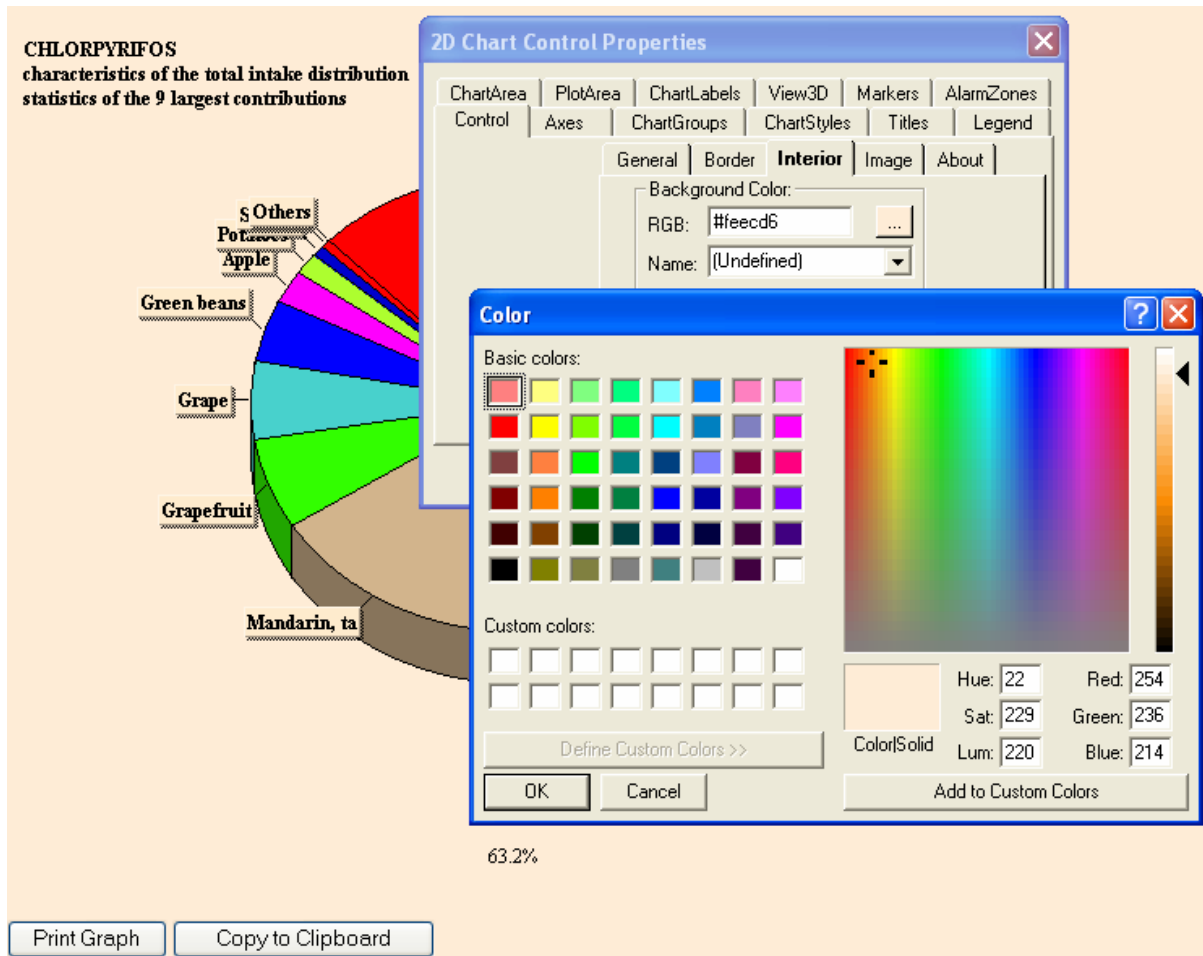


Figure 45: Property editor

### 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

This chapter describes the stochastic (or Monte Carlo) models behind the MCRA program. These models assess acute or chronic 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 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 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 product  $k$  (in g),  $c_{ijk}$  is the concentration of the chemical compound in product  $k$  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 products accounted for in the model. Note that the definition of ‘product’ 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 the quantities  $x_{ijk}$ ,  $w_i$  and  $c_{ijk}$  are assumed to arise from probability distributions describing the variability for food consumption and weight,  $p(x_1, \dots, x_p, w)$ , and for compound concentrations in each food product,  $p_k(c)$ . 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).

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 ( $x_1, \dots, x_p$ ), body weight ( $w$ ) and age ( $a$ ) are available for 6250 consumers on 2 consecutive days. Depending on the problem, Monte Carlo 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).

In general a recipe data base is necessary to convert the amounts of food as consumed (e.g. pizza) to amounts of products ( $x_1, \dots, x_p$ ) of primary agricultural products 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.

Given these probability distributions (or estimates thereof) Monte Carlo 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). 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 5.6).

## 5.2 How to deal with limited information

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 5.3.1.1 ) or a parametric (see 5.3.1.2 and 5.3.1.3 ) approach. In a *parametric approach* the data are modelled with an appropriate distributional form (e.g. lognormal with parameters  $\sigma$  and  $\mu$ ). 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 products there are far less data than for others, it may be sensible to consider *pooling* procedures for means and or variances of the concentration distributions (see 5.3.1.3 ).

If the amount of data is limited, this may lead to a relatively large *sampling uncertainty*. Bootstrap procedures may be used to assess the magnitude of this uncertainty (see 5.7 ).

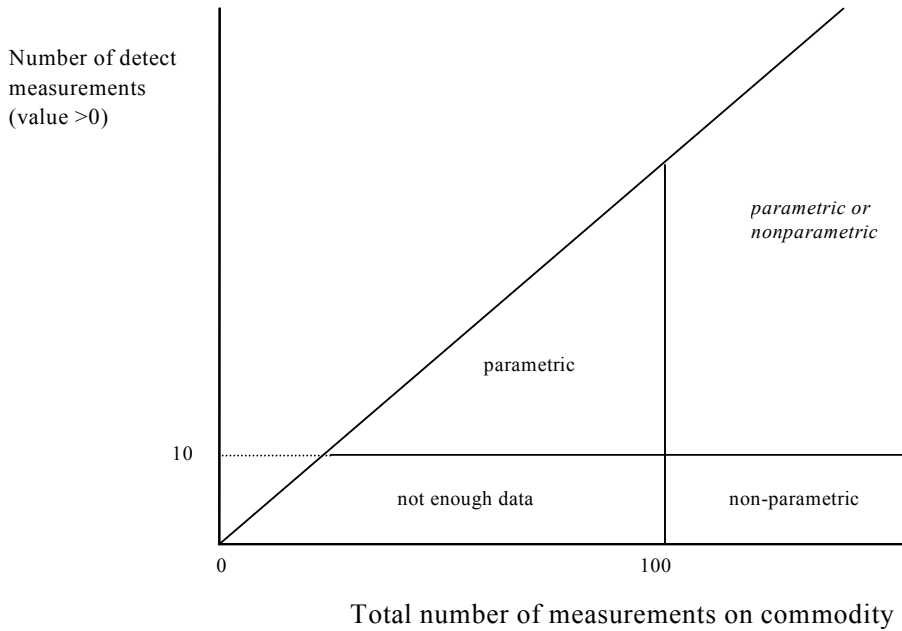
[Back to: Concentration model options](#)

### 5.2.1 The choice between a parametric and non-parametric approach

How many compound data are required for a sensible calculation of upper-tail percentiles in the exposure 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 95<sup>th</sup> percentile and at least 100 measurements to estimate the 99<sup>th</sup> percentile.

More generally, the number of measurements per food product ( $n$ ) should at least equal  $1/(1-p\%/100)$  to allow a rough empirical estimate of the  $p^{\text{th}}$  percentile of the compound concentration distribution to be made. Of course, the risk assessment is only coarse with this minimum amount of data and larger sample sizes per food product 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 46 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 product.



**Figure 46: Use of non-parametric or parametric modelling for estimating 99% exposure percentile in relation to sample size and number of positive 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  $n_1 \dots n_c$ , which fall into  $c$  classes with limits  $c_1 \dots c_c$ . The number  $n_1$  is the number of positive (*detect*) 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  $c_2, \dots$ ;  $n_c$  is the number of samples that fall in between limits  $c_{c-1}$  and  $c_c$ .

For histogram data, parameters  $\mu$  and  $\sigma$  of the lognormal distribution can be obtained by fitting a normal distribution to a set of observations  $n_1 \dots n_c$ . 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  $\frac{1}{2} \cdot \text{LOR}$ ), 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 ( $\frac{1}{2} \cdot \text{LOR}$ ). 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:

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

with  $\Phi(\cdot)$ , 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  $N$   $\alpha$  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  $\mu$  and  $\sigma$ , or
- 2) the reported statistics are sufficient to extract  $\mu$  and  $\sigma$ , or
- 3) the information is redundant so various estimates for  $\mu$  and  $\sigma$  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 product-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  $\mu$  and  $\sigma^2$  if  $Y = \ln X$  is normally distributed with mean  $\mu$  and variance  $\sigma^2$ . 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  $\mu$  and  $\sigma$  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  $\mu$  and  $\sigma$ .

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 50<sup>th</sup> 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:

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

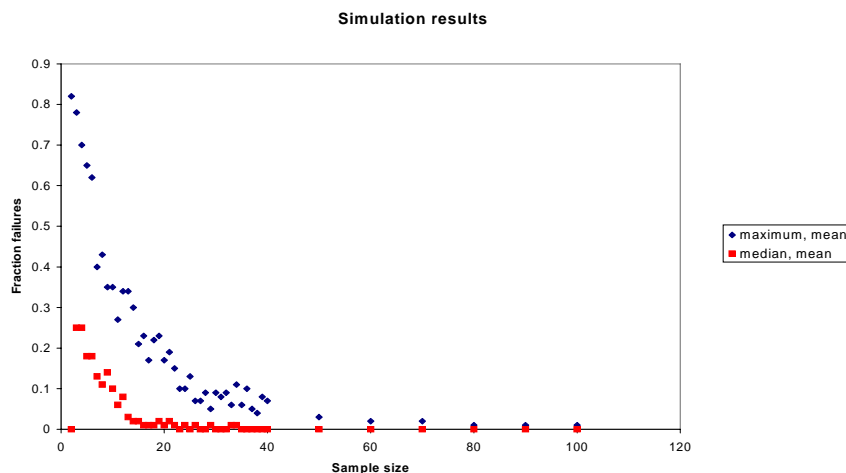
Approach 2: estimation based on a quantile and the mean.

Solving for  $\sigma$  using (1) and (4) gives:

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

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

For moderate to small sample sizes the estimation of  $\sigma$  fails because the discriminant is negative, i.e. 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 47 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  $\sigma$  is assessed by equating the discriminant to zero. Empirical results show that this works out very well for sample size  $n > 4$ , although  $\sigma$  is slightly biased upwards being a conservative estimate. In case of the median no solution to this problem is available so the estimate of  $\sigma$  is set to a missing value.



**Figure 47: Simulated fraction of failures versus sample size for estimation of  $\sigma$  based on the mean and respectively the maximum and median**

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

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

In case of the mean and maximum  $\sigma$  is estimated with:

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

Note that max 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,  $vc = \sqrt{(\text{variance})/\text{mean}}$ . Using (5), parameter  $\sigma$  is estimated with:

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

and  $\mu$  is estimated solving (1).

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

For estimation of  $\sigma$ , see approach 3. Using (4), parameter  $\mu$  is estimated with:

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

For the median, estimation of  $\mu$  simplifies to:

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

## 5.3 Concentration model: modelling of compound concentrations in consumed food

### 5.3.1 Distributional assumptions

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

#### 5.3.1.1 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 exposure values. To assess the risk-exposure, percentiles of the exposure distribution are estimated.

[Back to: Empirical](#)

#### 5.3.1.2 Parametric modelling of concentrations (binomial/lognormal no pooling)

In the parametric approach, compound concentrations per food product 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 product), 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  $\geq$  LOR on food products 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 product 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  $\mu$  and  $\sigma$  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 product  $k$  and  $I_{ijk}$  is sampled separately for each consumer  $i$  on occasion  $j$ .

[Back to: Binomial/lognormal \(no pooling\)](#)

### 5.3.1.3 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  $\mu$  and  $\sigma^2$ , representing mean and variance of the natural logarithm of the concentrations. However, estimation is often hampered because data on compounds in specific food products are sparse or even missing. In those cases, grouping of products into product 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 products. 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 product groups.** For each product the variance  $\sigma^2$  and mean  $\mu$  is estimated. Then, products are assigned to product groups which are composed of related products, e.g. a productgroup containing sorts of cabbages or a group containing all kind of berries. Products where agricultural use is allowed are remained separate from products where agricultural use is not allowed. The homogeneity of variances in the different product 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 product 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 products except the tested product 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 products into (sub) groups precedes the actual simulation of the intake distribution.

To summarise, actions are:

- calculate variances and means for each product
- classify products into groups
- test homogeneity of variances and equality of means within groups of products. Results are: not significant ( $p > 0.05$ ) or significant ( $p \leq 0.05$ ).
- take products(-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](#)

### 5.3.2 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 limit of reporting (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  $100p_{crop-treated}$ , then this knowledge may be used to infer that  $100(1-p_{crop-treated})\%$  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  $100(p_{non-detect} + p_{crop-treated} - 100)\%$  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](#)

### 5.3.3 Modelling of processing effects

Concentrations in the consumed food may be different from concentrations in the product as measured in monitoring programs (typically raw product) 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 product, 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 product  $k$  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 the information will be of limited quality, and this may be entered in the Monte Carlo modelling by specification of uncertainties. A practical proposal is to specify for each processing factor two values:

1.  $f_{k,nom}$ : the nominal value, typically some sort of mean 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, there will often be uncertainty due to the appropriateness of the processing study for the population of the risk assessment). The upper limit should be such that experts will easily agree that it is not set too low.

A typical data base entry might thus read:

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

and, confronted with the need to have processing factors for pears in a specific risk assessment, an expert may decide upon:

RAC	processing	$f_{k,nom}$	$f_{k,upp}$
pear	washing	0.5	0.8

In the Monte Carlo modelling, processing factors can be used in either of three ways (for each product  $k$  to be chosen by the user):

#### 5.3.3.1 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.



### 5.3.3.2 Processing (fixed factors)

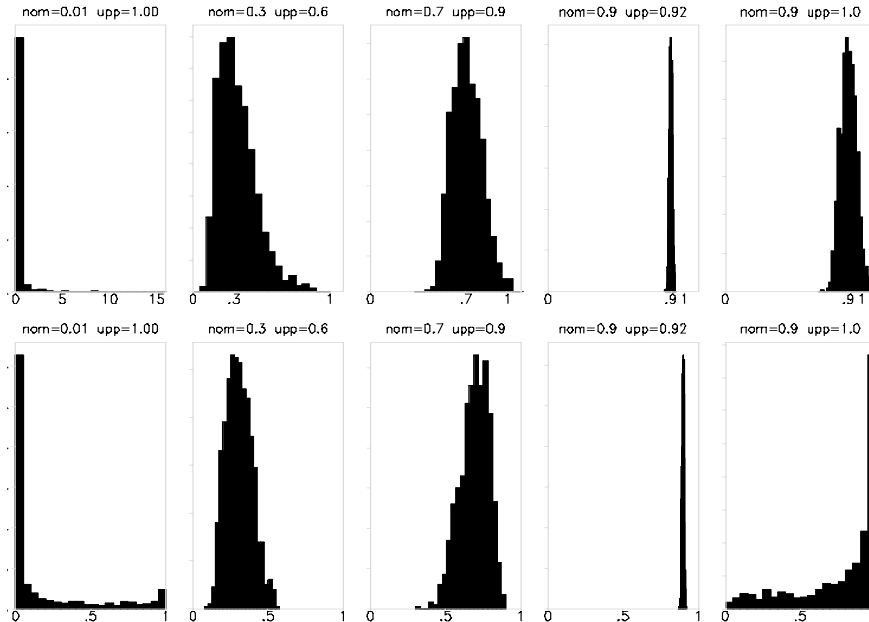
Use  $f_k = f_{k,upp}$ . Available information on specific processing effects is used, although still in a cautionary way (in accordance with the precautionary principle). Note that  $f_{k,nom}$  values need not to be specified.

### 5.3.3.3 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.

- 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 Monte Carlo 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 user-specified value (e.g. 0.01); this is useful computationally to avoid problems with logarithms.
- 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 Monte Carlo 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 user-specified values (e.g. 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 48) occasionally values above 1 are sampled (upper row, 1<sup>st</sup>, 3<sup>rd</sup> and 5<sup>th</sup> plot). A logit transformation should be considered instead.



**Figure 48: 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](#)

## 5.4 Modelling of unit variability

### 5.4.1 Introduction, variability in deterministic modelling

Variability in compound concentrations between individual units is a relevant factor in the assessment of short-term dietary exposure to 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 product  $k$  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 product 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 products 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 19: 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 97.5<sup>th</sup> 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](#)

#### 5.4.2 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 products 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 product 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 Monte Carlo 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 97.5<sup>th</sup> 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/products, 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 19. 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 ( $v = 97.5^{\text{th}} \text{ percentile}/\text{mean}$ ). 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](#)

### 5.4.2.1 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 product  $k$ , 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  $\mu$  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  $\mu$  and an externally specified value for  $cv_k$  the parameters  $a$  and  $b$  of the beta distribution are calculated as:

$$a = b(nu_k - 1)^{-1}$$

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

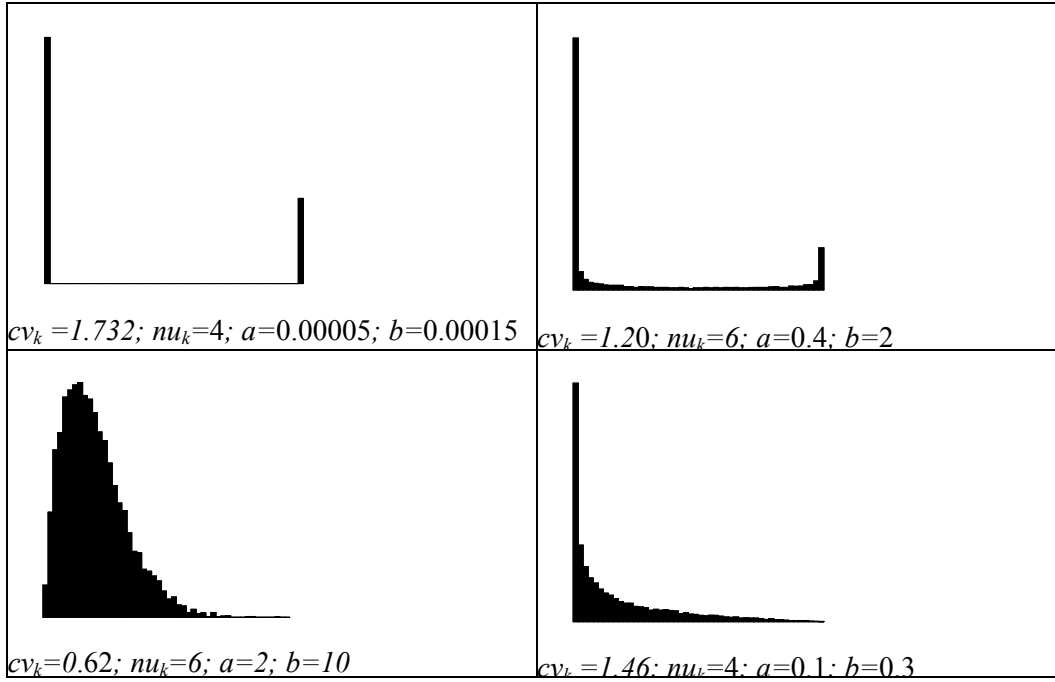
From the second formula it can be seen that  $cv_k$  should not be larger than  $\sqrt{nu_k - 1}$  in order to avoid negative values for  $b$ .

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 49 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 show an intermediate situation, moderate to high variability.



**Figure 49: 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](#)

#### 5.4.2.2 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  $\mu$  and  $\sigma$ , 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  $\sigma$  on the log-transformed scale with:

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

A variability factor  $v$  is converted into the standard deviation  $\sigma$  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  $\mu$  and  $\sigma$  representing the mean and standard deviation of the log-transformed concentrations. So

$$\ln(v) = 1.96\sigma - 1/2\sigma^2$$

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

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

The smallest positive root is taken as an estimate for  $\sigma$  (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](#)

#### 5.4.2.3 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](#)

#### 5.4.2.4 Estimation of intake values using the concept of unit variability

- For each iteration  $i$  in the Monte Carlo simulation, obtain for each product  $k$  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  $\kappa_{ikl}$  from a Beta or Bernoulli distribution. Calculate concentration values as  $c_{ikl} = \kappa_{ikl} * cm_k$ ,  $c_{ikl, max} = \kappa_{ikl} * cm_k * nu_k$ . Sum to obtain the simulated concentration in the consumed portion:

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

- 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  $\sigma$ . Back transform and sum to obtain the simulated concentration in the consumed portion:

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

#### 5.4.3 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 products.

MCRA calculates IESTI for comparison with Monte Carlo 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 97.5 <sup>th</sup> percentile of the distribution of consumed portions on days with positive consumption of the product (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 product 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 product, 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 product (unit weight is below 25 gr).

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

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 product is less than large portion weight.

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

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 product exceeds large portion weight.

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

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 products where bulking or blending means that the median represents the likely highest concentration.

