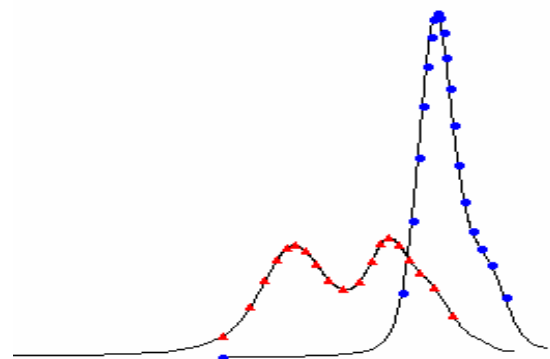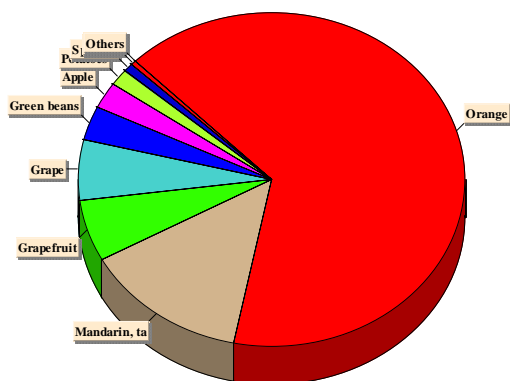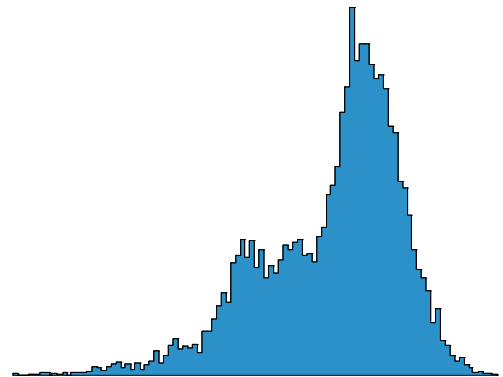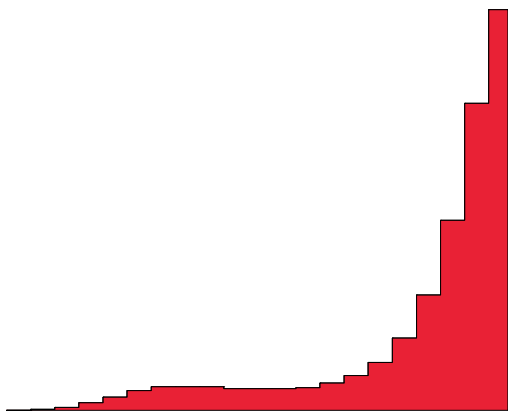# MCRA, Release 5
a web-based program for Monte Carlo Risk Assessment

Waldo J. de Boer
Hilko van der Voet

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

Biometris 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

1 Foreword

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

The methods for probabilistic modelling described here are implemented in the program Monte Carlo Risk Assessment (MCRA). MCRA is a computational tool for dietary risk assessment. MCRA can calculate intake distributions for both short-term (acute) and long-term (chronic) intakes. Basically, it simulates daily consumptions by sampling a food consumption database and combines these with a random sample from either a compound database (empirical distribution) or a parametric distribution of compound concentrations. The result is a full *distribution* of intakes, rather than traditional deterministic methods which only provide a point estimate. Percentiles of the intake distribution can be used to assess risks by relating them to e.g. an acute reference dose (ARfD). In a chronic risk assessment, MCRA calculates the distribution of the usual intakes 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 concentration below analytical reporting limits (LOR) can be assessed. Large portion consumption and the highest compound or median compound in case of bulking or blending in the composite sample is used in IESTI 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 5. It covers the current release 5.0 (release 5 version 0) and all future updates starting with the same release number. Major updates of the program, encompassing new or improved facilities will be released with an increased release number and a new manual.

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

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

2.2 Registration

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

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

Figure 1: Registration form

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

2.3 MCRA Main menu

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

Figure 2: MCRA Main Menu

- Click 'MCRA 5.0' to start preparations needed to do an MCRA analysis.
- Click 'Manage input/output' to upload (and download) your own data (see 2.4) or to download output from a former MCRA analysis.
- Click 'Older versions' to run old releases of the program. However, older versions are not supported anymore.
- Click 'MCRA 5.0 (field trial data and Dutch consumption data)' to run MCRA with features to edit your own compound concentration data.
- Click 'Registered user information' to view which information about you is stored in the user database.
- Click 'Help' for explanation about options in the Main menu.
- Click 'Logout from MCRA website' to leave the MCRA website. Your personal data files, latest output files and the latest input options remain stored for later use.

2.4 Data needed

What data are needed to run MCRA? All data for MCRA are stored in Microsoft Access database tables according to a fixed format. For use of own data, you need to prepare your database off-line and upload it to your personal user area on the MCRA website. The 'MCRA 5.0 (field trial data and Dutch consumption data)' option in the Main menu (see 2.3) offers some possibilities to edit data on-line. However, on-line editing is restricted to compound concentration values, unit variability, processing factors, acute reference dose (ARfD) and average daily intake (ADI) (see Ch. 7).

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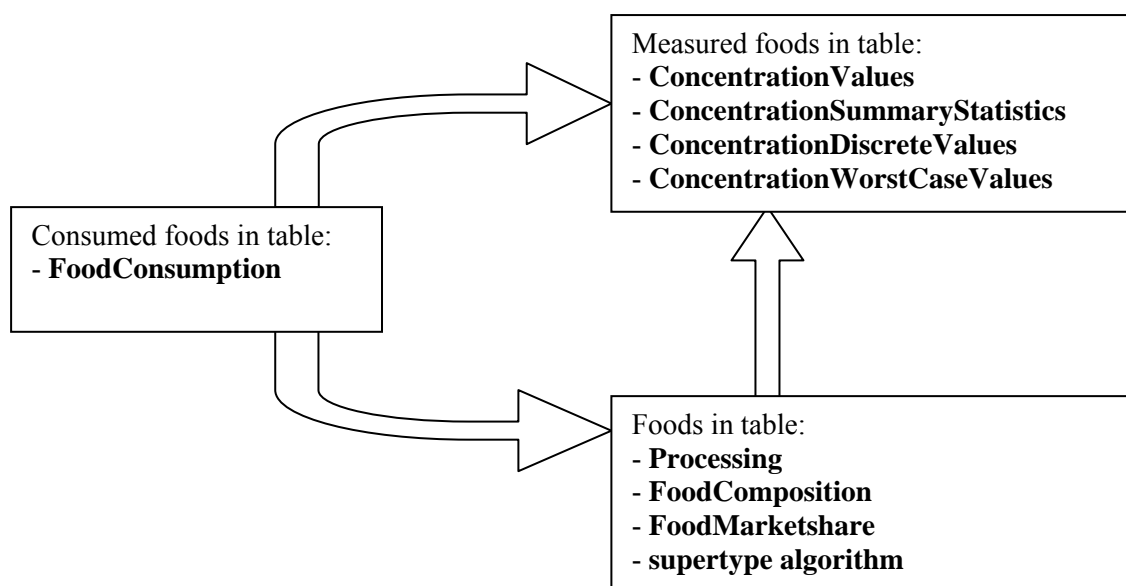
Figure 3: Links between consumption and compound concentration data

The MCRA system has a central database with example data. You may run example analyses without having data yourself. However, MCRA is primarily designed to work with user databases, or with a mixture of user data and centrally supplied data. For example, provide your own data on compound concentrations and combine these with the centrally supplied consumption survey data. Be careful when using tables from different databases: the codes of foods of the centrally supplied data and your own data should be consistent with each other.

Consumption data are portions of food (consumed at different days) of consumers of whom the age, weight and sex is known. The second type of data, compound concentration data, are the amounts of compound found on monitoring samples of food. The third category, additional tables, provides information that links consumption data to compound concentration data or store information for more sophisticated analyses like unit variability (see Figure 3).

Food composition data specifies the composition of foods. So, speaking about pizza, the composition specifies proportions for e.g. wheat, tomato, cheese etc. Food marketshare data specifies the proportion of subtypes, so for apple, marketshares are e.g. Jonagold, Granny Smith, Golden Delicious etc. Processing data specify the unprocessed food, the processed food and the corresponding processing factors, e.g. for grapes raisins are specified. The supertype of a food is, if needed, automatically determined. So the supertype of e.g. Granny Smith is apple.

As a registered MCRA user you have complete control over the file management in your personal area by starting 'Manage input/output' in the Main menu (see 2.5).

2.5 Manage input/output

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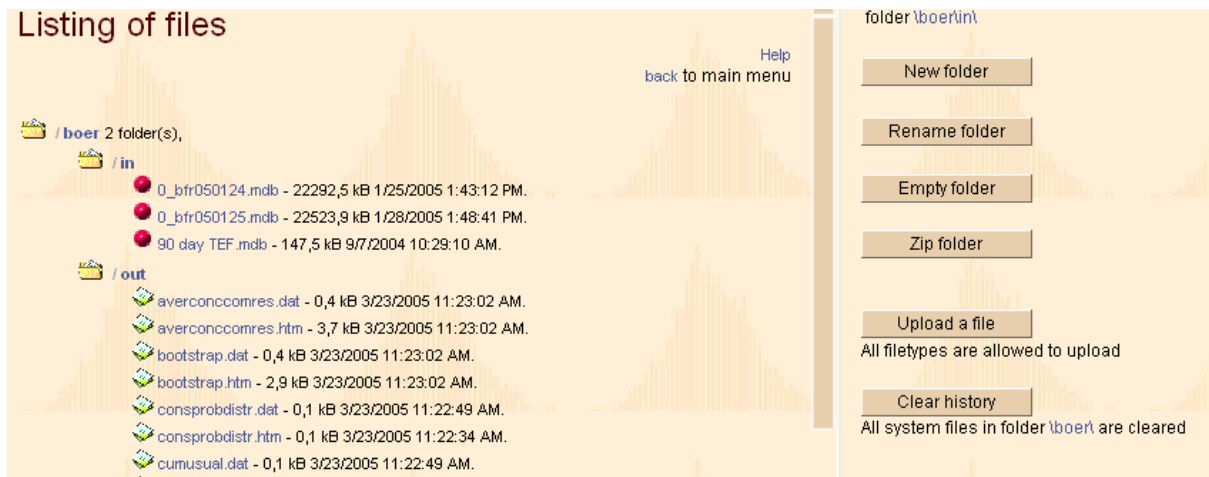
Figure 4: Manage input/output

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

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

Occasionally, system faults occur due to errors like incorrect database contents, queries giving empty subsets, subset selections combining inconsistent levels or scripting errors. The best way to proceed is to log out and enter the website again (login). Then, click the 'Manage input/output' link in the main menu, click your personal directory link or the 'IN' or 'OUT'-subdirectory links (a number of buttons appear) and click the 'Clear history'-button: all system files (files created by MCRA, but not visible) are deleted from the personal directory. Files on the 'IN' or 'OUT'-subdirectory are not cleared.
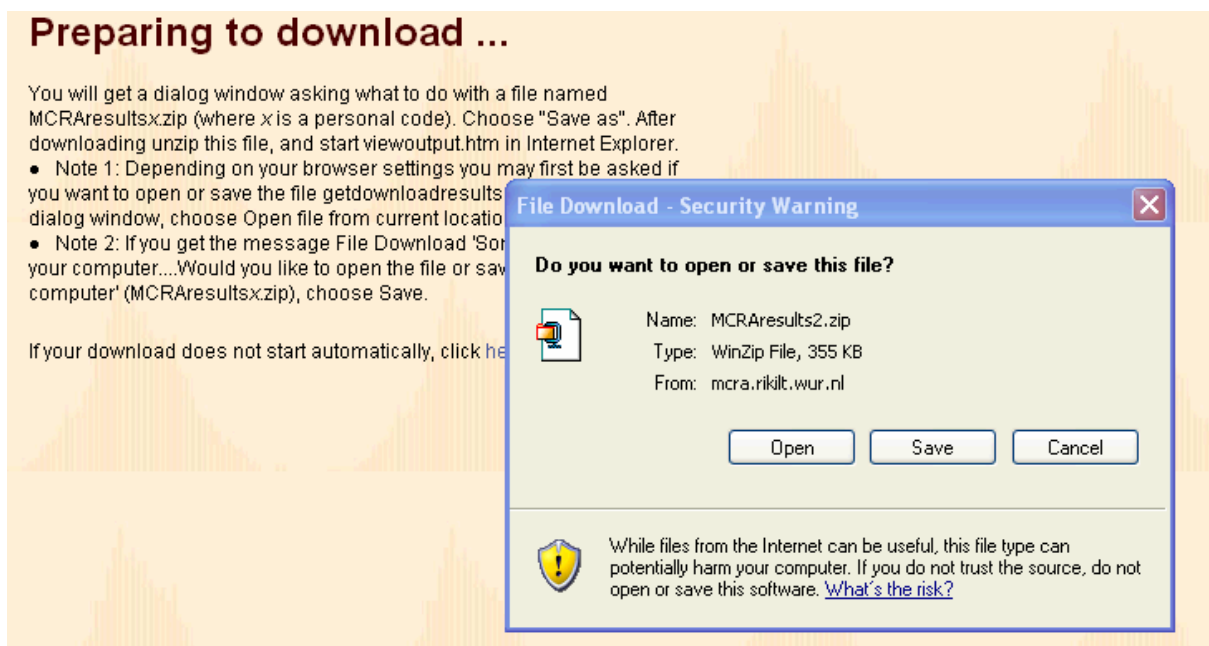
Figure 5: Download output for off-line viewing

2.6 Starting the MCRA program

To start MCRA, click the 'MCRA 5.0' option in the Main menu (see Figure 2). Then, depending on whether you are a new user or not the screen in Figure 6 or Figure 7 is shown. Each activity is started from the MCRA central menu window. After finishing the activity, the user returns to the MCRA central menu to start a new activity.

New users are automatically brought to the MCRA central menu in Figure 6. Click the 'go'-button to start selection of data.
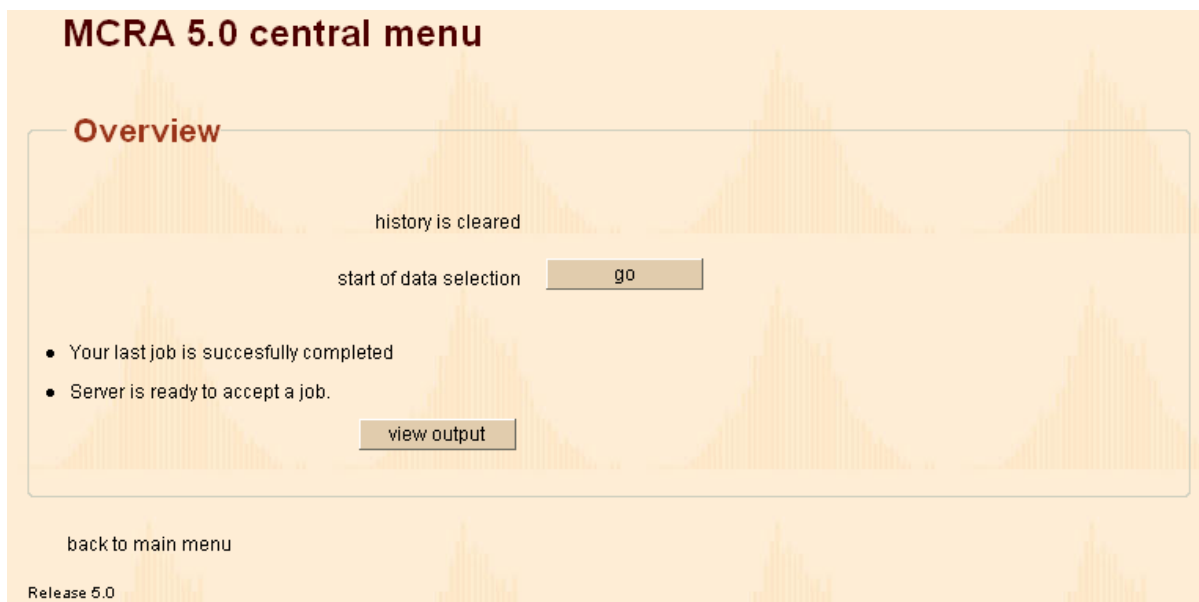
Figure 6: MCRA central menu, start selection of tables, survey and compound

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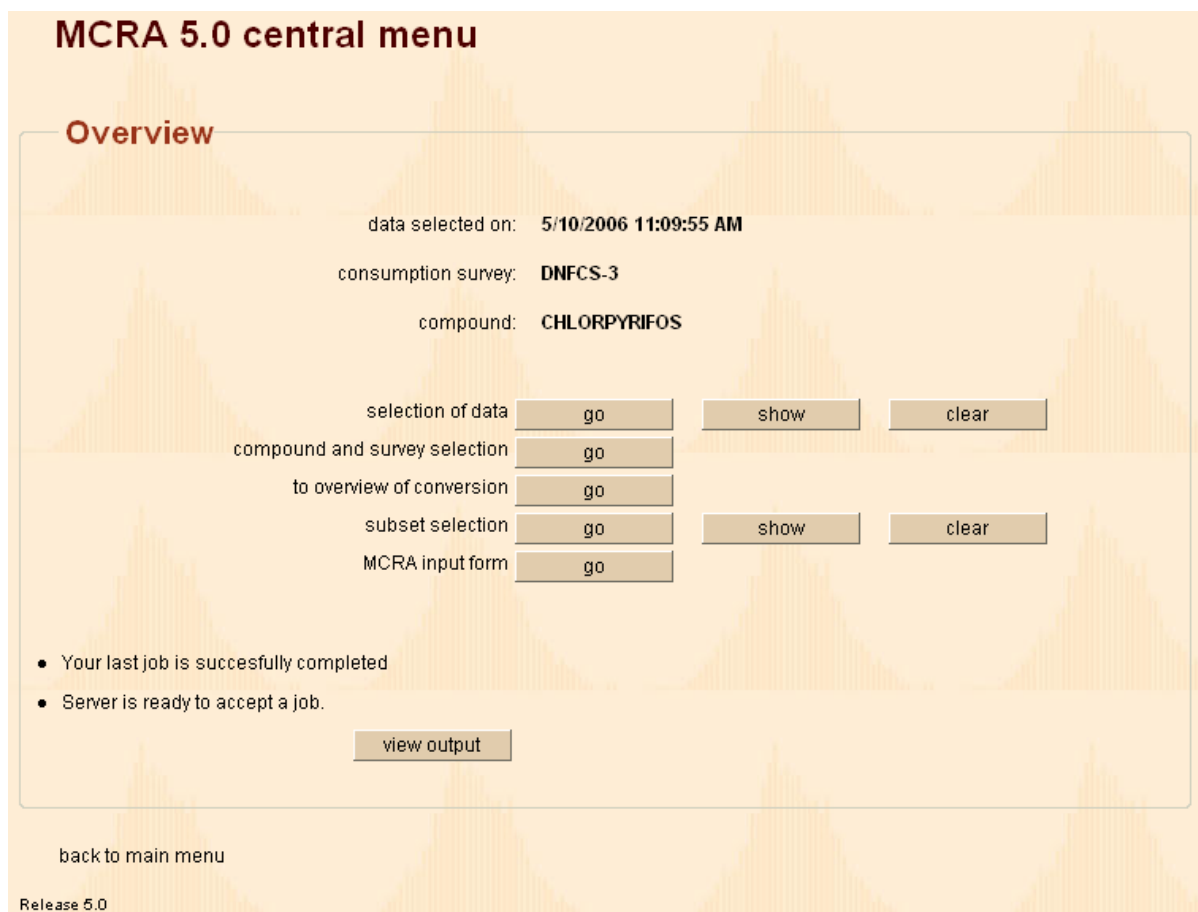
Figure 7: MCRA central menu

First, start selecting the data. In Figure 6, this is the only possibility. When a selection already exist, continue with the same selections as before by clicking the MCRA-input-form 'go'-button (see Figure 7). For selecting new data, click the selection-of-data 'clear'-button and enter the screen in Figure 6. Click the to-overview-of-conversion 'go'-button, to enter an overview of the latest conversion made.

2.6.1 Selection of the tables

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

Figure 8: Selection of data servers

In Figure 9 all databases that are available to you are shown. From here select databases for further use in the MCRA analysis. Click the compound-and-survey-selection 'go'-button for selection of a new compound or survey if data are retrieved on a earlier occasion.

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

On each of these sources, find some information by clicking the 'info'-button and, next screen (not shown), by clicking the buttons with country names.
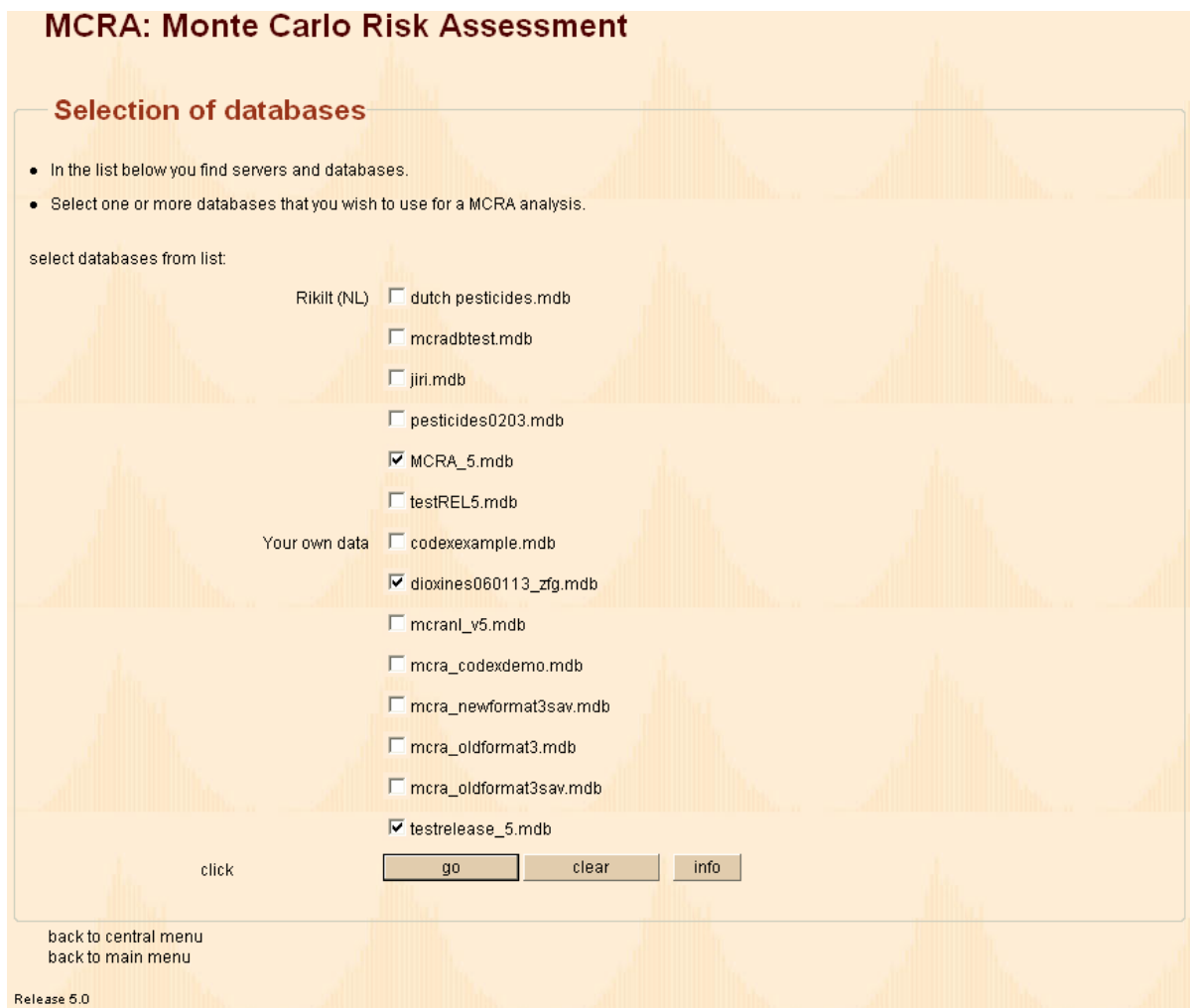
Figure 9: Selection of databases

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

By scrolling the mouse over the checkboxes, tablenames are displayed in red, green or grey textboxes.

