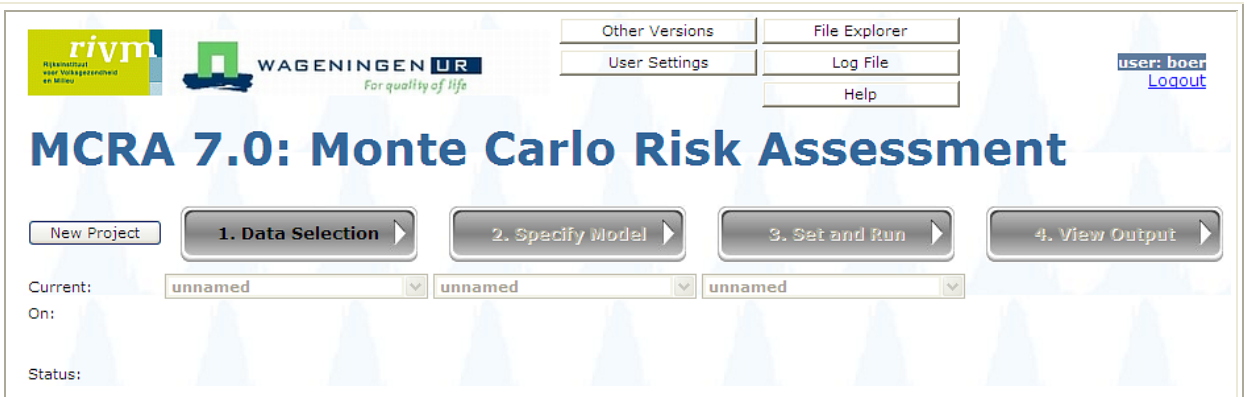
MCRA 7 Example: patulin (BBN, mdb data)

For a quick start in MCRA, the main tasks and steps of the interface are described using a case study as described in de Boer *et al.* (2009). After login, all tasks with corresponding actions are started from the MCRA central menu.



MCRA is only availabe for registered users. Register at https://mcra.rivm.nl

The central menu contains four main tasks:

- Data Selection (load data from Access or Excel files)
- Specify Model (specification of intake model options)
- Set and Run (specification of output options, start main analysis)
- View Output (managing output)

A main task is started by clicking the button. Then, a menu containing actions related to the main task is displayed. A main button can only be pressed when the name of the tasks is displayed in **black**. Names of main tasks that are not available or active at the moment, are displayed in grey. After clicking a main button, it turns into blue to indicate that the task is active. For a first time user, the figure above shows the central menu and only the Data Selection button is active. Otherwise, you may press New Project to clear all selections.

Design of the database

MCRA can use two types of data:

1. Simple tables, listing intake of a food or compound on multiple persondays can be entered as Excel [.xls] spreadsheets. Additional columns may list covariate information for the individuals, e.g. age or gender (these values have to be replicated for all days of the individual).
2. Access [.mdb] databases allow much more information to be handled, such as analysing the importance of all separate foods (both foods as eaten, e.g. pizza, and foods as measured, e.g. wheat), modelling concentration data, food processing effects, unit variability, brand loyalty. Data are stored in multiple tables in one or more Access [.mdb] files. For a detailed description of table format see the appendix.

MS Access database (mdb)

We consider the example of the patulin data (see also de Boer *et al*. 2009) illustrating the mdb database approach.

Patulin was only detected in two food ingredients: apple juice and canned apple sauce. Many of the measured samples were so-called non-detects, i.e. for these samples the concentration was reported only as being below a certain limit (which is consequently named in MCRA the Limit Of Reporting, or LOR). Concentration values are entered in a table ConcentrationValues, and non-detects should be entered as negative numbers –LOR.