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

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

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 product). These two different approaches handle the same problem, but it is undefined to which Monte Carlo 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 product. To allow a fair comparison, the Monte Carlo percentiles are calculated in the same way. Note, however, that in a multi-product Monte Carlo 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 products).

[Back to: Additional options concerning IESTI and consumption days](#)

## 5.5 Intake model

### 5.5.1 Only empirical estimates

This is an acute risk assessment using empirical compound data.

### 5.5.2 Empirical estimates and betabinomial/lognormal with age

An acute risk assessment may be followed by a parametric model where the intake frequency and logarithmically transformed intake amounts are related to a spline or polynomial function of age. The betabinomial distribution is used to estimate the intake frequency and a lognormal distribution to model the  $\ln(\text{intake})$  amounts.

[Back to: Empirical estimates and betabinomial/lognormal with age](#)

#### 5.5.2.1 Monte Carlo data generation of intake

Through Monte Carlo 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 products. 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/lognormal with age](#)

#### 5.5.2.2 Modelling the intake frequency distribution

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 age (and/or other consumer characteristics), using a betabinomial distribution with binomial totals  $n_i$  and overdispersion parameter  $\phi$  (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/lognormal with age](#)

#### 5.5.2.3 Modelling $\ln(\text{intake})$ amounts

For the positive intakes, consider the log-transformed values  $y_{ijk}$ . Average over replicates to obtain consumer day averages  $y_{ij}$ . These values are modelled in a REML 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/lognormal with age](#)



#### 5.5.2.4 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  $\sigma_{y'}^2$  of  $y'_{ijk}$ . We denote the estimated variance as  $\hat{\sigma}_{y'}^2$ . Now for each selected age  $x$  the logarithmic transformed positive intake distribution is modelled as normal with mean  $\hat{\mu}_x = f(x)$  and variance  $\hat{\sigma}_{y'}^2$ .

[Back to: Empirical estimates and betabinomial/lognormal with age](#)

#### 5.5.2.5 Estimating the acute intake distribution

The age-dependent acute intake distribution is obtained by integrating the betabinomial  $(\hat{\pi}_x, \hat{\phi})$  and the lognormal  $(\hat{\mu}_x, \hat{\sigma}_{y'}^2)$  distributions.

[Back to: Empirical estimates and betabinomial/lognormal with age](#)

### 5.6 Chronic risk assessment

In the MCRA program we have 3 models available to assess chronic risks: 1) the betabinomial/lognormal model (no age dependency); 2) in the extended version of this model, spline or polynomial functions are used to model the effect of age or any other explanatory variable, and; 3) the discrete/semi-parametric model according to Nusser (no age dependency).

#### 5.6.1 Introduction

In dietary risk assessment, usual intake is defined as the long-run average of daily intakes of a dietary component by a consumer. From a statistical point of view, assessing the usual intake can be reduced to the problem of estimating the distribution of a random variable  $y_i$  that is measured with error. A model for the relationship between the observations  $y_{ij}$  and the true random variable of interest  $y_i$  is:

$$y_{ij} = y_i + u_{ij}$$

where  $u_{ij}$  is an additive measurement error for consumer  $i$  on day  $j$ . For independent, normally distributed  $y_i$  and  $u_{ij}$ , estimation of the distribution of  $y_i$  is straightforward. When observations  $y_{ij}$  are non-normal and the measurement error variance is heterogeneous across sampling units, estimation is less simple.

#### 5.6.2 Modelling usual daily intake

Through the assumed independence of consumption data and compound concentration values (a most reasonable assumption) the modelling of usual intake simplifies to a univariate stochastic model for intakes.

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 conformity with our method of data entry for consumption (see 9.3.1). As a consequence, days without consumptions do have zero intake.

For all days of all consumers the daily intake on day  $j$  is calculated as the sum over products of consumption amount per kg body weight times average concentration. The average concentration of all available concentration measurements on a product 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.

The model for the usual intake distribution is:

$$y_{ij} = y_i + u_{ij}$$

with  $y_{ij}$  the observed intake of consumer  $i$  on day  $j$ ,  $y_i$  is the unobservable usual intake value for consumer  $i$ , and  $u_{ij}$  is the unobservable measurement error for consumer  $i$  with  $i = 1 \dots n$  on day  $j$  with  $j = 1 \dots o$ . In the normal scale,  $y_i \sim N(\mu, \sigma^2_{cons})$ ,  $u_{ij} \sim N(0, \sigma^2_{day})$ .

The daily intake data are separated in a set of zero daily intakes and non-zero daily intakes. In the following steps indices on  $y_{ij}$  are dropped and only the non-zero daily intakes are used (this prevents problems with a large spike of zero intake values).

[Back to: Chronic risk analysis](#)

### 5.6.3 Betabinomial/lognormal

The betabinomial model describes the probability of a consumer of having an intake. Log intakes are modelled with a random effects linear model using residual maximum likelihood (REML) estimating parameters of the usual intake distribution. By integrating the probability distribution the log amount distribution, inference is made about the percentiles of the usual intake distribution of the entire population.

[Back to: Betabinomial/lognormal](#)

#### 5.6.3.1 Modelling the intake frequency distribution

Let  $n$  and  $npos_i$  be the total number of days per consumer (for all consumers equal) and the number of days with a positive intake, respectively. Then  $npos_i$  is modelled using a betabinomial distribution with binomial totals  $n$  allowing for overdispersion. The mean and variance of the beta distribution are  $\alpha/(\alpha + \beta)$  and  $\alpha\beta(\alpha + \beta + n)/[(\alpha + \beta)^2(\alpha + \beta + 1)]$ , respectively.

Reparameterizing by  $\pi = \alpha/(\alpha + \beta)$  and  $\varphi = 1/(\alpha + \beta + 1)$  is a more stable and interpretable parameterization. The mean and variance of the beta distribution are then given by  $\pi$  and  $\pi(1 - \pi)[1 + (n - 1)\varphi]$ , respectively. Note that the first part of the variance  $\pi(1 - \pi)$  equals the binomial variance; the second part is the so-called overdispersion factor. The number of consumption days  $x$  out of  $n$  then follows a betabinomial distribution: the mean and variance of this distribution are simply  $n$  times the mean and variance given above.

[Back to: Betabinomial/lognormal](#)

#### 5.6.3.2 Modelling $\ln(\text{intake})$ amounts

For the positive intakes, consider the log-transformed values  $y_{ij}$  (the natural logarithm is used). These values are modelled in a REML analysis with random terms consumer and consumer.day to estimate the between-consumer and within-consumer variance component.

[Back to: Betabinomial/lognormal](#)

#### 5.6.3.3 Estimating usual intake distributions

The total variance of the daily intakes is partitioned into a between-consumer component and a within-consumer component. The between-consumer variance component parameterizes the dispersion of the usual intake distribution of the positives only. The usual intake distribution of the entire population is obtained by integrating the betabinomial  $(\hat{\pi}, \hat{\phi})$  and the lognormal  $(\hat{\mu}, \hat{\sigma}^2)$  distributions.

[Back to: Betabinomial/lognormal](#)

### 5.6.4 Betabinomial/lognormal with age

The betabinomial model describes the probability of a consumer of having an intake. The functional relation between the intake frequency and age is modelled using a spline or polynomial function.

Log intakes are modelled with a random effects linear model using residual maximum likelihood (REML) estimating parameters of the usual intake distribution. The relation between the log intake amounts and age is modelled using a spline or polynomial. By integrating both distributions, inference is made about age dependent percentiles of the usual intake distribution of the entire population.

[Back to: Betabinomial/lognormal \(with age\)](#)

#### 5.6.4.1 Modelling the intake frequency distribution

Let  $n$  and  $npos_i$  be the total number of days per consumer (for all consumers equal) and the number of days with a positive intake, respectively. Then  $npos_i$  is modelled as a function of age (and/or other consumer characteristics) using a betabinomial distribution with binomial totals  $n$  and overdispersion parameter independent of age. As seen in 5.6.3, the mean and variance of the beta distribution are given by  $\pi$  and  $\pi(1-\pi)[1+(n-1)\phi]$ , respectively.

The mean probability  $\pi$  can be related to age (or any explanatory variable) by employing a logit link, e.g.  $\text{logit}(\pi) = \theta_0 + \theta_1 \text{age}$ . Then, fitted binomial probabilities are age dependent  $\hat{\pi}_x = f(x_i)$ , where  $x_i$  is the age of consumer  $i$ .

The number of consumption days  $x$  out of  $n$  follows a betabinomial distribution: the mean and variance of this distribution are simply  $n$  times the mean  $\hat{\pi}_x$  and variance.

[Back to: Betabinomial/lognormal \(with age\)](#)

#### 5.6.4.2 Modelling ln(intake) amounts

For the positive intakes, consider the log-transformed values  $y_{ij}$ . These values are modelled in a REML analysis with random terms consumer and consumer.day as a function of age (and/or other consumer characteristics). The fitted values from the model are  $\hat{\mu}_x = f(x_i)$ , where  $x_i$  is the age of consumer  $i$ .

[Back to: Betabinomial/lognormal \(with age\)](#)

#### 5.6.4.3 Estimating usual intake distributions

The total variance of the daily intakes is partitioned into a between-consumer component and a within-consumer component. The between-consumer variance component parameterizes the dispersion of the usual intake distribution of the positives only. The usual intake distribution of the entire population is obtained by integrating the betabinomial  $(\hat{\pi}_x, \hat{\phi})$  and the lognormal  $(\hat{\mu}_x, \hat{\sigma}_y^2)$  distributions.

[Back to: Betabinomial/lognormal \(with age\)](#)

#### 5.6.5 Discrete/semiparametric (Nusser)

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 products which were consumed on an almost daily basis, e.g. dioxin in fish, meat or dairy 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 consumption 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
4. estimating the percentiles of the distribution of usual intakes in the original scale.

[Back to: Discrete/semiparametric \(Nusser\)](#)

### 5.6.5.1 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  $\{10, 2, 1, \frac{1}{1.5}, \frac{1}{2}, \frac{1}{2.5}, \frac{1}{3}, \frac{1}{3.5}, \dots, \frac{1}{100}\}$ , 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  $n^{\text{th}}$  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/semiparametric \(Nusser\)](#)

### 5.6.5.2 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/semiparametric \(Nusser\)](#)

### 5.6.5.3 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  $\sigma_1^2$  and the mean and the variance of the within-consumer variance component distribution ( $\sigma_B^2$  and  $\sigma_A^2$ ) 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/semiparametric \(Nusser\)](#)

#### 5.6.5.4 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 x 400 values representing the distribution of daily intakes around each of the 400 points;
2. These 9 x 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 ( $\theta_m$ ;  $m = 0, \dots, M$ ) 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 ( $\theta_0$ ) is estimated separately with no restriction to be similar to the other  $\theta_m$ . 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  $\theta_0$  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 ( $\hat{\theta}_0, \dots, \hat{\theta}_M$ ) 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/semiparametric \(Nusser\)](#)

## 5.7 Uncertainty of risk assessments: bootstrapping data sets

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 products. 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 exposure 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.

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* 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 bootstrap sample). By repeating this process  $B$  times, one can obtain  $B$  bootstrap samples, 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 95<sup>th</sup> percentile, etc.) can also be calculated from each of the  $B$  bootstrap samples. This generates a *bootstrap distribution* for the statistic under consideration. The bootstrap 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 Monte Carlo modelling of acute risks two datasets are combined: consumption data and compound concentration data. It makes sense to apply bootstrapping to both datasets separately, in order to characterise the uncertainty in the final exposure. In MCRA the bootstrap algorithm (when selected) is applied to:

1. the multivariate consumption patterns and associated body weights: actually the data set of consumer identifiers is bootstrapped, 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 bootstrapped independently for all products. In principle, the bootstrap 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 bootstrap sample 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 bootstrap distribution of percentiles of the exposure distribution is summarised by specifying empirical 2.5<sup>th</sup>, 25<sup>th</sup>, 75<sup>th</sup> and 97.5<sup>th</sup> percentiles. The outer percentiles constitute a central 95% confidence interval for the variability percentiles. However, for this it is necessary that the number of bootstrap samples  $B$  is high enough. The number of bootstrap samples should be chosen depending on the confidence level wanted for the uncertainty interval. Typically 500-2000 bootstrap sets will be reasonable for a 95 % confidence interval (Efron & Tibshirani 1993, pp. 14-15, 275).

The same bootstrap algorithm can also be applied to deterministic estimates which are calculated from data sets. For example the maximum concentration found in a bootstrap sample 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.

[Back to: Uncertainty analysis](#)

## 5.8 Binning

Binning is a method to summarise the simulated data (total intake, intake per product, consumption per product, concentration per product) in frequency intervals for further use in deriving the exposure 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 of the observations in the first chunk of the simulation (*mean*) 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  $mean \times 1.01^i$ . For values below the mean also 1100 bins are defined with lower limits defined by  $mean \times 1.01^{-i}$ . 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 ( $1.01^{2200} = 3.2 \times 10^9$ ), which should be amply sufficient for most practical problems.

To get accurate results, it is rather important that the mean value in the first chunk represents, approximately, the true mean of the sampled distribution. Therefore, chunk size (defined as the total number of simulations divided by the number of chunks) should not be chosen too small. During the simulation, the maximum of the sampled observations in each chunk is calculated. When this value is higher than the upper limit of the last bin, representing a potential maximum, this bin limit is replaced by the new maximum, and a warning is issued. 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 products with nonzero intakes. Also in this case a warning is given.

## 6 Selection of consumers, foods and/or products

In the next sections, the centrally supplied database 'dutch pesticides.mdb' is used. All tables are selected and data for survey DNFCS-3 and compound CHLORPYRIFOS are retrieved from the database. Find in Figure 50 the outcome of the selection. In total, 557 foods are eaten. Application of the conversion model shows that the number of primary products is 118. Note that the selected foods and primary agricultural products do not depend on the choice of the compound. The retrieved foods and products 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 25 primary agricultural products. This means that there is at least one sample of each of the 25 products with a positive concentration. On 70 products only non-detects are found. This means that for each of the 70 products all samples were negative, i.e. no chlorpyrifos found. For the remaining 23 products ( $118 - 25 - 70$ ) there are no samples with chlorpyrifos in the database.

By checking one of the three radio buttons you can specify which products must be used in the risk analysis. Pressing the 'go' button in Figure 50 brings you for each choice to the central menu. If you want to do the same analysis for one of the other 2 not chosen number of products you have to start again with the selection of consumption and compound tables by clicking the selected-consumption-and-compound-tables 'go' button.

**Table selection for MCRA analysis**

The name of the survey is: **DNFCS-3**

The selected compound is: **CHLORPYRIFOS**

The number of levels for foodname is: 557

The number of levels for productname is: 118

There are 118 products:

for 25 products positive concentration values are available.

for 70 products measurements have been made but only nondetects were found. Include these products in the analysis if you want to replace them with the limit of reporting (LOR).

for 23 products no measurements have been made. Include these products in the analysis if you have information on concentrations for those products for which agricultural use is allowed.

selected tables, survey and compound

current subset selection

Continue with 25 products (with positive concentrations)

Continue with 25 + 70 = 95 products (measured)

Continue with 25 + 70 + 23 = 118 products (all)

press after making your choice

**Figure 50: Selected products for survey DNFCS-3 and compound CHLORPYRIFOS**

In the case all (here 118) products are chosen products for which no concentration measurements are available and for which agricultural use is allowed (coded as useallowed=1) will be included in the analysis using a user-defined worst-case value.

In the case all measured (here 95) products are chosen products for which no concentration measurements are available will be included in the analysis using the Limit of Reporting (LOR) value. This value can be specified using the options in the option block Concentration Data in the right



section of the MCRA input form (see 3.2.2 and 3.2.2.1 ). If the option ‘no replacement of non-detects’ is specified the choice of the first and second radio button in Figure 50 is equivalent but the choice of the first button is computationally more efficient because less products are involved. The choices are postponed until after subset selection the start-MCRA ‘go’ button in the central menu is pressed (see Figure 51).

With the subset selection ‘go’ button in the MCRA central menu (see Figure 51) it is possible to restrict the MCRA analysis to a subset of consumers in the food survey or to a subset of foods and/or products.

The menus for subset selection utilise the relations between tables, offering a flexible tool to select data according to your own wishes. Two steps are distinguished. Through the use of scroll-down menus the user 1) selects ages, weights and sexes to define the population of consumers; 2) selects day(s) of consumption, foods, primary agricultural products and year, country and sampling type of the compound concentration data. The hierarchical structure of the product code may be helpful to select a group of primary agricultural products or foods at once. In Figure 51 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.

**MCRA 4.0 central menu**

**Overview**

compound: **CHLORPYRIFOS**

consumption survey: **DNFC3-3**

products: **only products with positive concentration measurements**

data selected on: **11/17/2005 11:13:02 AM**

selection of tables, survey and compound

subset selection

start MCRA

- Your last job is successfully completed
- Server is ready to accept a job.

[back to main menu](#)

**Figure 51: MCRA central menu, subset selection**

## 6.1 Subset selection: consumers

After clicking the subset-selection ‘go’ button in the MCRA central menu (see Figure 51) the first subset selection screen is shown (see Figure 52). In this screen you define the consumer population based on characteristics of the consumers (age, weight, sex). The first time, the current selection is equal to the information stored in the database, 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 an empty population e.g. the result of selecting ages within the range 1 to 8 and weights in

the range 70 kg to 150 kg is zero (0) consumers. So try to avoid making combinations of variables that yield empty subsets. The number of consumers currently selected is always mentioned in the screen (see Figure 52).

**MCRA subset selection**

**Consumer population**

Check individual characteristics for subset selection

variable	levels in current selection	check if you want to select levels	select from:	
age:	min=1; max=97	<input type="checkbox"/>	<input checked="" type="radio"/> current selection	<input type="radio"/> database (min=1; max=97)
weight:	min=8; max=150	<input type="checkbox"/>	<input checked="" type="radio"/> current selection	<input type="radio"/> database (min=8; max=150)
sex:	n = 2	<input type="checkbox"/>	<input checked="" type="radio"/> current selection	<input type="radio"/> database (n=2)

Number of individuals currently selected: 6250

show/select levels for checked variables

use total population (no consumer subset selection)

**Figure 52: MCRA subset selection: consumer population**

To select a population of consumers based on all variables, check all three checkboxes and click the upper 'go' button. A new screen appears (see Figure 53) where you can choose minimum and maximum values for age and weight (the unit depends on the survey you are currently using, see table Food consumption survey in 9.3.4 ) and levels of sex. 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.

**MCRA subset selection**

**Select levels for individuals**

age between  and

weight between  and

sex total number of levels: 2

subset selection based on chosen levels

back to previous screen (no selection in this step)

back to main menu

**Figure 53: MCRA subset selection: select levels for consumers**

For example, select a population between 1 and 6 years old. 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 of levels are selected that are not in correspondence with each other. So, in the children example, if you want to return to your original population check for both variables age and weight the radio button 'database' and 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

In the first step of the subset selection process the total population is selected. A second subset selection screen is reached that shows you the current selected levels of each variable (see Figure 54). Also the levels of each variable stored in the database are shown. In our example, the number of products with positive concentration values is 25 and these products are derived from 267 foods. Originally 118 products and 557 foods were consumed (see Figure 50). However in the database 265 products and 1574 foods are stored (see most right column of Figure 50). The reason for differences in these numbers is that the database consists of two surveys: DNFCS-3 and EU-BABY. Apparently the selected survey DNFCS-3 retrieves not all products and foods found in the EU-BABY survey.

variable	levels in current selection	check if you want to select levels	select from:
day:	n = 2	<input type="checkbox"/>	<input checked="" type="radio"/> current selection <input type="radio"/> database (n=2)
foodname:	n = 267	<input type="checkbox"/>	<input checked="" type="radio"/> current selection <input type="radio"/> database (n=1574)
foodgroupname:	n = 17	<input type="checkbox"/>	<input checked="" type="radio"/> current selection <input type="radio"/> database (n=25)
productname:	n = 25	<input type="checkbox"/>	<input checked="" type="radio"/> current selection <input type="radio"/> database (n=265)
productgroupname:	n = 3	<input type="checkbox"/>	<input checked="" type="radio"/> current selection <input type="radio"/> database (n=4)
productsubgroupname:	n = 11	<input type="checkbox"/>	<input checked="" type="radio"/> current selection <input type="radio"/> database (n=46)
year:	n = 3	<input type="checkbox"/>	<input checked="" type="radio"/> current selection <input type="radio"/> database (n=3)
samplingtype:	n = 1	<input type="checkbox"/>	<input checked="" type="radio"/> current selection <input type="radio"/> database (n=1)
countryname:	n = 2	<input type="checkbox"/>	<input checked="" type="radio"/> current selection <input type="radio"/> database (n=247)

show/select levels for subset selection

continue to central menu

**Figure 54: 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:

- food, food group: to restrict the analysis to specific foods or food groups;
- product, product group and product subgroup: to restrict the analysis to specific primary agricultural products and (sub)groups (see also 9.2.1 , 9.2.3 and 9.2.4 );
- (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 products originating from specific countries.

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 productname and country will give an empty subset. In this example it is quite clear why things go wrong. Often it is less clear what you can 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 foodnames and productnames and by clicking the upper button. A new screen appears with scroll-down menus for foods and products (see Figure 55) Choose the three BAMI GORENG foods in the first scroll-down menu and click the upper 'go' button.