- It is compulsory to select all tables for which red textboxes appear (Foodconsumption, Foods, Individual, Compounds and Country).
- It is also compulsory to make a choice between one of the tables containing compound concentration data (ConcentrationValues, ConcentrationSummaryStatistics and ConcentrationDiscreteValues). These tables have their name displayed in green textboxes (see also 3.2.1).
- All tables with grey textboxes are optional (Foodproperties, Processing, Processingtype, ConcentrationWorstCaseValues, AgriculturalUse, VariabilityProd, VariabilityCompProd, VariabilityProcCompProd).

Make sure to select all compulsory tables otherwise the menu keeps returning with warnings about missing tables. Make also sure that in the selection, each type of table is represented once.

To use an alternative language for labeling foods, check 'Use alternative foodnames (foodname2))'.

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 con sump tion	Food	Indivi dual	Com pound	Country	Concen tra tion values	Conc. Sum mary stat.	Conc. Dis crete values	Food com posi tion	Food Market share	Food proper ties	Agri cultural use	Proces sing type	Proces sing	Varia bility Prod	Varia bility Comp Prod	Varia bility ProcComp Prod	Conc. worst case values	All Tables
MCRA_5.mdb	☑	☑	☑	☑	☑	☑	-	-	☐	-	☑	☑	☑	☐	☑	-	-	-	☐
dioxines060113_zfg.mdb	☐	-	☐	-	☐	-	-	-	-	-	-	-	-	☑	-	-	-	-	☐
testrelease_5.mdb	☐	☐	☐	☐	☐	☐	-	-	☑	☑	-	-	-	☐	-	-	-	☐	☐

☐ Use alternative foodnames (foodname2)

select tables	go
clear all settings	clear

back to main menu

Release 5.0

Figure 10: 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 11 is shown. Checkboxes for 'allow marketshares not summing to 100% (step 4)' and 'allow worstcase concentrations (step 7)' are only shown when table FoodMarketshare and ConcentrationWorstCaseValues are selected. Here, table ConcentrationWorstCaseValues is not selected (checkbox not shown). When checkbox 'allow marketshares not summing to 100% (step 4)' is not checked, all derived foods not summing up to 100% are ignored in the analysis.

If more than one survey is available, select a surveyname from the scroll-down menu. MCRA works on single compounds. Select a compoundname from the scroll-down menu (see Figure 11).

Figure 11: Selection of consumption survey and of compound

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

Figure 12: Error report on missing columns

In the continuation of the example, all tables of the 'MCRA_5.mdb' database are selected.

2.6.3 Selection of foods

After selecting your tables, information on the number of consumed and derived foods is displayed. For consumed foods, three situations may occur:
1. on some foods only positive concentration values are measured,
2. on some foods only non-detects are found,
3. on some foods only worstcase values are found.

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

5. on some consumed foods no information is found.

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

Figure 13: Food selection for MCRA analysis

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

More information about selection of the consumer population and foods is found in Chapter 5 .

3 Specifying MCRA options and running a job

3.1 Overview

In this chapter, all options of the input form are discussed with references to paragraphs in Chapter 5 for more theoretical background.

Click the MCRA-input-form 'go'-button of the central menu (see Figure 7). When MCRA is entered for the first time, the screen in Figure 14 (default options) is entered.

Figure 14: MCRA input form.

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

- In the top section, 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 unit variability is not relevant for chronic risks.
- 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 14 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 15, default options).

After clicking any (radio) button or any item in a scroll-down menu, the screen is rebuilt to implement the choice. After changing a value in a text box, leave the field (move the mouse/replace pointer) to implement the value.

Each description (like 'risk type', 'uncertainty analysis', 'concentration model' etc. in Figure 14) has its own mouse-over function revealing a screen-tip. By double-clicking the description name, the On Line User Manual is opened for more information.

Save your settings by clicking the save-user-defaults-button in the top section.
The scroll-down menu Choose-default-option-setting has 4 settings: Current, Previous, System and User.
- 'Current' is shown when one of the input options is changed.
- 'Previous' recalls all program settings from the last performed MCRA analysis.
- 'System' sets all default settings.
- 'User' implements all user defaults.

When no MCRA analysis has been performed in the past, setting 'Previous','User' and 'System' are identical.

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

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

3.2 Acute risk analysis

If acute risk is specified, find the specifications in the left section of the screen (see Figure 16, default options). Scroll-down menus for the concentration model (see 3.2.1), the unit variability model (see 3.2.3) and the intake model (see 3.2.5) are shown together with some numerical specifications.

Figure 16: Left section of the MCRA input form if type of analysis is Acute

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 16, the total number of MC-iterations is 100,000 divided over 10 chunks (10.000 MC-iterations per chunk). Keep in mind that the number of foods, 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, MCRA is 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 MC-terations 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 33) 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 foods and/or the total number of combinations of foods, processing types and unit variability.
The number of MC-terations 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 MC-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, specify the number of bootstrap samples. In Figure 16 the number is 100.

Number of iterations per bootstrap sample
If an uncertainty analysis is chosen, specify the number of iterations per bootstrap sample. In Figure 16 the number of iterations per bootstrap sample equals the number of MC-iterations per chunk (is 10.000). The size of a bootstrap sample never exceeds the chunksize.

3.2.1 Concentration model options

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

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

In the probabilistic model, a distribution of compound data is used to sample from. A choice is 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 μ and σ has been selected as being both theoretically sensible and practically useful (Shimizu & Crow 1988, Van der Voet et al. 1999). The non-parametric approach requires more data to obtain a satisfying representation of the full distribution.

Summarizing, for concentration data we have:

Type of concentration data	full data	summary data	histogram data
Option settings in concentration model			
Non-parametric: empirical	x	-	-
Parametric: binomial/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

3.2.1.1 Empirical

In the non-parametric approach, choose 'empirical': concentrations are sampled at random from the available data and combined with food consumption data to generate the intake distribution of intake 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 are sampled from parametric distributions based on full, histogram or summary data. Parameters μ and σ of the lognormal distribution are estimated using the log-transformed non-zero compound concentrations (full data) or condensed data (summary or histogram data). Choose this setting only, when enough data are available to estimate μ's and σ's for all foods.

Estimation of the variance and/or mean may fail because compound measurements on specific foods are sparse or even missing. In case of missing parameters, a warning message is printed. Re-run MCRA with 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 foods are used to give sufficient data to base estimates upon. Foods are classified into groups of similar foods and missing or unreliable parameters are estimated using all concentration data in a group. This process of using concentration data on similar foods to base estimates for μ and σ upon is called pooling (see 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 foods are automatically assigned into groups and pooled. The identification of groups is based on the CODEX system, using the first 2 characters of the code. When the code contains information on supertypes e.g. indicated by the presence of symbol '$', then the supertype is used to form groups. Foods with equal first 2 characters or equal supertypes are placed into the same group.

Pooling is performed in a two step procedure following the next scheme:
1. **Test homogeneity of variances within the groups**
          if variances are homogeneous,
                    pool variances.
    test homogeneity of means within the groups
          if means are homogeneous,
                    pool means.
2. **Test homogeneity of variances of foods with df < 10 against overall-variance**
          if variances are homogeneous,
          replace variances with overall-variance.

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

An example of pooling is given in 10.7 .
See also: Parametric modelling of concentrations (binomial/lognormal with pooling)

3.2.2 Concentration data options

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

Figure 17: Option block Concentration data

3.2.2.1 Replacement of non-detects

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

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

then replacement by LOR is restricted to an appropriate fraction of the non-detects by specifying 'Replace non-detects based on crop treated'.

See also: Modelling of missing data and replacement of non-detects

3.2.2.2 Modelling processing effects

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

Concentrations in the consumed food may be different from the monitoring compound due to processing such as peeling, washing and cooking. Usually, processing lowers the concentration in the consumed food compared to the concentration in the unprocessed food. The effect of processing is modelled by multiplying the monitoring compound by a factor f_k which will typically be between 0 and 1. Occasionally, the processing factor may also be > 1, e.g. for drying. Often, processing factors 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 'Processing' ($f_{k,nom}$ = proc_nom, $f_{k,upp}$ = proc_upp) and processing codes and labels from table 'ProcessingType'. Note that specifying no processing is a worst case scenario ($f_k = f_{k,upp} = 1$).

The program multiplies concentrations with fixed processing factors (in which case the conservative value $f_{k = }f_{k,upp}$ is used), or with random values sampled from a normal distribution with parameters μ and σ. The mean and standard deviation are based on transformed values of $f_{k,upp}$ and $f_{k,nom}$. The type of transformation for each processing type is specified in the last column of table ProcessingType. Choose disttype = 1 for a logistic-normal distribution or disttype = 2 for a log-normal distribution. To process simultaneously some foods using fixed factors and others distribution based, choose 'Processing (distribution based)'. Now, fixed factors f_k are obtained by providing only $f_{k,upp}$ whereas random factors f_k are sampled when both $f_{k,upp}$ and $f_{k,nom}$ are given.

It is not necessary to fill out a complete list of processing factors for all foods. Missing values of $f_{k,nom}$ and $f_{k,upp}$ are, by default, replaced by the value 1.

See also: Modelling of processing effects

3.2.3 Unit variability model options

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

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

In MCRA, the following three models for unit variability are available:

1. **Beta model**, requires knowledge of the number of units in a composite sample, and of the variability between units (realistic or conservative estimates);
2. **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)	• no censoring at cm_k • no upper limit to the unit concentration	• no censoring at cm_k • unit values never higher than $nu_k \cdot cm_k$
Estimates of unit variability are conservative (C)	• unit values will be left-censored at cm_k • no upper limit to the unit concentration	• unit values will be left-censored at cm_k • unit values never higher than $nu_k \cdot cm_k$

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

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

3.2.3.2 Beta distribution

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

Figure 18: Option block Unit variability: Beta distribution

The parameter for unit variability is specified as a variability factor v or as a coefficient of variation cv of the unit values in the composite sample. Variability factors v (97.5[th] percentile divided by mean), coefficient of variation cv (standard deviation divided by mean) and number of units nu in the composite sample are retrieved from table VariabilityProd when unit variability is independent of the

compound and processing type. If the variability factor is dependent on compound and/or processing type data are expected in tables VariabilityCompProd or VariabilityProcCompProd, respectively.

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

The latter information is specified in field bulkingblending of table ProcessingType (see 9.4.4). If the parameter for variability is missing, zero variability is assumed, and the unit concentrations are equal to the sampled composite sample concentrations.
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 18, but can be modified by the user.
See also: Beta model for unit variability

3.2.3.3 Lognormal distribution

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

Figure 19: Option block: Unit variability: Lognormal distribution

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

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[th] 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

27

3.2.3.4 Bernoulli distribution

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

Figure 20: Option block Unit variability: Bernoulli distribution

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

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

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

3.2.4 Additional options concerning IESTI and consumption days

For the acute risk model, find in the right section of the input form, option block Additional that handles IESTI and consumption days (see Figure 21, default options).

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

The IESTI (International Estimated Short-Term Intake) is a prediction of the short-term intake of a compound on the basis of the assumptions of high daily food consumption per consumer and highest compounds and, in case of blending and bulking, the median compound from supervised trials. The IESTI is expressed in microgr/kg body weight/day and estimated per food.

IESTI estimates are requested when estimation of IESTI is set to yes. Standard unit variability factors and a standard body weight of 60 kg are specified. The IESTI is compared with estimates of a specified percentile (per foodt) of the MC-simulation.

In the output (not shown) two kinds of estimates of the MC-percentile are given: one for "All days" and one for "Consumption days only". Be aware that specification of option 'consumption days only' may alter the interpretation (and estimate) of the percentile for "All days". In the latter case the estimate refers to a smaller subset containing consumption days only. However, note that still not every food is eaten on every consumption day. The interpretation and estimate of the percentile for "Consumption days only" is not affected by setting option 'consumption days only'.

The IESTI calculations correspond to the definition of FAO/WHO (2002) that may be considered as the deterministic counterpart of the probabilistic approach used in MCRA.

In 10.3 output of IESTI calculations are shown.

See also: Additional: Comparison of probabilistic with deterministic estimates of acute risk

3.2.5 Intake model options

Settings are:
- **only empirical estimates** [default]
- **empirical estimates and betabinomial/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

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

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

Note that estimation of 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:
- an intake frequency function is estimated with a betabinomial model using a spline or polynomial function to model age effects;

- 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;
- 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 consumers of having an intake. Depending on the consumption pattern, we have regular, less regular and incidental consumers. So each consumer has his own probability of having an intake. For many foods, there may be a relation with age (or sex or, if available, other demographic data). 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 elimation 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(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(intake) distribution. 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 ln(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(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(intakes) and the estimated standard deviation of the second analysis, we now have all what is nessecary to estimate an acute intake distribution with a parametric age effect. The number of simulated age dependent intakes is set equal to the number or 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 top section (see Figure 14) chronic risk and uncertainty. Specifications in the left section (see Figure 23, default options) are:

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

In dietary risk assessment, 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. 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. Settings are the same as for the acute risk model. See 3.2

3.3.1 Concentration model options
Settings are:
- **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)

For options of Concentration data (see 3.2.2)

3.3.2 Intake model options
Settings are:
- **betabinomial/lognormal** [default]
- **betabinomial/lognormal (with age)**
- **discrete/semiparametric (Nusser)**

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
- an intake frequency function is estimated using the betabinomial model;
- non-zero intakes are log transformed, the transformed intake amounts are analysed using REML and parameters of the usual intake distribution are estimated;

- 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 consumers of having an intake. Depending on the consumption pattern, we have regular, less regular and incidental consumers. So each consumer has his own probability of having an intake. 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(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(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)

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

Figure 24: Option block Intake model if intake model is betabinomial/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

- an intake frequency function is estimated with a betabinomial model using a spline or polynomial function to model age effects;

- 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;
- 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 consumers of having an intake. Depending on the consumption pattern, we have regular, less regular and incidental consumers. So each consumer has his own probability of having an intake. For many foods, there may be a relation with age (or sex or, if available, other demographic data). 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 24, 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 elimation 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 24).

See also: Modelling the intake frequency distribution

Intake amount distribution

The non-zero intakes are logarithmically transformed and the ln(intakes) are analysed with REML using a spline or polynomial function to model age effects. The analysis provides us with age dependent ln(intakes) and estimates of the parameters of the non-zero usual intake distribution. The total variance of the non-zero ln(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 24, 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(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(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)

Find in the right section of the input form option block Intake model (see Figure 25, default options).

Figure 25: Option block Intake model if intake model setting is discrete/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 25 is 5 so the number of iterations is 5 x 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
Specify in the top section (see Figure 14) an uncertainty analysis.
See also: Uncertainty of risk assessments: bootstrapping data sets

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

Figure 26: Option block Uncertainty analysis if an acute 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 MC-samples, on the level of consumers and on the level of the compounds. To examine the uncertainty due to fresh MC-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 16). 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 99.99th percentile, for a bootstrap-sample containing 1000 values this is the 99.9th percentile. Note, that the number of iterations per bootstrap sample is never higher than the chunksize.

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

See uncertainty analysis for acute risks described in 3.4.1 .

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

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

Figure 27: Option block Uncertainty analysis if 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.

35

3.5 Output

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

3.5.1 Output options for acute risks: only empirical estimates

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

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

The next option is needed for summarising the contribution of foods 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

The upper part of Figure 29 (default options), see 3.5.1 . In the lower part, options related to age are shown. The minimum and maximum age are retrieved from the database, but are overruled by specifying own values. The steplenght 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.

Figure 29: 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

See Figure 30 (default options) and 3.5.1 .

Figure 30: 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

See Figure 31 (default options) and 3.5.1 and 3.5.2 .

Figure 31: 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)

See Figure 32 (default options) and 3.5.1 .

Figure 32: Option block Output if a chronic risk model is chosen with binomial/lognormal + discrete/semiparametric (Nusser)

3.6 Running an MCRA job

Click the 'MCRA-submit-job'-button in the top section to run a MCRA analysis. Check the radio button if you wish to be notified when the analysis is completed. After submitting the form, enter the MCRA central menu (see Figure 33).

3.7 Checking the processing time

After a submit, all model specifications are passed to the server and the analysis is initiated. First, all data are exported to system files located on the user directory. Export of data takes 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.

Click the 'show progress'-button to view the progress of the analysis (see Figure 33). 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. Click the 'update window'-button to refresh the screen. Click the 'kill job'-button to end a job (with a fatal error status, not shown).
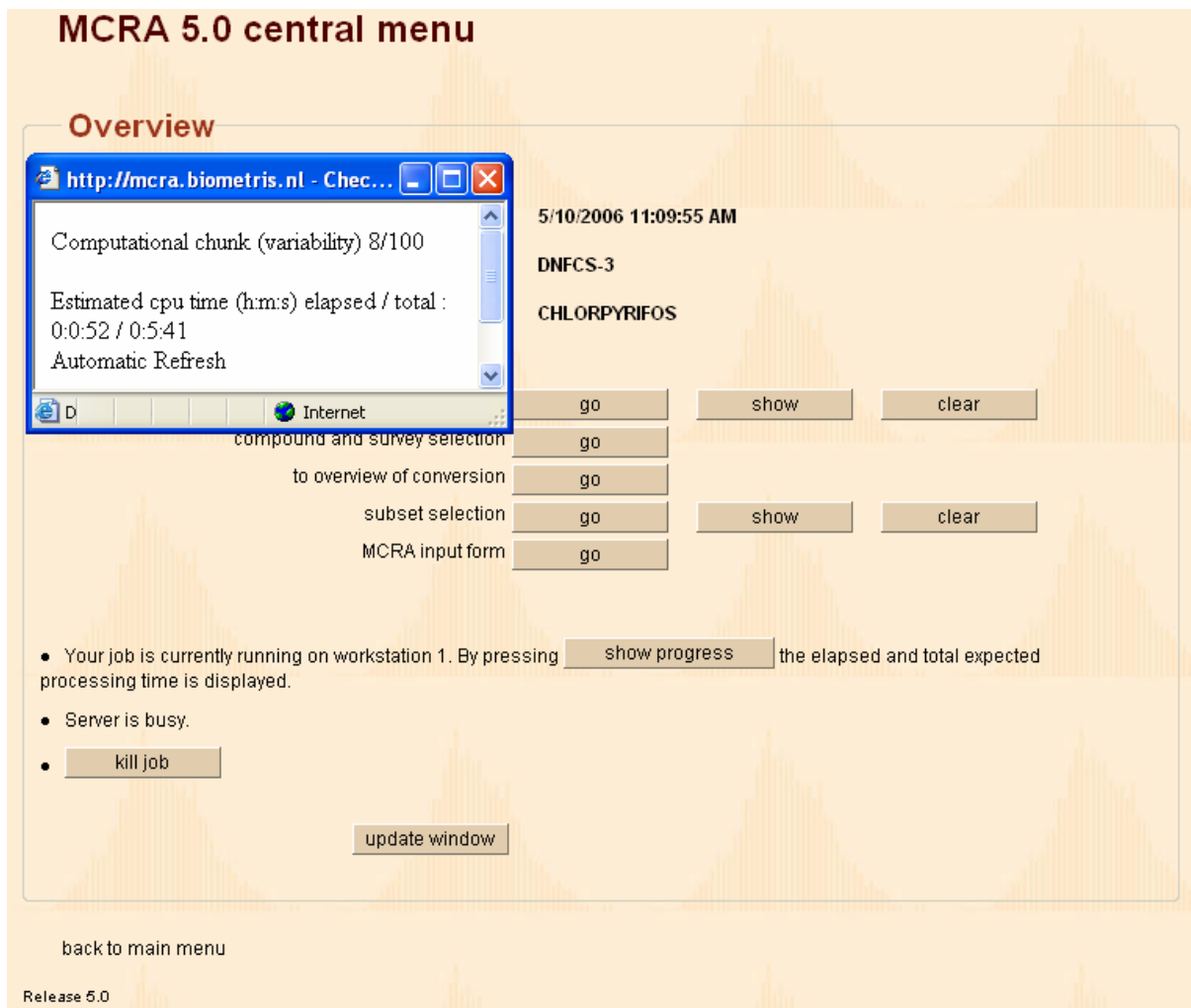
Figure 33: Central menu, show progress

When the job is completed and only after an update, a new button 'view output' appears (see Figure 7). Click the 'view output'-button to get the screen in Figure 35 (see 4.1 .

4 Output

Click the 'view output'-button in Figure 7 to view the results of the analysis. It is always possible to view output of the last job on a later moment. Both tabular and graphical results may be viewed. Depending on the type of analysis, the output window looks different (for example Figure 35 and Figure 40).

In 4.1 , the output is shown 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 . For downloading results for off-line viewing, see 2.5.

4.1 Acute risk assessment: basic analysis

This example shows output of an acute risk assessment for organo phosphate pesticide chlorpyrifos which has neurotoxic effects. The Dutch MCRA_5 database from Rikilt (NL) is used. Select 'All tables', survey 'DNFCS-3' and compound 'CHLORPYRIFOS'. Then, after subset selecting (see 6.2): STRAWBERRY, TABLE_GRAPES, MANDARINE, ORANGE SWEET, APPLE, NECTARINE, PEACH and PEPPER SWEET, the input form is reached (see Figure 34).

Figure 34: MCRA input form for basic acute risk analysis

Table 3 lists the main options 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       95
```

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, view your results using the 'view output'-button in the central menu. This gives Figure 35:
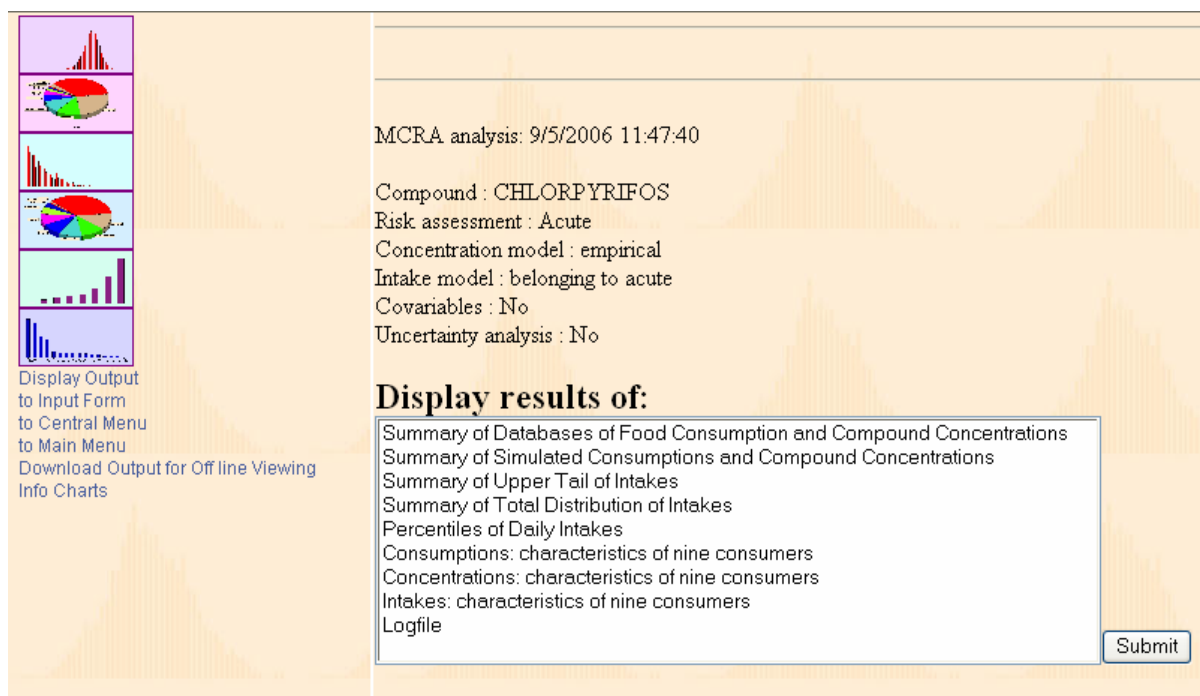
Figure 35: View output window for basic acute risk analysis

Choose 'Logfile' and click the 'Submit'-button in Figure 35. In Table 4 you find the main characteristics concerning this analysis taken from the logfile.