ConcentrationValues : Table

Compound	FoodMeasured	Year	Month	SamplingType	Country	NumberOfSamples	Value
031004005	&NL$02$0383	2006	99	M	99	1	1
031004005	&NL$02$0383	2006	99	M	99	3	-50
031004005	&NL$02$0383	2006	99	M	99	15	25
031004005	&NL$02$0383	2006	99	M	99	1	13.8
031004005	&NL$02$0383	2006	99	M	99	1	4.6

The column NumberOfSamples shows that certain values occur more than once (here 3 non-detects are shown for LOR 50, note that these data are peculiar because some of the other values are below 50, suggesting that the LOR was lower for those measurements). The codes for Compound and Foodmeasured are linked to appropriate names in separate tables Compound and Food. Columns Year, Month, Country and SamplingType specify further information, '99' is used if the information is missing.
Food consumption was recorded on two days in the Dutch Food Consumption Survey for young children (Ocké et al. 2008). These data are entered in a table FoodConsumption:

FoodConsumption : Table

Individual	DayOfSurvey	FoodConsumed	AmountConsum	FoodSurvey
332100	1	&NL$03$0230	45	VCP-kids
332100	1	&NL$03$0248	90	VCP-kids
332100	1	&NL$11$0513	7	VCP-kids
332100	1	&NL$23$1152	10	VCP-kids
332100	1	&NL$20$0436	35	VCP-kids

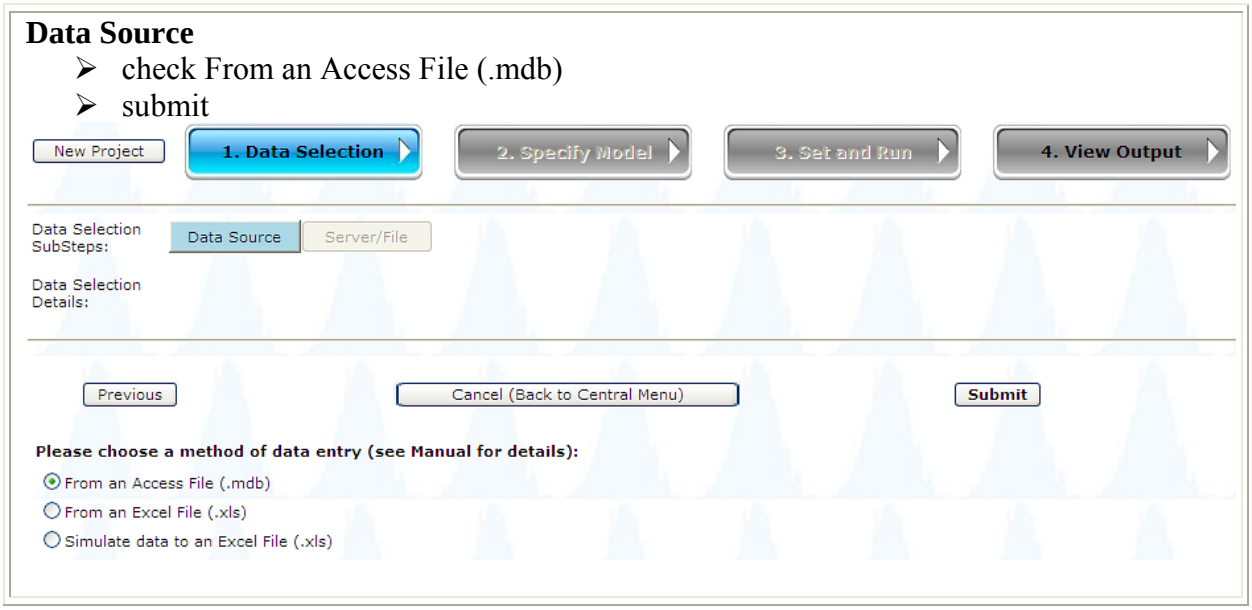
In this table, persondays with zero consumption can be omitted. Covariate values of individuals are specified in a separate table Individual, and appropriate food names are stored in the Food table.
If measurements are made on a different food coding level as the food classification used in the consumption survey, the link between foods as consumed and foods as measured can be specified in a table FoodComposition:

FoodComposition : Table

	food	ingredient	proportion%
	&NL$02$2148	&NL$02$0383	100
	&NL$02$0383	&NL$02$0383	100
	&NL$02$2144	&NL$02$0383	100
	&NL$02$6026	&NL$06$0179	16.7
	&NL$02$6221	&NL$06$0179	0.211
	&NL$02$6215	&NL$06$0179	0.264
	&NL$02$6214	&NL$06$0179	0.299

Selection of data from MS Access database (mdb)

Clicking the Data Selection button presents the user with three different main data sources: (1) relational databases as stored in MS Access database files (MDB), (2) simple table data as stored in an Excel spreadsheet, (3) or one can simulate data. Since the example data are stored in MDB files the radio button "From an Access File (.mdb)" is checked followed by clicking the "Submit" button.