The screenshot shows a web interface titled "Select levels" with a light orange background. It contains two scroll-down menus. The first menu, labeled "foodname" with "total number of levels: 267", lists various food items, with three items highlighted in blue: "BAMI GORENG CHINESE NOODLE DISH CANNED", "BAMI GORENG CHINESE NOODLE DISH W/O EGG", and "BAMI GORENG COOLFRESH MEAL". The second menu, labeled "productname" with "total number of levels: 25", lists various fruits and vegetables. At the bottom, there are two buttons labeled "go": one for "subset selection based on chosen levels" and one for "back to previous screen (no selection in this step)".

**Figure 55: MCRA subset selection: select levels**

The system automatically adjusts the levels of all variables: all products 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 productname checkbox and press upper button). In Figure 56, the selected products are shown: CARROT, GREEN BEANS (FRESH), LEMON and SWEET PEPPER are ingredients of BAMI GORENG foods.

**Figure 56: MCRA subset selection: select levels**

Find in Figure 57, the selected levels of each variable for this selection.

variable	levels in current selection	check if you want to select levels	select from
day:	n = 2	<input type="checkbox"/>	<input checked="" type="radio"/> current selection <input type="radio"/> database (n=2)
foodname:	n = 3	<input checked="" type="checkbox"/>	<input type="radio"/> current selection <input checked="" type="radio"/> database (n=1574)
foodgroupname:	n = 1	<input type="checkbox"/>	<input checked="" type="radio"/> current selection <input type="radio"/> database (n=25)
productname:	n = 4	<input checked="" type="checkbox"/>	<input checked="" type="radio"/> current selection <input type="radio"/> database (n=265)
productgroupname:	n = 3	<input type="checkbox"/>	<input checked="" type="radio"/> current selection <input type="radio"/> database (n=4)
productsubgroupname:	n = 4	<input type="checkbox"/>	<input checked="" type="radio"/> current selection <input type="radio"/> database (n=46)
year:	n = 3	<input type="checkbox"/>	<input checked="" type="radio"/> current selection <input type="radio"/> database (n=3)
samplingtype:	n = 1	<input type="checkbox"/>	<input checked="" type="radio"/> current selection <input type="radio"/> database (n=1)
countryname:	n = 2	<input type="checkbox"/>	<input checked="" type="radio"/> current selection <input type="radio"/> database (n=247)

**Figure 57: MCRA subset selection: consumption data and concentrations**

After an MCRA analysis you can find details on the selected food consumption. In Table 20 some output of the analysis is shown. By doing three subset selections followed by a MCRA analysis for the three BAMI GORENG foods separately you can find that BAMI GORENG CHINESE NOODLE DISH CANNED (with GREEN BEANS, CARROT and SWEET PEPPER) is consumed 4 times, BAMI GORENG CHINESE NOODLE DISH W/O EGGS (with LEMON) 131 times and BAMI GORENG COOLFRESH MEAL (with GREEN BEANS and LEMON) 11 times. Note that not all ingredients of BAMI GORENG are shown, only products with positive concentration measurements were requested (by specifying the first radio button in Figure 50).

FOOD CONSUMPTION DATA							
Code	Product	MeanCons (g)	MeanConsDays (g)	NConsDays	NDays	%PosCons (%)	
17111.10	GREEN BEANS	0.02	15.3	15	12500	0.1	
18402.20	CARROT	0.01	15.7	4	12500	0.0	
18506.10	SWEET PEPPER	0.00	10.9	4	12500	0.0	
19501.10	LEMON	0.02	2.2	142	12500	1.1	

**Table 20: Consumption of primary products based on 3 BAMI GORENG foods**

The example is continued by selection of all foods that according to the foodconversion model contain one or more of the primary agricultural products CARROT, GREEN BEANS (FRESH), LEMON and SWEET PEPPER. To implement this, check variable productname and foodname, check database (n = 1574) for foodname (see Figure 57) and press the upper button.

In Figure 58 all levels for foods stored in the database (n = 1574) and the current selection of products (n = 4) are shown. Select CARROT, GREEN BEANS (FRESH), LEMON and SWEET PEPPER from the scroll-down menu (do not select levels of foods) and press the upper button.

**Select levels**

foodname total number of levels: 1574

- AA DRINK MULTI NINE SPORTS DRINK RAAK
- ADVOCAAT EGGNOG
- AGDEMAYIWIRI LEAFY VEGETABLE SURINAM
- ALMONDS PEELED
- AMSOI MUSTARD LEAF SURINAM
- ANCHOVY IN OIL
- ANTROEWA AFRICAN EGGPLANT SURINAM
- APPLE JUICE CONCENTRATE DILUTED
- APPLE JUICE CONCENTRATED
- APPLE JUICE UNSWEETENED
- APPLE PIE DUTCH
- APPLE PUFF PASTRY
- APPLE PUREE W/D SUGAR CANNED
- APPLE SAUCE CANNED
- APPLE SAUCE LIGHT

productname total number of levels: 4

- CARROT
- GREEN BEANS (FRESH)
- LEMON
- SWEET PEPPER

subset selection based on chosen levels

back to previous screen (no selection in this step)

continue to MCRA central menu (no selection in this step)

**Figure 58: MCRA subset selection: select levels**

In Figure 59 the resulting selection is shown. The total number of selected products is 4 and these products are found in 22 different foods.

**Select levels**

Check variables for subset selection

variable	levels in current selection	check if you want to select levels	select from
day:	n = 2	<input type="checkbox"/>	<input checked="" type="radio"/> current selection <input type="radio"/> database (n=2)
foodname:	n = 22	<input checked="" type="checkbox"/>	<input checked="" type="radio"/> current selection <input type="radio"/> database (n=1574)
foodgroupname:	n = 1	<input type="checkbox"/>	<input checked="" type="radio"/> current selection <input type="radio"/> database (n=25)
productname:	n = 4	<input checked="" type="checkbox"/>	<input type="radio"/> current selection <input checked="" type="radio"/> database (n=265)
productgroupname:	n = 3	<input type="checkbox"/>	<input checked="" type="radio"/> current selection <input type="radio"/> database (n=4)
productsubgroupname:	n = 4	<input type="checkbox"/>	<input checked="" type="radio"/> current selection <input type="radio"/> database (n=46)
year:	n = 3	<input type="checkbox"/>	<input checked="" type="radio"/> current selection <input type="radio"/> database (n=3)
samplingtype:	n = 1	<input type="checkbox"/>	<input checked="" type="radio"/> current selection <input type="radio"/> database (n=1)
countryname:	n = 2	<input type="checkbox"/>	<input checked="" type="radio"/> current selection <input type="radio"/> database (n=247)

show/select levels for checked variables

continue to MCRA central menu

undo last selection step

**Figure 59: MCRA subset selection: consumption data and concentrations**

In Table 21 you find details on the selected food consumption as produced by MCRA. Compare this table with Table 20, where consumption is based on 3 foods.

FOOD CONSUMPTION DATA						
Code	Product	MeanCons (g)	MeanConsDays (g)	NConsDays	NDays	%PosCons (%)
17111.10	GREEN BEANS	0.07	9.7	87	12500	0.7
18402.20	CARROT	0.27	26.2	131	12500	1.0
18506.10	SWEET PEPPER	0.57	13.7	523	12500	4.2
19501.10	LEMON	0.03	1.8	212	12500	1.7

**Table 21: Consumption of primary products based on 22 foods**

If you are interested in all primary products that can be derived from these 22 foods, check the checkboxes for variables foodname and productname and check database ( $n = 265$ ) for productname (see Figure 59), then select all 22 foods from the scroll-down menu (do not select levels of productnames) and press the upper button (screens are not shown). The resulting number of products is 5, the extra selected product is POTATOE.

### 6.3 User-defined worst-case values

As mentioned before, after selection of tables, compoundname and surveyname the user is confronted with three choices concerning the presence of concentration measurements on consumed primary agricultural products. In Figure 60 (see also Figure 50) the user decides to continue with all 118 products (products with positive concentration values = 25, products with non-detects only = 70 and products without measurements = 23). Press the ‘go’ button to reach the central menu.

Continue with 25 products (with positive concentrations)  
 Continue with 25 + 70 = 95 products (measured)  
 Continue with 25 + 70 + 23 = 118 products (all)

press after making your choice

**Figure 60: Table selection for MCRA analysis, continue with all products**

After pressing the ‘start MCRA’ button (see Figure 6) a menu is reached to include user-defined worst-case values for products without measurements (see Figure 61): only products for which agricultural use is allowed (useallowed = 1) are included in the analysis while all other products are deleted from the analysis. Information on agricultural use is stored in table **Agriculturaluse** (see 9.4.9).

## MCRA worst case concentration

### Products without residue measurements

Products have been selected for which no concentration measurements are available.  
Please specify a country and year of agricultural use to search for default worst case values  
(these values can be changed in the next screen).

countryname    total number of levels: 2

year    total number of levels: 2

select levels for worst case concentrations   

**Figure 61: MCRA worst case concentration, scroll-down menus**

In Figure 61, a choice is made from the available levels as specified in table Agriculturaluse for countries and years. Default worst-case values are selected according to the selection criteria. In the next window, the selected worst-case values are shown for country = FRANCE and year = 2003: only for product SUNFLOWER SEEDS and ALMOND default worst-case value are found, e.g. 15 and 25 respectively (see Figure 62). The default worst-case values may be replaced by a user-defined values.

## MCRA worst case concentration

### Specify worst case value

In the selected consumption data there are 118 products in total.  
Only 95 of these products have measured concentrations (detects and/or nondetects).  
Among the other 23 products:

- agricultural use is assumed to be not allowed for 19 products, and these products are ignored in the further MCRA analysis.
- agricultural use is allowed for 4 products (according to information in the AgriculturalUse table).

The selected products according to your criteria for year and country are listed below.  
Please specify worst case values to be used in MCRA.

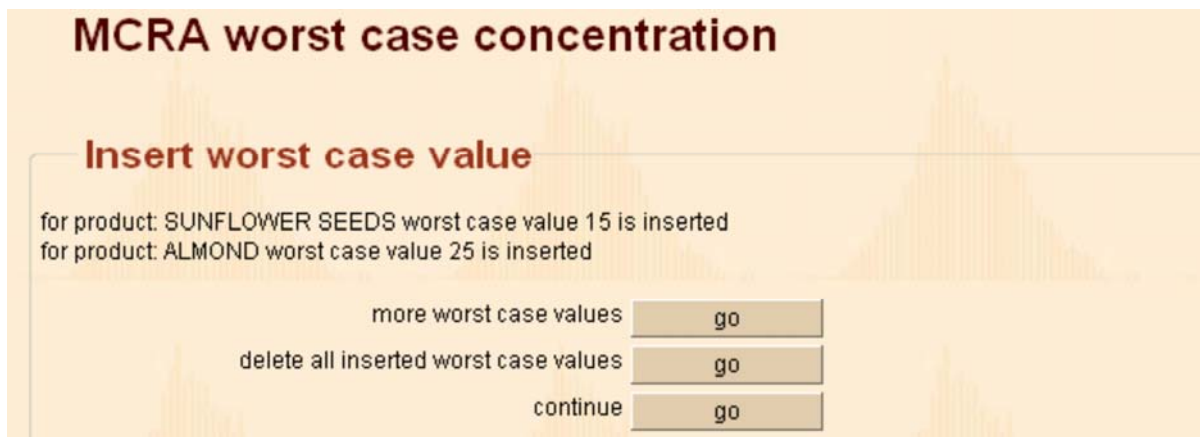
productname	worst case value
SUNFLOWER SEEDS	<input type="text" value="15"/>
ALMOND	<input type="text" value="25"/>

continue   

**Figure 62: MCRA worst case concentration, specify worst case value**

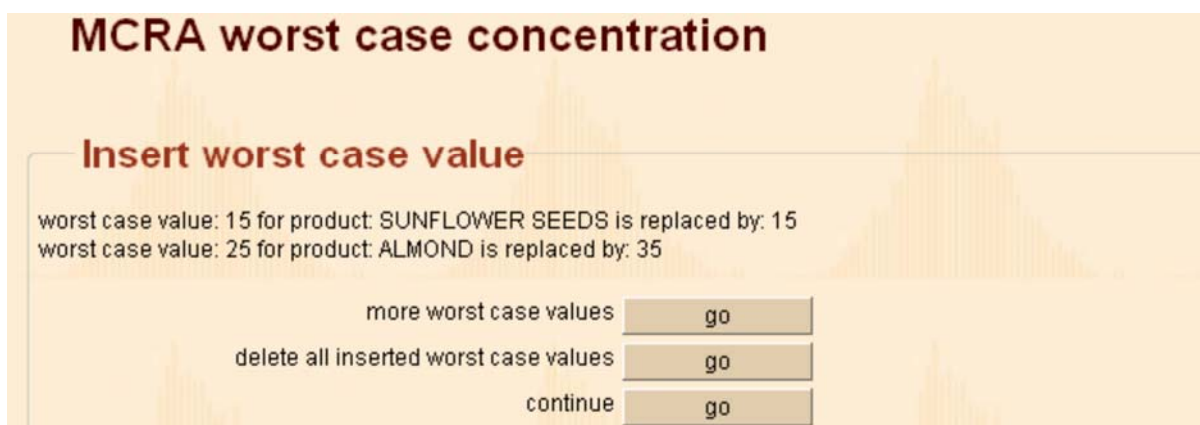
By pressing the 'go' button the default worst-case values are inserted and the result is shown in Figure 63.





**Figure 63: MCRA worst case concentration, insert worst case value**

By choosing more-worst-case-values, new worst-case values may be inserted for other products or already inserted (default) values may be changed. To change an already inserted value, specify the same levels for country and year (see Figure 61) as before (country = FRANCE and year = 2003). The same screen as in Figure 62 appears and the worst-case value 25 for ALMOND (from table Agriculturaluse) is changed in 35. The resulting screen is shown in Figure 64 and differs slightly from the screen in Figure 63.



**Figure 64: MCRA worst case concentration, change worst case value**

Worst-case values are deleted by pressing the delete-all-inserted-worst-case-values 'go' button. If you choose this option, all information on worst-case values is lost and insertion of new values has to be repeated for all selected products without concentration measurements. By pressing the lowest 'go' button (see Figure 64) the input form is reached (see Figure 12).

## 7 On-line editing of data

Instead of choosing your data from already formed databases as discussed in 2.6.1 you can edit data on-line. After choosing ‘MCRA 4.0 (field trial data and Dutch consumption data)’ in the MCRA main menu (see Figure 2) a series of menus is reached from where you can add and edit on-line your own concentration data instead of using centrally supplied data or data stored in your own user directory. This is a safe way to combine your concentration data with centrally supplied food consumption data because product codes for the concentration data are automatically generated from the food consumption data.

From the central menu (see Figure 5), click the start-selection-of-consumption-and-compound-tables ‘go’ button. Then the available databases are shown (see Figure 65).

The screenshot shows a web interface titled "MCRA for fieldtrial data". Below the title is a section "Selection of databases" with two bullet points: "In the list below you find servers and databases." and "Select one or more databases that you wish to use for a MCRA analysis." Below this, it says "select databases from list:". There is a list of databases: "Rikilt (NL)" with a checked checkbox for "dutch pesticides.mdb", and unchecked checkboxes for "mcradbtest.mdb", "jiri.mdb", and "pesticides0203.mdb". At the bottom of the list are three buttons: "press", "go", "clear", and "info". Below the list are two links: "back to central menu" and "back to main menu".

Figure 65: Database selection menu, MCRA for field trial data

After checking one or more databases and clicking the ‘go’ button the table selection menu appears (see Figure 66). Note that its layout is being changed displaying only tables related to consumption e.g. Foodconsumption, Foodconversionmodel, Products and Individual. It is compulsory to select these four tables (see also 2.6.1 ). Tables related to concentration data are not being displayed. After checking ‘All Tables’ here the menus for on-line adding and editing data are reached.

The screenshot shows a web interface titled "MCRA for fieldtrial data". Below the title is a section "Table selection for MCRA analysis" with the instruction "Select data tables by checking the boxes below, then click the select-tables 'go' button" and a note "Do not select more than one table with the same name." Below this, there is a table with columns: "Database:", "Food consumption", "Food conversion model", "Products", "Individual", and "All Tables". The "dutch pesticides.mdb" row has checkboxes for "Food consumption", "Food conversion model", "Products", and "Individual", and a checked checkbox for "All Tables". Below the table is a checked checkbox for "check columnnames". At the bottom are two buttons: "select tables" and "go", and "clear all settings" and "clear". Below the buttons is a link "back to main menu".

Figure 66: Table selection menu, MCRA for field trial data

The first time a user has chosen the field trial menu a new database has to be created. Select create-new-database from the scroll-down menu (see Figure 67). In all subsequent cases, on-line created databases 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 products and a proper functioning of the menu is not guaranteed anymore.



**Figure 67: Select database for concentration, create new database**

Specify the name of the database (see Figure 68 ) and press the 'go' button. Note you should only use letters and digits, no other characters are allowed.



**Figure 68: Select database for concentration, specify a new database name**

This brings you to Figure 69 if you entered concentrations values in the past. If not the 'list or add variability factors', 'list or add processing factors' and 'list or edit ARfD or ADI' buttons are not available.

By clicking the list-or-add-concentrations button brings you to a screen with an add-records link and a message '\*\*\* Concentration database is empty, add records \*\*\*' if you did not enter concentration values before.

## MCRA fieldtrial data menu

### Select data

prepare data for MCRA

**Figure 69: Select data**

### 7.1 List or add concentrations

For adding concentration values click the ‘list or add concentrations’ button. You are requested to specify the name of a compound (see Figure 70) if you are entering values for the first time.

## MCRA for fieldtrial data

### Specify compound

Database: **testdata.mdb**

Acute Reference Dose (ARfD, microgr/kg bw/d)  9999 indicates missing value

Acceptable Daily Intake (ADI, microgr/kg bw/d)

compound name

press

clear

[back to main menu](#)

**Figure 70: Specify compound name**

If your selected database contains concentration values, the menu for specifying a compound name is skipped and you are brought directly to a screen where you can choose your primary agricultural products. Choose one of products in the scroll-down menu for which you want to add values (see Figure 71 ) and click the enter-product ‘go’ button.

## MCRA for fieldtrial data

### Select product

- Products are selected one by one from the pull-down menu

database: **testdata.mdb**  
compound: **CHLORPYRIFOS**  
ORANGE

enter product

reset list

**Figure 71: Select product**

In Figure 72, for product ORANGE and compound CHLORPYRIFOS, the number of samples and the concentration value is entered. After saving this record a message ‘\*\*\* one record added \*\*\*’ appears and another concentration value may be entered. Alternatively, click the select-new-product ‘go’ button for entering values for a new product.

## MCRA for fieldtrial data

### Concentration data

Compound: **CHLORPYRIFOS**  
Product: **ORANGE**

number of samples  enter frequency if concentration occurs more than once

concentration (mg/kg)  enter compound concentration

save this record

select new product

[list records](#)

[back to fieldtrial data menu](#)

**Figure 72: Enter concentration data**

By clicking the list-records link (see Figure 72) an overview of the concentration values in the database is displayed (see Figure 73).

## MCRA for fieldtrial data

Click button to modify a record.

no		frequency	concentration
<input type="button" value="1"/>	ORANGE	1	0.43
<input type="button" value="2"/>	ORANGE	1	1.23
<input type="button" value="3"/>	ORANGE	4	0.56
<input type="button" value="4"/>	ORANGE	1	0.23

[add records](#)

[back to fieldtrial data menu](#)

**Figure 73: List concentration values**

By clicking one of the grey buttons displaying the primary key of the rows in the database, the record is retrieved and the value for the number of samples or the concentration value may be changed. Save or delete the record or choose a new product (see Figure 74).

## MCRA for fieldtrial data

### Edit concentration value for: ORANGE

Record id: 2

number of samples  enter frequency if concentration occurs more than once

concentration (mg/kg)  enter compound concentration

save record

delete record

select new product

[list records](#)

[add records](#)

**Figure 74: Edit concentration value**

## 7.2 List or add variability factors

After supplying all concentrations values and clicking the back-to-fieldtrial-data-menu link (see Figure 72), the user is brought back to the select data menu (see Figure 69). At this stage, all products are known and adding of variability factors, processing factors and ARfD/ADI doses may be started. After entering all values, it is still possible to supply new concentrations values for old and/or new products. The process of entering variability factors and/or processing factors for new products is repeated as well.

Click the 'list or add variability factors' button. If you did not enter variability factors before, a message '\*\*\* Variability database contains no records yet \*\*\*' is displayed. By clicking the add-records link, the menu for adding variability factors is displayed (see Figure 75). The variability factor itself and the number of units of the composite sample should be specified. For each product in the concentration table, values should be supplied. Only products 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 products present in the concentration table \*\*\*' or '\*\*\* For all products, variability factors are already specified \*\*\*' is displayed. Records may be changed afterwards by clicking one of the grey buttons in table displaying a listing of all values in the database.

**MCRA for fieldtrial data**

**Variability factors**

- For every product in the concentration table a variability factor and the number of units of the composite sample should be supplied.
- This menu generates the names of the products for which concentrations are edited automatically. After editing a product, the name of the next product will be displayed and can be edited.
- This process is repeated until variability factors for all products are supplied.

compound: **CHLORPYRIFOS**  
 add variability factors for: **ORANGE**

variability factor  defined as p97.5/mean  
 number of units  number of units in the composite sample

save this record

clear value

**Figure 75: Variability factors**

### 7.3 List or add processing factors

In Figure 76 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. Then, click the add-records link. Choose 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 ORANGE are specified' appears and click the select-next-product 'go' button.