```
Log-file
Survey name                        DNFCS-3
Compound code                      CHLORPYRIFOS
Number of foods                    8
Acute reference dose (ARfD)        100
Acceptable daily intake (ADI)      10
Number of detects                  100
Number of non-detects              938
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       10718
```

Table 4: Information taken from the log-file

Choose 'Summary of Databases of Food Consumption and Compound Concentrations ' and click 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 food you will find the code and label.

In the upper part the average consumption over all consumers and all days (MeanConsum) together with the average consumption on consumption days only (MeanConsDays) is given. In addition, the number of consumption days (NconsDays) and total number of days (Ndays) are displayed also expressed as percentage consumption days (%ConsDays). So, there are 12500 days (2 days for 6250 consumers). Strawberry is consumed on 6095 days (48.8%). The average consumption of Strawberry on these days is 17.45 x 12500/6095 = 35.8 g.

```
SUMMARY OF DATABASES OF CONSUMPTIONS AND COMPOUND CONCENTRATIONS WITH RESPECT TO:
**********************************************************************

  Compound:  CHLORPYRIFOS

**********************************************************************

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


FOOD CONSUMPTION DATA

      Code          Food MeanConsum MeanConsDays NConsDays   NDays  %ConsDays
                          (g)          (g)                             (%)
      FB0275 STRAWBERRY     17.45        35.8      6095     12500      48.8
      FB1235 TABLE-GRAPES   69.51       179.7      4836     12500      38.7
      FC0206 MANDARIN, SE   11.12        51.9      2677     12500      21.4
      FC0208 ORANGE, SWEE   74.25       129.6      7163     12500      57.3
      FP0226 APPLE         100.57       162.5      7737     12500      61.9
      FS0245 NECTARINE       0.80       119.6        84     12500       0.7
      FS0247 PEACH          11.88        42.0      3538     12500      28.3
      VO0445 PEPPERS, SWE   13.03        64.3      2533     12500      20.3

COMPOUND CONCENTRATION DATA

      Code          Food MeanConcen  MeanPosConc NSamplPos NSamples   %SamplPos
                          (mg/kg)       (mg/kg)                          (%)
      FB0275 STRAWBERRY     0.0006        0.0433        3       225       1.3
      FB1235 TABLE-GRAPES   0.0090        0.0795       20       176      11.4
      FC0206 MANDARIN, SE   0.0564        0.1479       29        76      38.2
      FC0208 ORANGE, SWEE   0.0232        0.1090       30       141      21.3
      FP0226 APPLE          0.0016        0.0440        5       136       3.7
      FS0245 NECTARINE      0.0112        0.1160        5        52       9.6
      FS0247 PEACH          0.0033        0.0350        4        42       9.5
      VO0445 PEPPERS, SWE   0.0013        0.0600        4       190       2.1
                                                     ----      ----
                            number of detects =       100      1038  = the total number of
samples

                            the number of non-detects =   938  ( 1038  -   100 )
```

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 (MeanConcen) and of positives only (MeanPosConc) are calculated. The number of positives (NSamplPos) and the total number of concentration measurements (NSamples) are given as well as the percentage of positive concentrations (%SamplPos). In this example for Strawberry 3 positive concentrations are found out of 225 samples (1.3%). The mean concentration of the positive samples is 0.0006 x 225/3 = 0.0433 mg/kg.

Choose 'Summary of Simulated Consumptions and Compound Concentrations'. In Figure 35 you find essentially the same kind of information but all statistics are calculated using simulated data (see Table 6). The simulation is performed by sampling N (in Table 6 200,000) times from the days in the food consumption database (with replacement) irrespective of consumer. For each consumed food per day a concentration is sampled from the distribution of concentrations of that food. Multiplying consumptions and concentrations gives the intake per food. Summing up the intakes over the foods 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.

```
SUMMARY OF SIMULATIONS OF CONSUMPTIONS AND COMPOUND CONCENTRATIONS WITH RESPECT TO:
**********************************************************************

 Compound:  CHLORPYRIFOS


**********************************************************************


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

FOOD CONSUMPTION DATA

        Code         Food MeanConsum   DeltaC MeanConsDays   NConsDay    NDays %ConsDays
                           (g)     (%)       (g)                                 (%)
     FB0275 STRAWBERRY      17.41 ( -0.2)    35.75      97397   200000      48.7
     FB1235 TABLE-GRAPES    69.35 ( -0.2)   179.47      77282   200000      38.6
     FC0206 MANDARIN, SE    10.96 ( -1.4)    51.47      42580   200000      21.3
     FC0208 ORANGE, SWEE    73.96 ( -0.4)   129.31     114387   200000      57.2
     FP0226 APPLE          100.49 ( -0.1)   162.41     123748   200000      61.9
     FS0245 NECTARINE        0.88 (  9.0)   120.45       1454   200000       0.7
     FS0247 PEACH           11.88 (  0.0)    42.05      56512   200000      28.3
     VO0445 PEPPERS, SWE    12.93 ( -0.8)    63.78      40535   200000      20.3

COMPOUND CONCENTRATION DATA

        Code         Food MeanConcen   DeltaR NSamplPos NSamples %SamplPos ProcFact
                           (mg/kg)    (%)                          (%)
     FB0275 STRAWBERRY      0.0006 (  3.3)     1337    97397     1.4     1.00
     FB1235 TABLE-GRAPES    0.0089 ( -1.3)     8647    77282    11.2     1.00
```

42

```
        FC0206 MANDARIN, SE      0.0571 (  1.1)     16350     42580     38.4      1.00
        FC0208 ORANGE, SWEE      0.0234 (  1.0)     24536    114387     21.4      1.00
        FP0226 APPLE             0.0016 ( -1.1)      4513    123748      3.6      1.00
        FS0245 NECTARINE         0.0098 (-12.3)       123      1454      8.5      1.00
        FS0247 PEACH             0.0034 (  1.7)      5477     56512      9.7      1.00
        VO0445 PEPPERS, SWE      0.0012 ( -2.7)       845     40535      2.1      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 5%, corresponding with an intake of 0.324 microgr/kg_bw/day. The acute reference dose (ARfD) for Chlorpyrifos is equal to 100 microgr/kg bw/day.
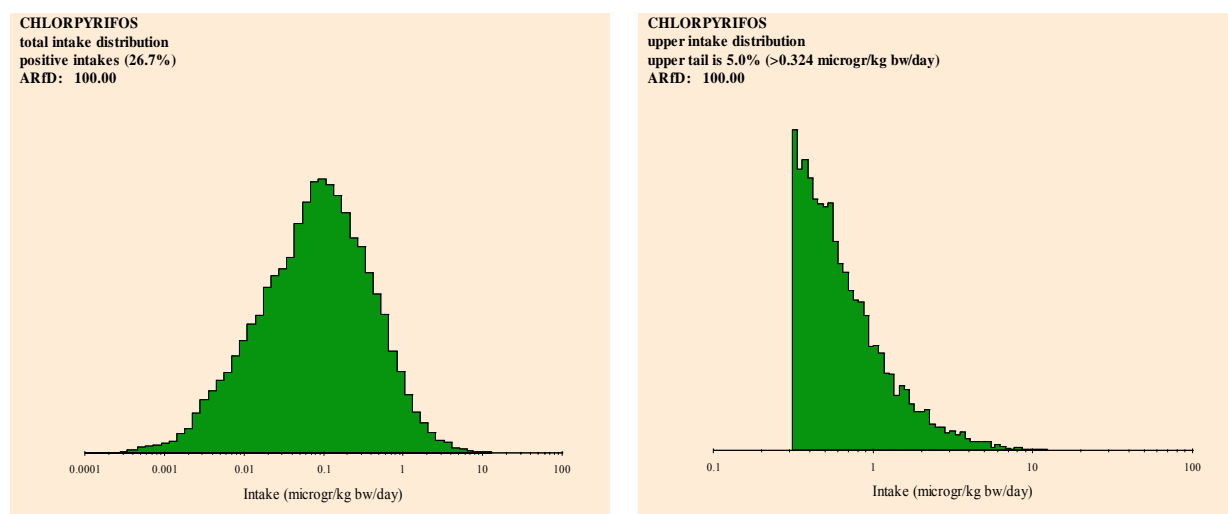
Figure 36: Total intake distribution and upper tail (1%)

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.

Choose 'Summary of Total Distribution of Intakes' and see Table 7. Here the contributions are expressed as percentages (RelContr) of each food to the total intake distribution. A pie chart of the foods 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 food are given in Table 7. The last column (%Zeros) shows the number of zero intakes per food. If %Zeros is greater than 97.5% the p97.5% is lower than the mean!
Let's take Orange as an example. It contributes 53.1% to the total intake distribution and its average concentration is 0.033 microgr/kg bw/day. The p2.5%, median, p97.5% and %Zeros of the total intake distribution of Orange are: 0.000, 0.000, 0.344 microgr/kg bw/day and 87.7% respectively.

```
Compound:  CHLORPYRIFOS


*****************************************************************************
Contribution (%) per food to the total distribution
RelContr: relative contribution (%) per food
p2.5%  : 2.5% perc. of intake distr. per food(microgr/kg bw/day)
Mean   : mean of intake distr.per food (microgr/kg bw/day)
Median : median of intake distr.per food(microgr/kg bw/day)
```

```
p97.5%  : 97.5% perc. of intake distr. per food(microgr/kg bw/day)
%Zeros  : percentage zeros per food
***************************************************************
Food          RelContr    Mean        Median      p2.5%       p97.5%      %Zeros
                 %       (microgr/kg  (microgr/kg  (microgr/kg  (microgr/kg     %
                          bw/day)     bw/day)     bw/day)     bw/day)
STRAWBERRY       0.3      0.000       0.000       0.000       0.000       99.3
TABLE-GRAPES    18.3      0.011       0.000       0.000       0.051       95.7
MANDARIN, SE    20.6      0.013       0.000       0.000       0.140       91.8
ORANGE, SWEE    53.1      0.033       0.000       0.000       0.344       87.7
APPLE            5.6      0.003       0.000       0.000       0.000       97.7
NECTARINE        0.2      0.000       0.000       0.000       0.000       99.9
PEACH            1.4      0.001       0.000       0.000       0.001       97.3
PEPPERS, SWE     0.4      0.000       0.000       0.000       0.000       99.6
```

Table 7: Contribution to the total intake distribution

Choose 'Summary of Upper Tail of Intakes' and see Table 8. Here you find the contributions expressed as percentages (RelContr) of each food to the upper tail according to the specified percentage (here 5%). A pie chart of the foods with the 9 largest contributions to the upper 5% 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 food are given in Table 8. This table shows the same statistics for that part of the intakes per food that correspond with the upper 5% of the intake distribution. So this table represents a drilldown of the upper 5% of the intake distribution and shows statistics for the intakes per foods that constitutes the upper 5%. See also Figure 38, where the mean intake for each food in the upper tail is graphically displayed. The last column (%Zeros) shows the number of zero intakes per food.

Let's take Orange again as an example. It contributes 57.5% to the upper 5% of the intakes and its average concentration is 0.499 microgr/kg_bw/day. The p2.5%, median, p97.5% and %Zeros of the upper 5% of the intakes are: 0.000, 0.343, 2.637 microgr/kg bw/day and 38.7% respectively. Note that Mandarin has the second largest contribution to the total intake distribution (see Table 7) whereas Table-Grapes has the second largest to the upper 5% of the intake distribution.

```
Compound:  CHLORPYRIFOS

****************************************************************************
Characteristics per food of the upper 5.0% of the distribution,
corresponding with a total intake higher than 0.3240(microgr/kg bw/day)
RelContr: relative contribution (%) per food
p2.5%   : 2.5% perc. of intake distr. per food(microgr/kg bw/day)
Mean    : mean of intake distr.per food(microgr/kg bw/day)
Median  : median of intake distr.per food(microgr/kg bw/day)
p97.5%  : 97.5% perc. of intake distr. per food(microgr/kg bw/day)
%Zeros  : percentage zeros per food
****************************************************************************
Food          RelContr    Mean        Median      p2.5%       p97.5%      %Zeros
                 %       (microgr/kg  (microgr/kg  (microgr/kg  (microgr/kg     %
                          bw/day)     bw/day)     bw/day)     bw/day)
STRAWBERRY       0.1      0.001       0.000       0.000       0.000       98.8
TABLE-GRAPES    20.7      0.180       0.000       0.000       1.440       75.7
MANDARIN, SE    16.5      0.143       0.000       0.000       1.036       70.9
ORANGE, SWEE    57.5      0.499       0.343       0.000       2.637       38.7
APPLE            4.4      0.038       0.000       0.000       0.480       91.8
NECTARINE        0.2      0.002       0.000       0.000       0.000       99.6
PEACH            0.4      0.003       0.000       0.000       0.020       95.8
PEPPERS, SWE     0.2      0.002       0.000       0.000       0.000       99.3
```

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.
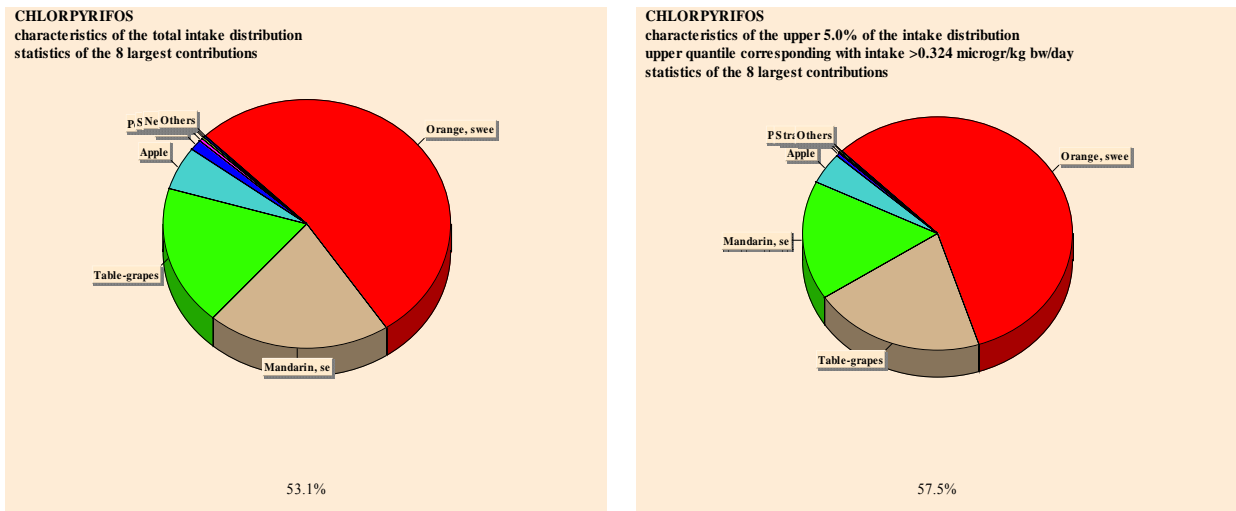
Figure 37: Relative contribution of foods to the total intake distribution and upper tail (5%)

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

Choose 'Percentiles' and see Table 9. The percentiles of the acute intake distribution for the percentages specified in the first line of the Output options in the MCRA input form (see Figure 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         0.13733
      95.00         0.32396
      99.00         1.07677
      99.90         3.65245
      99.99         7.61810
       mean         0.06194
    maximum        12.62322
```

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 food in the upper tail (the third column of Table 8).
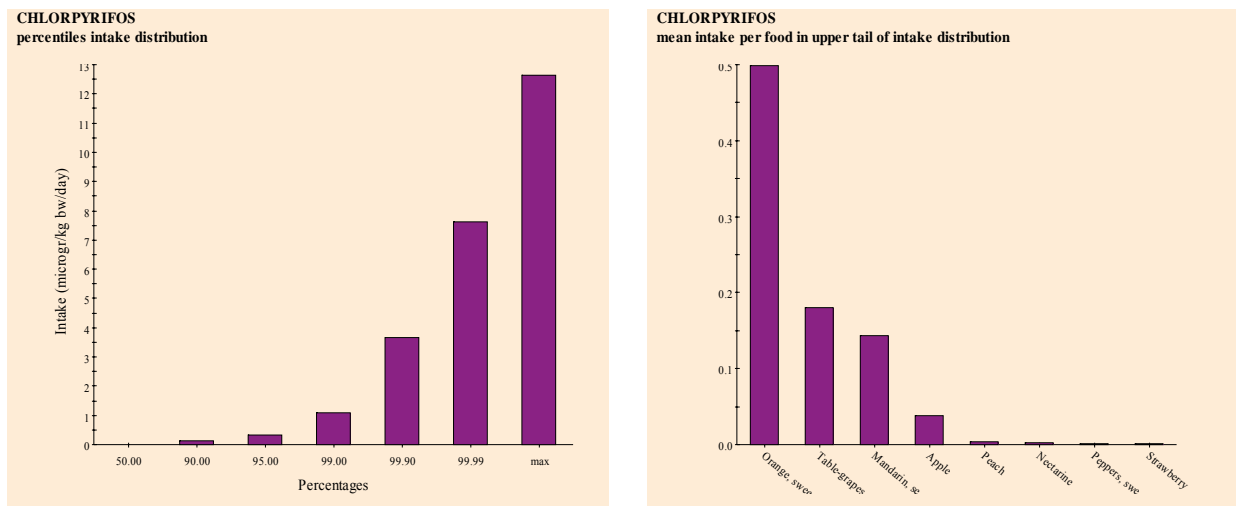
Figure 38: Percentiles and mean intake per food of the acute intake distribution

Choose 'Consumption: characteristics of nine consumers', 'Concentration: characteristics of nine consumers' or 'Intake: characteristics of nine consumers', respectively. This gives the summaries displayed in Table 10, Table 11 and Table 12. 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 foods (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 x 0.380/86 = 0.664 microgr/kg bw/day.

```
Compound:   CHLORPYRIFOS

Drill down: consumption (g/day) per food of the 9 consumers
around the specified percentage (97.50%) for the intake distribution
                -4       -3       -2       -1        0       +1       +2       +3       +4
food
 STRAWBERRY      .        .        .        .      34.2    30.36      .        .        .
 TABLE-GRAPES    .        .        .        .      29.1    31.80    418.8      .        .
 MANDARIN, SE    .        .        .        .      55.0      .        .        .        .
 ORANGE, SWEE  278.0    278.0    248.0    140.0    34.7    29.04      .      240.0    200.00
 APPLE         145.9    145.9      .        .     569.7   315.79    200.0    108.0     24.50
 NECTARINE       .        .        .        .        .        .        .        .        .
 PEACH           .        .        .        .      29.8    30.10      .        .        .
 PEPPERS, SWE    .        .        .        .        .        .      49.0      .       28.80
```

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

```
Compound:   CHLORPYRIFOS

Drill down: compound concentrations (mg/kg) per food of the 9 consumers
around the specified percentage (97.50%) for the intake distribution
                -4       -3       -2       -1        0       +1       +2       +3       +4
food
 STRAWBERRY      .        .        .        .     0.000    0.000      .        .        .
 TABLE-GRAPES    .        .        .        .     0.000    0.000    0.040      .        .
 MANDARIN, SE    .        .        .        .     0.340      .        .        .        .
 ORANGE, SWEE  0.040    0.040    0.210    0.280   0.000    0.000      .      0.190    0.190
 APPLE         0.000    0.000      .        .     0.000    0.050    0.000    0.000    0.000
 NECTARINE       .        .        .        .        .        .        .        .        .
 PEACH           .        .        .        .     0.040    0.000      .        .        .
 PEPPERS, SWE    .        .        .        .        .        .      0.040      .      0.000
```

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     370605    370605     70281    349112    274043    313764    362114    192562    178881
  Body Weight        19        19        89        67        34        27        32        78        65
          Age         2         2        27        55        11         7        11        54        23
        Total     0.585     0.585     0.585     0.585     0.585     0.585     0.585     0.585     0.585
Food
 STRAWBERRY           .         .         .         .     0.000     0.000         .         .         .
 TABLE-GRAPES         .         .         .         .     0.000     0.000     0.524         .         .
 MANDARIN, SE         .         .         .         .     0.550         .         .         .         .
 ORANGE, SWEE     0.585     0.585     0.585     0.585     0.000     0.000         .     0.585     0.585
 APPLE            0.000     0.000         .         .     0.000     0.585     0.000     0.000     0.000
 NECTARINE            .         .         .         .         .         .         .         .         .
 PEACH                .         .         .         .     0.035     0.000         .         .         .
 PEPPERS, SWE         .         .         .         .         .         .     0.061         .     0.000
```

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 antibioticum 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:

Input form	
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. View your results using the 'view output'-button in the central menu. This gives the screen of Figure 40:
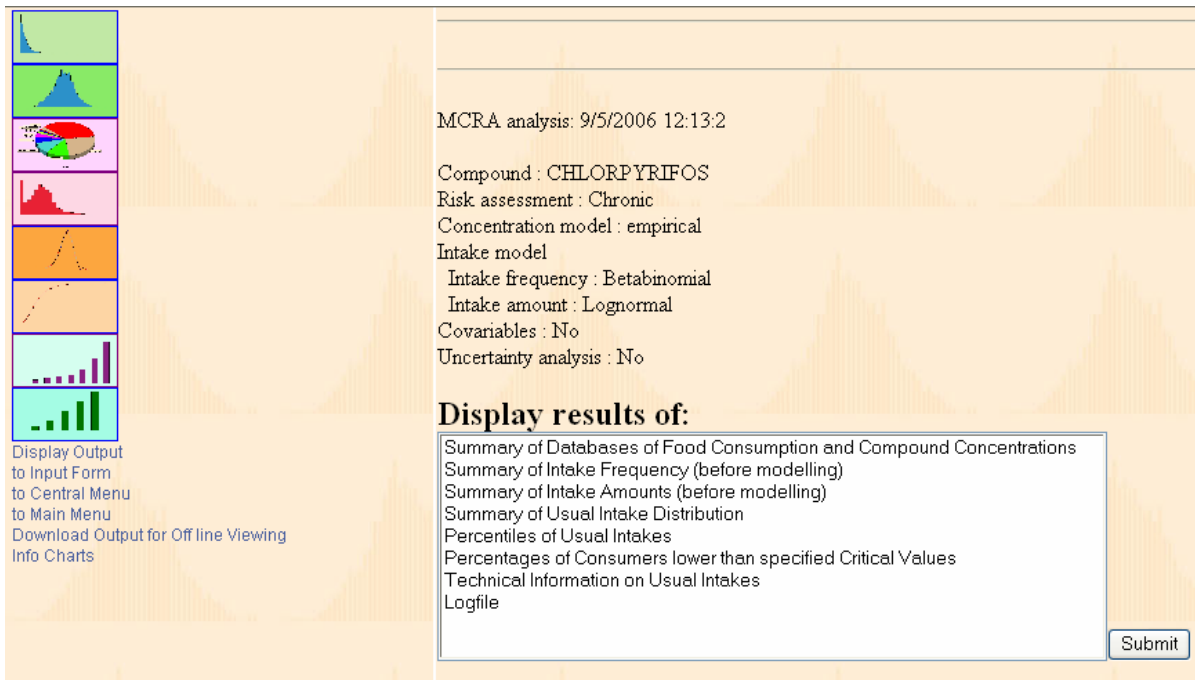
Figure 40: View output window for chronic risk analysis betabinomial/lognormal