For MDB files the following buttons become visible

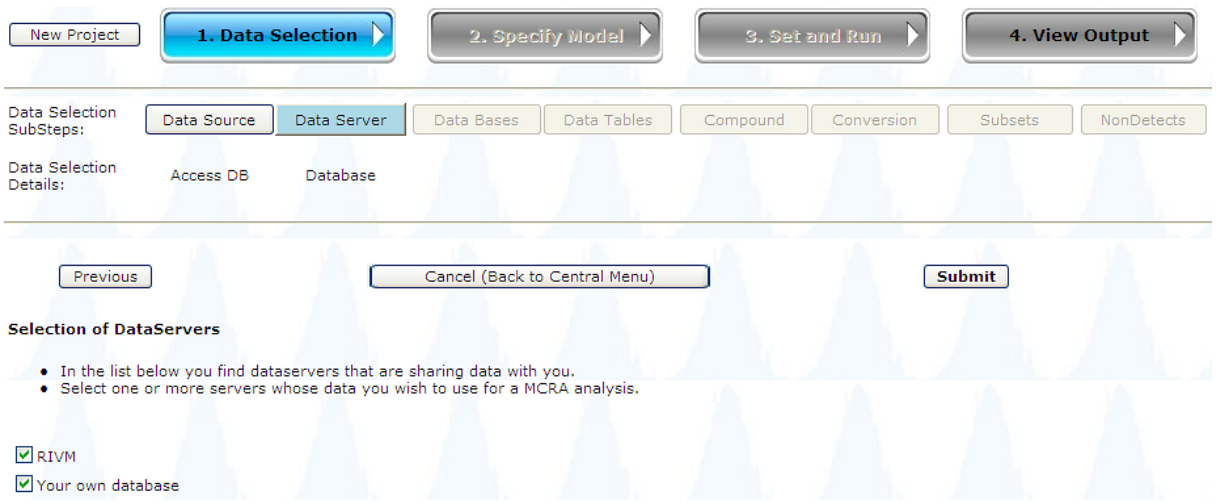
1. Data Selection	
Data Source	Selection of data source (mdb, xls or simulated data)
Data Server	Selection of data server or file (own data or centrally supplied data from RIVM)
Data Bases	Selection of databases
Data tables	Selection of tables
Compound	Selection of compound, survey and/or covariates
Conversion	Start conversion of food as eaten to food as measured
Subsets	Subset selection of individuals and foods

NonDetects	Estimation of parametric distributions for concentration values

These buttons are processed in sequence:

Data Server

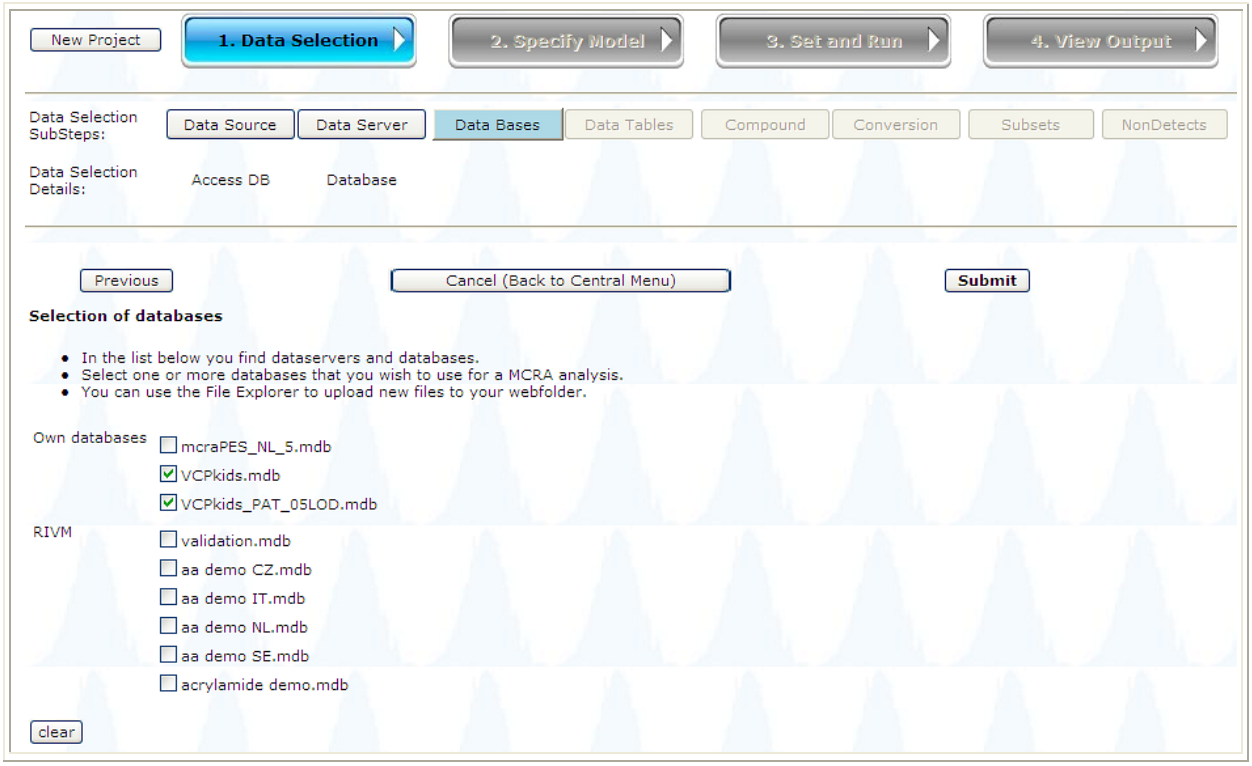
- check RIVM
- check Your own database
- submit



RIVM supplies a central database with example data. To use your own data, upload your data to your user directory using the Filemanager or combine own and centrally supplied data.