## MCRA for fieldtrial data

### Processing factors

- For all products in the concentration file, the processing type and processing factor may be specified.
- Processing types are selected from the list. Not all processing types should be selected.
- After selecting all processing types for a product, the next product is displayed automatically, and editing is repeated.
- By clicking the select-next-product 'go' button selection of processing types is ended and the next product is displayed. After supplying values for all products, the editing menu is stopped automatically.
- When no products or processing types are required, press repeatedly the select-next-product 'go' button until the message 'processing is ended' appears.

compound: **CHLORPYRIFOS**

add processing factors for: **ORANGE**

processing type

processing factor  processing factor is used as fixed value

save this record

clear value

select next product

[back to fieldtrial data menu](#)

**Figure 76: Processing factors**

If you wish to finish on-line editing for a product, click the select-next-product button and continue entering values for the next product. There is no need to enter values for all combinations of products and processing types: missing processing factors are by default replaced by value 1 (no processing).

If you wish to enter processing factors for only a few number of products and processing type combinations, enter your values and click repeatedly the select-next-product button until the message '\*\*\* Processing factors for products 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 product and contain only those processing types for which no values are supplied. Records may be changed afterwards by clicking one of the grey buttons in table displaying a listing of all values in the database.

### 7.4 List or edit ARfD or ADI

Click the 'list or edit ARfD or ADI' button in Figure 69 if you wish to enter values for the acute reference dose (ARfD) or average daily intake (ADI). This brings you to Figure 77.



## MCRA for fieldtrial data

Click button to modify record.

	acute reference dose	average daily intake	coumpoundname
1	100	10	CHLORPYRIFOS

[back](#) to fieldtrial data menu

**Figure 77: 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>
- 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, eighth 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 Dooren, MMH , Boeijen, I, van Klaveren, JD and van Donkersgoed G (1995). Conversie van consumeerbare voedingsmiddelen naar primaire agrarische producten. 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

The internet version of MCRA requires that all information needed for an exposure assessment is stored in a database, residing in one or more MS Access database files. The database organises data into a number of tables. Besides the products tables the tables can be divided into 2 groups: tables with information from the food consumption survey and those from the concentration of compounds survey. See Table 22 for an overview. To run a Monte Carlo Risk Assessment, tables called ‘required’ should always be supplied. The needs to supply ‘additional’ tables depend on subset requirements (see Ch. 6 ) and on the model specifications (see Ch. 3 ).

---

#### Required tables (minimal configuration)

---

Products	product and food codes and labels, and product specific information (see 9.2.1 )
<b><u>Food consumption</u></b>	<b><u>Description</u></b>
Foodconsumption	consumption of foods (see 9.3.1 )
Foodconversionmodel	conversion factors from foods into primary agricultural products (see 9.3.2 )
Individual	consumer characteristics (see 9.3.3 )
<b><u>Concentration of compounds</u></b>	<b><u>Description</u></b>
Compounds	compound codes, labels, agricultural and toxicological limits (see 9.4.1 )
Country	country codes, labels (see 9.4.2 )
<b>One of:</b>	
Concentration	compound concentration data (full data) (see 9.4.3 )
Summarydata	compound concentration data (as summary statistics) (see 9.4.4 )
Histogramdata	compound concentration data (table of frequency counts) (see 9.4.5 )

---

#### Additional tables (for querying or specific options)

---

Productcategory	product and food category codes and labels (see 9.2.2 )
Productgroup	product and food group codes and labels (see 9.2.3 )
Product_subgroup	product and food subgroup codes and labels (see 9.2.4 )
<b><u>Food consumption</u></b>	<b><u>Description</u></b>
Foodsurvey	Name of survey (see 9.3.4 )
Processing	processing codes and labels (see 9.3.5 )
Processingfactor	processing factors (see 9.3.6 )
<b><u>Concentration of compounds</u></b>	<b><u>Description</u></b>
VariabilityProd	unit variability factors (see 9.4.6 )
VariabilityCompProd	unit variability factors, compound-specific (see 9.4.7 )
VariabilityProcCompProd	unit variability factors, processing- and compound-specific (see 9.4.8 )
Agriculturaluse	information on the agricultural use of compounds (e.g. use allowed, percent crop treated, maximum compound limits) (see 9.4.9 )

---

**Table 22: Overview of tables**

A table in a database file is a collection of data about a specific subject or topic. The data in the tables are organised into columns (called fields) and rows (called records).

In the next paragraphs, the format of tables needed for a Monte Carlo Risk Assessment is described.

General remarks:

- Table and column (field) names should be exactly as indicated in the sections below except for tables **Summarydata** (see 9.4.4 ) and **Histogramdata** (see 9.4.5 ).

- In tables Productcategory, Productgroup, Product\_subgroup and Products, the columnnames for primary agricultural products are identical to the columnnames for foods as eaten as such (e.g. productcategory, productgroup, productsubgroup, product, productq and product<category/group/subgroup>name). The distinction between primary agricultural products and foods as eaten as such is made in column productcategory by using respectively a 1 or 9.
- In table Foodconsumption foods as eaten as such are coded using columnnames foodcategory (= 9), foodgroup, foodsubgroup, food and foodq.
- In table Foodconversion foods as eaten as such are coded using columnnames foodcategory (=9), foodgroup, foodsubgroup, food and foodq; primary agricultural products are coded using columnnames productcategory (=1), productgroup, productsubgroup, product and productq.
- Missing values are indicated with code 9999, unless stated otherwise. In general, an empty cell is also interpreted as a missing value. However, occasionally the use of empty cells leads to errors in retrieving data. Therefore, it is advised to use the code 9999 instead to indicate missing values
- Product codes and compound codes are entered in multiple columns, corresponding to a hierarchical coding system.
- Food codes are entered in multiple columns (not hierarchical).
- The use of quotes (‘) in product, compound, country or survey names is not allowed

In the next sections, the heading displays the table name, followed by block with field names and a short description of each field. A table should contain all fields except for tables **Summarydata** (see 9.4.4 ) and **Histogramdata** (see 9.4.5 ). For each field, numbers are assumed unless stated otherwise (in parentheses). Each section ends with some notes and a short example of the contents of the table.

## 9.2 Products tables

The products table is the bridge between the food consumption data and the compound concentration data. It contains coding for both primary agricultural products and foods as eaten as such. Both items can be distinguished by their productcategory number (1 = primary agricultural products: 9 = foods as eaten as such).

### 9.2.1 Products

field name	description
productcategory	product category number
productgroup	product group number
productsubgroup	product subgroup number
product	product number
productq	product quality number
productname (text)	product label
unitweight	nominal weight of a unit (gr)
edibleportion	edible portion (corrected large portion weight, gr)
largeportion	weight of a large portion (gr)

- When the nominal unit weight is unknown the value 0 is used.
- For foods (productcategory = 9), the value for unitweight, edibleportion and largeportion is not relevant, fill out all empty cells with value 0.
- Missing values for edibleportion and largeportion: 9999.

Example:

Products : Table									
productcategory	productgroup	productsubgroup	product	productq	productname	unitweight	edibleportion	largeportion	
1	9	89	996	1	HARD FRUIT	30	20	90	
1	9	89	997	1	SOFT FRUIT	50	9999	100	
1	9	89	998	1	OTHER FRUIT, NUTS	40	35	50	
1	9	89	999	1	OTHER DRIED FRUIT	50	45	35	
9	0	0	5039	0	MILK HALF FULL VITAMIN PLUS	9999	9999	9999	
9	0	0	9600	0	QUORN	9999	9999	9999	
9	0	0	9601	0	MILK DRINK MILK&FRUIT ORANGE	9999	9999	9999	
9	0	0	9603	0	FRIS & FIT YOGHURT	9999	9999	9999	
9	0	0	9604	0	BREAKFAST PRODUCT WAKE-UP BRINTA	9999	9999	9999	

## 9.2.2 Productcategory

field name	description
productcategory	product category number
productcategoryname (text)	product category label

- Table Productcategory contains coding for both primary agricultural products and foods as eaten as such. Both items can be distinguished by their productcategory number (1 = primary agricultural products: 9 = foods as eaten as such).
- No missing values allowed.

Example:

Productcategory : Table	
	productcategory
▶ +	1 PRIM AGRICULTURAL COMMODITY
+ +	9 FOOD
* *	0

## 9.2.3 Productgroup

field name	description
productcategory	product category number
productgroup	product group number
productgroupname (text)	product group label

- Table Productgroup contains coding for both primary agricultural products and foods as eaten as such. Both items can be distinguished by their productcategory number (1 = primary agricultural products: 9 = foods as eaten as such).
- No missing values allowed.

Example:

Productgroup : Table		
	productcategory	productgroup
▶ +	1	6 GRAINS AND GRAIN PRODUCTS
+ +	1	7 PULSES, SEEDS, KERNELS AND NUTS
+ +	1	8 VEGETABLES, POTATOES, BEET AND TUBERS
+ +	1	9 FRUIT
+ +	9	0 ENRICHED FOODS
+ +	9	1 POTATOES
+ +	9	2 ALCOHOLIC AND NON-ALCOHOLIC DRINKS
+ +	9	3 BREAD
+ +	9	4 OTHERS
+ +	9	5 EGGS
+ +	9	6 FRUIT AND FRUIT PRODUCTS

## 9.2.4 Product\_subgroup

field name	description
productcategory	product category number
productgroup	product group number
productsubgroup	product subgroup number
productsubgroupname (text)	product subgroup label

- Table Product\_subgroup contains coding for both primary agricultural products and foods as eaten as such. Both items can be distinguished by their productcategory number (1 = primary agricultural products: 9 = foods as eaten as such).
- No missing values allowed.

Example:

	productcategory	productgroup	productsubgroup	productsubgroupname
+	1	8	2	CABBAGES
+	1	8	1	LEAF/STALK VEGETABLES
+	1	7	1	PULSES
+	1	9	1	FRUIT, MICELLANOUS
+	9	2	0	ALCOHOLIC AND NON-ALCOHOLIC DRINKS
+	9	12	0	SPICES
+	9	23	0	MEAT AND POULTRY
+	9	22	0	FISH
+	9	21	0	FATS, OIL AND SAUCES
+	9	20	0	SUGAR, CANDY

## 9.3 Food consumption tables

### 9.3.1 Foodconsumption

field name	description
individualid	Consumer identification number
day	day (sequential number in food consumption survey)
foodcategory	food category number
foodgroup	food group number
foodsubgroup	food subgroup number
food	food number
foodq	food quality number
amountconsumed	consumed portion of food (g)
foodsurvey (text)	name of survey

- This table contains data on consumed foods (foodcategory=9). 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.
- Each food is characterised by a food code built from 5 numbers (not hierarchically):
  - 1 – food category number (9 indicates food as eaten as such)
  - 2 – food group number
  - 3 – food subgroup number
  - 4 – food number
  - 5 – food quality number
- Compound concentration data are usually measured on primary agricultural products. This means that conversion from consumed food to primary agricultural products should be done before performing an MCRA analysis (see van Dooren *et al.* 1995). Conversion factors should be provided in table Foodconversionmodel (see 9.3.2).

- No missing values allowed.

Example:

individualid	day	foodcategory	foodgroup	foodsubgroup	food	foodq	amountconsumed	foodsurvey
320491	2	9	10	0	455	0	8	DNFCS-3
320491	2	9	20	0	433	0	8	DNFCS-3
320662	1	9	20	0	377	0	8	DNFCS-3
320662	1	9	20	0	522	0	8	DNFCS-3
320662	2	9	13	0	271	0	8	DNFCS-3
320662	2	9	20	0	377	0	8	DNFCS-3
320731	1	9	15	0	207	0	8	DNFCS-3

### 9.3.2 Foodconversionmodel

field name	description
foodcategory	food category number
foodgroup	food group number
foodsubgroup	food subgroup number
food	food number
foodq	food quality number
productcategory	primary agricultural product category number
productgroup	primary agricultural product group number
productsubgroup	primary agricultural product subgroup number
product	primary agricultural product number
productq	primary agricultural product quality number
percentage	percentage of primary agricultural product each food consists of
proccode	processing code

- For each food (foodcategory=9), the composition is given in percentages of primary agricultural products (productcategory=1).
- Each product is characterised by a code built hierarchically from 5 numbers:
  - 1 – product category number (1 indicates primary agricultural product)
  - 2 – product group number
  - 3 – product subgroup number
  - 4 – product number
  - 5 – product quality number
 European product codes are a concatenation of 2, 3 and 4.
- No missing values allowed.

Example:

foodcategory	foodgroup	foodsubgroup	food	foodq	productcategory	productgroup	productsubgroup	product	productq	percentage	proccode
9	18	0	5119	0	1	7	1	11	1	8.5	3
9	18	0	5119	0	1	7	1	1	1	3.3	3
9	18	0	5115	0	1	8	4	1	1	30	3
9	18	0	5115	0	1	7	1	11	1	17	3
9	18	0	5115	0	1	7	1	1	1	3.3	3
9	18	0	5115	0	1	6	7	999	1	5	99
9	18	0	5105	0	1	8	4	2	2	22.2	3

### 9.3.3 Individual

field name	description
individualid	consumer identification number



age	age (e.g. in years, months or days)
weight	body weight (e.g. in kg or g)
sex (text)	gender
foodsurvey (text)	name of survey

- In table Foodsurvey the unit for age and weight is defined.
- No missing values allowed.

Example:

	individualid	age	weight	sex	foodsurvey
+	102	343	8.32	male	EU-BABY
+	101	347	9.23	male	EU-BABY
+	100	317	9.94	female	EU-BABY
+	901312	89	70	male	DNFCS-3
+	901311	87	60	female	DNFCS-3
+	901231	80	48	female	DNFCS-3
+	901192	68	63	male	DNFCS-3
+	901191	64	68	female	DNFCS-3

### 9.3.4 Foodsurvey

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

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

Example:

foodsurvey	year	country	agein	weightin
DNFCS-3	1997	NL	Y	KG
EU-BABY	1998	NL	D	KG

### 9.3.5 Processing

field name	description
proccode	code of processing type
proctype (text)	description of processing type
disttype	indicator (1/2): <ul style="list-style-type: none"> <li>• simulated processing factors are restricted to the interval (0,1) using a logistic-normal distribution (1),</li> <li>• or simulated processing factors are restricted to positive values using a log-normal distribution (2)</li> </ul>
bulkingblending	indicator (0/1) for types of processing applied on large batches, e.g. juicing, sauce/puree (obligatory), <ul style="list-style-type: none"> <li>• 0 = no bulking/blending ;</li> <li>• 1 = bulking/blending</li> </ul>

- 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:

Processing : Table				
	proccode	proctype	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
*	0		0	0

### 9.3.6 Processingfactor

field name	description
compoundgroup	compound group number
compoundsubgroup	compound subgroup number
compound	compound number
productcategory	primary agricultural product category number
productgroup	primary agricultural product group number
productsubgroup	primary agricultural product subgroup number
product	primary agricultural product number
productq	primary agricultural product quality number
proccode	code of processing type
procnom	nominal value (best estimate) of processing factor
procupp	upper value (estimate of 97.5 <sup>th</sup> percentile or “worst case” estimate) of processing factor

- The value for procupp should be higher than the value for procnom.

Example:

Processingfactor : Table											
	compoundgroup	compoundsubgroup	compound	productcategory	productgroup	productsubgroup	product	productq	proccode	procnom	procupp
	13	6	6	1	8	4	1	1	3	0.11	0.11
	13	6	6	1	8	4	1	1	5	0.23	0.34
	11	5	1	1	3	2	3	1	11	0.5	0.56
▶	11	2	1	1	4	2	3	1	2	0.2	0.77

## 9.4 Concentration of compounds tables

### 9.4.1 Compounds

field name	description
compoundgroup	compound group number (first level in hierarchy)
compoundsubgroup	compound subgroup number (second level in hierarchy)
compound	compound number(third level in hierarchy)
arfd	ARfD (acute reference dose), in microgr/kg bw/day
adi	ADI (acceptable daily intake), in microgr/kg bw/day
compoundname (text)	compound label (name of compound)
unit	coding for specification of units for compound and intake

- Each compound is characterised by a code built hierarchically from 3 numbers.
- 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	compoundgroup	compoundsubgroup	arfd	adi	compoundname	unit
56	12	7	2	0.6	MONOCROTOPHOS	-6
83	12	7	2	9999	OXYDEMETON-METHYL	-6
34	12	7	3	0.8	FENAMIPHOS	-6
28	12	7	3	0.3	DISULFOTON	-6
6	12	99	3	0.2	FIPRONIL	-9
55	12	7	3	0.8	MEVINPHOS	-9
44	19	99	3	0.3	PROPYLENE THIOUREA	-9

### 9.4.2 Country

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

- No missing values allowed

Example:

country	countryname
99	UNKNOWN
AA	ARUBA
AB	AZERBEIJAN
AD	ANDORRA
AE	UNITED ARAB EMIRATES
AF	AFGANISTAN
AG	ANTIGO
AK	AFRICA
AL	ALBANIA
AN	THE NETHERLANDS ANTILLES

### 9.4.3 Concentration

field name	description
compoundgroup	compound group number
compoundsubgroup	compound subgroup number
compound	compound number
productcategory	primary agricultural product category number
productgroup	primary agricultural product group number
productsubgroup	primary agricultural product subgroup number
product	primary agricultural product number
productq	primary agricultural product quality number
year	sampling year
month	number of month
country (text)	country of sample
samplingtype (text)	type of sampling (monitoring)
numberofsamples	count of the number of times the specified concentration or limit of reporting (LOR) occurs
value	concentration (mg/kg) or LOR (see below)

- Concentration values are measured on primary agricultural products (productcategory=1).
- The limit of reporting (mg/kg) 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.
- The number of non-detects for product 1 6 1 1 1 is 10, the total number of samples taken is equal to 11 (see example, first and second row).
- 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:

	compoundgroup	compoundsubgroup	compound	productcategory	productgroup	productsubgr
	12	99	5	1	9	
	13	2	5	1	9	
	13	6	5	1	9	
	13	89	5	1	9	
	11	13	5	1	9	
	11	89	5	1	9	
	11	99	5	1	9	
	11	99	5	1	9	

productgroup	productsubgroup	product	productq	year	month	samplingtype	country	numberofsamples	value
9	2	3	1	2002	99	M	IMP	3	-0.05
9	2	3	1	2002	99	M	IMP	3	-0.05
9	2	3	1	2002	99	M	IMP	3	-0.05
9	2	3	1	2002	99	M	IMP	3	-0.05
9	4	1	1	2001	99	M	NL	3	-0.05
9	4	1	1	2001	99	M	NL	3	-0.05
9	4	1	1	2001	99	M	IMP	3	0.08
9	4	1	1	2001	99	M	IMP	3	0.14

### 9.4.4 Summarydata

field name	description
compoundgroup	compound group number
compoundsubgroup	compound subgroup number
compound	compound number
productcategory	primary agricultural product category number
productgroup	primary agricultural product group number
productsubgroup	primary agricultural product subgroup number
product	primary agricultural product number
productq	primary agricultural product quality number
country (text)	code for country
limitofreporting	limit of reporting (mg/kg)
numberofsamples	size of sample (detects and non-detects)
numberofpositives	number of positive concentration values (detects)
<i>the mean</i> : mean or meanall	statistic for the mean
<i>the median</i> : med or medall	statistic for the median
max	statistic for the maximum
<i>the variance</i> : var or varall	statistic for the variance
<i>the percentile</i> : perc or perc all	statistic for the percentile
percentile	specifies the percentage of the statistics <b>perc</b> and <b>perc all</b>

- Summary statistics are derived from primary agricultural products (productcategory=1).
- MCRA release 4 has an option to input summary data for compound concentrations. These data are then used for estimating parameters of a parametric model (see 3.2.1 ).
- Field names mean, meanall, med, medall, max, var, varall, perc, perc all 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 product 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 **perc all** 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:

	compoundgroup	compoundsubgroup	compound	productcategory	productgroup	productsubgroup	product	productq	country
▶	11	4	1	1	9	4	3	1	FI
	11	5	1	1	8	1	5	4	NL
	11	4	1	1	8	5	2	1	DK
	11	4	1	1	9	2	1	1	NL
	11	5	1	1	9	4	1	1	NL

arydata : Table										
try	limitofreporting	numberofsamples	numberofpositives	mean	meanall	med	max	var	perc	percentile
	0.01	20	12	2.1	9999	0.2	4.1	9999	9999	9999
	0.3	42	13	9999	9999	1.6	4	9999	3.2	95
	5	90	15	0.4	9999	0.34	1	0.754	9999	9999
	0.3	25	20	0.934	9999	9999	3.13	9999	9999	9999
	0.3	120	25	9999	3	9999	5.6	1.43	9999	9999

### 9.4.5 Histogramdata

field name	description
compoundgroup	compound group number
compoundsubgroup	compound subgroup number
compound	compound number
productcategory	primary agricultural product category number
productgroup	primary agricultural product group number
productsubgroup	primary agricultural product subgroup number
product	primary agricultural product number
productq	primary agricultural product quality number
country (text)	code for country
limitofreporting	limit of reporting (mg/kg)
numberofsamples	size of sample (detects and non-detects)
c%01	<ul style="list-style-type: none"> <li>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).</li> </ul>
c%02	
c%05	
c%1	
c%2	
c%5	<ul style="list-style-type: none"> <li>classes (<i>i.e.</i> columnnames) are free to choose so <i>c&lt;xxx&gt;...</i> may be replaced with any appropriate concentration e.g. c5, c10 etc.</li> </ul>
c1	
c2	
c<xxx>	
c<xxx>	
c<xxx>	
cE10	

- Histogramdata are derived from primary agricultural products (productcategory=1).
- MCRA release 4 has an option to input frequency counts (histogram data) for compound concentrations. These data are then used for estimating parameters of a parametric model (see 3.2.1 ).
- Field names for columns 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  $1 \cdot 10^{10}$ .
- 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 product, delete the entire row.