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.

```
Log-file
Number of foods                   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(*)
```

48

```
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 samples, 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. Choose 'Summary of database' and see 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)                           (%)
    &NL001  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)                        (%)
    &NL001  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 consumption of 26.1 g on 8630 consumption days and 0 g on 12500 - 8630 = 3870 non-consumption days. The average concentration of the 31 positive lasalocide samples and 219 non-detects (=0) samples is 0.4825 mg/kg.
Choose 'Summary of intake days' and see 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   Positive
                      intakes   intakes     %
   Number of observations   12500     8630   69.0
      Number of consumers    6250     5409   86.5
```

Table 16: Summary of intake days

Choose 'Summary of intake distribution' and see 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 78.
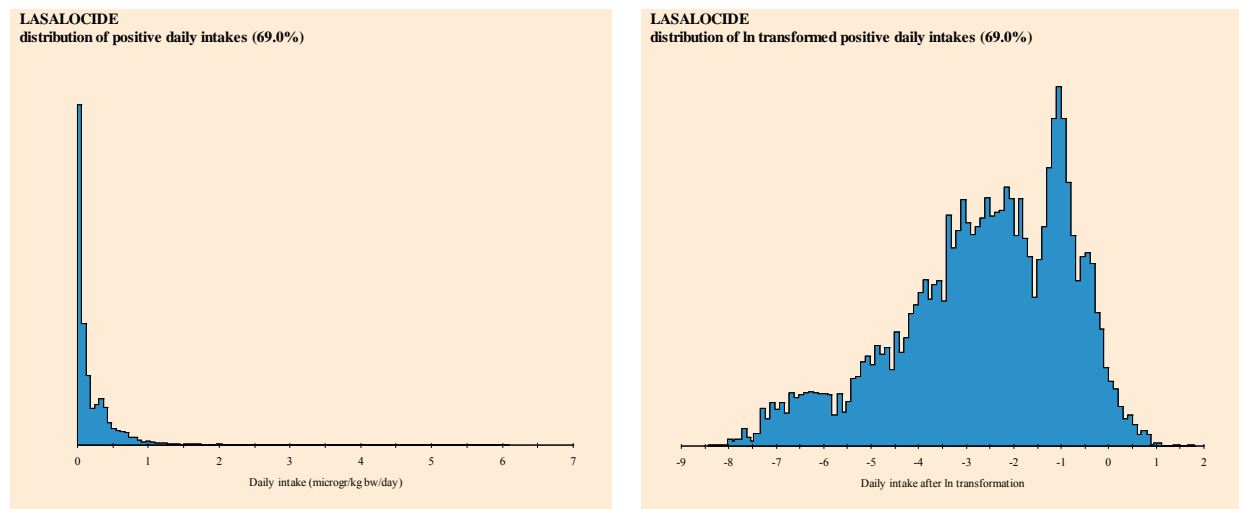
Figure 41: Untransformed and ln transformed intake distribution

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.
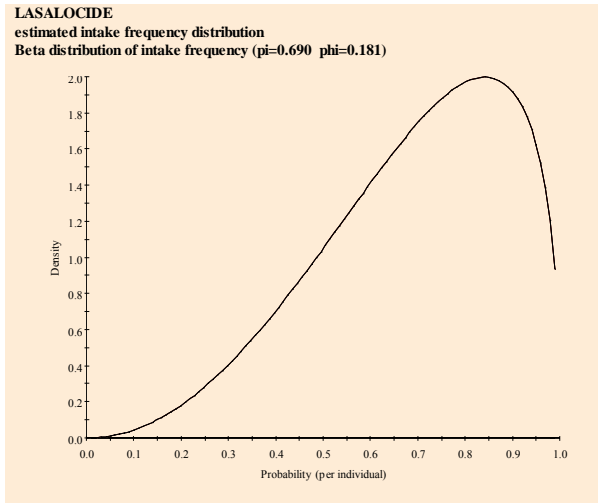
Figure 42: Intake frequency

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.
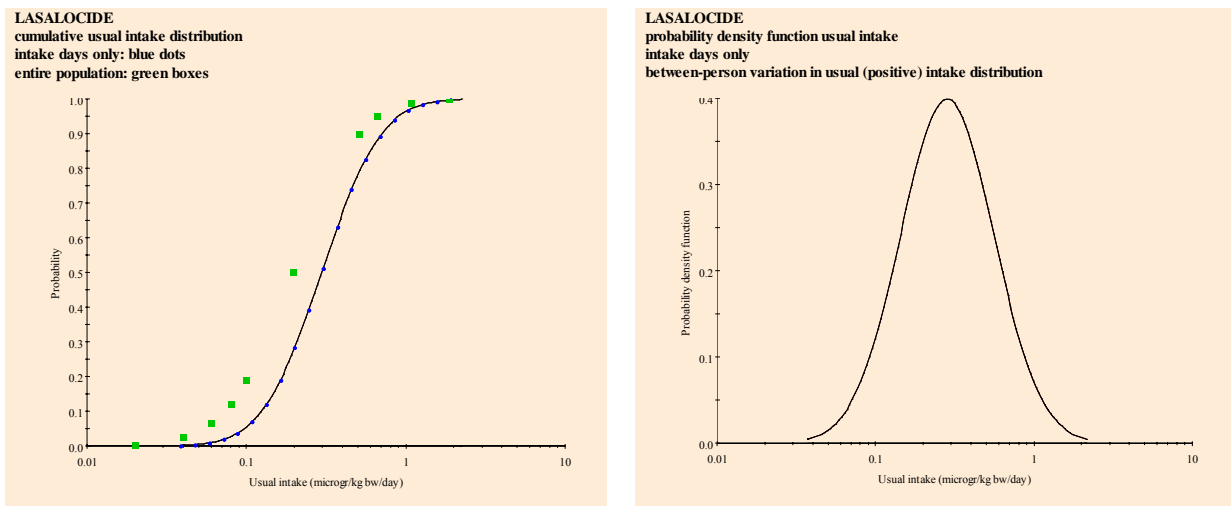
Figure 43: Cumulative usual intake and density function

Choose 'Percentiles'. The percentiles of the usual intake distribution for the entire population are shown (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.
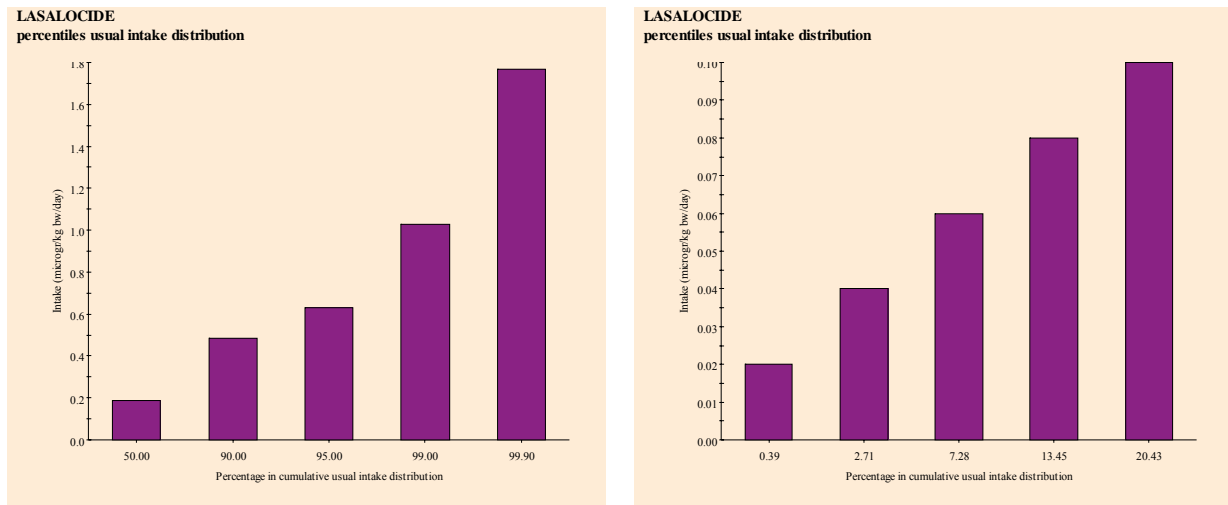
Figure 44: Percentiles and percentages

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).
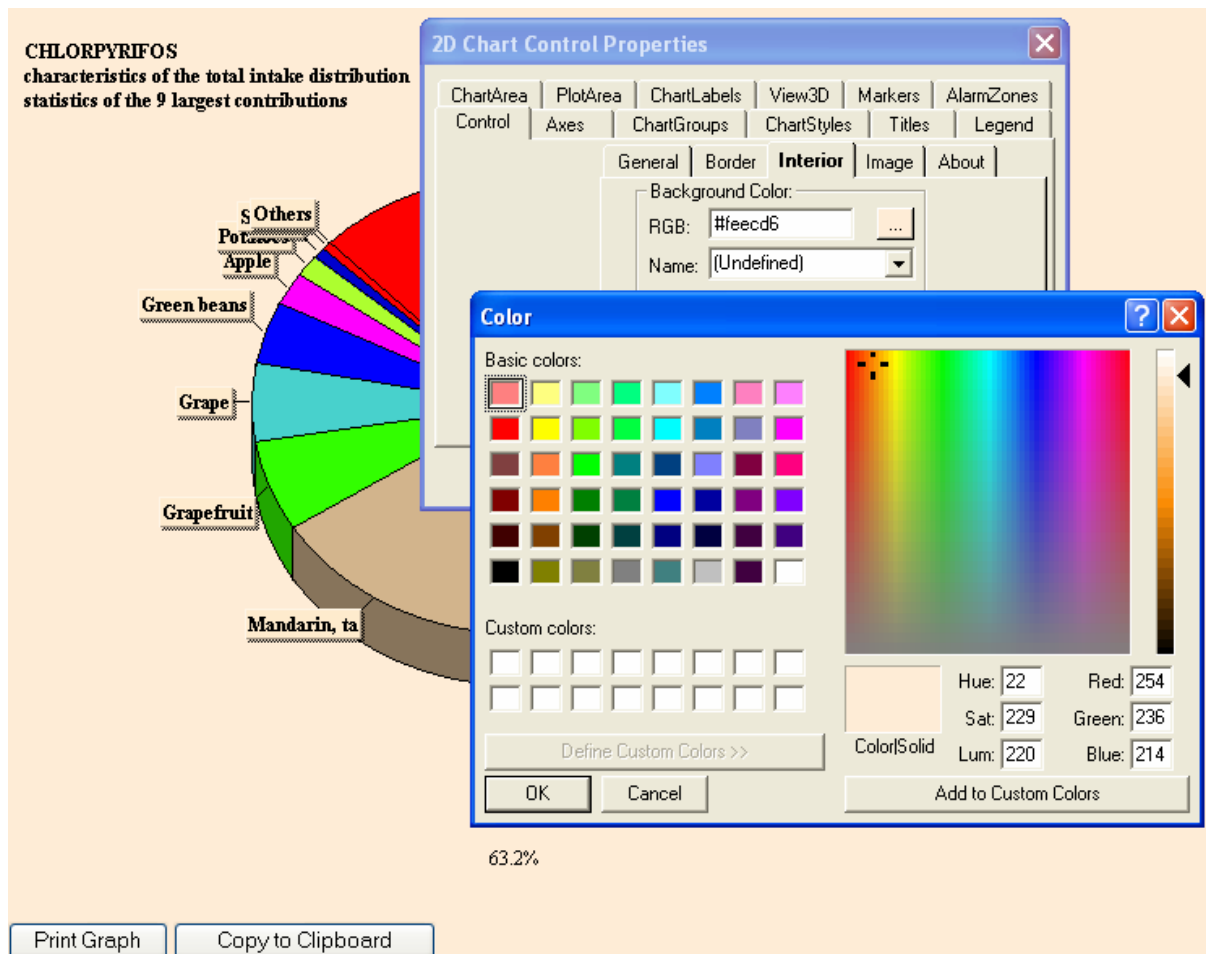
Figure 45: Property editor

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

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

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

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

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

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

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

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

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

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

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

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

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

5.2 How to deal with limited information on compound concentration data

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

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

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

Back to: Concentration model options

5.2.1 The choice between a parametric and non-parametric approach

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

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

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

measurements. Figure 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.
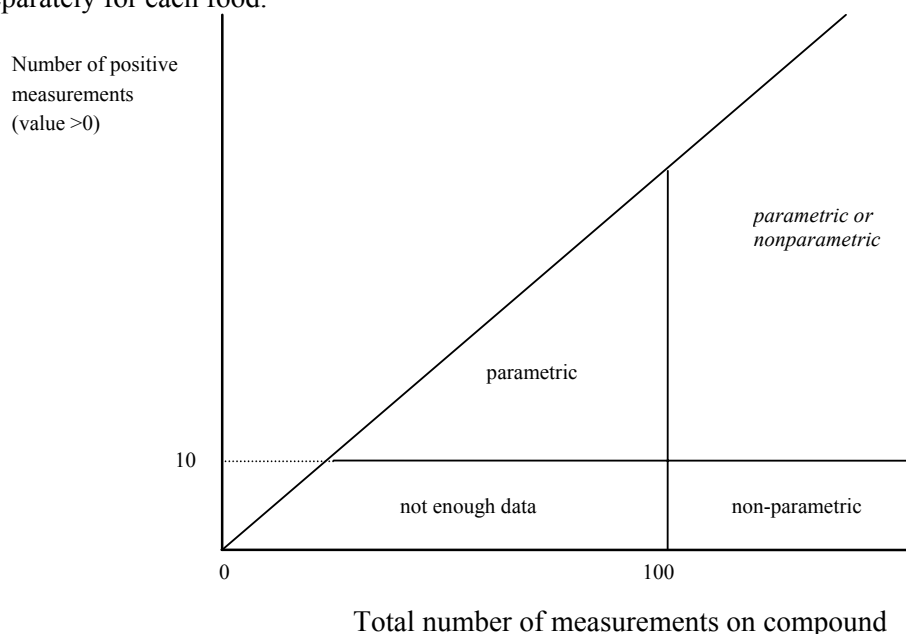
Figure 46: Use of non-parametric or parametric modelling for estimating the 99% percentile of the intake in relation to number of positive measurements and the total number of measurements.

5.2.2 Estimation based on histogram data

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

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

5.2.3 Estimation based on summary data

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

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

with $\Phi(.)$, the cumulative probability function of the standard normal distribution, N, the total number of samples and n_0, the number of zeroes. n0=n1 of 5.2.2 ???

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

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

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

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

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

5.2.3.1 Moments and other characteristics

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

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

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

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

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

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

5.2.3.2 Estimation

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

Using (4) gives:

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

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

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

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

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

For moderate to small sample sizes the estimation of σ fails because the discriminant is negative, i.c. the argument of the square root function. Empirical simulations show that a negative discriminant happens more often for small sample sizes and for estimation based on extreme quantiles like the maximum. Figure 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 σ is assessed by equating the discriminant to zero. Empirical results show that this works out very well for sample size n > 4, although σ is slightly biased upwards being a conservative estimate. In case of the median no solution to this problem is available so the estimate of σ is set to a missing value.
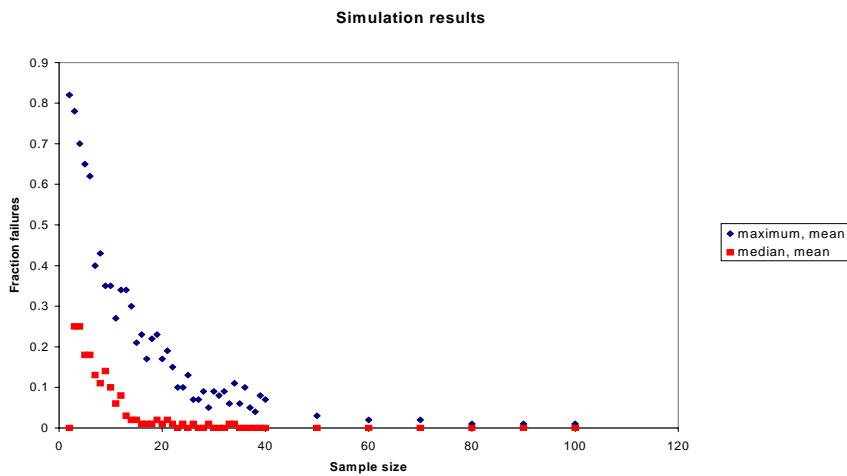
Figure 47: Simulated fraction of failures versus sample size for estimation of σ based on the mean and respectively the maximum and median

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

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

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

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

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 = $\sqrt{\text{(variance)}}$/mean. Using (5), parameter σ is estimated with:

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

and μ is estimated solving (1).

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

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

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

For the median, estimation of μ simplifies to:

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

5.3 Acute risk assessment

5.3.1 Concentration model: modelling of compound concentrations in consumed food

5.3.1.1 Distributional assumptions

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

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

5.3.1.1.2 Parametric modelling of concentrations (binomial/lognormal no pooling)
In the parametric approach, compound concentrations per food are sampled from parametric distributions. A special feature of compound data is that the large majority of measured concentrations (often more than 80%) is recorded as zero (non-detects). These values may correspond to true zero concentrations (for example because the compound is never used in the specific food), or they may correspond to low concentrations which are below a pre-established reporting limit (LOR). In any case, the compound concentration distribution is very skew, with a large spike at zero and an extended tail to higher values. For statistical modelling a two-step procedure is chosen. First, the presence of a concentration $\geq$ LOR on food is modelled with a binomial distribution with a parameter p representing the probability of a reported concentration. Probability p depends on the chemical compound and the food and is estimated as the fraction of detects. Secondly, the non-zero compounds are modelled with the lognormal distribution. After consideration of several possibilities using the program *BestFit*, this distribution has been selected as being both theoretically sensible and practically useful. The parameters μ and σ are the mean and standard deviation of the log-transformed non-zero compound concentrations.
In the basic model (see 5.1)

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

with I_{ijk} indicating whether a compound concentration is sampled ($I_{ijk}=1$) or not ($I_{ijk}=0$), and $cpos_{ijk}$ the compound concentration in the subpopulation of positive values. The probability of I_{ijk} being 1 or 0 depends on the number of detects found for food k and I_{ijk} is sampled separately for each consumer i on occasion j.

Back to: Binomial/lognormal (no pooling)

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

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

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

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

Back to: Binomial/lognormal with pooling

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

total number of measurements made together with the LOR. We use LOR to mean exactly what the term says: measurements below LOR are not reported, whereas values equal to or higher than LOR are represented by numerical values in the database.

In the analytical and food risk fields analytical limits are often indicated as LOD (limit of detection) or LOQ (limit of quantification). Unfortunately, it is not always clear what is meant with these terms. In any case official recommendations are to always report any available numerical values even if they are below LOD or LOQ limits (IUPAC 1995).

For legal applications of compounds data may be available about the percentage of the crop which receives treatment. When a chemical compound can enter the food chain only via crop treatment, and when the percentage of crop treated is (approximately) known to be $100 p_{crop-treated}$, 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.

5.3.1.3 Modelling of processing effects

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

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

where cr_{ijk} is the concentration in the raw food, and where f_k is a factor for a specific combination k of RAC and processing. Values will typically be between 0 and 1, although occasionally the processing factor may also be >1 (e.g. drying as applied for grapes and figs).

The user of the model will have to specify processing factors for each food k as defined in the food consumption data base. For this purpose, it is advised to maintain a data base of processing factors, indexed by chemical compound, RAC and processing type (e.g. washing, peeling or other processing). Before running the model, it may then be necessary to specify how the necessary processing factors are derived from the data base entries and/or other information. Example: if there are no processing factors known for captan in pears, it may be decided to use the corresponding factors for apples instead.

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

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

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

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

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

b) The logits of $f_{k,nom}$ and $f_{k,upp}$ are equated to the mean and the 95% one-sided upper confidence limit of a normal distribution. This normal distribution thus is specified by a mean $\text{logit}(f_{k,nom})$ and a standard deviation $\{\text{logit}(f_{k,upp}) - \text{logit}(f_{k,nom})\}/1.645$. Values are drawn from this distribution in the MC-simulations. Processing factors f_k will be between 0 and 1. Note: $f_{k,upp}$ and $f_{k,nom}$ values equal to 0 and 1 are replaced by 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^{rst}, 3rd and 5th plot). A logit transformation should be considered instead.
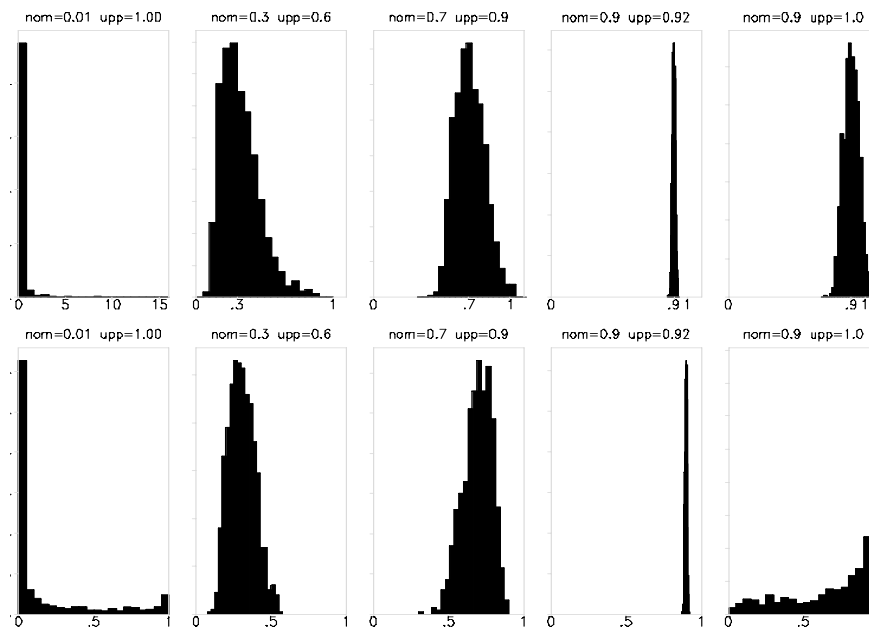
Figure 48: Lognormal (upper row) and logistic (lower row) distributions for various values of $f_{k,nom}$ (= nom) and $f_{k,upp}$ (= upp)

5.3.2 Modelling of unit variability

5.3.2.1 Introduction, variability in deterministic modelling

Variability in compound concentrations between individual units is a relevant factor in the assessment of short-term dietary 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 food k is composed of nu_k units with nominal unit weight wu_k each. The weight of a composite sample is therefore $wm_k = nu_k \times wu_k$. This weight is often larger than a consumer portion, e.g. a typical composite sample of 20 sweet peppers weighs 3.2 kg, whereas daily consumer portion weights in the Dutch Food Consumption Survey 1997 ranged from 0.08 g to 458 g.

How should monitoring data be used to estimate the raw food concentration levels cr_{ijk} in consumer portions? Although the mean level of cm_k may be a fair estimate of the mean level of cr_{ijk}, the variability of cm_k is not appropriate to estimate the variability of cr_{ijk}. In smaller portions more extreme values may occur more readily, and thus acute risks may be higher than would follow from a direct use of the composite sample data.