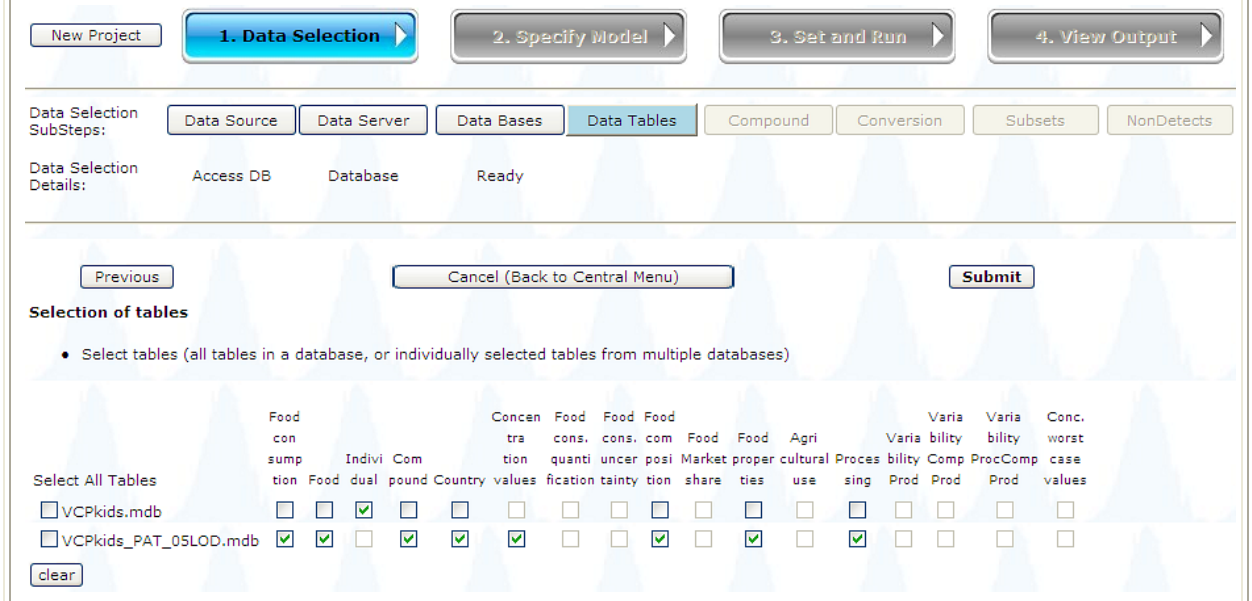
Data Bases

- check Own database VCPkids.mdb
- check Own database VCPkids_Pat_05LOD.mdb
- submit



Data Tables

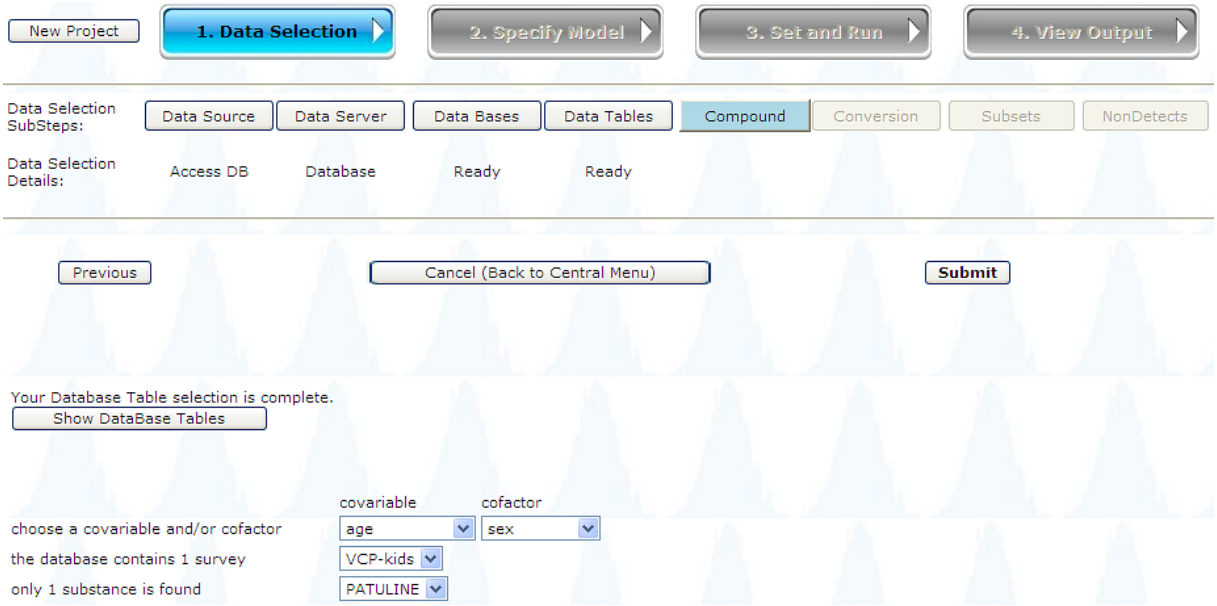
- check the tables as given below
- submit



In the simplest case all tables reside in one mdb file, and it is sufficient to put a checkmark in the first column (Select All Tables). Here two mdb files have been used. Be careful when tables from different databases are selected. Because of the relational structure of the database, data should be consistent, *e.g.* individual ID's in table Foodconsumption should correspond to the supplied ID's in table Individual, foodcodes used should be the same in all databases.