Example:

Histogramdata : Table									
	compoundgroup	compoundsubgroup	compound	productcategory	productgroup	productsubgroup	product	productq	coun
▶	11	4	1	1	8	5	2	1	NL
	11	4	1	1	9	4	3	1	NL
	11	4	1	1	9	2	1	1	NL
	11	5	1	1	8	1	5	4	NL
	11	5	1	1	9	4	1	1	NL

Histogramdata : Table																
	country	limitofreporting	numberofsamples	c%01	c%02	c%05	c%1	c%2	c%5	c1	c2	c5	c10	c20	c50	cE10
▶	NL	-9999	109	0	0	0	0	0	0	0	0	0	0	0	0	0
	NL	0.03	128	0	0	4	6	4	5	0	0	0	0	0	0	0
	NL	0.03	129	0	6	17	9	6	7	6	0	0	0	0	0	0
	NL	-9999	134	0	0	0	0	0	0	0	0	3	12	34	23	0
	NL	0.03	181	0	0	0	0	0	0	0	2	4	8	6	3	0

### 9.4.6 VariabilityProd

field name	description
productcategory	primary agricultural product category number
productgroup	primary agricultural product group number
productsubgroup	primary agricultural product subgroup number
product	primary agricultural product number
productq	primary agricultural product quality number
varfac	variability factor
coefvar	coefficient of variation
nounitcomp	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.
- All columns need to be present: use code 9999 when data are not available.
- When the parameter for unit variability is a coefficient of variation and the number of units equals 1, unit variability is ignored for this product.

Example:

VariabilityProd : Table								
	productcategory	productgroup	productsubgroup	product	productq	varfac	coefvar	nounitcomp
	1	8	2	1	1	5	1.375	12
	1	8	2	2	1	4	1.9	12
▶	1	8	2	3	1	3	1.5	12
	1	8	2	4	1	5	1.5	12
	1	8	2	5	1	1	9999	12
	1	8	2	6	1	7	1.86	12

### 9.4.7 VariabilityCompProd

field name	description
compoundgroup	compound group number
compoundsubgroup	compound subgroup number
compound	compound number
productcategory	primary agricultural product category number
productgroup	primary agricultural product group number
productsubgroup	primary agricultural product subgroup number

product	primary agricultural product number
productq	primary agricultural product quality number
varfac	variability factor
coefvar	coefficient of variation
nounitcomp	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:

compoundgroup	compoundsubgroup	compound	productcategory	productgroup	productsubgroup	product	productq	varfac	coefvar	nounitcomp
11	5	3	1	3	2	3	3	3	0.34	45
11	5	3	1	3	2	3	4	4	0.25	56
13	4	3	1	2	2	1	13	5	0.5	30
12	1	1	1	9	9	2	15	3.4	1.2	23

### 9.4.8 VariabilityProcCompProd

field name	description
compoundgroup	compound group number
compoundsubgroup	compound subgroup number
compound	compound number
productcategory	primary agricultural product category number
productgroup	primary agricultural product group number
productsubgroup	primary agricultural product subgroup number
product	primary agricultural product number
productq	primary agricultural product quality number
proccode	processing type code
varfac	variability factor
coefvar	coefficient of variation
nounitcomp	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:

compoundgroup	compoundsubgroup	compound	productcategory	productgroup	productsubgroup	product	productq	proccode	varfac	coefvar	nounitcomp
11	5	1	1	3	4	3	1	3	4	0.25	20
11	7	7	1	2	2	1	18	5	2.4	0.34	25
2	2	1	1	4	4	0	19	5	3.1	0.5	10
7	7	1	1	0	2	2	11	4	2.5	0.3	20

### 9.4.9 Agriculturaluse

field name	description
compoundgroup	compound group number
compoundsubgroup	compound subgroup number
compound	compound number
productcategory	primary agricultural product category number
productgroup	primary agricultural product group number
productsubgroup	primary agricultural product subgroup number



product	primary agricultural product number
productq	primary agricultural product quality number
country (text)	code for country
year	year
useallowed	indicator (0/1) whether use of the compound for the product is allowed (1) or not (0)
percproptreated	maximum percentage of the product that is treated with the compound
maximumresiduelimit	worst case value

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

Example:

Agriculturaluse : Table							
	compoundgroup	compoundsubgroup	compound	productcategory	productgroup	productsubgroup	product
	11	5	1	1	7		4
	12	4	2	1	1		4
	11	5	1	1	4		4

se : Table							
group	product	productq	country	year	useallowed	percproptreated	maximumresiduelimit
4	2	1	NL	2002	1	50	567
4	2	12	UK	2003	1	34	23
4	2	1	FR	2001	1	80	230

## 10 APPENDIX B: Other output depending on specifications of options

In 4.1 and 4.2 the output of two examples of an MCRA analysis are shown. In this appendix the output of more types of analysis (specified by the options in Ch. 3 ) are demonstrated.

### 10.1 Acute risk assessment: processing fixed factors

In this example, output is shown for an acute risk assessment and processing using fixed factors for organo phosphate pesticide chlorpyrifos. Table 23 lists the main options:

<b>Input form</b>	
risk type	acute
uncertainty analysis	no
concentration model	empirical
number of Monte Carlo iterations	200000
number of computational chunks	10
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 23: Input form options: processing**

In Table 24 you find the main characteristics concerning this analysis taken from the logfile. Note that the number of products and processing combinations is 52 whereas the number of products is 12.

<b>Logfile</b>	
Number of products	12
Acute reference dose (ARfD)	100
Average daily intake (ADI)	10
Number of detects	364
Number of non-detects	2408
Number of products and processing type combinations	52
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	12191

**Table 24: Information in logfile**

For a summary of the data, see Table 5. In Table 25 you find a summary of the simulated intakes. Compared to Table 6, this table contains a second section with information on processed products only. New in the lower part is 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.

Compound: CHLORPYRIFOS

\*\*\*\*\*

SUMMARY SIMULATION

Code : product code  
 Product : product label  
 MeanCons : 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  
 N : total number of iterations  
 %PosCons : percentage consumption days  
 MeanConc : mean concentration in simulations with positive  
           amount consumed (after processing)  
 DeltaR : difference (%) compared to average concentration  
           in database  
 NPosConc : number of positive concentrations in simulations  
           with positive amount consumed.  
 NConc : total number of concentration measurements  
           (detects and non-detects) in simulations with  
           positive amount consumed.  
 %PosConc : percentage positive concentrations  
 ProcFact : mean processing factor

\*\*\*\*\*

FOOD CONSUMPTION DATA

Code	Product	MeanCons (g)	DeltaC (%)	MeanConsDays (g)	NConsDay	N	%PosCons (%)
17111.10	GREEN BEANS	11.73	( -0.5)	116.13	20209	200000	10.1
18115.20	SPINACH	8.19	( 1.4)	226.00	7252	200000	3.6
18401.10	POTATOES	138.40	( -0.1)	172.95	160042	200000	80.0
18402.20	CARROT	12.19	( 0.6)	36.35	67069	200000	33.5
18506.10	SWEET PEPPER	3.36	( 0.7)	16.53	40620	200000	20.3
19201.10	APPLE	62.04	( 0.9)	100.14	123909	200000	62.0
19304.10	PEACH	2.07	( -0.4)	7.32	56502	200000	28.3
19401.10	GRAPE	13.27	( 0.3)	34.29	77403	200000	38.7
19501.10	LEMON	1.51	( -1.8)	4.15	73069	200000	36.5
19502.10	ORANGE	56.85	( -0.4)	99.35	114436	200000	57.2
19503.10	MANDARIN TA	8.77	( -0.6)	40.96	42843	200000	21.4
19507.10	GRAPEFRUIT	4.09	( 0.5)	28.77	28404	200000	14.2

COMPOUND CONCENTRATION DATA

Code	Product	MeanConc (mg/kg)	DeltaR (%)	NPosconc	NConc	%PosConc (%)	ProcFact
17111.10	GREEN BEANS	0.0121	( 0.8)	182	20360	0.9	1.00
18115.20	SPINACH	0.0038	( 1.3)	12	7340	0.2	0.93
18401.10	POTATOES	0.0005	( -3.4)	559	206500	0.3	1.00
18402.20	CARROT	0.0014	( -0.1)	240	76793	0.3	1.00
18506.10	SWEET PEPPER	0.0023	( -1.6)	109	46799	0.2	1.00
19201.10	APPLE	0.0011	( -0.9)	924	172223	0.5	0.82
19304.10	PEACH	0.0065	( -1.2)	734	60336	1.2	1.00
19401.10	GRAPE	0.0162	( 0.2)	1467	89382	1.6	0.92
19501.10	LEMON	0.0105	( 0.2)	3492	81571	4.3	1.00
19502.10	ORANGE	0.0347	( 0.6)	11825	144400	8.2	0.73
19503.10	MANDARIN TA	0.0350	( -0.4)	4016	45082	8.9	0.60
19507.10	GRAPEFRUIT	0.0697	( 1.7)	3872	28690	13.5	1.00

=====++=====

Product & Processing	MeanCons (gr)	DeltaC (%)	MeanConsDays (gr)	NConsDays	N	%PosCons (%)	ProcFact
GREEN BEANS:cooking in w	7.66	( -0.4)	11777.00	130	200000	0.1	1.00
GREEN BEANS:canned/conse	4.07	( -0.8)	8565.00	95	200000	0.0	1.00
GREEN BEANS:washing/clea	0.00	( 12.5)	18.00	50	200000	0.0	1.00
SPINACH:raw	0.00	( -8.5)	128.00	3	200000	0.0	0.70
SPINACH:cooking in water	1.77	( -0.5)	2373.00	149	200000	0.1	0.80
SPINACH:canned/conserved	6.41	( 2.0)	4794.00	267	200000	0.1	1.00
SPINACH:washing/cleaning	0.01	( -14.1)	45.00	61	200000	0.0	1.00
POTATOES:cooking in wate	89.55	( 0.0)	110821.00	162	200000	0.1	1.00
POTATOES:frying/baking i	45.08	( -0.3)	50825.00	177	200000	0.1	1.00
POTATOES:unknown	3.76	( 1.0)	44854.00	17	200000	0.0	1.00
CARROT:raw	0.08	( -1.1)	3560.00	4	200000	0.0	1.00
CARROT:cooking in water	8.45	( 0.9)	33324.00	51	200000	0.0	1.00
CARROT:canned/conserved	2.47	( -0.4)	35943.00	14	200000	0.0	1.00
CARROT:washing/cleaning	1.18	( 0.7)	3966.00	60	200000	0.0	1.00
SWEET PEPPER:raw	0.08	( -1.1)	3560.00	4	200000	0.0	1.00
SWEET PEPPER:cooking in	2.08	( 1.5)	18992.00	22	200000	0.0	1.00

SWEET PEPPER:canned/cons	0.68	( 1.7)	19729.00	7	200000	0.0	1.00
SWEET PEPPER:sauce/puree	0.02	( 2.5)	972.00	4	200000	0.0	1.00
SWEET PEPPER:washing/cle	0.50	( -3.3)	3546.00	28	200000	0.0	1.00
APPLE:raw	0.12	( -2.6)	6921.00	3	200000	0.0	0.70
APPLE:peeling	19.80	( 0.6)	42641.00	93	200000	0.0	0.80
APPLE:drying	0.21	( 0.6)	9190.00	5	200000	0.0	1.00
APPLE:juicing	17.38	( 1.2)	69397.00	50	200000	0.0	0.70
APPLE:sauce/puree	6.67	( 0.6)	17640.00	76	200000	0.0	1.00
APPLE:washing/cleaning	17.86	( 1.0)	26434.00	135	200000	0.1	1.00
PEACH:raw	0.04	( -0.3)	1308.00	7	200000	0.0	1.00
PEACH:canned/conserved	0.48	( 5.0)	5622.00	17	200000	0.0	1.00
PEACH:juicing	0.59	( -0.5)	47792.00	2	200000	0.0	1.00
PEACH:marmelade/jam	0.00	( 3.9)	116.00	1	200000	0.0	1.00
PEACH:washing/cleaning	0.95	( -2.9)	5498.00	34	200000	0.0	1.00
GRAPE:raw	0.06	( -0.3)	1308.00	8	200000	0.0	1.00
GRAPE:canned/conserved	0.12	( 6.7)	2369.00	10	200000	0.0	1.00
GRAPE:drying	9.08	( -0.1)	45836.00	40	200000	0.0	1.00
GRAPE:juicing	1.54	( 6.5)	35464.00	9	200000	0.0	0.80
GRAPE:sauce/puree	0.01	( -3.9)	141.00	16	200000	0.0	1.00
GRAPE:washing/cleaning	2.46	( -2.1)	4264.00	115	200000	0.1	1.00
LEMON:raw	0.07	( -0.4)	38595.00	0	200000	0.0	1.00
LEMON:peeling	0.03	( 0.6)	398.00	15	200000	0.0	1.00
LEMON:juicing	1.41	( -1.9)	42578.00	7	200000	0.0	1.00
ORANGE:peeling	13.58	( -0.1)	29811.00	91	200000	0.0	0.70
ORANGE:drying	0.03	( 1.1)	1912.00	3	200000	0.0	1.00
ORANGE:juicing	43.06	( -0.5)	76065.00	113	200000	0.1	0.60
ORANGE:marmelade/jam	0.16	( 1.7)	36471.00	1	200000	0.0	1.00
ORANGE:sauce/puree	0.02	( -2.8)	141.00	22	200000	0.0	1.00
MANDARIN, TA:peeling	8.55	( -0.6)	32219.00	53	200000	0.0	0.60
MANDARIN, TA:canned/cons	0.10	( 1.4)	2019.00	10	200000	0.0	1.00
MANDARIN, TA:juicing	0.08	( 1.2)	10107.00	2	200000	0.0	0.50
MANDARIN, TA:marmelade/j	0.00	( 3.9)	116.00	1	200000	0.0	1.00
MANDARIN, TA:washing/cle	0.04	( -2.6)	621.00	14	200000	0.0	1.00
GRAPEFRUIT:peeling	1.30	( 1.3)	2203.00	118	200000	0.1	1.00
GRAPEFRUIT:canned/conser	0.02	( 2.7)	1407.00	3	200000	0.0	1.00
GRAPEFRUIT:juicing	2.77	( 0.2)	25080.00	22	200000	0.0	1.00

**Table 25: Summary of simulation, including processed products**

For processing, all tables as found in 4.1 are extended with the additional information on the processed product. Note that the percentiles of the acute intake distribution after applying processing factors, see Table 26, are slighter lower than without processing (see Table 9)

```

*****
Random sampling is based on seed :          0
Number of iterations (consumers):  200000 out of 6250
CHLORPYRIFOS (microgr/kg bw/day)  consumption: 200000 out of 200000

*****
Compound:  CHLORPYRIFOS
*****
Percentiles, maximum and average intake
*****
Percentage      Percentiles of CHLORPYRIFOS      (microgr/kg bw/day)
50.00           0.00000
90.00           0.13440
95.00           0.26667
99.00           0.78129
99.90           2.80839
99.99           6.05685
mean            0.05109
maximum         20.40600

```

**Table 26: 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 27 lists the main options:

**Input form**

risk type	Acute
uncertainty analysis	No
concentration model	Empirical
number of Monte Carlo iterations	200000
number of computational chunks	10
unit variability model	beta distribution
random seed	0
intake model	only empirical estimates
concentration data	system defaults
Additional	system defaults
Output	system defaults

**Table 27: Input form options: unit variability**

In Table 28 you find the main characteristics concerning this analysis taken from the logfile. The logfile contains some information on unit variability like missing variability factors and the maximum number of units found in a consumption.

**Logfile**

```

Number of products                12
Acute reference dose (ARfD)      100
Average daily intake (ADI)       10
Number of detects                 364
Number of non-detects            2408
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     12191
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.
Some variability factors are missing.
For product(s):      code      product
                   18402.20  CARROT
unit variability is ignored.
Maximum number of units in a consumption:  13
for product(s):      APPLE

```

**Table 28: Information in logfile**

In Table 29, the percentiles of the intake distribution after applying unit variability are shown. Note that the percentiles are much higher (p99.99 = 116.41) than without unit variability (p99.99 = 7.11, see Table 9) and after applying processing (p99.99 = 6.06, see Table 26).

```

*****
Random sampling is based on seed :      0
Number of iterations (consumers):  200000 out of  6250
CHLORPYRIFOS (microgr/kg bw/day)  consumption:  200000 out of 200000
*****
Compound:  CHLORPYRIFOS

```

```

*****
Percentiles, maximum and average intake
*****
Percentage      Percentiles of CHLORPYRIFOS      (microgr/kg bw/day)
50.00           0.00000
90.00           0.06731
95.00           0.31501
99.00           1.99204
99.90           12.34006
99.99           116.41577
mean            0.13768
maximum         414.39878

```

**Table 29: 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 30 lists the main options:

```

Input form
risk type                Acute
uncertainty analysis     No
concentration model      Empirical
number of Monte Carlo iterations  200000
number of computational chunks  10
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
Output                  system defaults

```

**Table 30: Input form options: IESTI**

In Table 31 you find the main characteristics concerning this analysis taken from the logfile.

```

Logfile
Number of products      12
Acute reference dose (ARfD)  100
Acceptable daily intake (ADI)  10
Number of detects       364
No of consumers         6250
Number of non-detects   2408
Population characteristics,
  minimum age           1
  maximum age           97
  minimum weight        8
  maximum weight        150
  Sex                   female, male
Total no of consumption days  12191

```

**Table 31: Information in logfile**

In Table 32 IESTI estimates are displayed together with the estimate expressed as percentage of the ARfD (%ARfD). Also a comparison is made with the Monte Carlo percentile per product for positive consumption days only (ConsPos) as well as all consumption days (AllDays). For Orange the IESTI estimate is 1.067 microgr/kg bw/day which is slighter lower than the Monte Carlo percentiles for positive consumption days only (1.341 microgr/kg bw/day) and slighter higher than the Monte Carlo percentiles for all consumption days (0.978 microgr/kg bw/day).

```

*****
Compound:  CHLORPYRIFOS
*****
IESTI estimates (microgr/kg bw/day)
%ofARfD : estimates expressed as % of Acute Reference Dose
Cons_Pos: percentiles per product for positive consumptions only
Alldays : percentiles per product for all days (including days without consumption)

Note that option Consumption days only restricts the total number of days to a subset.
If one restricts attention to Consumption days Only, percentiles are typically
based on consumption data which are partly zero (days with consumption zero for
some but not all products)

IESTI compared with Monte Carlo Percentiles for Positive Consumption Only and All days

```

Product	IESTI	%ofARfD	ConsPos p99.00%	AllDays p99.00%
GREEN BEANS	2.145	2.145	0.215	0.000
SPINACH	0.960	0.960	0.000	0.000
POTATOES	0.099	0.099	0.000	0.000
CARROT	0.187	0.187	0.002	0.000
SWEET PEPPER	0.960	0.960	0.003	0.000
APPLE	0.131	0.131	0.087	0.043
PEACH	0.752	0.752	0.013	0.002
GRAPE	2.360	2.360	0.232	0.069
LEMON	0.288	0.288	0.014	0.005
ORANGE	1.067	1.067	1.341	0.978
MANDARIN TA	0.950	0.950	0.634	0.256
GRAPEFRUIT	7.067	7.067	0.690	0.071

**Table 32: 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 33 lists the main options:

Input form	
risk type	acute
uncertainty analysis	yes
concentration model	empirical
number of Monte Carlo iterations	200000
number of computational chunks	10
unit variability model	no unit variability
random seed	0
intake model	only empirical estimates
number of bootstrap samples	1000
number of iterations per bootstrap sample	20000
concentration data:	system default
Additional	system defaults
Output	system defaults

**Table 33: Input form options: uncertainty**

In Table 34 you find the main characteristics concerning this analysis taken from the logfile. Consumptions and concentration data are bootstrapped.

<b>Logfile</b>	
Number of products	12
Acute reference dose (ARfD)	100
Average daily intake (ADI)	10
Number of detects	364
Number of non-detects	2408
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	12191
Total no of simulations	20000000
(uncertainty)	
Bootstrap consumptions	yes
Bootstrap concentration data	yes

**Table 34: Information in logfile**

In Table 35, 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 (7.1128 microgr/kg bw/day) is (4.7934, 11.6237).

```

*****
Random sampling is based on seed :          0
Number of iterations (consumers):  200000 out of  6250
CHLORPYRIFOS (microgr/kg bw/day)  consumption:  200000 out of 200000
*****
Compound:  CHLORPYRIFOS
*****
Percentiles, maximum and average intake
*****
Percentage      Percentiles of CHLORPYRIFOS      (microgr/kg bw/day)
50.00           0.00000
90.00           0.19603
95.00           0.39706
99.00           1.16024
99.90           3.51131
99.99           7.11281
mean            0.07303
maximum        20.40600

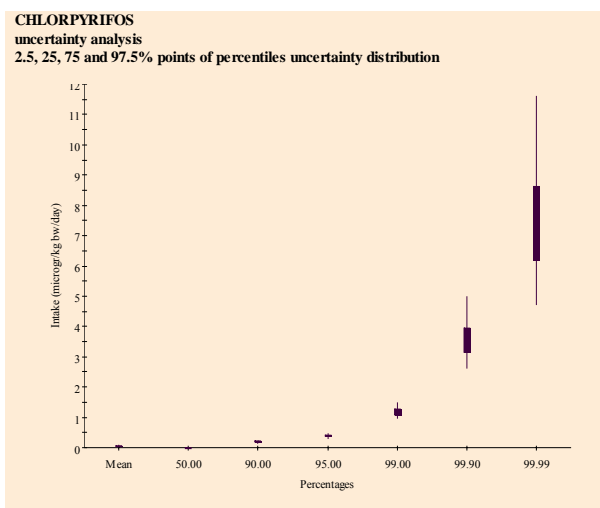
Uncertainty of percentiles distribution
Percentage      2.5%           25%            75%            97.5%
50.00          0.0000         0.0000         0.0000         0.0000
90.00          0.1570         0.1800         0.2059         0.2330
95.00          0.3271         0.3687         0.4167         0.4605
99.00          0.9524         1.0915         1.2614         1.4790
99.90          2.5699         3.1447         3.9837         4.8936
99.99          4.7934         6.2099         8.5555         11.6237
mean           0.0589         0.0678         0.0779         0.0881
maximum        5.6234         8.0000         11.5054         20.4060

```

**Table 35: Percentiles for the intake distribution and uncertainty**

The information of Table 35 is plotted in Figure 78. The 2.5 and 97.5% points are displayed by the small line segments, the thick bar indicates the 25 and 75% points of the percentile uncertainty distribution. As seen, percentiles for high percentages are uncertain and have a large uncertainty interval.