Therefore the FAO/WHO Geneva Consultation recommended to include a *variability factor* (v) in the non-probabilistic calculation of an international estimate of short-term intake (*IESTI*) (FAO/WHO 1997). The *IESTI* has been adopted by the Joint Meeting of FAO and WHO experts on Pesticide Compounds in food in 1999, and was modified in 2000 to reflect that the supply for actual consumption on a given day is likely to be derived from a single lot (JMPR 1999, 2000). In both the original and the modified definition, the variability factor is used in a similar way. The basic idea is that the compound concentration for the first unit eaten is multiplied by v, whereas this factor is not applied for any remaining part of the daily consumption.

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

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

Table 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[th] percentile of the unit levels divided by the sample mean (Harris et al. 2000), and this is used in the current version of MCRA as the defining relation.

Back to: Unit variability model options

5.3.2.2 Approaches to unit variability in probabilistic modelling: specifying distributions

How should variability between units be incorporated in probabilistic modelling of acute risks? In probabilistic modelling we generate consumption amounts and compound concentrations which will be multiplied, summed over foods and divided by body weight to estimate the intake. However, the compound concentration cm_k will usually be derived from a distribution based on measurements on composite samples. Assume that a batch of food contains N units (N large, for the statistics we assume infinite). The monitoring measurement cm_k is made on a composite sample of nu_k units (for example, $nu_k = 5$). These units are assumed to be representative of the batch. Unit concentrations cr_{ijk} are to be simulated for one or more units from this batch that will be part of a consumption portion in the MC-simulation. Basically, there are three possibilities depending on the availability of data:
1. use actual measurement data on individual units;
2. use variability factors or other summary statistics based on measured individual units;
3. use conservative assumptions.

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

In MCRA the following three models, discussed below in more detail, are implemented:
1. **Beta model**, requires knowledge of the number of units in a composite sample, and of the variability between units (realistic or conservative estimates);
2. **Bernoulli model**, requires only knowledge of the number of units in a composite sample (results are always conservative);
3. **Lognormal model**, requires only knowledge of the variability between units (realistic or conservative estimates).

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

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

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

Back to: Estimated parameters for unit variability

5.3.2.2.1 *Beta model for unit variability*
With this model MCRA will generate values for individual unmeasured units of a measured composite sample. If cm_k is the concentration measured (or simulated) for the composite sample in

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

The standard beta distribution is defined on the interval $(0, 1)$ and is usually characterised by two parameters a and b, with $a>0$, $b>0$ (see $e.g.$ Mood $et\ al.$ 1974). Alternatively, it can be parameterised by the mean $\mu=a/(a+b)$ and the variance $\sigma^2=ab(a+b+1)^{-1}(a+b)^{-2}$, or, as applied in MCRA, by the mean μ and the squared coefficient of variation $cv^2=ba^{-1}(a+b+1)^{-1}$. Note that the coefficient of variation is the same for the unscaled and the scaled distributions.

For the simulated unit values in each iteration of the program we require an expected value cm_k. This scales down to a mean value $\mu = cm_k/c_{max} = 1/nu_k$ in the (standard) beta distribution. From this value for μ and an externally specified value for cv_k the parameters a and b of the beta distribution are calculated as:

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

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

When the unit variability is specified by a variability factor $v_k = \dfrac{p97.5_k}{cm_k}$ 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 shows an intermediate situation, moderate to high variability.

$cv_k =1.732;\ nu_k=4;\ a=0.00005;\ b=0.00015$ $cv_k =1.20;\ nu_k=6;\ a=0.4;\ b=2$

cv_k=0.62; nu_k=6; a=2; b=10   cv_k =1.46; nu_k=4; a=0.1; b=0.3
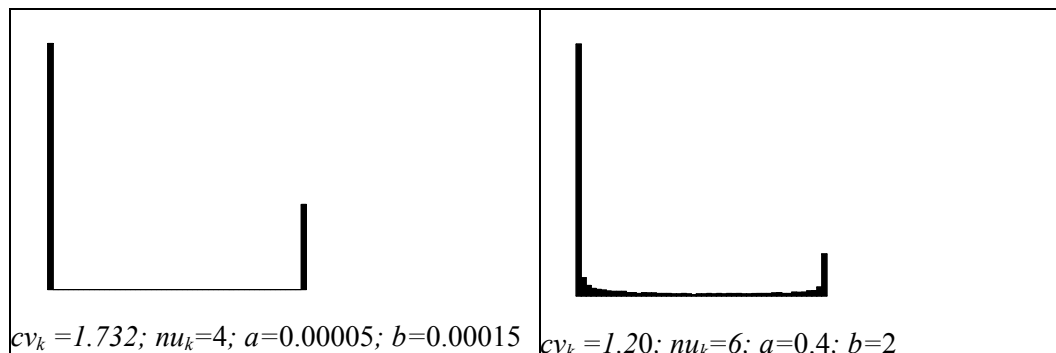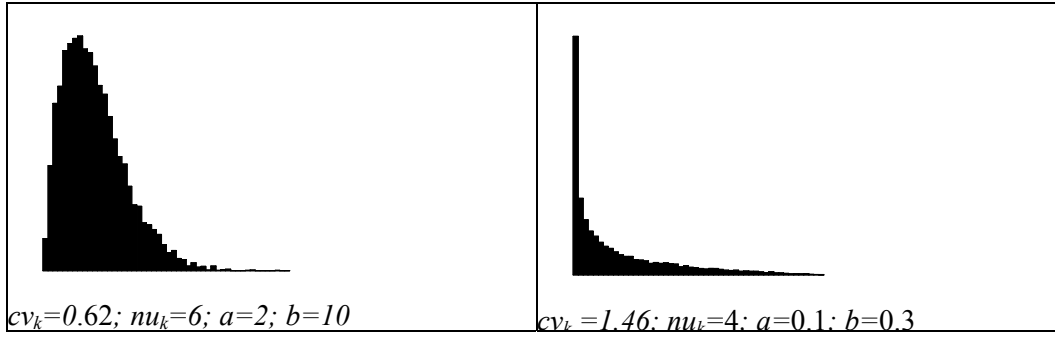
Figure 49: Standard Beta distribution for different values of the coefficient of variation cv_k and number of units nu_k in the composite sample. x axis from 0 to 1.

Back to: Beta distribution

5.3.2.2.2 Lognormal model for unit variability
With the Beta and Bernoulli models, MCRA simulates concentrations for units in the composite sample, such that the concentration of an individual unit can never be higher than the monitoring measurement multiplied by the number of units in the composite sample $c_{max} = nu_k * cm_k$.
With the Lognormal model for unit variability MCRA simulates concentrations for new units in the batch from which the composite sample was taken. Effectively the number of units in a batch is very large, so in this case there is no practical upper limit to the concentration that can be present.
The lognormal distribution is considered as an appropriate model for many empirical positive concentration distributions. With the Lognormal model MCRA assumes a lognormal distribution for unit compound concentrations. Let this distribution be characterised by μ and σ, which are the mean and standard deviation of the log-transformed concentrations. The unit log-concentrations are drawn from a normal distribution with mean $\mu = \ln(cm_{ik})$.

Also for the Lognormal model MCRA allows two choices to specify the parameter for the unit variability. The parameter is specified as a coefficient of variation (cv_k) or as a variability factor (v_k). The coefficient of variation cv is turned into the standard deviation σ on the log-transformed scale with:

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

A variability factor v is converted into the standard deviation σ as follows:

$$v = \frac{p97.5}{mean} = \frac{e^{\mu+1.96\sigma}}{e^{\mu+1/2\sigma^2}} = e^{1.96\sigma-1/2\sigma^2}$$

with μ and σ representing the mean and standard deviation of the log-transformed concentrations. So

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

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

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

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

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

5.3.2.2.3 Bernoulli model for unit variability
The Bernoulli model is a limiting case of the Beta model, which can be used if no information on unit variability is available, but only the number of units in a composite sample is known. As a worst case approach we may take cv_k as large as possible. When cv_k is equal to the maximum possible value $\sqrt{nu_k - 1}$, the (unstandardised) Beta distribution simplifies to a Bernoulli distribution with probability $(nu_k - 1)/nu_k$ for the value 0 and probability $1/nu_k$ for the value $c_{max} = nu_k * cm_k.$.
In MCRA values 0 are actually replaced by cm_k, to keep all values on the conservative side. For example, with $nu_k = 5$, there will be 80% probability at $cr_{ijk} = cm_k$ and 20% probability at $cr_{ijk} = c_{max.}$.
Back to: Bernoulli distribution

5.3.2.2.4 Estimation of intake values using the concept of unit variability
- For each iteration i in the MC-simulation, obtain for each food k a simulated intake x_{ik} , and a simulated composite sample compound concentration cm_{ik} .
- Calculate the number of unit intakes nux_{ik} in x_{ik} (round upwards) and set weights w_{ikl} equal to unit weight wu_k, except for the last partial intake, which has weight $w_{ikl} = x_{ik} - (nux_{ik} - 1)wu_k$.
- For the Beta or Bernoulli distribution: draw nux_{ik} simulated values κ_{ikl} from a Beta or Bernoulli distribution. Calculate concentration values as $c_{ikl} = \kappa_{ikl} * cm_{k, max} = \kappa_{ikl} * cm_k * nu_k$. Sum to obtain the simulated concentration in the consumed portion:

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

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

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

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

The IESTI (International Estimated Short-Term Intake) is a deterministic estimate of the short-term intake of a compound on the basis of the assumptions of high daily food consumption per consumer and highest compounds from supervised trials. The IESTI is expressed per kg body weight and has only been defined for single foods.
MCRA calculates IESTI for comparison with MC-percentiles.
Calculations of IESTI (according to FAO 2002) recognise four different case (1, 2a, 2b and 3). In cases 1 to 3 the following definitions are used:
LP: Highest large portion reported, calculated as the 97.5[th] percentile of the distribution of consumed portions on days with positive consumption of the food (kg food/day)
HR: Highest residue (=compound) in composite sample, mg/kg
bw: Mean body weight, kg; in MCRA values may be input by the user, or weighted means are calculated over consumers with the number of days on which they consumed the food as weights
U: Unit weight of the edible portion, kg.
v: Variability factor – the factor applied to the composite compound to estimate the concentration in a high-compound unit
MR: Median residue (=compound) in food, mg/kg
Although the FAO Manual refers to supervised trials only, MCRA calculates concentrations (HR or MR) from any compound concentration data set supplied (may also be monitoring data).
Concentrations (HR or MR) may be multiplied with a processing factor on beforehand, in MCRA this depends on the options chosen for processing.

Case 1:
The compound in a composite sample reflects the concentration in meal-sized portion of the food (unit weight is below 25 gr).

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

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

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

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

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

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

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

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

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

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

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 food). These two different approaches handle the same problem, but it is undefined to which MC-percentile the IESTI value should be compared. In MCRA the user is free to choose a percentage point for this comparison.

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

5.3.3 Intake model

5.3.3.1 Only empirical estimates

This is an acute risk assessment using empirical compound data.

5.3.3.2 Empirical estimates and betabinomial/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(intake) amounts.

5.3.3.2.1 Monte Carlo data generation of intake

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

5.3.3.2.2 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 ϕ (independent of age). The fitted binomial probabilities are $\hat{\pi}_x = f(x_i)$, where x_i is the age of consumer i, and the estimated overdispersion parameter is $\hat{\phi}$.

5.3.3.2.3 Modelling ln(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.

5.3.3.2.4 Estimating the acute risk variability of positive intake amounts

Correct the full set of simulated positive intakes by $y'_{ijk} = y_{ijk} - \hat{\mu}_{x(i)}$. Estimate the variance $\sigma_{y'}^2$ 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 $\hat{\sigma}_{y'}^2$.

5.3.3.2.5 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.

5.4 Chronic risk assessment

In the MCRA program we have 3 models available to assess chronic risks:
1) the beta-binomial/lognormal model without covariable (see 5.4.2 , 5.4.3 and 5.4.4);
2) in the extended version of this model, spline or polynomial functions are used to model the effect of a covariable like age (see 5.4.5) and;
3) the discrete/semi-parametric model according to Nusser without covariable (see 5.4.6).

5.4.1 Introduction

In a chronic risk assessment the main interest goes to the fraction of consumers with a usual intake per day higher than an exposure limit. Usual intake is defined here as the long-run average of daily intakes of a (chemical) compound by a consumer.

Usually, food consumption data are available for consumers on 2 (or more) consecutive days. We assume an equal number of days for each consumer. This is in confirmity with our method of data entry for consumption (see 9.4.1). As a consequence, days without consumptions do have zero intake. Through the assumed independence of consumption data and compound concentration values (a most reasonable assumption) the daily intake of consumer i on day j can be calculated as the sum over foods of consumption amount per kg body weight times average concentration. See the basic model in 5.1. The average concentration of all available concentration measurements on a food is taken, with non-detect measurements entered as zero, $\frac{1}{2} LOR$ or LOR , or any other fraction of LOR as specified in the input options.

To calculate the usual intake in the first 2 available models two aspects will be taken into account: the number of days that a consumer eats that compound and if so, the amount of that compound. The number of days will be assumed to have a binomial distribution with a probability p different for each consumer. These probabilities will be assumed to come from a beta distribution. The eaten compound will be assumed to have a lognormal distribution. In the next paragraphs the modelling will be further discussed.

5.4.2 Modelling the intake frequency distribution

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

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

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

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

This distribution is known as the betabinomial distribution.

The mean and variance of a beta distribution are:

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

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

$$n\,\pi$$

and $n\,\pi(1-\pi)[1 + (n-1)\varphi]$, respectively.

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

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

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

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

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

Back to: Chronic risk analysis

5.4.3 Modelling the positive ln(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 interaction *consumer.day* to estimate the between-consumer and within-consumer variance component:

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

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

It can be shown that the expectation and variance of the positive intake per consumption day of a random consumer $\underline{y}$ are:

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

5.4.4 Modelling usual daily intake

To obtain the usual intake (that is the average daily intake over both consumption and non-consumption days) the $E(\underline{y})$ from 1.3 has to be multiplied by the probability π from 1.2. If π was constant for all consumers the usual intake would have a lognormal distribution with mean $\ln(\pi) + \mu + \tfrac{1}{2}\,\sigma^2_{within}$ and variance $\sigma^2_{between}$. But because we have assumed in 1.2 that consumers have different p's coming from a beta distribution, the probability that a consumer has a usual intake lower than say a exposure limit z equals:

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

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

where Φ is the cumulative normal distribution. In the MCRA program this integral is calculated with a Fortran-program using 2 ISML routines DZREAL and DQDAGS.

Back to: Betabinomial/lognormal

5.4.5 Covariables in the models

5.4.5.1 Intake frequencies depending on covariable(s)

Intake frequencies can be modelled to be dependent of a covariable, like age. In that case the following regression model can be specified:

$$\text{logit}(\pi) = \gamma_0 + \gamma_1\, f(x_1; df),$$

where x_1 is the covariable, f is a spline or polynomial function and df the degrees of freedom of the spline or the degree of the polynomial. In the program the optimal df can be calculated by a backward or forward search.

After the analysis the fitted probabilities $\hat{\pi}$ can be plotted against the covariable.

5.4.5.2 Positive amounts depending on covariable(s)

Also the positive amounts can be modelled to be dependent of a covariable. The model for the ln(amount) is then:

$$\ln(\underline{y_{ij}}) = \gamma_0 + \gamma_1\, f(x_1; df) + \underline{y_i} + \underline{u_{ij}},$$

with the the same notation as in 1.2 and 1.5.1.

The fitted values from the model are $\hat{\mu}_x = f(x_i)$, where x_i is the age of consumer i.

5.4.5.3 Usual intakes depending on covariable(s)

When one (or both) of the two models of intake frequency depends on a covariable the usual intake depends on that covariable too. The probability that a consumer has a usual intake lower than say a exposure limit z will in that case also plotted against the covariable.

The percentage of consumers

Back to: Betabinomial/lognormal (with age)

5.4.6 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 foods which were consumed on an almost daily basis, e.g. dioxin in fish, meat or diary products. The estimation of usual intake from data sets with a substantial amount of zero intakes became feasible by modelling separately zero intake on part or all of the days via the estimation of 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
1. estimating the percentiles of the distribution of usual intakes in the original scale.

Back to: Discrete/semiparametric (Nusser)

5.4.6.1 Power or log transformation

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

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

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

5.4.6.2 Spline fit

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

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

5.4.6.3 Estimation of the parameters of the usual intake distribution

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

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

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

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

$$z_p = \Phi^{-1}(p).$$

5.4.6.4 Back transformation and estimation of usual intake

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

1. First the 400+ values x_i are expanded in a set of 9 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 ($\theta_m; \quad m = 0,..., M$) is estimated by optimising the fit of the predicted proportions of consumers with 0, 1, 2, ... intake days to the observed proportions. The number of parameters to be estimated is usually higher than the number of possible outcomes for a consumer (e.g. 3 when there are two days per consumer), and therefore a smooth approximation is made using a modified minimum chi-squared estimator. See Dodd (1996) for details. Only the fraction of non-consumers (θ_0) is estimated separately with no restriction to be similar to the other θ_m. It can be noted that the distribution of individual intake probabilities can be better estimated when the number of days per consumer in the consumption survey becomes higher. With only 2 days per consumer the procedure gives a rather artificial distribution, often with an estimated θ_0 of zero

This step can be time-consuming. Therefore, the number of iterations in the estimation procedure can be limited by the user. In our experience it is not generally necessary to use 50,000 iterations as in Dodd (1996).

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

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

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

5.5 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 foods. However, these calculations do not consider the amount of *uncertainty* that is due to the limited size of the underlying datasets. Typically, in a large number of simulations very many different combinations of consumption and compound concentrations are made. This leads to a smooth distribution of simulated intakes, and the impression of a very precise estimation of 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 95th 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 MC-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 foods. 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[th] , 25th, 75th and 97.5[th] 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

5.6 Binning

Binning is a method to summarise the simulated data (total intake, intake per food, consumption per food, concentration per food) in frequency intervals for further use in deriving the 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* x 1.01^i. For values below the mean also 1100 bins are defined with lower limits defined by *mean* x 1.01^{-i}. After the process of binning is completed, the quantile value of a specific percentile is determined by linear interpolation between the bin limits. These 2200 bins together provide efficient storage for numbers spanning more than 9 decades ($1.01^{2200}=3.2\times10^9$), 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 foods with nonzero intakes. Also in this case a warning is given.

6 Selection of consumers and foods

In the next sections, the centrally supplied database 'mcra_5.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, 1194 foods are eaten. Application of the conversion algorithm shows that the number of derived foods is 19 + 54 = 73. Note that the selected foods do not depend on the choice of the compound. The retrieved foods are entirely determined by consumption patterns, e.g. the choice of a survey.

The information in the lower part of the screen is related with the choice of the compound. Here, chlorpyrifos is selected and this compound is found on 19 derived foods. This means that there is at least one sample of each of the 19 foods with a positive concentration. On 54 derived foods only non-detects are found. This means that for each of the 54 foods all samples were negative, i.e. no chlorpyrifos found. No worstcase values were found in the database.

Checking one of the two radio buttons to specify which foods must be used in the risk analysis. Press the 'go'-button in Figure 50 to enter the central menu. If you want to do the same type of analysis for one of the other options, click the to-overview-of-conversion 'go'-button in the centralmenu.

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

If non-detects are not replaced by the LOR (see input form), the first and second radio button in Figure 50 are equivalent. Note that the first button is computationally more efficient because less foods are involved.

With the subset-selection 'go'-button in the MCRA central menu (see Figure 51) subset selections are made on the consumer population or on the foods. Through the use of scroll-down menus the user 1) selects ages, weights and sexes to define the population of consumers; 2) selects day(s) of consumption, foods, year, country and sampling type of the compound concentration data. In Figure 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.

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). Here, the consumer population is defined using characteristics on consumers (age, weight, sex). The first time, the characteristics of the total population are shown, so for age the minimum and maximum values are 1 and 97 year (both current selection as database). Note that combining levels of variables occasionally results in empty subsets e.g. the combination of ages within the range 1 to 8 and weights in the range 70 kg to 150 kg is an empty subset. So avoid making combinations of variables that yield empty subsets. The number of consumers currently selected is always mentioned in the screen, e.g. 6250 (see Figure 52).

Figure 52: MCRA subset selection: consumer population

To make a selection of consumers, check one or more of the checkboxes and click the upper 'go'-button. The screen in Figure 53 appears. Choose values for age and weight (the unit depends on the survey you are currently using, see table FoodSurvey in 9.4.3) 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.

Figure 53: MCRA subset selection: select levels for consumers

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

for both age and weight the radio button 'database', check both checkboxes and press the 'go'-button again. Then, your original population with 6250 consumers is selected again.

6.2 Subset selection: consumption and concentration data

Select the total population. The next screen shows you the current selected levels of each variable (see Figure 54). Also the levels of each variable stored in the database are shown. In our example, the number of derived foods with positive concentration values is 19 and these foods are derived from 197 consumed foods.

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

In general, do not select combinations of levels which may result in empty subsets, e.g. grapefruits are not grown in the Netherlands. So, combination of these levels for foodname and country results in empty subsets. In this example, it is quite clear why things go wrong. Often it is less clear what you might expect from a combination of levels, especially when 'current selection' is combined with 'database' in order to retrieve the original levels for a variable, so be aware of empty subsets.
The example is continued by checking the checkboxes for derived and consumed foods.Click the upper 'go'button. A new screen appears with scroll-down menus for derived and consumed foods(see Figure 55) Choose the three BAMI GORENG foods in the first scroll-down menu and click the upper 'go'-button.

MCRA subset selection

Select levels

derived food: total number of levels: 19

STRAWBERRY	FB0275
TABLE-GRAPES	FB1235
GRAPEFRUIT, SEE ALSO SUBGROUP 0005 SHADDOCKS OR PO	FC0203
LEMON, SEE ALSO SUBGROUP 0002 LEMONS AND LIMES	FC0204
LIME, SEE ALSO SUBGROUP 0002 LEMONS AND LIMES	FC0205
MANDARIN, SEE ALSO SUBGROUP 0003 MANDARINS	FC0206
ORANGE, SWEET, SEE ALSO SUBGROUP 0004 ORANGES, SWE	FC0208
BANANA	FI0327
MANGO	FI0345
APPLE	FP0226
PEAR	FP0230
CHERRY, SWEET	FS0244
NECTARINE	FS0245
PEACH	FS0247
CELERY LEAVES	HH0624

consumed food: total number of levels: 197

BAMI GORENG CHINESE NOODLE DISH CANNED	&NL$18$0368
FRIED RICE INDONESIAN STYLE WITH MEAT CANNED/FROZE	&NL$18$0371
CHILDRENS MEAL GREEN AVERAGED	&NL$18$0372
BABI PANGANG PORK CHINESE STYLE IN SPICY SAUCE	&NL$18$0469
BAMI GORENG CHINESE NOODLE DISH W/O EGG	&NL$18$0470
CHOP-SUEY INDONESIAN STYLE	&NL$18$0472
SALAD RUSSIAN	&NL$18$0577
SMALL SALAD RUSSIAN	&NL$18$0614
SPAGHETTI BOLOGNESE FROZEN	&NL$18$0646
GADO GADO W/O RICE AND EGG	&NL$18$0853
INFANT MEAL GREEN AVERAGE	&NL$18$1278
INFANT MEAL COMPLETE	&NL$18$1279
TURLU VEGETABLE DISH WITH MEAT READY TO SERVE TURK	&NL$18$1377
DOLMA PATLICA STUFFED AUBERGINE TURKISH	&NL$18$1378
PIE W VEGETABLES/CHEESE/HAMFILLING	&NL$18$1493

subset selection based on chosen levels [go]

back to previous screen (no selection in this step) [go]

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

Figure 55: MCRA subset selection: select levels

The system automatically adjusts the levels of all variables: all foods that are not an ingredient of a BAMI GORENG food are removed as you can check by pressing the 'show'-button in the central menu. Here, the steps above are repeated (check foodname checkbox and press upper button). In Figure 56, the selected foods are shown: LEMON, CELERY and SWEET PEPPER are ingredients of BAMI GORENG foods.