Compound

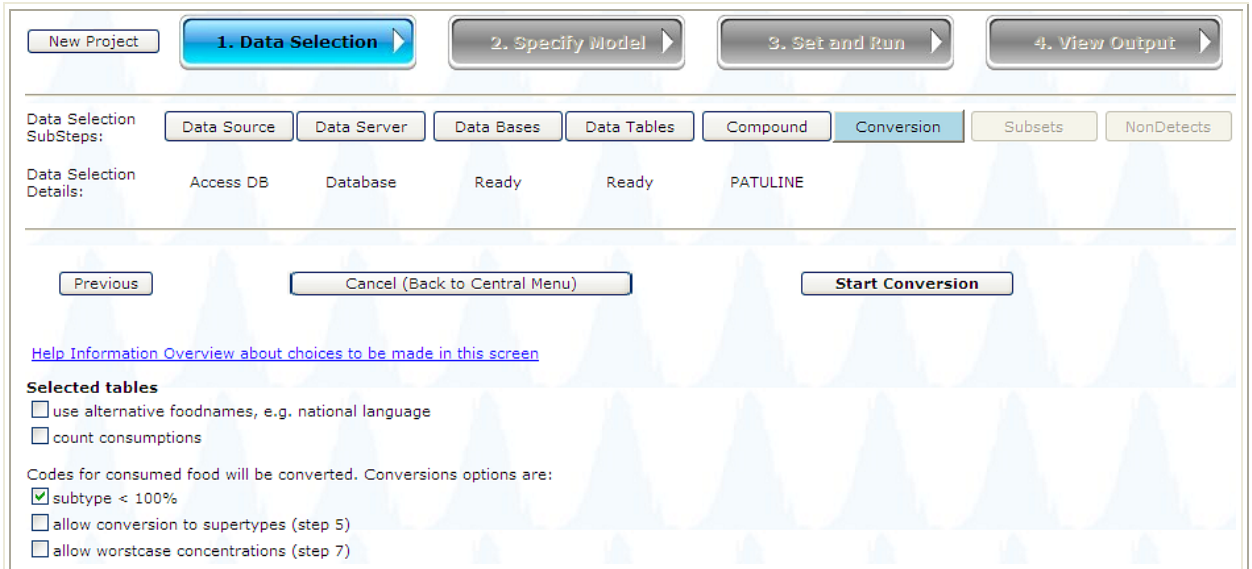
- select covariable age
- select cofactor sex
- select survey VCP-kids
- select compound patuline
- submit



A covariate (covariable or cofactor) is a property of an individual. Covariates can be used for subset selection and/or for modelling intake as a function of the selected covariates. Selection of covariables or cofactors is optional. A cofactor, unlike a covariable, takes only a limited set of values, one for each group. MCRA 7.0 allows the selection of one covariable and one covariate.
If the data tables contain information from multiple food consumption surveys, select one survey.
If the data tables contain information for multiple compounds, select one compound.

Conversion

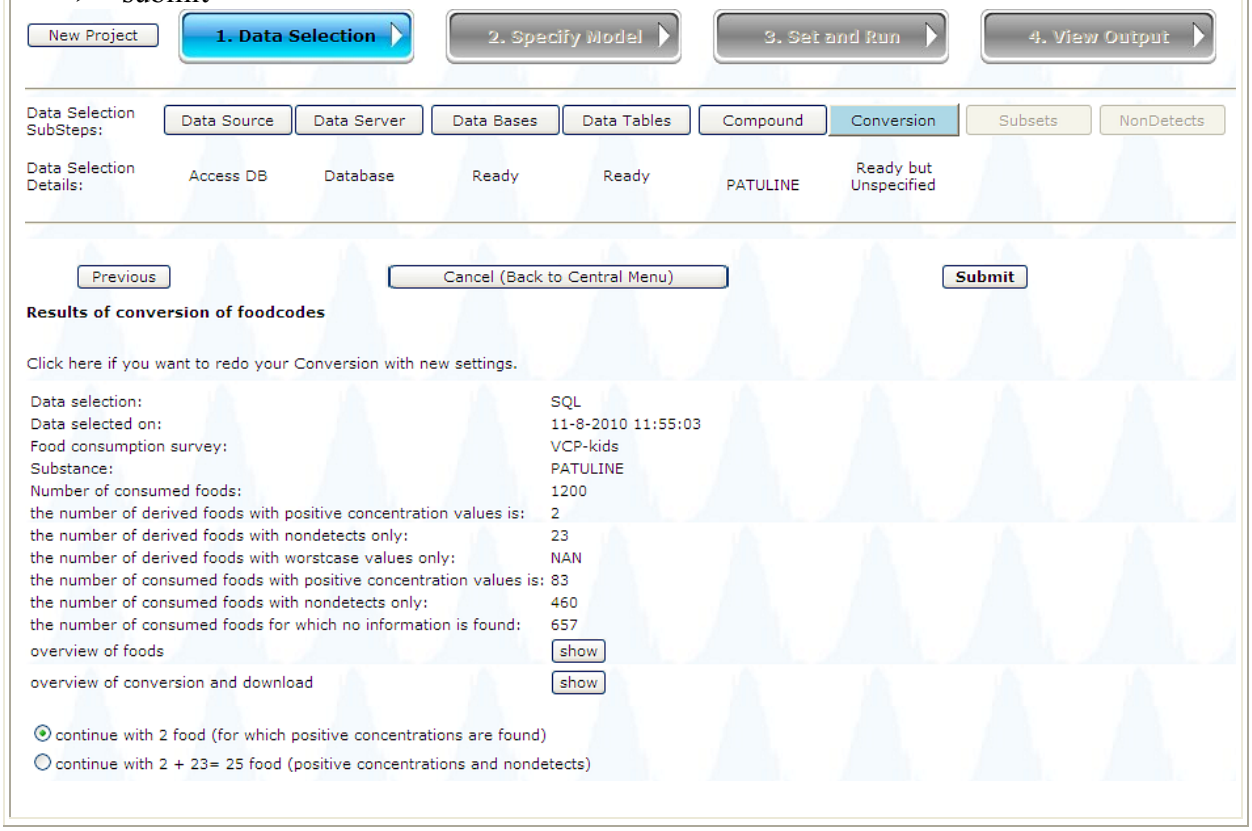
- check subtype <100%
- Start Conversion