**Figure 78: Uncertainty analysis**

## 10.5 Acute risk assessment: diagnostics

In this example, output is shown for an acute risk assessment and an uncertainty analysis focussing on the diagnostics for organo phosphate pesticide chlorpyrifos. For input options and information in logfile, see Table 33 and Table 34, respectively.

Diagnostics are only available after running an acute risk analysis in combination with uncertainty analysis. The diagnostic tools provided, focus on the stability of the percentiles, or, re-phrasing, quantify 1) the amount of Monte Carlo variability and, 2) the amount of variability due to bootstrapping 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 Monte Carlo iterations, the number of iterations per bootstrap and the number of bootstrap samples used. Recall that we ran the analysis with 200.000 Monte Carlo iterations and 1000 bootstrap samples with 20.000 iterations each.

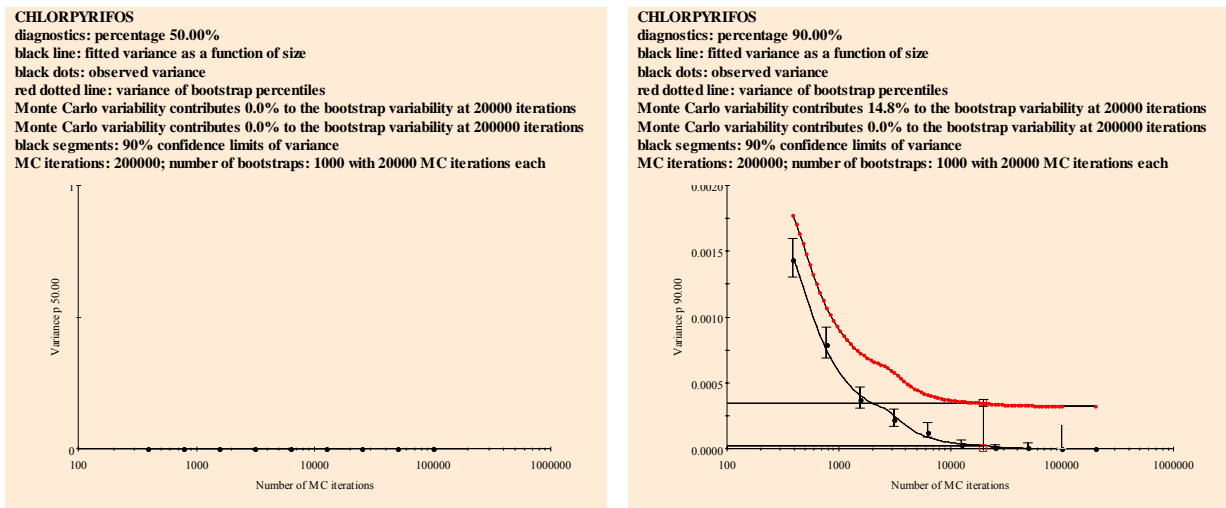
To make inference, we divide the total number of MC-iterations in 2 samples of 100.000 iterations each, 4 samples of 50.000 each, 8 samples of 25.000 each, ..., etc. By doing so, we get  $n$  partitions of samples and in each partition we have  $2^n$  samples of size  $200.000/2^n$ . In each partition, 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 100.000; in partition  $n = 2$  the estimate of the variance is based on 4 percentiles derived from samples of size 50.000, ..., etc. The estimated variances of each partition are plotted against the number of MC-iterations per sample of each partition. We expect the variance to decrease as function of sample size, so for larger sample sizes Monte Carlo 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 1000 bootstrap samples of sample size 20.000.

Now the fitted spline function is interpolated to estimate the amount of MC-variability at 20.000 iterations and to calculate the contribution of MC-variability to the total bootstrap variability at 20.000 iterations. The fitted spline function can be used to estimate the MC-contribution to the total bootstrap variability for any arbitrary number of MC-iterations.

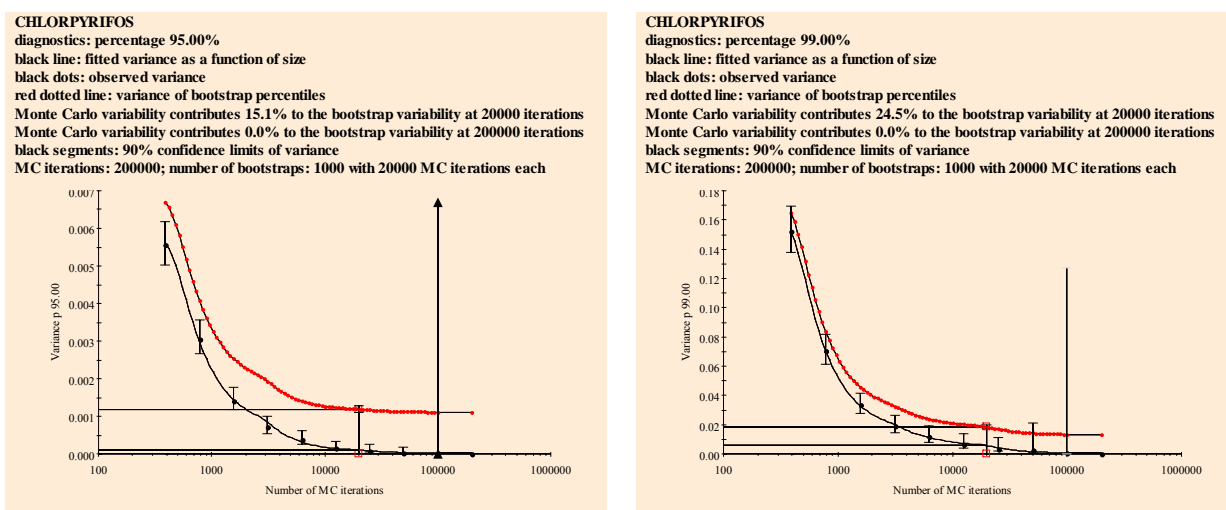
In the left plot of Figure 79 diagnostics are displayed for percentage point 50%. We can skip this plot because the percentile is 0.0000 microgram/kg bw/day. For percentage point 90% in the right plot we have an estimate of the percentile (0.19603 microgr/kg bw/day) and we are able to make inferences about the stability of the estimate.

On the x-axis the number of MC-iterations 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 iterations in each bootstrap sample. Note the horizontal black line with the open red boxes at 20.000 iterations, which is the estimate of the variance of the bootstrap percentiles. The black dot at 200.000 iterations is the extrapolated value for the MC-variability using the spline fit. At 20.000 iterations the MC-variability contributes 14.8% to the total bootstrap variability. For a theoretical bootstrap sample of size 200.000 the MC-contribution would be 0.0%.



**Figure 79: Diagnostic graphs for percentage 50% and 90%**

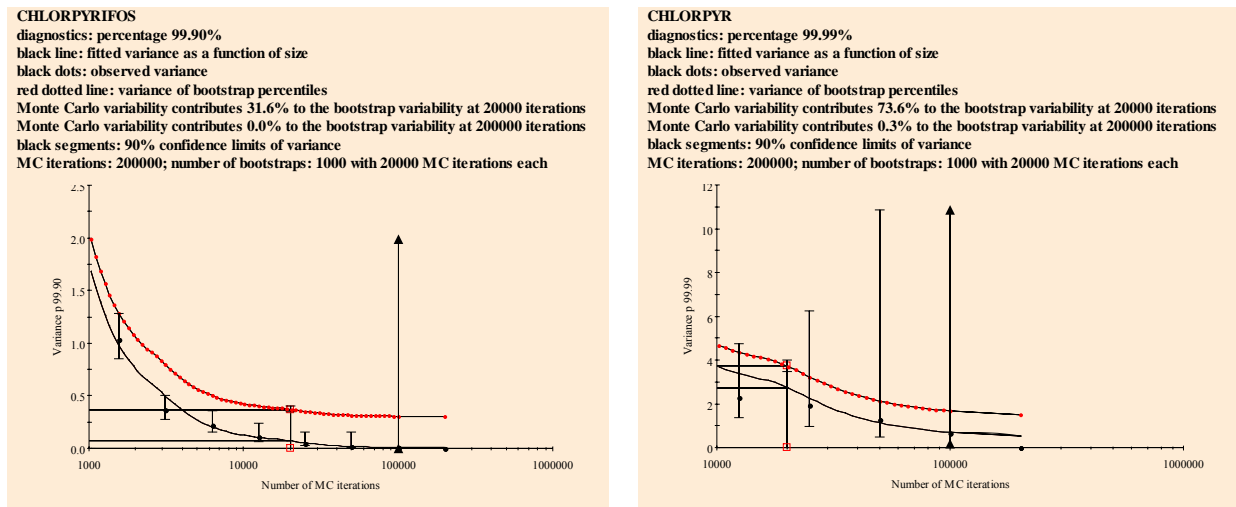
In Figure 80, the diagnostics for percentage point 95% and 99% are displayed. The contribution of the MC-variability for p95 and p99 is 15.1% and 24.5%, respectively, indicating that a higher percentage has a less stable percentile than a lower percentage. The confidence interval at 100.000 MC-iterations is displayed as an arrow for graphical reasons and indicates that a cut off is used. The real confidence interval is much higher.



**Figure 80: Diagnostic graphs for percentage 95% and 99%**

In Figure 81 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 31.6% and 73.6.5%, respectively, indicating that these percentiles are unstable. Extrapolation to 200.000 iterations shows that MC-variability contributes

0.3% to the total bootstrap variability (p99.99 right plot). Note that in Figure 81 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:  $(100 - 100/\text{size})$ ).



**Figure 81: Diagnostic graphs for percentage 99.9% and 99.99%**

## 10.6 Acute risk assessment: betabinomial distribution with age

In this example, output is shown for an acute risk assessment followed by a betabinomial/lognormal 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 36 lists the main options:

### Input form

```

risk type                acute
uncertainty analysis      no
concentration model       empirical
number of Monte Carlo iterations 200000
number of computational chunks 10
unit variability model    no unit variability
random seed              0
intake model              empirical estimates +
                           betabinomial/lognormal with age
concentration data        system defaults
additional                system defaults
intake model:
age effect for intake frequency  yes
model
function to model effect    spline
testing method             backward
minimum degrees of freedom  0
maximum degrees of freedom  4
testing at level          0.01
age effect for intake amount  yes
model
function to model effect    spline
testing method             backward
minimum degrees of freedom  0
maximum degrees of freedom  4
testing at level          0.01
Output                    system defaults
  
```

**Table 36: Input form options: acute risk and betabinomial/lognormal with age**

In Table 37 you find the main characteristics concerning this analysis taken from the logfile. The intake frequency function is estimated using a betabinomial model and a spline 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,  $\ln(\text{intake})$ , is based on REML and a spline function to model age effects.

**Logfile**

```

Number of products                12
Acute reference dose (ARfD)      100
Acceptable daily intake (ADI)    10
Number of detects                 364
No of consumers                  6250
Number of non-detects            2408
Population characteristics,
  minimum age                     1
  maximum age                     97
  minimum weight                  8
  maximum weight                  150
  Sex                             female, male
Total no of consumption days     12191
EXPOSURE SECTION
Acute Risk Assessment
  Betabinomial/Lognormal model
    Intake frequency model is based on Beta Binomial model.
    Include effect of covariable (age).
    Function of covariable        : spline
    Minimum degrees of freedom   : 0
    Maximum degrees of freedom   : 4
    DF selection                  : backward
    Testing at                    : 0.01
  Model for intake amounts is based on REML
    Include effect of covariable (age)
    Function of covariable        : spline
* backward selection of degrees of freedom for spline model with Prob=0.01.
  Df  Ncycle    Phi    _2Loglik    ResDf    Chi    ChiDf    ChiProb
  4   1    0.0003  30284.047    6244    0.60     1    0.439
  3   2    0.0003  30284.646    6245    0.54     1    0.464
  2   2    0.0003  30285.183    6246    1.20     1    0.273
  1   2    0.0003  30286.386    6247    0.01     1    0.906
  0   1    0.0003  30286.400    6248    -        -        -
* Degrees of freedom of spline according to backward deviance testing is 0.

Regression analysis
=====
Estimates of parameters
-----
Parameter                estimate        s.e.        t(*)
Constant                  -0.38910      0.00458     -84.99
Estimate of overdispersion parameter phi = 0.0002787 with s.e. = 0.0004977

Intake Amounts: REML model with age
=====
* Automatic selection of spline model.
Estimates of parameters
-----
Parameter                estimate        s.e.        t(*)
Constant                  -3.791        0.009174    -413.24
Age                       -0.009126     0.009633    -0.95
Spline(age) variance      0.0002838
Between consumer variance 0.001003      0.01196     0.08

```

Within consumer variance	6.602	0.1139	57.99
Variance of ln(intake) distribution is:		6.726	

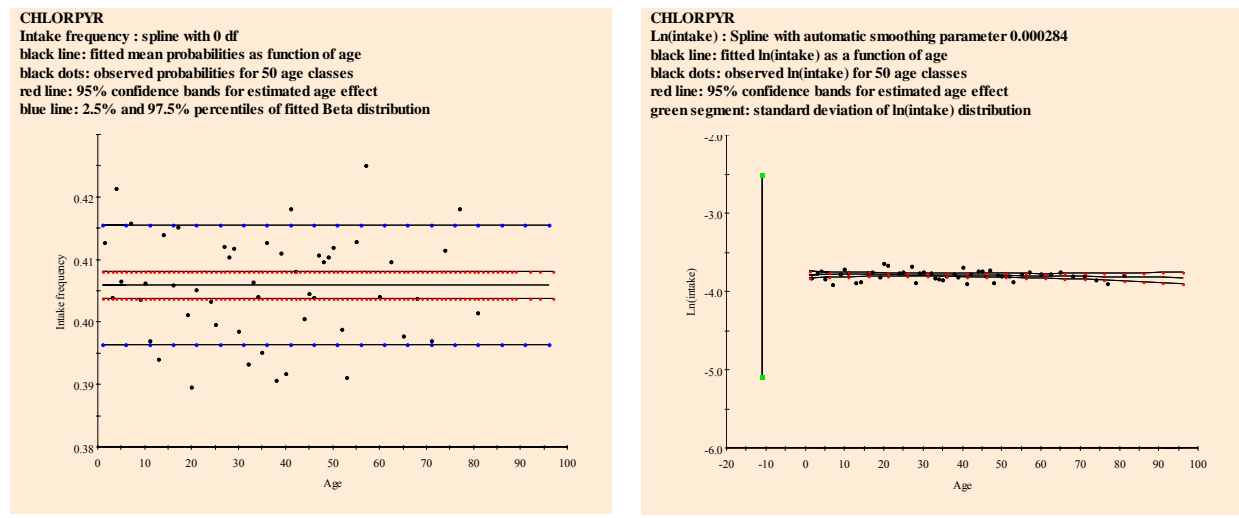
**Table 37: Information in logfile**

As seen in Table 37, no effect for age is found for the intake frequency. A spline with 0 degrees of freedom is fitted which is equivalent to fitting a constant. The overdispersion parameter phi is equal to 0.0002787, representing between consumer variation. The overdispersion is almost zero meaning that variation between consumers is negligible small and each consumer has the same probability of having an intake.

The ln(intake) amounts are modelled using REML with a spline function to model the effect of age. The degree of the spline is automatically selected. A very small age effect is found (-0.009126) meaning that the ln(intake) of chlorpyrivos decreases with age but the effect is almost negligible. The variance of the ln(intake) distribution is 6.726.

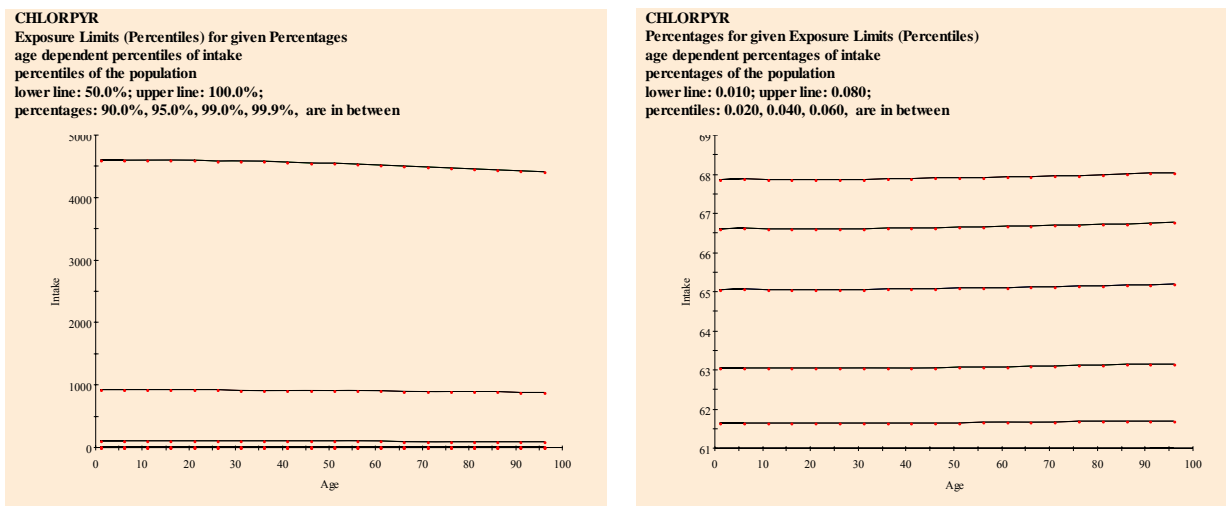
In the left plot of Figure 82 the intake frequency is shown. For 50 age classes, the mean intake frequencies are displayed (black dots). There is no age effect so a horizontal line is plotted through the dots. The red dotted line indicates the 95% confidence interval for the fitted age effect. The blue lines are the 2.5 and 97.5% percentiles of the fitted betabinomial distribution and indicate that the probability of having an intake is between 0.388 and 0.421 with a 0.95 probability.

The right plot displays the ln(intake) distribution. The age effect is very small and the 95% confidence interval is indicated by a red dotted line. 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. It is obvious that the dispersion of the ln(intakes) is large.



**Figure 82: Age dependent intake frequency and ln(intake) distribution**

In Figure 83, age dependent percentiles (derived from percentages) and percentages (derived from percentiles) are shown. The same information is found as tabular output (not shown).



**Figure 83: Age dependent percentiles and percentages**

Find in Table 38 a short summary of an acute risk assessment using age as explanatory variable.

```

Compound:  CHLORPYRIFOS

SHORT TERM EXPOSURE, AGE DEPENDENT PERCENTILES  (ACUTE INTAKE)
Function of covariable is spline
Intake frequency model is based on Beta Binomial model
    with overdispersion parameter phi = 0.0003
Degrees of freedom of spline according to 4 testing is 0
The significance level is backward
Model for consumption (intake) amounts is based on REML
Function of covariable is spline
Variance components (transformed scale)
    spline is 0.00
    within consumers is 6.60
    between consumers is 0.00
    Variance ln(intake) is 6.73
  
```

**Table 38: 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 which can have neurotoxic effects. Table 39 lists the main options:

Input form	
risk type	acute
uncertainty analysis	no
concentration model	binomial/lognormal (with pooling)
number of Monte Carlo iterations	200000
number of computational chunks	10
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 39: Input form options: binomial with pooling**

For information in the logfile, 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 product, the positives samples are taken to estimate the variance and mean on the lognormal scale. If pooling is requested, product groups need to be formed. Each product is characterised by a product code built hierarchically from 5 numbers. The first 3 columns (productcategory, productgroup and productsubgroup) in combination with factor Allowed define a product group used in pooling. Factor Allowed indicates whether a chemical substance is allowed on a product or not

Table 40 illustrates the pooling procedure. There are 8 product groups (ProdGr): 6 groups with a single product and two groups with 2 and 4 products: group 184000 contains products POTATOES and CARROT; group 195000 contains LEMON, ORANGE, MANDARIN and GRAPEFRUIT. Within a productgroup variances and means of products 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  $\mu$  and  $\sigma$  of the lognormal distribution; the pooled number of degrees of freedom is in column Df.