Figure 56: MCRA subset selection: select levels

7 On-line editing of data (not implemented for Release 5)

Instead of choosing data from already formed databases as discussed in 2.6.1 , it is also possible to edit data on-line. Choose 'MCRA 5.0 (field trial data and Dutch consumption data)' in the MCRA main menu (see Figure 2). A series of menus is reached where you can add and edit on-line your own concentration data. This is a safe way to combine your concentration data with centrally supplied food consumption data because food codes for the concentration data are automatically generated from the food consumption data.

From the central menu (see Figure 6), click the start-selection-of-consumption-and-compound-tables 'go'-button. Then the available databases are shown (see Figure 57).

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

Check one or more databases and click the 'go'-button. The table selection menu appears (see Figure 58). Note that its layout is being changed displaying only tables related to consumption e.g. Foodconsumption, Foods and Individual. It is compulsory to select these three tables (see also 2.6.1).

83

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.
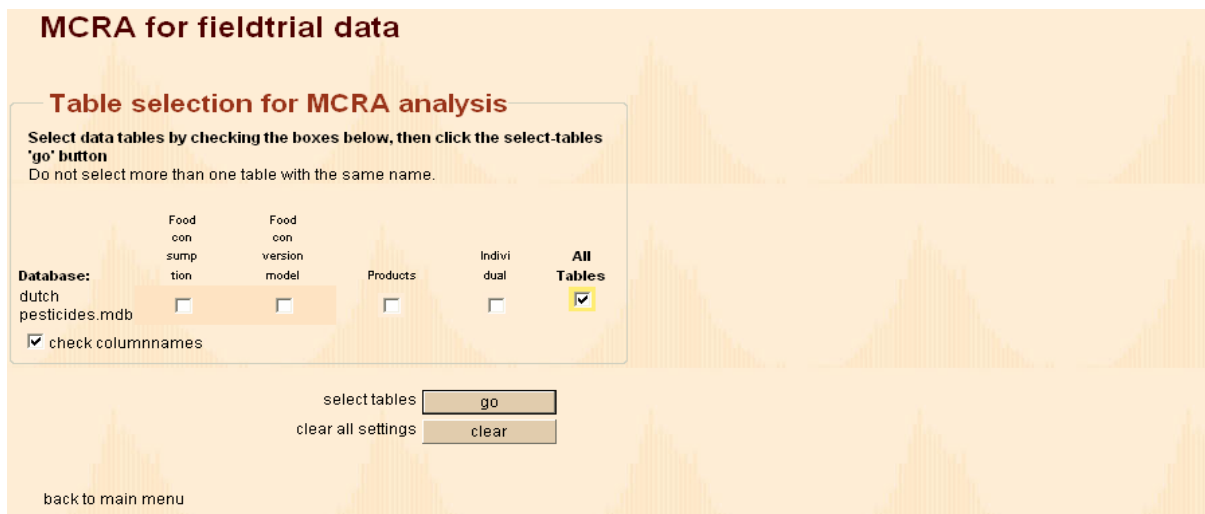
Figure 58: Table selection menu, MCRA for field trial data

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 59). 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 foods and a proper functioning of the menu is not guaranteed anymore.
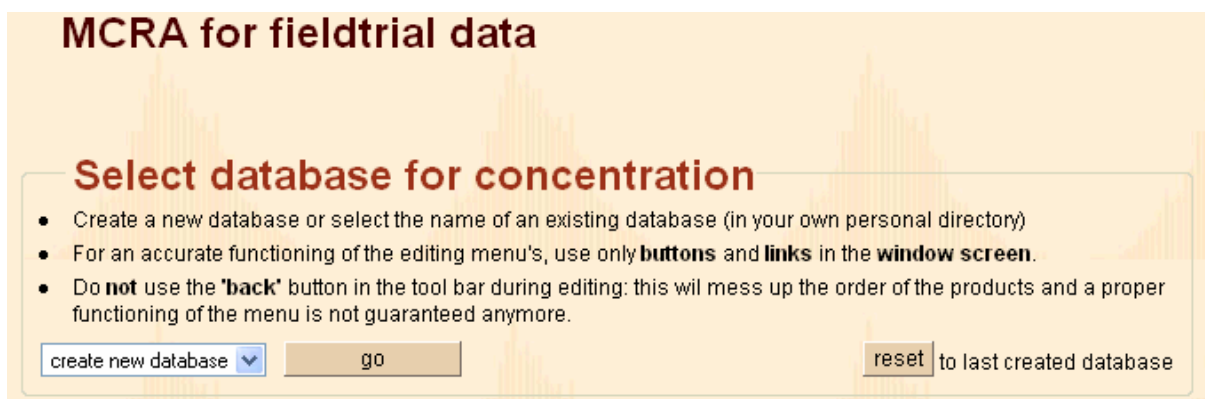
Figure 59: Select database for concentration, create new database

Specify the name of the database (see Figure 60) and press the 'go'-button. Note you should only use letters and digits, no other characters are allowed.

Figure 60: Select database for concentration, specify a new database name

This brings you to Figure 61 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.

Figure 61: 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 62) if you are entering values for the first time.

Figure 62: 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 foods. Choose one of foods in the scroll-down menu for which you want to add values (see Figure 63) and click the enter-food 'go'-button.

Figure 63: Select food

In Figure 64, for food 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-food 'go'-button for entering values for a new food.

Figure 64: Enter concentration data

Click the list-records link (see Figure 64) for an overview of the concentration values in the database (see Figure 65).

Figure 65: List concentration values

Click 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 food (see Figure 66).

Figure 66: Edit concentration value

7.2 List or add variability factors

After supplying all concentrations values, click the back-to-fieldtrial-data-menu link (see Figure 64), the user is brought back to the select data menu (see Figure 61). At this stage, all foods 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 foods. The process of entering variability factors and/or processing factors for new foods 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. Click the add- records link, to enter the menu for adding variability factors (see Figure 67). The variability factor itself and the number of units of the composite sample should be specified. For each food in the concentration table, values should be supplied. Only foods for which no values are supplied are displayed and the process of supplying values is repeated until all necessary values are given. Then and only then, the message '*** Variability factors are specified for all foods present in the concentration table ***' or '*** For all foods, variability factors are already specified ***' is displayed. Records may be changed afterwards by clicking one of the grey buttons in table displaying a listing of all values in the database.

Figure 67: Variability factors

7.3 List or add processing factors

In Figure 68 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-food 'go'-button.

Figure 68: Processing factors

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

To enter processing factors for only a few number of foods and processing type combinations, enter your values and click repeatedly the 'select-next-food'-button until the message '*** Processing factors for foods are specified ***' is displayed. If you want to enter values at a later stage, just click the 'list or add processing factors'-button in the select data menu. A list of processing factors found in the database is given and after clicking the add-records link, new processing values may be entered. Note that the scroll-down menus are dependent on the food and contain only those processing types for which no values are supplied. Records may be changed afterwards by clicking one of the 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 61 to enter values for the acute reference dose (ARfD) or average daily intake (ADI). This brings you to Figure 69.

Figure 69: 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, eigth edition, VSN International Ltd., Oxford.
- Hamey, PY (2000). A practical application of probabilistic modelling in assessment of dietary exposure of fruit consumers to pesticide residues. *Food Additives and Contaminants*, 17: 601-610.
- Harris, C *et al.* (2000). Summary report of the International Conference on pesticide residues variability and acute dietary risk assessment. *Food Additives and Contaminants*, 17: 481-485.
- Harter, HL. Expected values of normal order statistics. Biometrika 48: 151-165.
- IUPAC (1995). Nomenclature in evaluation of analytical methods including detection and quantification capabilities (IUPAC Recommendations 1995). *Pure and Applied Chemistry* 67: 1699-1723.
- JMPR (1999, 2000). Reports of the joint FAO/WHO meetings of experts on Pesticide residues in food.
- Kistemaker C, Bouman M and Hulshof KFAM (1998). De consumptie van afzonderlijke producten door Nederlandse bevolkingsgroepen - Voedselconsumptiepeiling 1997-1998. Zeist, TNO-Voeding (Report No: 98.812).
- Nusser SM, Carriquiry AL, Dodd KW & Fuller WA (1996). A semi-parametric transformation approach to estimating usual daily intake distributions. *Journal of the American Statistical Association*, 91: 1440-1449.
- Nusser SM, Fuller WA, and Guenther PM (1997). Estimating usual dietary intake distributions: adjusting for measurement error and nonnormality in 24-hour food intake data. In: Lyberg L, Biemer P, Collins M, DeLeeuw E, Dippo C, Schwartz N, and Trewin D (editors), *Survey Measurement and Process Quality*, Wiley, New York. p. 689-709.

- Pearson, ES and Hartley, HO. Biometrika tables for statisticians (1977). Vol II.
- Shimizu, K, and Crow, EL (eds). (1988). Lognormal distributions: theory and applications. Marcel Dekker, INC. New York.
- Snedecor, GW & Cochran, WG (1980). Statistical Methods (7th edition). Iowa State University Click, Ames, Iowa.
- van der Voet, H, de Boer, WJ & Keizer, LCP (1999). Statistical instruments for dietary risk assessment concerning acute exposure to residues and contaminants. Report August 1999, Centre for Biometry Wageningen, Wageningen.
- van der Voet, H, de Boer, WJ & Boon, P (2001). Modelling exposure to pesticides. Note HVT-2001-03, Centre for Biometry Wageningen, Wageningen.
- van Dooren, MMH , Boeijen, I, van Klaveren, JD and van Donkersgoed G (1995). Conversie van consumeerbare voedingsmiddelen naar primaire agrarische produkten. RIKILT-report. Wageningen, RIKILT-DLO (Report No: 95.17).
- van Klaveren, JD (1999). Quality programme for agricultural products. Results residue monitoring in the Netherlands. RIKILT Institute of Food Safety, Wageningen.

9 APPENDIX A: Preparing the data

9.1 Overview

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

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

See Table 20 for an overview. To run MCRA, tables called 'required' should always be supplied. Selection of 'additional' tables depends on subset requirements (see Ch. 5) and model specifications (see Ch. 3).

Required tables (minimal configuration)

Food	food codes and labels (see 9.3.1)

Food consumption	**Description**
FoodConsumption	consumption of foods (see 9.4.1)
Individual	consumer characteristics (see 9.4.2)

Concentration of compounds	**Description**
Compound	compound codes, labels, agricultural and toxicological limits (see 9.5.1)
Country	country codes, labels (see 9.5.2)

One and only one of:

ConcentrationValues	compound concentration data (full data) (see 9.5.3)
ConcentrationSummaryStatistics	compound concentration data (as summary statistics) (see 9.5.4)
ConcentrationDiscreteValues	compound concentration data (table of frequency counts) (see 9.5.5)

Additional tables (for querying or specific options)

FoodProperties	food codes and labels, and food specific information(see 9.3.2)
FoodComposition	food codes and labels, compositions (see 9.3.3)
FoodMarketShare	food codes and labels, marketshares (see 9.3.4)

Food consumption	**Description**
FoodSurvey	Name of survey (see 9.4.3)
ProcessingType	processing codes and labels (see 9.4.4)
Processing	processing factors (see 9.4.5)

Concentration of compounds	**Description**
VariabilityProd	unit variability factors (see 9.5.6)
VariabilityCompProd	unit variability factors, compound-specific (see 9.5.7)
VariabilityProcCompProd	unit variability factors, processing- and compound-specific (see 9.5.8)
AgriculturalUse	information on the agricultural use of compounds (e.g. use allowed, percent crop treated) (see 9.5.9)
ConcentrationWorstcaseValues	information on worstcase values (e.g. compound and food specific worstcase values) (see 9.5.10)

Table 20: Overview of tables

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

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

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

9.2 Harmonised CODEX codes

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

9.2.1 Definitions

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

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

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

The first symbol should be:
- a letter (indicates a CODEX code or a code derived from a CODEX code), or
- & followed by a 2-letter country code (indicates a national food code)

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

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

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

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

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

9.2.2 Conversion of codes

We distinguish 3 types of food code conversion, which may be provided in three different tables:
1. Food processing (table **Processing**)
    Processing factors will be applied to concentration data.
2. Food composition (table **FoodComposition**)
    Composition percentages will be used to transform the consumed amounts.
3. Subtypes/ Market share (table **FoodMarketShare**)
    Market share percentages will be used for as probabilities of selecting concentration data for each of the subtypes.

9.2.3 Conversion rules

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

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

9.2.4 Scheme to link food consumption and compound concentration data

Find in Figure 70, an schematic outline of the search for food codes. After a successful search, the code is found in one of the concentration values tables. If the code is not found, searching starts in one of the link tables. If the code is found, the search starts again in one of the concentration values tables and the search is repeated. If a code is not found at all, the search for a code is unsuccessful.
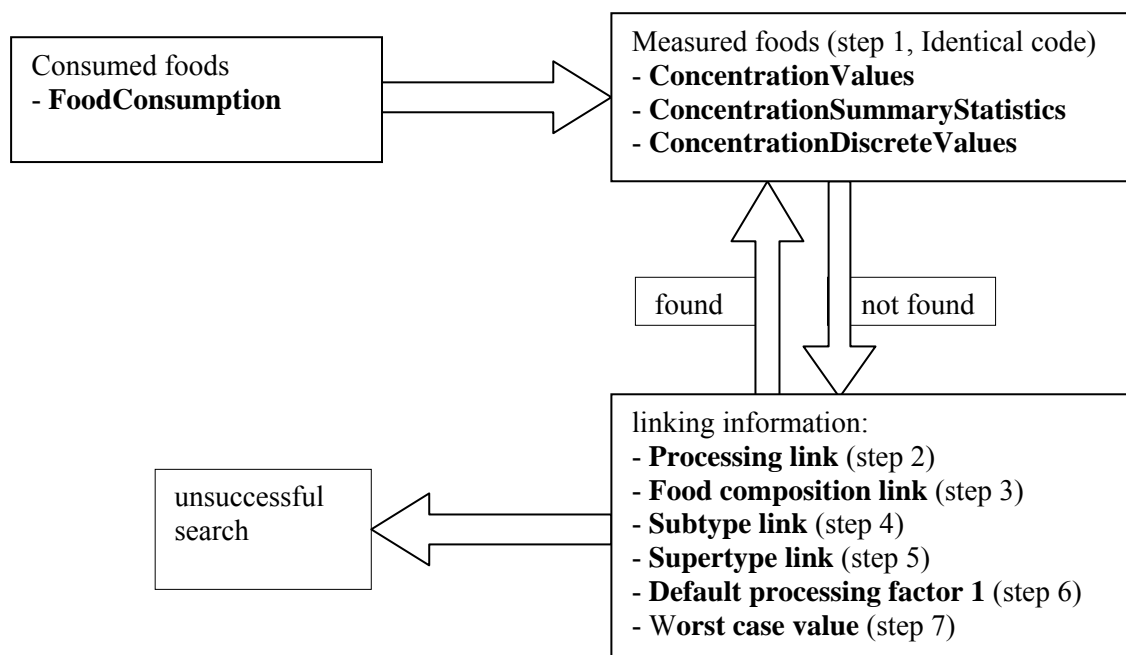
Figure 70: Link scheme to find CODEX codes

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

Step 2*. (Processing link)
    a. Try to find *code* in the field **foodprocessed** of the **Processing** table (9.4.5), and convert to the code specified in the field **foodunprocessed**.
    b. (wildcard match) Try to find a wildcard match code in the field **foodprocessed** of the **Processing** table. Wildcard match codes consist of an initial string (*startcode*, may be empty), an asterisk (*), and possibly a processing part (-*processingtype*). * may be any string *endcode* (not containing a -) such that *code* equals *startcodeendcode* or *startcodeendcode-processingtype*.
        a. If *code* contains a processing part (-*processingtype*), then the wildcard match code should also end with -*processingtype*. Convert to the code specified in the field **foodunprocessed**, where *endcode* is substituted for any * in the new code.
        b. If *code* contains no processing part, then the wildcard match code should also contain no processing part. Convert to the code specified in the field **foodunprocessed**, where *endcode* is substituted for any * in the new code.

Step 3. (Food composition link)
    a. Try to find *code* in the field **food** of the **Foodcomposition** table (9.3.3), and convert to one or more ingredient codes found in the field **ingredient**
    b. If *code* contains a processing part (*maincode-processingtype*), then try to find *maincode* in the field **food** of the **Foodcomposition** table. Convert to one or more ingredient codes and add -*processingtype* to the new codes.

Step 4. (Subtype link)

Starting from *code*, try to find subtype codes equal to *code$** in the field **foodtype** of the **FoodMarketShare** table (9.3.4 , where the strings represented by the wildcard are not allowed to contain a $ themselves (meaning that we look only for codes one level down in the type hierarchy).

Check that for the selected codes the market share percentages in the field **marketshare%** sum to 100%.

> If this is not OK, then the result depends on the user option "Allow conversion to subtypes not summing to 100% and rescale".
>
> > 1. If this option is allowed, then the marketshare percentages are rescaled to a sum of 100 %.
> > 2. If this option is not allowed, then the search in step 4 is considered unsuccessful.

Step 5. (Supertype link)

This step will only be taken if the user has explicitly allowed this option.

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

Step 6. (Default processing factor 1)

If *code* contains a processing part (*-processingtype*), then remove it.

Step 7. (Worst case value)

This step will only be taken if the user has explicitly allowed this option.

Try to find *code* in the field **food** of the **ConcentrationWorstcaseValues** table (9.5.10). If found, the search is ended successfully.

9.2.5 Example of use

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

individual	dayofsurvey	foodconsumed	amountconsumed	foodsurvey
1012	1	&NL070251	150	DNFCS

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

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

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

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

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

where the field **proctype** specifies explicitly the type of processing (2 = peeling in this case), and fields **procnom** and **procupp** are processing factor nominal and upper values.

The **FoodMarketShare** table (9.3.4) may contain market shares for subtypes of apple:

foodtype	marketshare%
FP0226$Elstar	30

FP0226$JonaGold	70

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

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

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

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

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

9.3 Food tables

9.3.1 Food (compulsory)

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

- Foodname2 is used for alternative foodnames.

Example:

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

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

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

Example:

9.3.3 FoodComposition (optional)

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

- Specifies the composition of foods and corresponding proportions.

Example:

9.3.4 FoodMarketShare (optional)

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

- Specifies food marketshares of subtypes.

Example:

9.4 Food consumption tables

9.4.1 FoodConsumption (compulsory)

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

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

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

Example:

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

FoodConsumption : Table

9.4.2 Individual(compulsory)

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

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

Example:

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

Individual : Table

9.4.3 FoodSurvey (optional)

field name	description
foodsurvey (text)	name of survey
year (number)	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 : Table

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

9.4.4 ProcessingType (optional, (for processing compulsory))

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

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

Example:

ProcessingType : Table

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

9.4.5 Processing (optional)

field name	Description
compound (text)	compound code

foodprocessed (text)	food code processed
foodunprocessed (text)	food code unprocessed
proctype (number)	code of processing type
procnom (number)	nominal value (best estimate) of processing factor
procupp (number)	upper value (estimate of 97.5th percentile or "worst case" estimate) of processing factor

- Procupp should be higher than procnom.

Example:

compound	foodprocessed	foodunprocessed	proctype	procnom	procupp
C01001001	&NL1-2	&NL1	2	0.51	0.8
C01001001	&NL42-15	&NL42	15	0.25	0.73
C01001002	AA*	BB*	11	0.114	0.167

9.5 Compound concentration tables

9.5.1 Compound (compulsory)

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

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

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

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

Example:

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

9.5.2 Country (compulsory)

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

- No missing values allowed

Example:

9.5.3 ConcentrationValues (optional)

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

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

Example:

9.5.4 ConcentrationSummaryStatistics (optional)

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

(number)	statistic for the percentile
percentile (number)	specifies the percentage of the statistics **perc** and **percall**

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

Example:

ConcentrationSummaryStatistics : Table

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

9.5.5 ConcentrationDiscreteValues (optional)

field name	description
compound (text)	compound code
foodmeasured (text)	food code
country (text)	code for country
limitofreporting (number)	limit of reporting (mg/kg)
numberofsamples (number)	size of sample (detects and non-detects)
c%01 (number)	• 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).
c%02 (number)	
c%05 (number)	
c%1 (number)	
c%2 (number)	
c%5 (number)	
c1 (number)	
c2 (number)	• classes (*i.e.* columnnames) are free to choose so *c<xxx>*… may be replaced with any appropriate concentration e.g. c5, c10 etc.
c<xxx> (number)	
c<xxx> (number)	
c<xxx> (number)	
cE10 (number)	

- Field names representing the number of frequency counts are constructed as follows:
  - c = indicates class limit,
  - % = represents the decimal point (if necessary),
  - *xx* = is the value of the class limit.
- Thus: field name c%02 specifies class limit 0.02, field name c2 specifies class limit 2, field name cE10 specifies class limit $1*10^{10}$.

- The number of non-detects measurements is given as the difference between the numberofsamples and the sum of frequency counts, e.g. see example first record 377 – 1 = 376,.
- Missing LORs are reported as –9999. The MCRA program replaces missing LORs with 1) the maximum LOR found in the database, 2) if all LORs are missing…... A warning is generated when 1) and 2) are not possible.
- For columns numberofsamples, c%02…cE10 no missing values is allowed: classes without frequency counts are reported as 0. When no data are available for a food, delete the entire row.

Example:

ConcentrationDiscreteValues : Table

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

9.5.6 VariabilityProd (optional)

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

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

Example:

VariabilityProd : Table

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

9.5.7 VariabilityCompProd (optional)

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

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

Example:

9.5.8 VariabilityProcCompProd (optional)

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

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

Example:

9.5.9 AgriculturalUse (optional)

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

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

Example:

9.5.10 ConcentrationWorstcaseValues (optional)

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

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

Example:

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

10 APPENDIX B: Other output depending on specifications of options

Appendix B shows model output (see options in Ch. 3). See also 4.1 and 4.2 .

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 21 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:
processing factors              processing (fixed factors)
additional                      system defaults
output                          system defaults
```

Table 21: Input form options: processing

In Table 22 you find the main characteristics concerning this analysis taken from the logfile. Note that the number of foods and processing combinations is 52 whereas the number of foods is 12.

```
Logfile
Number of foods                 8
Acute reference dose (ARfD)     100
Average daily intake (ADI)      10
Number of detects               100
Number of non-detects           938
Number of foods and processing
type combinations               10
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    10718
```

Table 22: Information in logfile

For a summary of the data, see Table 5. In Table 23 you find a summary of the simulated intakes. Compared to Table 6, this table contains a second section with information on processed foods 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.

```
SUMMARY OF SIMULATIONS OF CONSUMPTIONS AND COMPOUND CONCENTRATIONS WITH RESPECT TO:
*********************************************************************

 Compound:  CHLORPYRIFOS
```

```
*********************************************************************

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

FOOD CONSUMPTION DATA

       Code          Food MeanConsum   DeltaC MeanConsDays  NConsDay    NDays %ConsDays
                             (g)        (%)        (g)                           (%)
     FB0275 STRAWBERRY      17.37    ( -0.4)     35.56      97700    200000     48.9
     FB1235 TABLE-GRAPES    68.84    ( -1.0)    178.21      77258    200000     38.6
     FC0206 MANDARIN, SE    11.20    (  0.8)     51.87      43195    200000     21.6
     FC0208 ORANGE, SWEE    73.47    ( -1.1)    128.29     114530    200000     57.3
     FP0226 APPLE          100.25    ( -0.3)    162.27     123560    200000     61.8
     FS0245 NECTARINE        0.78    ( -2.6)    117.81       1329    200000      0.7
     FS0247 PEACH           11.77    ( -1.0)     41.56      56622    200000     28.3
     VO0445 PEPPERS, SWE    12.96    ( -0.5)     64.40      40247    200000     20.1




COMPOUND CONCENTRATION DATA

       Code         Food   MeanConc   DeltaR  NSamplPos NSamples %SamplePos ProcFact
                           (mg/kg)    (%)                          (%)
     FB0275 STRAWBERRY      0.0006   ( -3.8)     1250      97700     1.3      0.99
     FB1235 TABLE-GRAPES    0.0092   (  3.1)     8919      77258    11.5      0.99
     FC0206 MANDARIN, SE    0.0442   ( -0.2)     5783      45435    12.7      0.78
     FC0208 ORANGE, SWEE    0.0230   (  0.0)    24326     114530    21.2      0.99
     FP0226 APPLE           0.0016   ( -0.5)     4512     123560     3.7      0.99
     FS0245 NECTARINE       0.0113   (  1.9)      121       1329     9.1      0.99
     FS0247 PEACH           0.0034   (  1.8)     5487      56622     9.7      0.99
     VO0445 PEPPERS, SWE    0.0012   ( -5.4)      830      40247     2.1      0.99




=====================================================================================================

FOOD CONSUMPTION DATA

       Food & Processing   MeanCons  DeltaC MeanConsDays NConsDays   NDays %ConsDays ProcFact
                            (gr)      (%)       (gr)                          (%)
 MANDARIN, SE:canned/cons   0.39    ( 1.3)     38.31       2012    200000     1.0      0.77
 MANDARIN, SE:juicing       0.76    ( 1.6)     14.93      10154    200000     5.1      0.11
 MANDARIN, SE:unknown      10.06    ( 0.7)     60.47      33269    200000    16.6      0.99
=====================================================================================================
```

Table 23: Summary of simulation, including processed foods

For processing, all tables as found in 4.1 are extended with the additional information on the processed food. Note that the percentiles of the acute intake distribution after applying processing factors, see Table 24, 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 24: 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 25 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 25: Input form options: unit variability

In Table 26 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 foods                8
Acute reference dose (ARfD)    100
Average daily intake (ADI)     10
Number of detects              100
Number of non-detects          938
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   10718
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.


Maximum number of units in a consumption:   23
  for product(s):                  TABLE-GRAPES
```

Table 26: Information in logfile

In Table 27, the percentiles of the intake distribution after applying unit variability are shown. Note that the percentiles are much higher (p99.99 = 14.793) than without unit variability (p99.99 = 7.618, see Table 9) and after applying processing (p99.99 = 6.06, see Table 24).

```
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         0.05363
       95.00         0.24068
       99.00         1.32902
       99.90         5.71742
       99.99        14.79337
        mean         0.06141
     maximum        24.26852
```

Table 27: 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 28 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 28: Input form options: IESTI

In Table 29 you find the main characteristics concerning this analysis taken from the logfile.

```
Logfile
Number of foods              8
Acute reference dose (ARfD)  100
Acceptable daily intake (ADI) 10
Number of detects            100
No of consumers              6250
Number of non-detects        938
Population characteristics,
      minimum age            1
      maximum age            97
      minimum weight         8
      maximum weight         150
      Sex                    female, male
Total no of consumption days 10718
```

Table 29: Information in logfile

In Table 30 IESTI estimates are displayed together with the estimate expressed as percentage of the ARfD (%ARfD). Also a comparison is made with the MC-percentile per food for positive consumption days only (ConsPos) as well as all consumption days (AllDays). For Orange the IESTI estimate is 1.067 microgr/kg bw/day which is slighter lower than the MC-percentiles for positive consumption days only (1.341 microgr/kg bw/day) and slighter higher than the MC-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
ConsPos: percentiles per food for positive consumptions only
AllDays : percentiles per food 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 foods)


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


                                            ConsPos    AllDays
Food                      IESTI     %ofARfD  p99.00%    p99.00%

STRAWBERRY                0.278     0.278    0.006      0.000
TABLE-GRAPES             1.190     1.190    0.753      0.318
MANDARIN, SE             1.435     1.435    0.767      0.320
ORANGE, SWEE             2.550     2.550    1.068      0.731
APPLE                    0.316     0.316    0.154      0.096
NECTARINE                2.160     2.160    0.624      0.000
PEACH                    0.153     0.153    0.080      0.027
PEPPERS, SWE             0.249     0.249    0.030      0.000
```

Table 30: IESTI estimates

10.4 Acute risk assessment: uncertainty

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

```
Input form
risk type                    acute
uncertainty analysis         yes
```

```
concentration model              empirical
number of Monte Carlo 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      100
number of iterations per
bootstrap sample                 20000
concentration data:              system default
Additional                       system defaults
Output                           system defaults
```

Table 31: Input form options: uncertainty

In Table 32 you find the main characteristics concerning this analysis taken from the logfile. Consumptions and concentration data are bootstrapped.

```
Logfile
Number of foods                  8
Acute reference dose (ARfD)      100
Average daily intake (ADI)       10
Number of detects                100
Number of non-detects            938
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     10718
Total no of simulations          2000000
(uncertainty)
Bootstrap consumptions           yes
Bootstrap concentration data     yes
```

Table 32: Information in logfile

In Table 33, 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 (7.61810 microgr/kg bw/day) is (3.3433, 14.3831).

```
************************************************************************
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       0.13733
      95.00       0.32396
      99.00       1.07677
      99.90       3.65245
      99.99       7.61810
      mean        0.06194
```

```
        maximum          12.62322



                Uncertainty of percentiles distribution
                      2.5%           25%           75%         97.5%
          50.00       0.0000        0.0000        0.0000        0.0000
          90.00       0.0960        0.1198        0.1628        0.2073
          95.00       0.2138        0.2871        0.3777        0.4627
          99.00       0.6728        0.9086        1.2553        1.6024
          99.90       1.7681        2.6031        4.0354        5.9873
          99.99       3.3433        5.7470        8.8052       14.3831
           mean       0.0412        0.0522        0.0714        0.0899
        maximum       4.4471        7.1227       12.8137       17.8488
```

Table 33: Percentiles for the intake distribution and uncertainty

The information of Table 33 is plotted in Figure 71. 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.
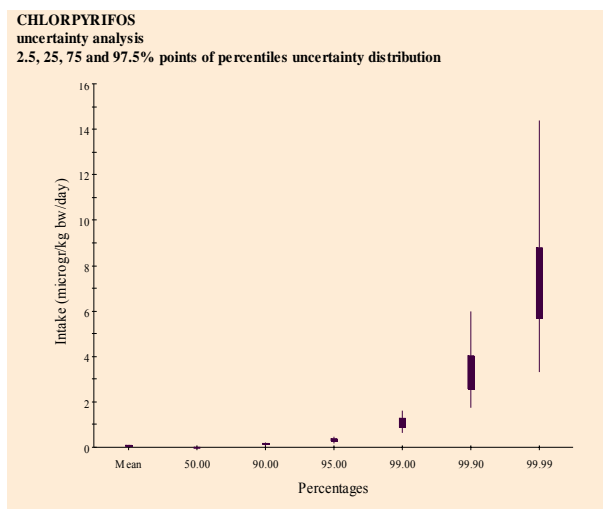
Figure 71: Uncertainty analysis

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 31 and Table 32, 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 MC-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 MC-iterations, the number of iterations per bootstrap and the number of bootstrap samples used. Recall that we ran the analysis with 200.000 MC-iterations and 100 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 MC-variability decreases, Therefore, through the observed variances a monotone decreasing spline function is fitted. For each variance the 90% confidence limits are calculated.

The uncertainty analysis provides an estimate of the variance of the percentiles derived from the 100 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 72 diagnostics are displayed for percentage point 50%. We can skip this plot because the percentile is 0.0000 microgram/kg bw/day. For percentage point 90% in the right plot we have an estimate of the percentile (0.13733 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 4.1% to the total bootstrap variability. For a theoretical bootstrap sample of size 200.000 the MC-contribution would be 0.0%.
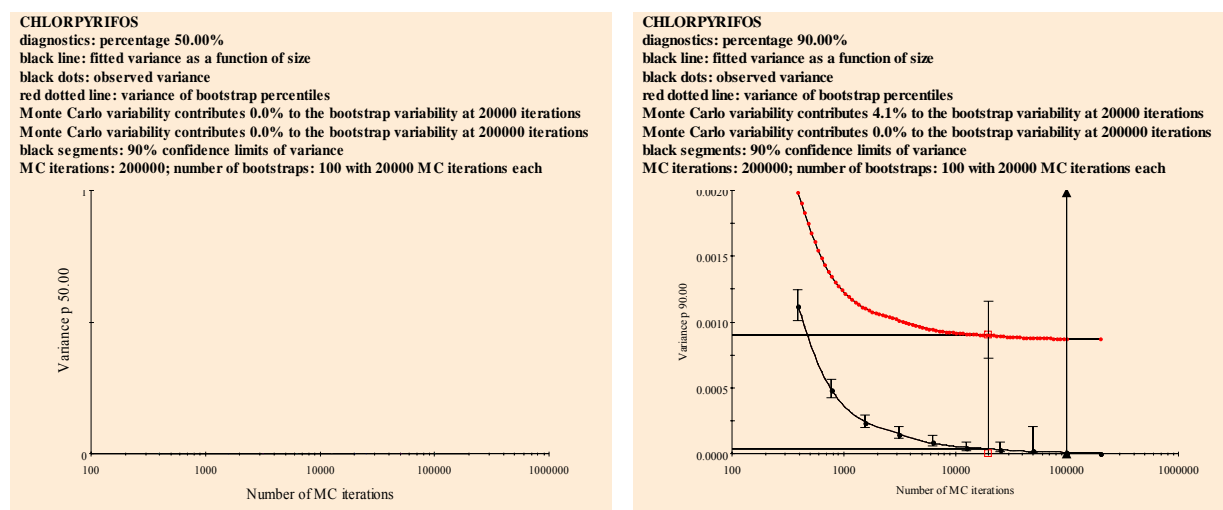
Figure 72: Diagnostic graphs for percentage 50% and 90%

In Figure 73, the diagnostics for percentage point 95% and 99% are displayed. The contribution of the MC-variability for p95 and p99 is 3.5% and 2.3%, respectively, indicating that these higher percentages have stable percentiles. 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.
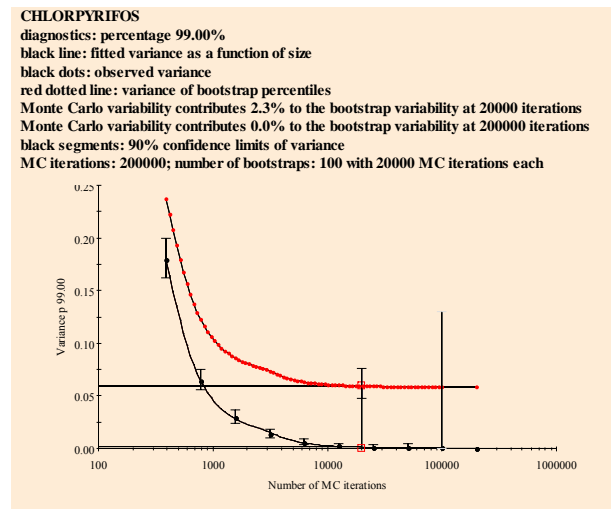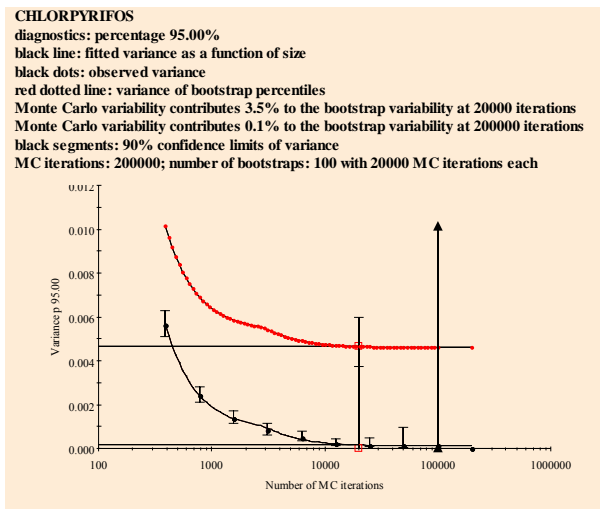
Figure 73: Diagnostic graphs for percentage 95% and 99%

In Figure 74 the diagnostics for percentage point 99.9% and 99.99% are displayed. The contribution of the MC-variability for p99.9 and p99.99 is 5.3% and 31.2%, respectively, indicating that, especially the last percentile is 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 74 not all estimated variances are displayed. As the number of samples in a partition increases, sample size decreases. This restricts the number of available percentiles (maximum possible percentage: (100 – 100/size)).
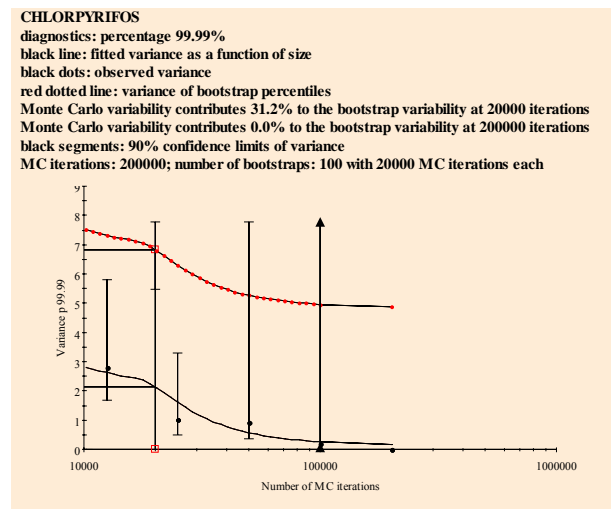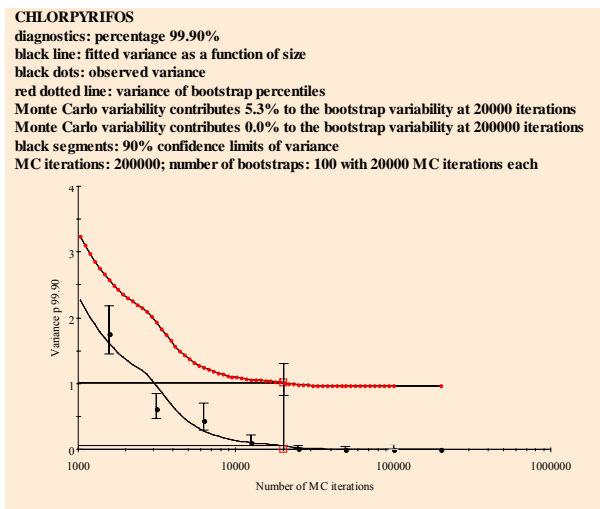
Figure 74: Diagnostic graphs for percentage 99.9% and 99.99%

10.6 Acute risk assessment: betabinomial distribution 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 34 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 34: Input form options: acute risk and betabinomial/lognormal with age

In Table 35 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(intake), is based on REML and a spline function to model age effects.

```
Logfile
Number of foods                   8
Acute reference dose (ARfD)       100
Acceptable daily intake (ADI)     10
Number of detects                 100
No of consumers                   6250
Number of non-detects             938
Population characteristics,
        minimum age               1
        maximum age               97
        minimum weight            8
        maximum weight            150
        Sex                       female, male
Total no of consumption days      10718
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    0.0001   29064.531     6244    0.98       1      0.323
```

```
    3        2    0.0001    29065.508    6245    1.25    1    0.264
    2        2    0.0001    29066.755    6246    1.61    1    0.204
    1        2    0.0001    29068.368    6247    0.03    1    0.854
    0        1    0.0001    29068.401    6248    -       -    -
* Degrees of freedom of spline according to backward deviance testing is 0.

Regression analysis
===================
Estimates of parameters
-----------------------
Parameter                        estimate         s.e.      t(*)
Constant                         -1.01026      0.00506   -199.53
Estimate of overdispersion parameter phi = 0.0001000  with s.e. = 0.0005139


Intake Amounts: REML model with age
===================================
* Automatic selection of spline model.
Estimates of parameters
-----------------------

Parameter                        estimate         s.e.      t(*)
Constant                           -2.586     0.007037   -367.44
Age                               0.006304    0.007050      0.89
Spline(age) variance          0.0000002643

Between person variance       0.0000002643   3.3669E-09     78.51
Within person variance               2.643      0.03367     78.51

Variance of ln(intake) distibution is:          2.623
```

Table 35: Information in logfile

As seen in Table 35, 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 0.0001, 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 (-0.0063) meaning dat the ln(intake) of chlorpyrifos increases with age but the effect is almost negligible. The variance of the ln(intake) distribution is 2.623.

In the left plot of Figure 75 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 0.258 and 0.275 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.
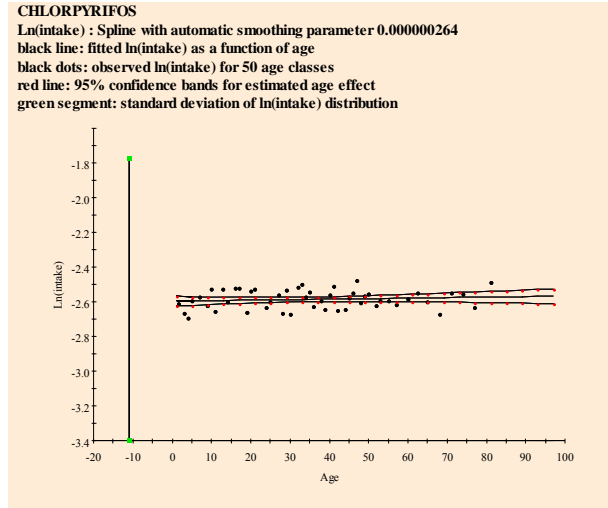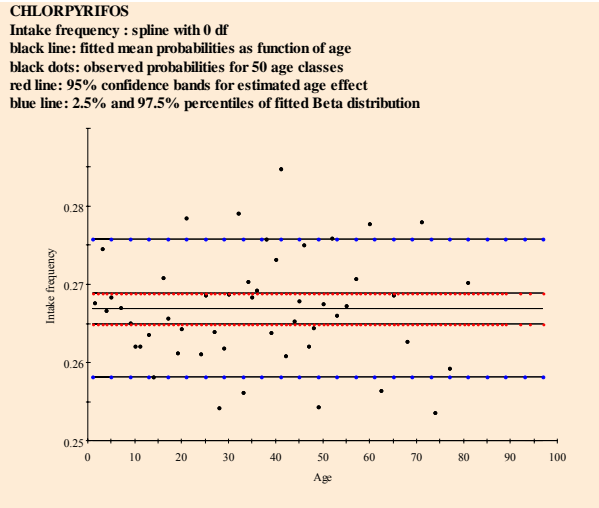
Figure 75: Age dependent intake frequency and ln(intake) distribution

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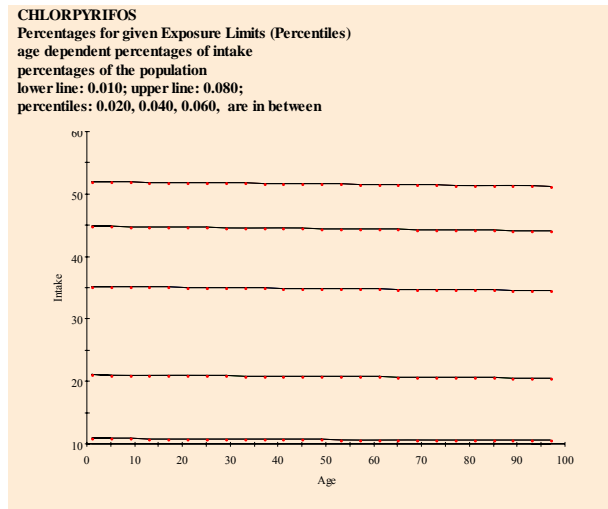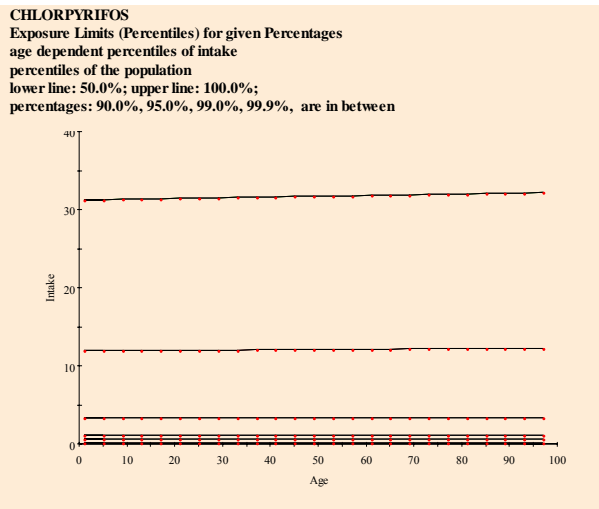
Figure 76: Age dependent percentiles and percentages

Find in Table 36 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 = 0.0001
Degrees of freedom of spline according to backward testing is 0
The significance level is 0.01
--------------------------------------------------------------------------
Model for log intake amounts is based on REML
Function of covariable is spline
    spline is 0.00
--------------------------------------------------------------------------
Variance components (transformed scale)
    within consumers is 2.64
    between consumers is 0.00
    Variance ln(intake) is 2.62
```

Table 36: 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 37 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 37: 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 food, the positives samples are taken to estimate the variance and mean on the lognormal scale. If pooling is requested, food groups need to be formed. Each food is characterised by a hierarchical food code. For CODEX codes, the first 2 characters in combination with factor Allowed define a group used in pooling. Factor Allowed indicates whether a chemical substance is allowed on a food or not. If the code is a subtype (e.g. X$y, the supertype is taken to form foodgroups (e.g. X).

Table 38 illustrates the pooling procedure. There are 8 food groups (ProdGr): 6 groups with a single food and two groups with 2 and 4 foods: group VR contains foods POTATOES and CARROT; group FC contains LEMON, ORANGE, MANDARIN and GRAPEFRUIT. Within a foodgroup variances and means of foods are pooled.

In the example, the original mean (Mean) and sigma (StdDev) are displayed together with the number of observations (nos). In columns Mu and Sigma, you find the result of pooling: parameters μ and σ of the lognormal distribution; the pooled number of degrees of freedom is in column Df.

In group 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 foods with less than 10 degrees of freedom are replaced by the overall variance if the test of homogeneity is not signifificant. For these foods, 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    : foodgroup
Allwd     : code if residue is allowed on product (1) or not (0)
Food      : Foodlabels
***********************************************************************

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

Table 38: Pooling information

In Table 39 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 39: Percentiles for the acute intake distribution using the binomial/lognormal with pooling

Example 2

The variances and means of foods are pooled within a food group. Food groups are formed based on the first two characters in combination with factor Allowed (see example, factor Levels (= food code) and Allowed). Factor Allowed indicates whether a chemical compound is allowed on a product or not. Foodgroup 10101 contains foods BEAN and SPERZIEBOON. Foodgroup 10201 is split into two subgroups (according to factor Allowed): one group with CHICORY, ENDIVE, CABBAGE LETTUCE and CURLY LETTUCE and a group with ROODLOF and SPINACH. On the last two foods the use of a chemical compound is not allowed.

In the example, the original mean and sigma are displayed together with the number of observations. The last three columns show the parameters μ and σ of the lognormal distribution together with the degrees of freedom after pooling. For BEAN and SPERZIEBOON, a pooled μ (= -1.67) and σ (= 1.31) are used. CHICORY, ENDIVE, CABBAGE LETTUCE and CURLY LETTUCE only sigma is pooled (σ = 1.47), the original means are maintained. For ROODLOF (1 observation) the overall sigma (= 1.36) is used to estimated the variance.