In group 195000, standard deviations are 0.63, 0.63, 0.62 and 0.81. A test of homogeneity of variances is not significant so the pooled standard deviation becomes 0.63. A test of homogeneity of means is significant so means are not pooled. For POTATOE (and SPINACH) no standard deviation is estimated because only one positive sample is available. In a second step, standard deviations of products with less than 10 degrees of freedom are replaced by the overall variance if the test of homogeneity is not significant. For these products, the pooled or original standard deviation is replaced by the value 0.75. Note, that for PEACH (StdDev = 0.98, n = 6) the standard deviation is replaced by the overall standard deviation, but for LEMON (Sigma = 0.65, StdDev = 0.63) 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 = 260). 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 product before pooling
Sigma     : parameter sigma of the lognormal distribution
StdDev    : original st.dev. per product 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 productgroup and allowed
ProdGr    : productgroup
Allwd     : code if residue is allowed on product (1) or not (0)
Product   : productlabels
*****

```

>0	nos	Frpos	Mu	Mean	Sigma	StdDev	AD	ER	Df	Group	ProdGr	Allwd	Product
6	218	0.03	-1.90	-1.90	1.63	1.63	0.70	s	5	1	171000	1	GREEN BEANS
1	158	0.01	-0.45	-0.45	0.75	.	.	.	352	2	181000	1	SPINACH
1	119	0.01	-2.81	-2.81	0.75	.	.	.	352	3	184000	1	POTATOES
2	161	0.01	-2.25	-2.25	0.75	0.40	-0.81	ns	352	3	184000	1	CARROT
3	256	0.01	-1.79	-1.79	0.75	0.75	-0.09	ns	352	4	185000	1	SWEET PEPPER
13	397	0.03	-3.17	-3.17	0.25	0.25	0.75	s	12	5	192000	1	APPLE
6	101	0.06	-2.72	-2.72	0.75	0.99	0.47	ns	352	6	193000	1	PEACH
68	689	0.10	-2.26	-2.26	0.98	0.98	0.90	s	67	7	194000	1	GRAPE
8	63	0.13	-2.70	-2.70	0.65	0.63	0.35	ns	260	8	195000	1	LEMON
139	340	0.41	-2.34	-2.34	0.65	0.63	0.44	ns	260	8	195000	1	ORANGE
84	188	0.45	-2.23	-2.23	0.65	0.62	0.43	ns	260	8	195000	1	MANDARIN TA
33	82	0.40	-2.05	-2.05	0.65	0.81	0.48	ns	260	8	195000	1	GRAPEFRUIT

**Table 40: Pooling information**

In Table 41 the percentiles for a parametric model are displayed. Note that the p99.99 changed from 7.1128 to 9.5988 microgr/kg bw/day.

```

*****
Random sampling is based on seed :      0
Number of iterations (consumers): 200000 out of 6250
CHLORPYRIFOS (microgr/kg bw/day) consumption: 200000 out of 200000
*****
Compound: CHLORPYRIFOS
*****
Percentiles, maximum and average intake
*****
Percentage      Percentiles of CHLORPYRIFOS      (microgr/kg bw/day)
50.00           0.00000
90.00           0.19622
95.00           0.39365
99.00           1.21115
99.90           3.81640
99.99           9.59882
mean            0.07540
maximum         37.59130

```

**Table 41: Percentiles for the acute intake distribution using the binomial/lognormal with pooling**

### Example 2

The variances and means of products are pooled within a product group. Product groups are formed, based on columns productcategory, productgroup and productsubgroup in combination with factor Allowed (see example, factor Levels (= product code) and Allowed). Each product is characterised by a product code built hierarchically from 5 numbers. The first 3 numbers define a product group used in pooling. In the example, only the first 3 numbers (= 5 digits) are printed. Factor Allowed indicates whether a chemical compound is allowed on a product or not.

Productgroup 10101 contains products BEAN and SPERZIEBOON. Productgroup 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 products 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  $\mu$  and  $\sigma$  of the lognormal distribution together with the degrees of freedom after pooling. For BEAN and SPERZIEBOON, a pooled  $\mu$  (= -1.67) and  $\sigma$  (= 1.31) are used. CHICORY, ENDIVE, CABBAGE LETTUCE and CURLY LETTUCE only sigma is pooled ( $\sigma$  = 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			Automatically pooling		
			Sigma	Mean	Nobs	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/semiparametric (Nusser)

In this example output is shown for a chronic risk assessment for antibioticum lasalocide. The example concerns the intake via eggs. Table 42 lists the main options:

Input form	
risk type	chronic
uncertainty analysis	no
concentration model	empirical
intake model	discrete/semiparametric (Nusser)
concentration data	system defaults
intake model:	
transformation	power
spline fit	yes
number of iterations to estimate	10
frequency distribution	
number of bins for discretisation	20
Output	system defaults

**Table 42: Input form options: discrete/semiparametric (Nusser)**

In Table 43 you find the main characteristics concerning this analysis taken from the logfile.

Logfile	
Number of products	1
Acute reference dose (ARfD)	*
Acceptable daily intake (ADI)	5
Number of detects	31
Number of non-detects	219
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	8630
Replace all non-detects	
Multiplicationfactor for LOR	0.5
EXPOSURE SECTION	
Chronic Risk Assessment	
Discrete/semiparametric (Nusser)	
Power transformation	
followed by spline transformation	
No. of iterations to estimate theta: 10000	
No. of binomial proportions (M): 20	

**Table 43: Information in logfile**

The Dutch national Food Consumption Survey is used, which has 2 days for 6250 consumers, so 12500 days in total. Lasalocide measurements in eggs were made in 250 cases, 31 of which had a positive value. The other 219 non-detects were in this analysis replaced by  $\frac{1}{2}LOR = \frac{1}{2}0.05 = 0.025$  mg/kg. Find in Table 44 a summary of the database:

```

Compound: LASALOCIDE

*****
SUMMARY DATABASE
Code      : product code
Product   : product label
MeanCons  : 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
%PosCons  : percentage consumption days
MeanConc  : mean concentration all samples (after processing)
MeanPosConc : mean concentration positives (after processing)
NPosConc  : number of positive concentrations
NConc     : total number of concentration measurements
%PosConc  : percentage positive concentrations
*****

FOOD CONSUMPTION DATA

Code      Product   MeanCons MeanConsDays NConsDays   NDays %PosCons
          (g)          (g)          8630    12500    (%)
11311.00 WHOLE EGG C      18.00      26.1
          69.0

COMPOUND CONCENTRATION DATA

Code      Product   MeanConc  MeanPosConc  NPosConc   NConc %PosConc
          (mg/kg)  (mg/kg)          31      250    (%)
11311.00 WHOLE EGG C      0.4825      3.8910
          12.4

```

**Table 44: Summary of the database, consumptions and compounds**

So on average everyone consumes 18 g of egg per day. This is an average of 26.1 g on 8630 consumption days and 0 g on 12500 - 8630 = 3870 non-consumption days. The average of the 31 positive lasalocide concentrations and 219 times 0.025 mg/kg is 0.4825 mg/kg.

In Table 45 some counting information about the intake calculated from the data is listed.

```

Compound: LASALOCIDE
LONG TERM EXPOSURE (USUAL INTAKE)
Summary of intake days in the data
proportion and number of consumers with positive intake on:
          0 days  0.1346  841
          1 days  0.3501  2188
          2 days  0.5154  3221

```

	All intakes	Positive intakes	%
Number of observations	12500	8630	69.0
Number of consumers	6250	5409	86.5

**Table 45: Summary of intake days**

Find in Table 45 some numerical statistics.

```

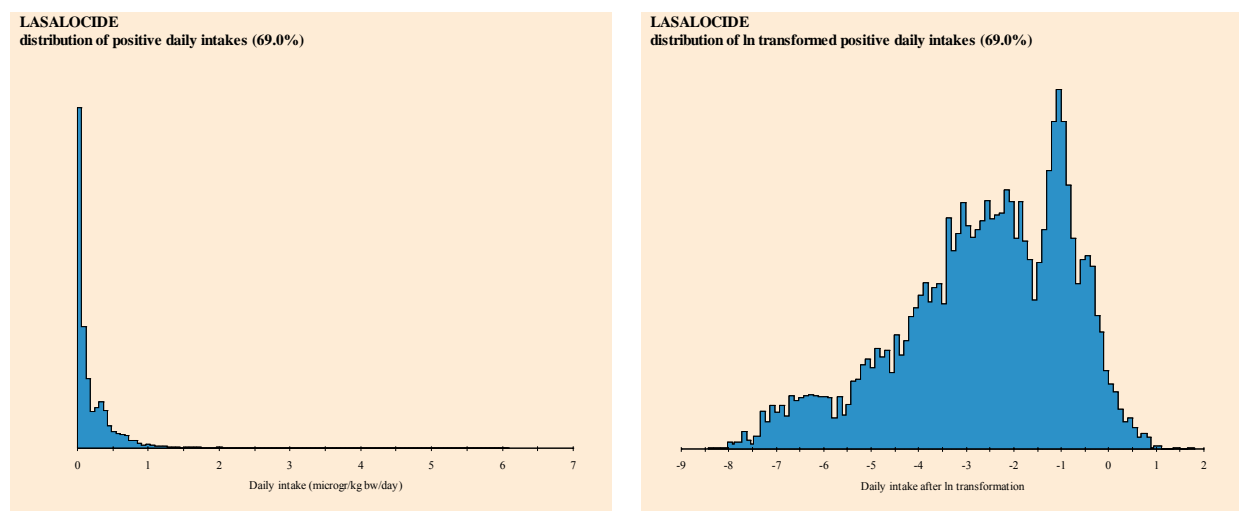
Compound: LASALOCIDE
LONG TERM EXPOSURE (USUAL INTAKE)
Summary of all intake data (zeros included)
  Number of observations = 12500
      Mean = 0.154
      Median = 0.031
      Minimum = 0.000
      Maximum = 6.368
  Lower quartile = 0.000
  Upper quartile = 0.181

Summary of positive intake data (without zeros)
  Number of observations = 8630
      Mean = 0.223
      Median = 0.095
      Minimum = 0.000
      Maximum = 6.368
  Lower quartile = 0.026
  Upper quartile = 0.322

```

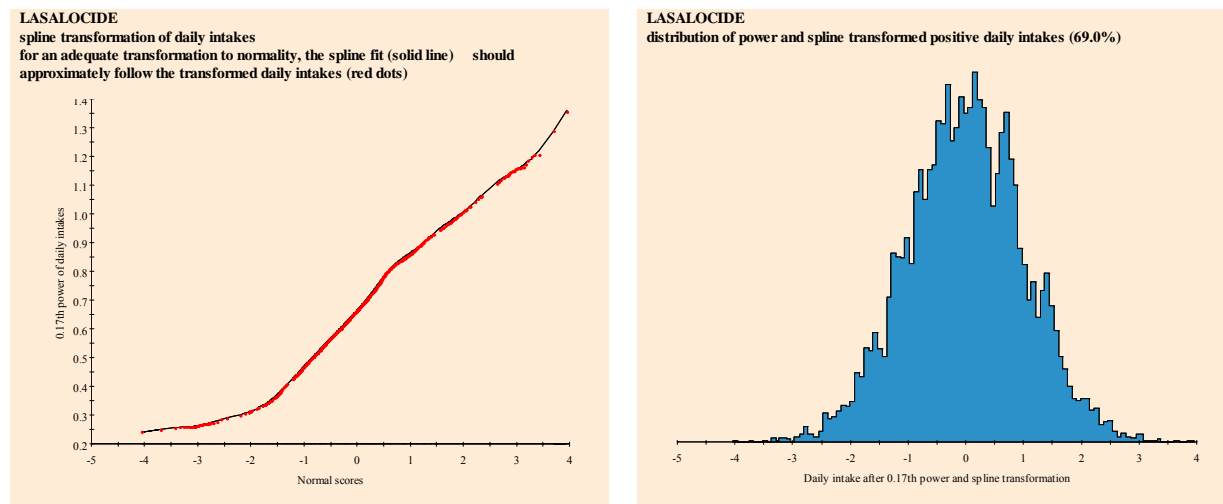
**Table 46: Summary statistics of the calculated daily intakes**

In the left plot of Figure 84 the empirical distribution of the non-zero daily intake values, calculated as daily consumption values times average concentration (0.4825 in this case) divided by body weight is shown. Note the spike near 0 (note that true zero intakes are already excluded here), and the bimodal character of the rest of the distribution (possibly related to the consumption of one or two whole eggs?). This distribution is definitely non-normal. In this case a power transformation and a spline transformation were requested for the non-zero daily intakes. The right plot shows the intake distribution after a power transformation. The optimal power was 0.169. There is a better symmetry, but clearly the bi- or even trimodal character of the data can not be removed by simple power or logarithmic transforms.



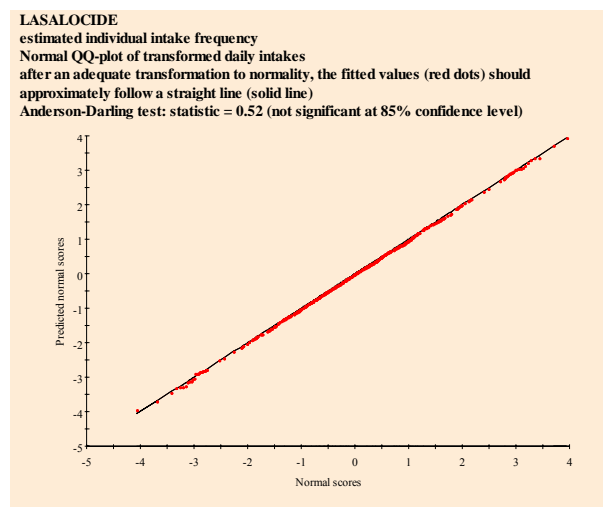
**Figure 84: Untransformed and power transformed intake distribution**

Therefore, subsequently a spline function is fitted to the power transformed values as a function of the normal Blom scores, see the left plot of Figure 85. The result of a power and spline transformation of non-zero daily intakes is shown in the right plot.



**Figure 85: Transformation plot and power and spline transformed intakes**

The fit seems adequate, and indeed this is confirmed by the QQ plot, which shows nicely a straight line and the Anderson-Darling test on the transformed non-zero daily intakes, see



**Figure 86: Normal probability plot**

In Table 47 you find details on the spline functions (16 knots) and the Anderson-Darling test result (0.52), which shows no significant deviation from normality for the transformed values ( $p$ -value = 0.25). In the same output section you also find a test whether there is evidence for heterogeneity of variance between consumers,  $MA4=3.2620$ ,  $p$ -value=0.2523.

Based on the daily intake the variance components are estimated, reported as within consumers is 0.84 and between consumers is 0.16. Therefore in this case the day-to-day variation in lasalocide intake is much higher than the variation between individual consumers.

Compound:	LASALOCIDE
LONG TERM EXPOSURE (USUAL INTAKE)	
Transformation to normality	
Power transformation	: 0.169
Number of knots spline function	: 16

```

Anderson-Darling test for normality
Test statistic      :    0.52
p-value            :    0.25

Heterogeneity of variance between consumers
Test statistic (MA4) :    3.2620    (3 for homogeneous variances)
p-value            :    0.2523

Variability within and between consumers
Variance components (transformed scale)
  within consumers (average) :    0.84
  between consumers          :    0.16
  ratio                      :    5.26

Intake frequency
Estimated fraction non-consumers (theta_0): 0.0000

```

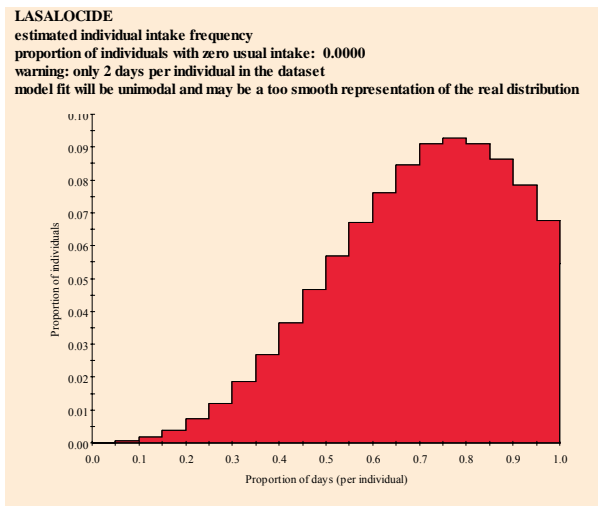
**Table 47: Technical information on usual intake**

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 0.000 in this case. In output file `theta.txt` (which you get after downloading results) you find the full results, see Table 48.

p	theta
0.00	0.000000000 (= theta0)
0.05	0.00020538
0.10	0.00102690
0.15	0.00287533
0.20	0.00636681
0.25	0.01191210
0.30	0.01971657
0.35	0.02854796
0.40	0.03881701
0.45	0.04867529
0.50	0.05812282
0.55	0.06674882
0.60	0.07475868
0.65	0.08256315
0.70	0.08892997
0.75	0.09139454
0.80	0.08934073
0.85	0.08605463
0.90	0.07968782
0.95	0.06880263
1.00	0.05545287

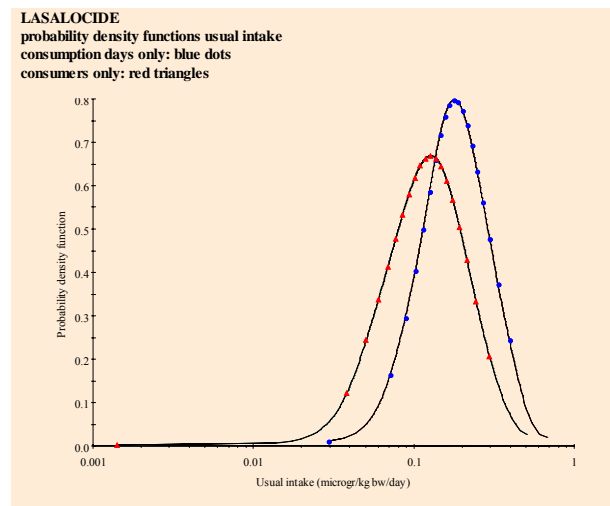
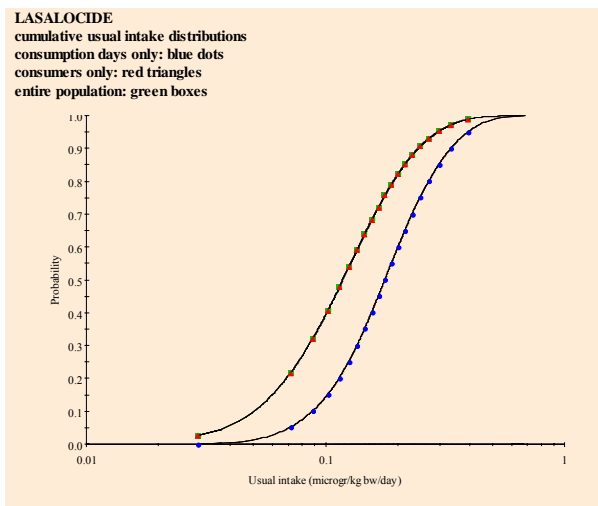
**Table 48: Estimates of the intake frequency distribution, theta**

Graphical results are shown in Figure 87. In this case the estimate is that a consumer never eats an egg ( $\theta_0$  is estimated as 0): the occurrence of non-consumers in the data set (with only 2 days per consumer) is easily explained as belonging to the classes of people who eat eggs in a low percentage (e.g. 10 % - 50 %) of days.



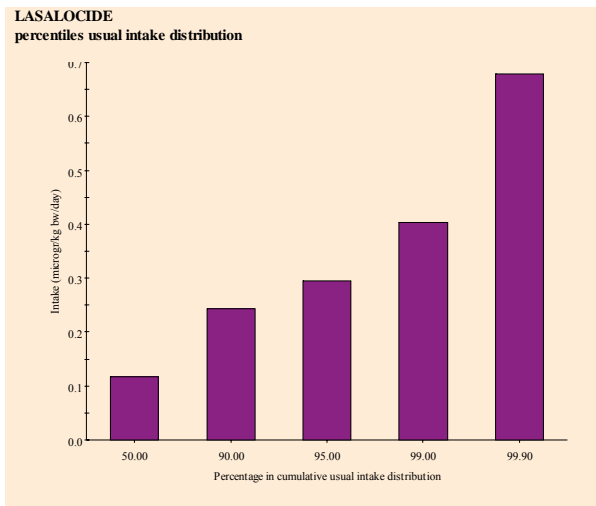
**Figure 87: Intake frequency distribution**

Finally three cumulative distributions are estimated: for consumption days only, for consumers only, and for the entire population. In this case, due to the estimate  $\theta_0 = 0.000$ , the two latter distributions coincide (there are no non-consumers), see the left plot of Figure 88. The same three distributions are also shown as probability densities in the right plot.



**Figure 88: Cumulative usual intake and density functions**

Finally, for the entire population the percentiles are calculated and shown (see Figure 89) graphically as:



Compound: LASALOCIDE

LONG TERM EXPOSURE (USUAL INTAKE)

Percentage	Entire population
50.00	0.12341
90.00	0.25536
95.00	0.30748
99.00	0.42032
99.90	0.56824
99.99	0.71030

**Figure 89: Percentiles usual intake distribution: entire population**

In Table 49 you find information on selected percentiles for the other types of distributions.

Compound: LASALOCIDE			
LONG TERM EXPOSURE (USUAL INTAKE)			
Percentage	Positives only	Consumers only	
50.00	0.18474	0.12341	
90.00	0.34882	0.25536	
95.00	0.41085	0.30748	
99.00	0.53754	0.42032	
99.90	0.71030	0.56824	
99.99	0.92162	0.71030	

**Table 49: Technical information on usual intake: percentiles for usual intake distribution, positives only and consumers only**

## 10.9 Chronic risk assessment: betabinomial/lognormal with age

In this example, output is shown for a chronic risk assessment for antibioticum lasalocide. The example concerns the intake via eggs. Table 50 lists the main options:

Input form	
risk type	chronic
uncertainty analysis	no
concentration model	empirical
intake model	betabinomial/lognormal with age
concentration data	system defaults
intake model:	
age effect for intake frequency	yes
model	
function to model effect	spline
testing method	backward
minimum degrees of freedom	0
maximum degrees of freedom	4
testing at level	0.01
age effect for intake amount	yes
model	
function to model effect	spline
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/lognormal with age**

In Table 51 you find the main characteristics concerning this analysis taken from the logfile. The intake frequency function is estimated with a betabinomial model using a spline function with 4 degrees of freedom to model age effects and smoothing parameter phi. Backward selection is applied meaning that model selection is started with a spline of the highest degree. The model for the logarithmically transformed intake amounts is based on REML using a spline function to model age effects.

Logfile							
Number of products			1				
Acute reference dose (ARfD)			*				
Acceptable daily intake (ADI)			5				
Number of detects			31				
Number of non-detects			219				
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			8630				
Replace all non-detects							
Multiplicationfactor for LOR			0.5				
EXPOSURE SECTION							
Chronic Risk Assessment							
Betabinomial/Lognormal model							
Intake frequency model is based on Beta Binomial model.							
Include effect of covariable (age).							
Function of covariable	:		spline				
Minimum degrees of freedom	:		0				
Maximum degrees of freedom	:		4				
DF selection	:		backward				
Testing at	:		0.01				
Model for intake amounts is based on REML							
Include effect of covariable (age)							
Function of covariable	:		spline				
Intake Frequency: Betabinomial model with age							
=====							
* backward selection of degrees of freedom for spline model with Prob=0.01.							
Df	Ncycle	Phi	_2Loglik	ResDf	Chi	ChiDf	ChiProb
4	3	0.1776	12175.564	6244	6.31	1	0.012
3	3	0.1781	12181.874	6245	8.54	1	0.003
2	3	0.1787	12190.414	6246	-	-	-
* Degrees of freedom of spline according to backward deviance testing is 3.							
Regression analysis							
=====							
Estimates of parameters							
-----							
Parameter		estimate	s.e.	t(*)			
Constant		0.6602	0.0420	15.72			
age Lin		0.00386	0.00103	3.75			
Estimate of overdispersion parameter phi = 0.1781 with s.e. = 0.01298							