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

10.8 Chronic risk assessment: discrete/semiparametric (Nusser)

In this example output is shown for a chronic risk assessment for antibioticum lasalocide. The example concerns the intake via eggs. Table 40 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 40: Input form options: discrete/semiparametric (Nusser)

In Table 41 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 41: 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 samples, 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 42 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)                          (%)
   &NL001  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)                         (%)
       &NL001    WHOLE EGG C       0.4825      3.8910         31      250      12.4
```

Table 42: Summary of the database, consumptions and compounds

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

In Table 43 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   Positive
                         intakes   intakes      %
   Number of observations   12500      8630   69.0
     Number of consumers     6250      5409   86.5
```

Table 43: Summary of intake days

Find in Table 43 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 44: Summary statistics of the calculated daily intakes

In the left plot of Figure 77 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.
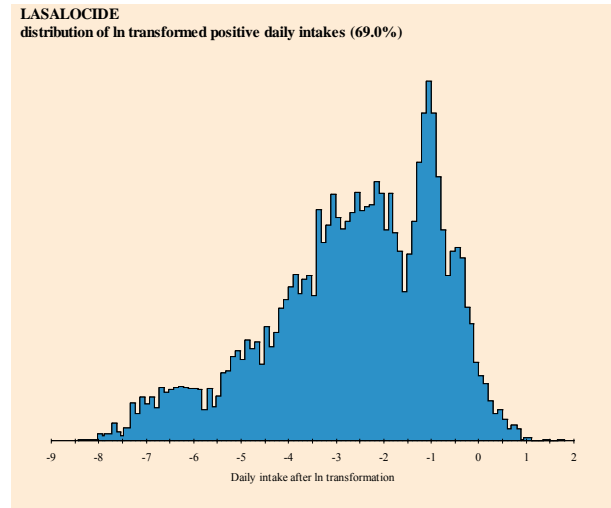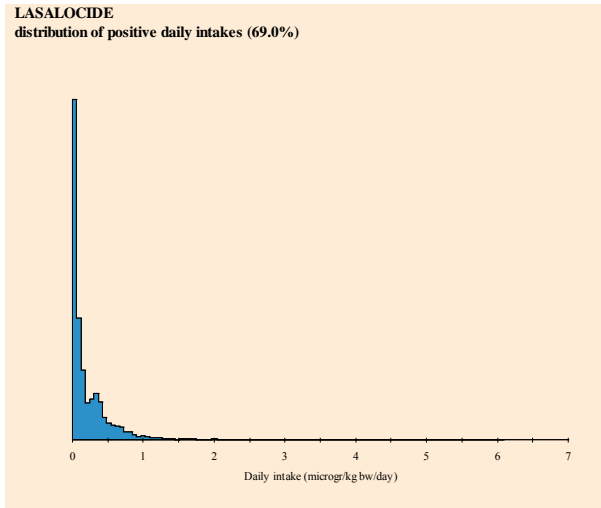
Figure 77: Untransformed and power transformed intake distribution

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 78. The result of a power and spline transformation of non-zero daily intakes is shown in the right plot.
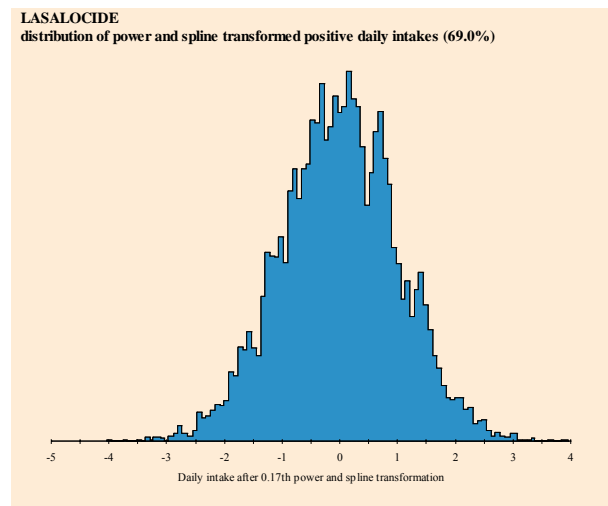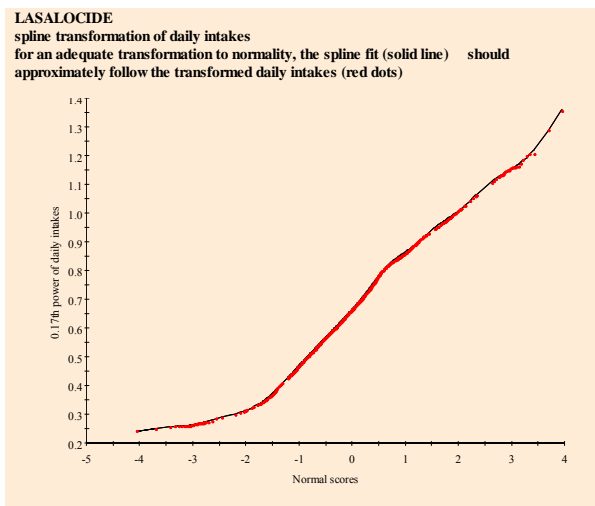
Figure 78: Transformation plot and power and spline transformed intakes

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
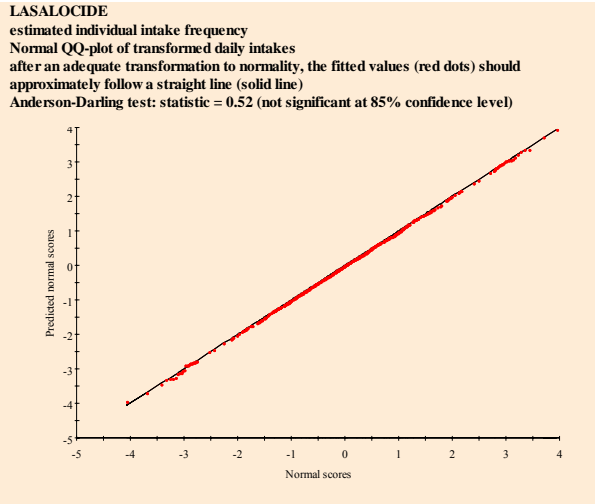
Figure 79: Normal probability plot

In Table 45 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 45: 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 46.

```
      p        theta
    0.00   0.00000000 (= 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
```

127

```
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 46: Estimates of the intake frequency distribution, theta

Graphical results are shown in Figure 80. 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.
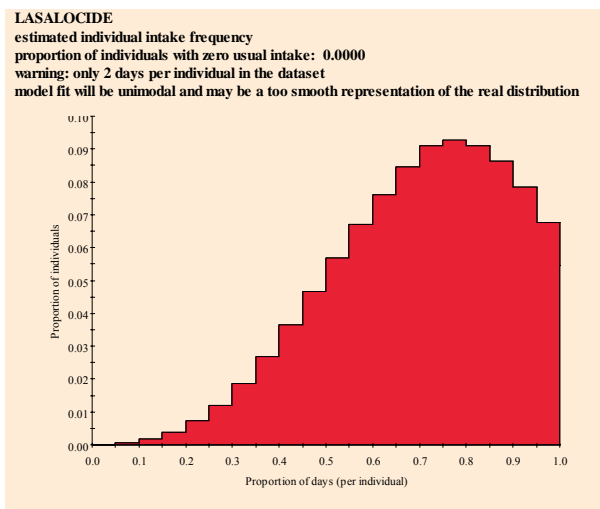
Figure 80: Intake frequency distribution

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 81. The same three distributions are also shown as probability densities in the right plot.
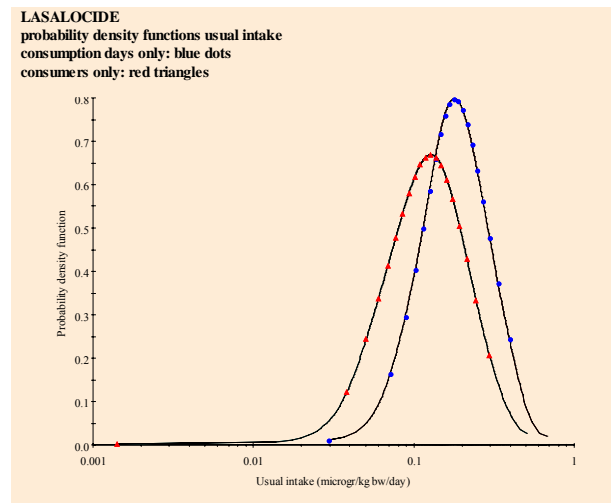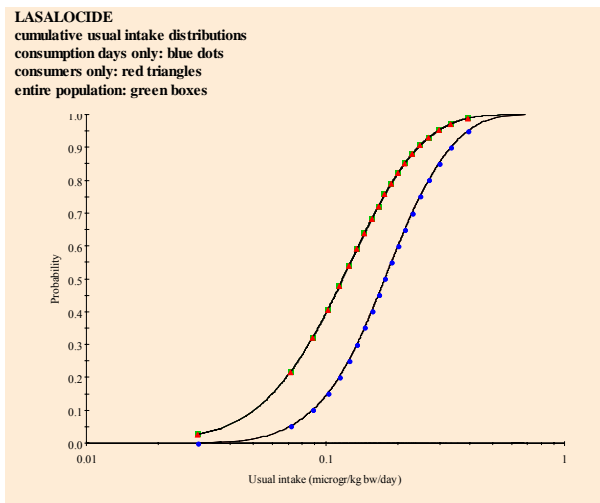
Figure 81: Cumulative usual intake and density functions

Finally, for the entire population the percentiles are calculated and shown (see Figure 82) 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 82: Percentiles usual intake distribution: entire population

In Table 47 you find information on selected percentiles for the other types of distributions.

```
Compound:  LASALOCIDE

LONG TERM EXPOSURE  (USUAL INTAKE)
   Percentage      Positives      Consumers
                       only            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
```
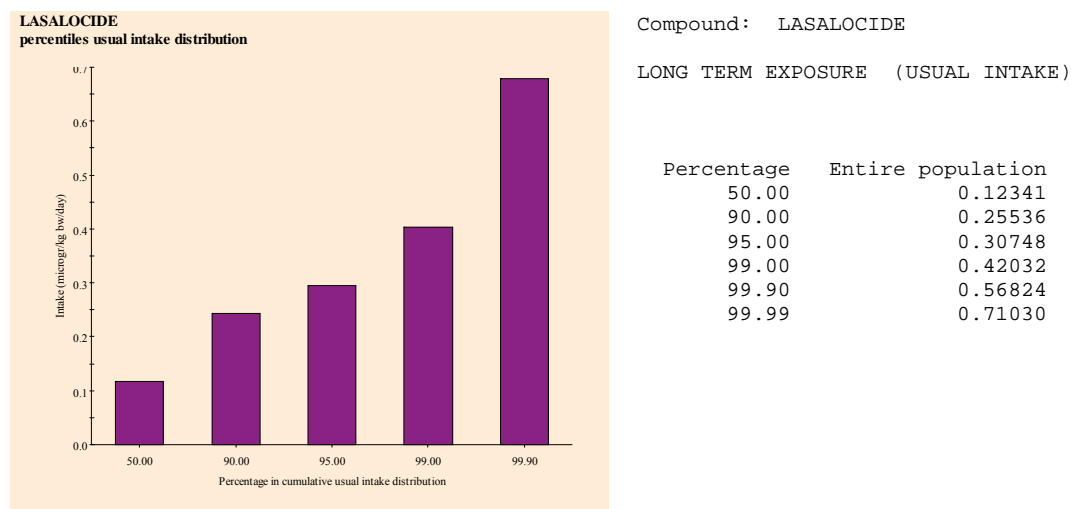
Table 47: Technical information on usual intake: percentiles for usual intake distribution, positives only and consumers only

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 48 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 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        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 48: Input form options: betabinomial/lognormal with age

In Table 49 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 49: 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 50 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)                          (%)
    &NL001   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)                       (%)
    &NL001   WHOLE EGG C   0.4825    3.8910        31       250   12.4
```

Table 50: 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 51 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  Positive
                          intakes   intakes      %
   Number of observations    12500      8630   69.0
    Number of consumers       6250      5409   86.5
```

Table 51: Summary of intake days

Find in Table 52 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 52: Summary statistics of the calculated daily intakes

In the left plot of Figure 83 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 77.
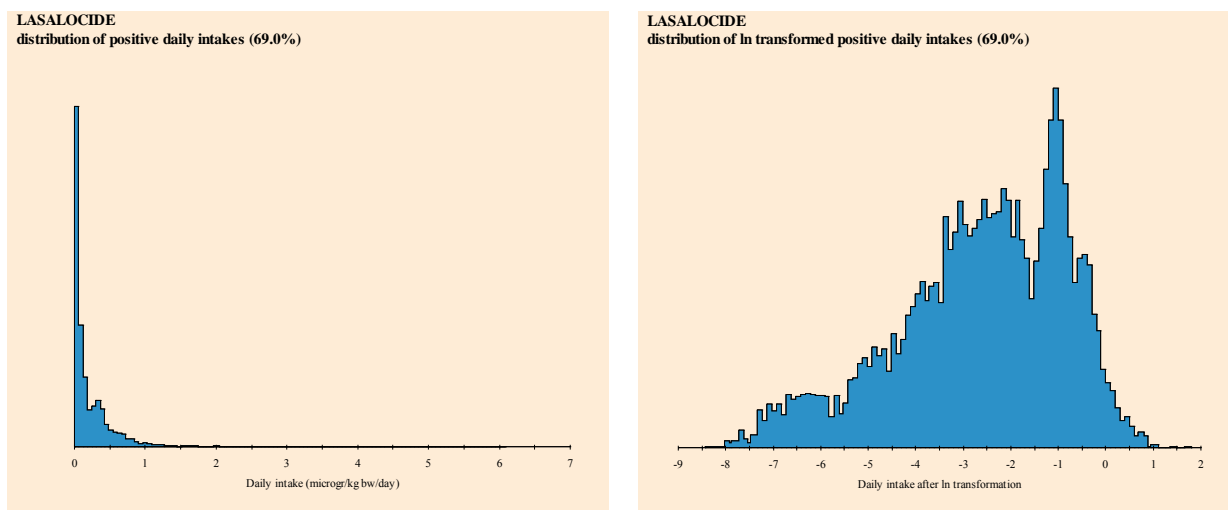
Figure 83: Untransformed and ln transformed intake distribution

As seen in Table 49, 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 84 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(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 dat the ln(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 84 displays the ln(intake) distribution. The age effect is plotted as a black line through the observed mean ln(intake) amounts per age class. The 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.
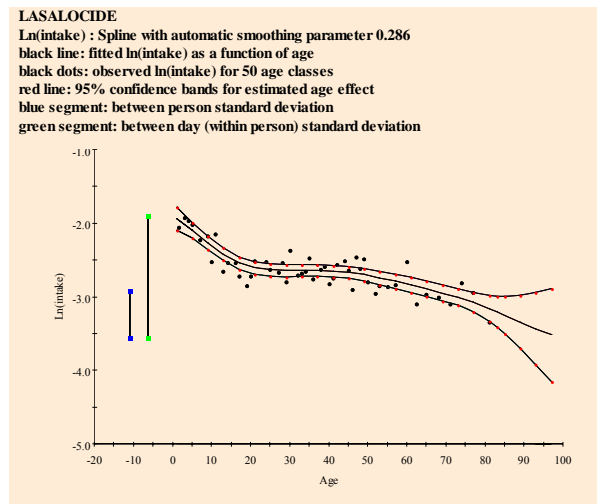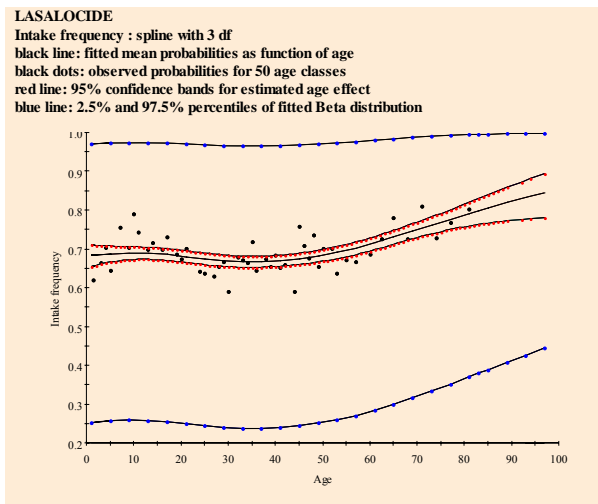
Figure 84: Age dependent intake frequency and ln(intake) distribution

In Table 36, a repport of technical information on the analysis is shown.

```
Compound:  LASALOCIDE

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

Table 53: Technical information on usual intake

Figure 85 shows the age dependent percentiles and percentage for the usual intake distribution.
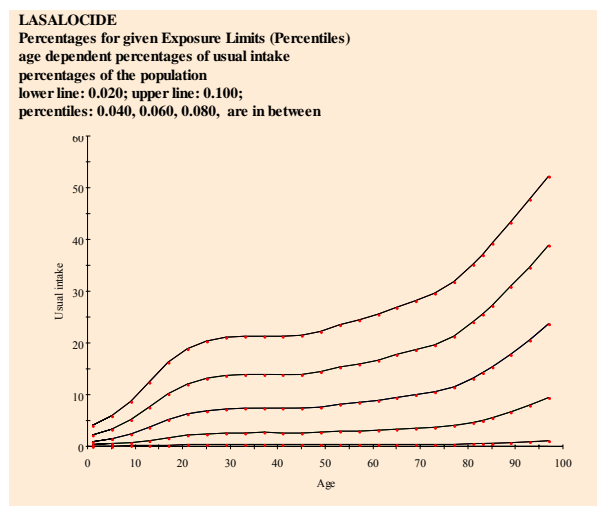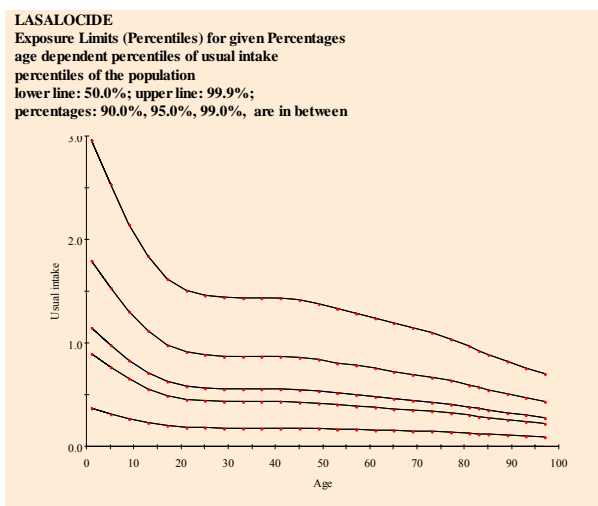
Figure 85: Age dependent percentiles derived from percentages and percentages derived from percentiles

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 54 lists the main options:

```
Input form
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 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 54: Input form options: betabinomial/lognormal with age and uncertainty

In Table 55 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.

```
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
```
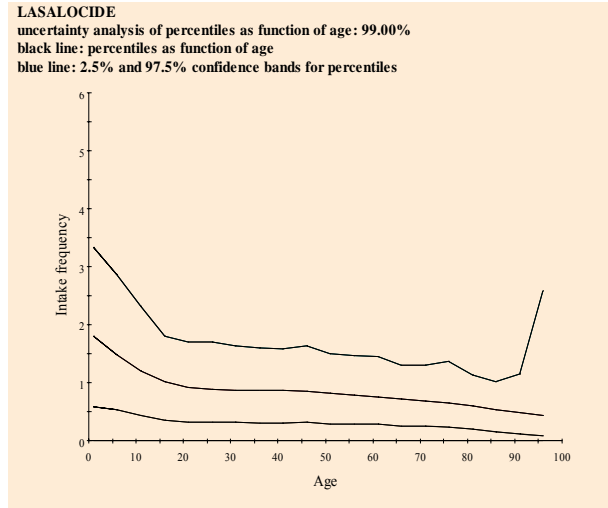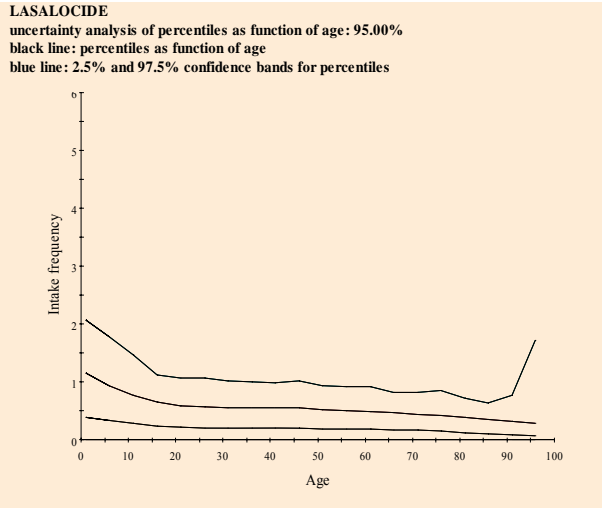
Table 55: Information in logfile

Find in Figure 86 and Figure 87 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 86: Age dependent percentiles derived from percentages
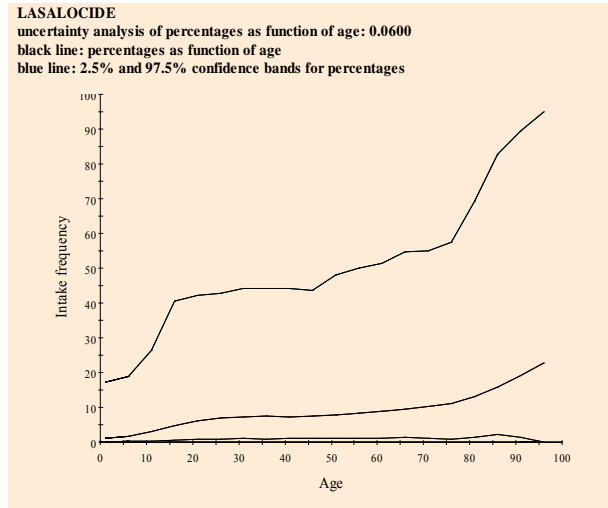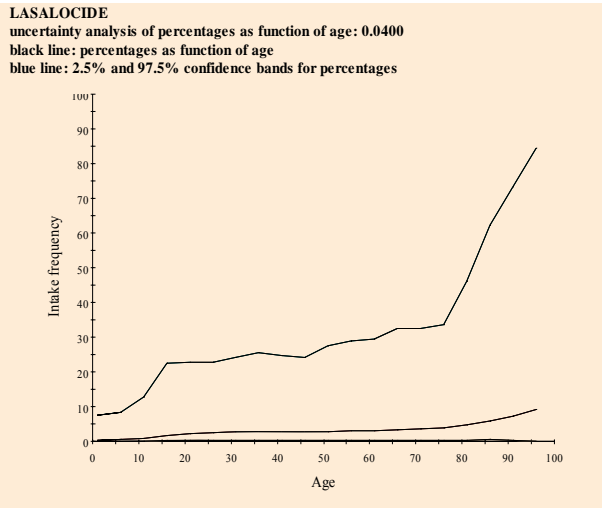
Figure 87: Age dependent percentages derived from percentiles

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 4).

If you cannot reach the MCRA main menu:
- close the internet explorer,
- login to the website again,
- go to manage input/output in the MCRA main menu (see Figure 2),
- move your mouse to directory 'IN' or 'OUT' or any other directory and left click,
- clear your history first by clicking the **'**Clear history'-button (see Figure 4).