```

Intake Amounts: REML model with age
=====
* Automatic selection of spline model.
Estimates of parameters
-----
Parameter                estimate          s.e.           t(*)
Constant                 -2.583          0.02144       -120.48
Age                     -0.2497        0.03105        -8.04
Spline(age) variance     0.2859
Between consumer variance 0.3884         0.05505        7.06
Within consumer variance 2.727          0.06468       42.17

```

**Table 51: Information in logfile**

The Dutch national Food Consumption Survey is used, which has 2 days for 6250 consumers, so 12500 days in total. Lasalocide measurements in eggs were made in 250 cases, 31 of which had a positive value. The other 219 non-detects were in this analysis replaced by  $\frac{1}{2}LOR = \frac{1}{2}0.05 = 0.025$  mg/kg. Find in Table 52 a summary of the database.

```

Compound: LASALOCIDE

*****
SUMMARY DATABASE
Code       : product code
Product    : product label
MeanCons   : 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
%PosCons   : percentage consumption days
MeanConc   : mean concentration all samples (after processing)
MeanPosConc : mean concentration positives (after processing)
NPosConc   : number of positive concentrations
NConc     : total number of concentration measurements
%PosConc   : percentage positive concentrations
*****

FOOD CONSUMPTION DATA

Code      Product    MeanCons  MeanConsDays  NConsDays  NDays  %PosCons
          (g)          (g)
11311.00  WHOLE EGG C    18.00     26.1          8630       12500  69.0

COMPOUND CONCENTRATION DATA

Code      Product    MeanConc  MeanPosConc  NPosConc  NConc  %PosConc
          (mg/kg)   (mg/kg)
11311.00  WHOLE EGG C    0.4825    3.8910       31         250    12.4

```

**Table 52: Summary of the database, consumptions and compounds**

So on average everyone consumes 18 g of egg per day. This is an average of 26.1 g on 8630 consumption days and 0 g on 12500 - 8630 = 3870 non-consumption days. The average of the 31 positive lasalocide concentrations and 219 times 0.025 mg/kg is 0.4825 mg/kg.

In Table 53 some counting information about the intake calculated from the data is listed.

```

Compound: LASALOCIDE

LONG TERM EXPOSURE (USUAL INTAKE)

Summary of intake days in the data

```

proportion and number of consumers with positive intake on:			
	0 days	0.1346	841
	1 days	0.3501	2188
	2 days	0.5154	3221
	All intakes	Positive intakes	%
Number of observations	12500	8630	69.0
Number of consumers	6250	5409	86.5

**Table 53: Summary of intake days**

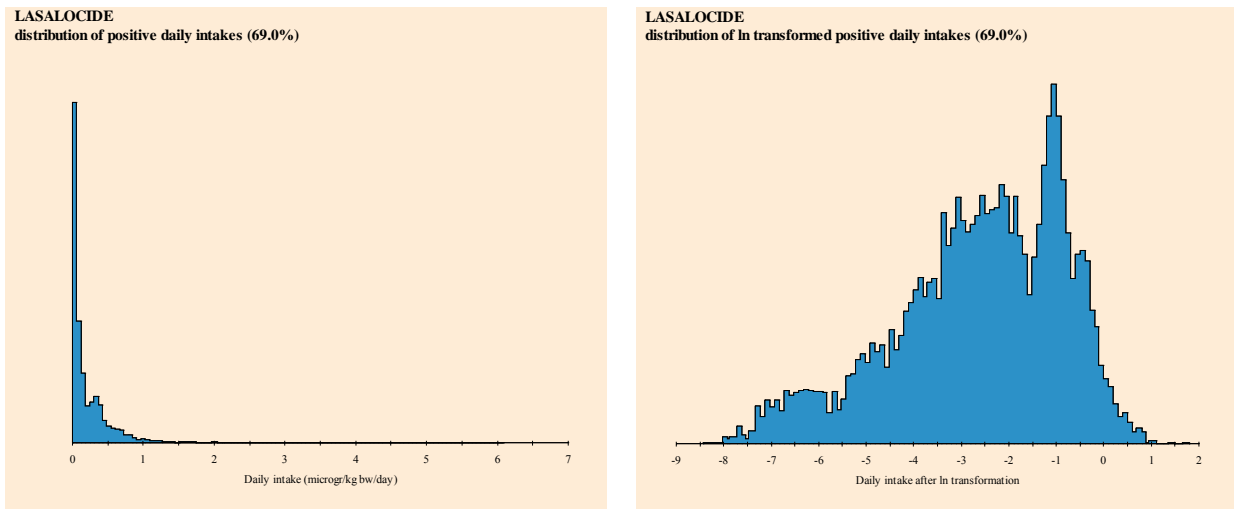
Find in Table 54 some numerical statistics.

Compound: LASALOCIDE	
LONG TERM EXPOSURE (USUAL INTAKE)	
Summary of all intake data (zeros included)	
Number of observations =	12500
Mean =	0.154
Median =	0.031
Minimum =	0.000
Maximum =	6.368
Lower quartile =	0.000
Upper quartile =	0.181
Summary of positive intake data (without zeros)	
Number of observations =	8630
Mean =	0.223
Median =	0.095
Minimum =	0.000
Maximum =	6.368
Lower quartile =	0.026
Upper quartile =	0.322

**Table 54: Summary statistics of the calculated daily intakes**

In the left plot of Figure 90 the empirical distribution of the non-zero daily intake values, calculated as daily consumption values times average concentration (0.4825 in this case) divided by body weight is shown. Note the spike near 0 (note that true zero intakes are already excluded here), and the bimodal character of the rest of the distribution (possibly related to the consumption of one or two whole eggs?). This distribution is definitely non-normal. A logarithmic transformation is requested for the non-zero daily intakes. The right plot shows the intake distribution after a ln transformation.

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 84.



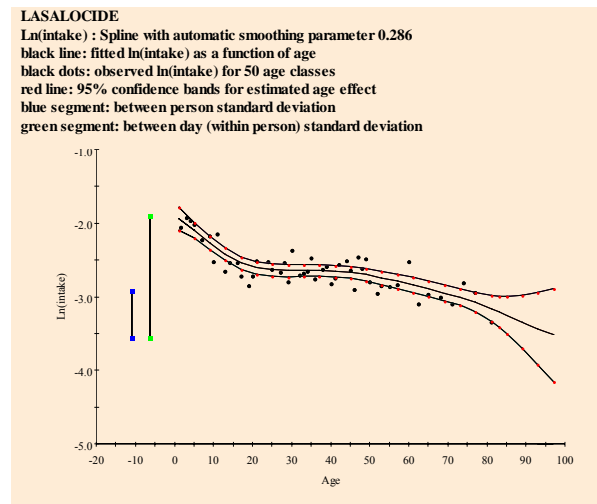
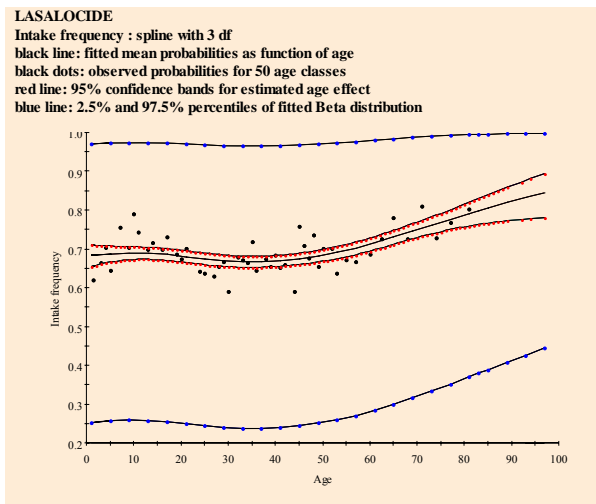
**Figure 90: Untransformed and ln transformed intake distribution**

As seen in Table 51, an age effect is found for the intake frequency. A spline with 3 degrees of freedom is fitted and overdispersion parameter  $\phi$  is equal to 0.1781. 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  $\phi$ .

In the left plot of Figure 91 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 increases 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 the effect 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 0.25 till almost 1 for young people and about 0.4 till 1 for old people.

The  $\ln(\text{intake})$  amounts are modelled using REML with a spline function to model age effects. The degree of the spline is automatically selected. A negative estimate for age is found (-0.2497) meaning that the  $\ln(\text{intake})$  amount of lasalocide decreases with age. The components of variance for the between consumers and within consumers variation are 0.3884 and 2.727, respectively.

The right plot of Figure 91 displays the  $\ln(\text{intake})$  distribution. The age effect is plotted as a black line through the observed mean  $\ln(\text{intake})$  amounts per age class. The 95% confidence interval is indicated by a red dotted line. The standard deviation of the between and within consumer variation are indicated by the vertical blue and green line segments.



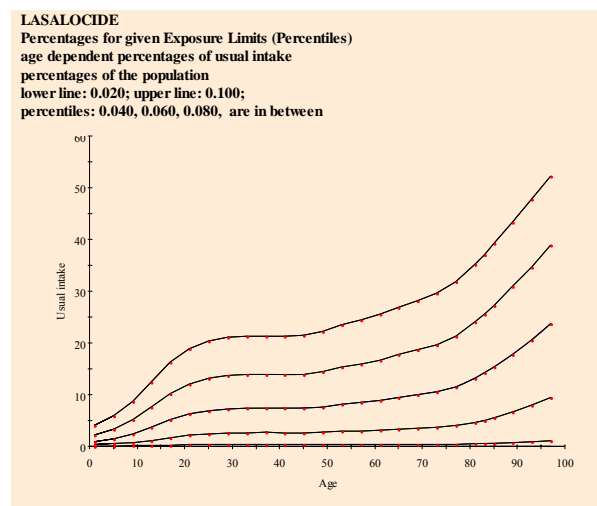
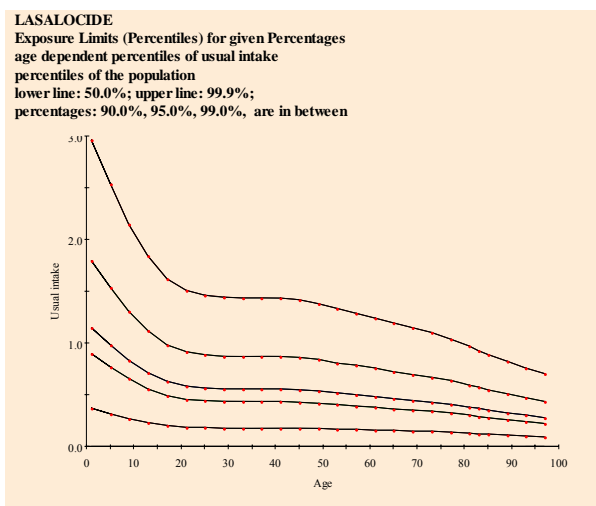
**Figure 91: Age dependent intake frequency and ln(intake) distribution**

In Table 38, a report of technical information on the analysis is shown.

<p>Compound: LASALOCIDE</p> <p>LONG TERM EXPOSURE, AGE DEPENDENT PERCENTILES (USUAL INTAKE)</p> <p>Intake frequency model is based on Non Consumer model.</p> <p>Model for consumption (intake) amounts is based on REML</p> <p>Function of covariable is spline</p> <p>Variance components (transformed scale)</p> <p>spline is 0.09</p> <p>within consumers is 0.88</p> <p>between consumers is 0.12</p> <p>ratio is 7.02</p>
---

**Table 55: Technical information on usual intake**

Figure 92 shows the age dependent percentiles and percentage for the usual intake distribution.



**Figure 92: Age dependent percentiles derived from percentages and percentages derived from percentiles**

## 10.10 Chronic risk assessment: betabinomial/lognormal with age and uncertainty

In this example, output is shown for a chronic risk assessment for antibioticum lasalocide. The example concerns the intake via eggs. Table 56 lists the main options:

<b>Input form</b>	
risk type	chronic
uncertainty analysis	yes
number of bootstrap samples	1000
random seed	0
concentration model	empirical
intake model	betabinomial/lognormal with age
concentration data	system defaults
intake model:	
age effect for intake frequency model	yes
function to model effect	spline
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
function to model effect	spline
testing method	backward
minimum degrees of freedom	0
maximum degrees of freedom	4
testing at level	0.01
Output	system defaults

**Table 56: Input form options: betabinomial/lognormal with age and uncertainty**

In Table 57 you find the main characteristics concerning this analysis taken from the logfile. The intake frequency function is estimated with a beta binomial model using a spline function with 4 degrees of freedom to model the age effect and smoothing parameter phi. Backward selection is applied meaning that model selection is started with a spline of the highest degree. The model for the intake amounts is based on REML, also using a spline function to model the age effect.

<b>Logfile</b>	
Number of products	1
Acute reference dose (ARfD)	*
Acceptable daily intake (ADI)	5
Number of detects	31
Number of non-detects	219
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	8630
Replace all non-detects	
Multiplicationfactor for LOR	0.5
EXPOSURE SECTION	
Chronic Risk Assessment	
Betabinomial/Lognormal model	

```

Intake frequency model is based on Beta Binomial model.
  Include effect of covariable (age).
  Function of covariable      : spline
  Minimum degrees of freedom  : 0
  Maximum degrees of freedom  : 4
  DF selection                 : backward
  Testing at                   : 0.01
Model for intake amounts is based on REML
  Include effect of covariable (age)
  Function of covariable      : spline
Intake Frequency: Betabinomial model with age
=====
* backward selection of degrees of freedom for spline model with Prob=0.01.
  Df  Ncycle   Phi      _2Loglik    ResDf      Chi      ChiDf    ChiProb
  4    3  0.1776   12175.564    6244      6.31      1      0.012
  3    3  0.1781   12181.874    6245      8.54      1      0.003
  2    3  0.1787   12190.414    6246      -         -         -
* Degrees of freedom of spline according to backward deviance testing is 3.

Regression analysis
=====
Estimates of parameters
-----
Parameter                estimate      s.e.      t(*)
Constant                  0.6602      0.0420    15.72
age Lin                   0.00386     0.00103   3.75

Estimate of overdispersion parameter phi = 0.1781 with s.e. = 0.01298

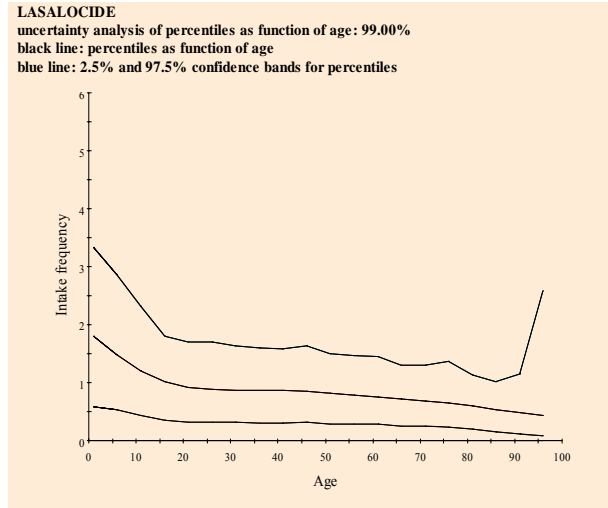
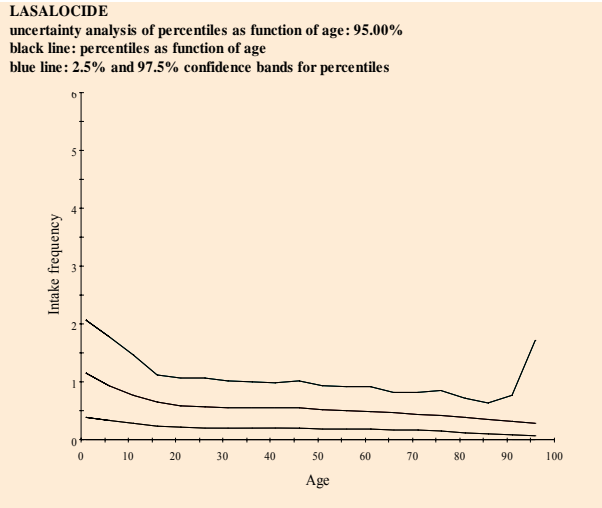
Intake Amounts: REML model with age
=====
* Automatic selection of spline model.
Estimates of parameters
-----
Parameter                estimate      s.e.      t(*)
Constant                 -2.583      0.02144  -120.48
Age                      -0.2497     0.03105   -8.04
Spline(age) variance     0.2859

Between consumer variance 0.3884      0.05505    7.06
Within consumer variance  2.727       0.06468   42.17

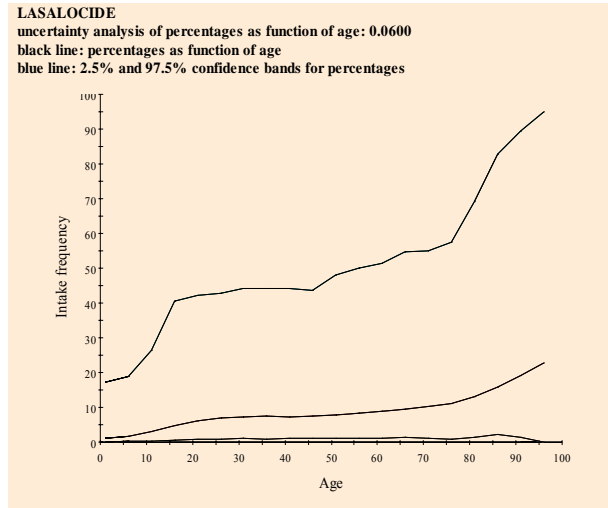
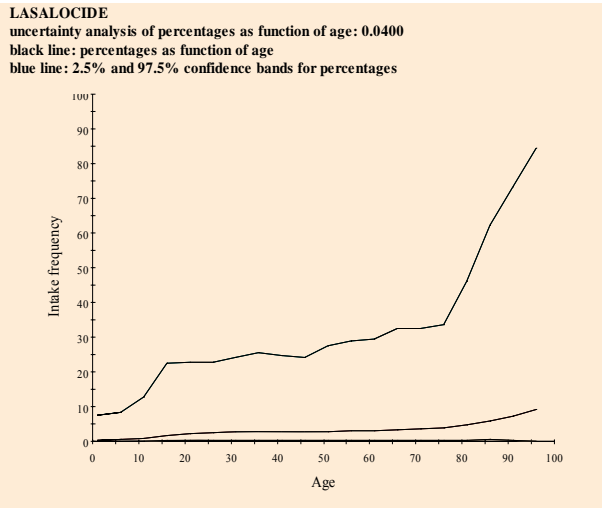
```

**Table 57: Information in logfile**

Find in Figure 93 and Figure 94 the percentiles and percentages for the 95% and 99% points and 0.04 and 0.06 percentiles of the uncertainty distribution. The blue dotted lines indicate the 2.5% and 97.5% confidence bands. Information on percentiles and percentages is also available as tabular output (not shown).



**Figure 93: Age dependent percentiles derived from percentages**



**Figure 94: Age dependent percentages derived from percentiles**

## 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 columnnames (see 9.1 ), and combinations of levels during subset selection that lead to empty subsets.

In general, the internet explorer error message is:



The page cannot be displayed

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 3).

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 3).