The conversion algorithm converts food as eaten to food as measured. So, speaking about pizza and assuming that no concentration measurements on whole pizza are available, it is converted to *e.g.* wheat, tomato, cheese etc with corresponding proportions. The conversion will take some time depending on the size of the databases.
Checkbox 'subtype < 100%' is only relevant for marketshare data. Occasionally, marketshares do not sum to 100%. Uncheck 'subtype < 100%' to ignore these foods in the analysis (not shown).

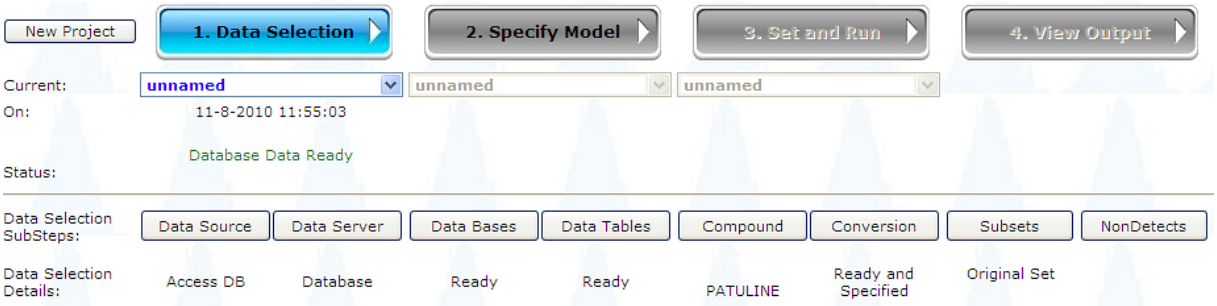
Conversion continued

- check continue with 2 foods (for which positive concentration are found)
- submit



After finishing the conversion, some details are displayed. A complete overview can be found pressing the show buttons. In this example, the number of consumed foods is 1200. After conversion, only 83 foods remain containing the 2 ingredients (apple juice and canned apple sauce) for which patulin was measured (positive values). A large number of ingredients, here 23, was measured but the corresponding samples were all nondetect (<LOR). A number of 460 foods as eaten contained only ingredients that were either not measured or for which only nondetects were found. A number of 657 foods as eaten contained only ingredients trhat were not measured.

After a successful Data Selection, the central menu indicates which steps were performed together with some short information. The selected data is displayed as **unnamed** and can be saved with a chosen name for future use.



- select scrolldown box under Data Selection
- save as…Patuline

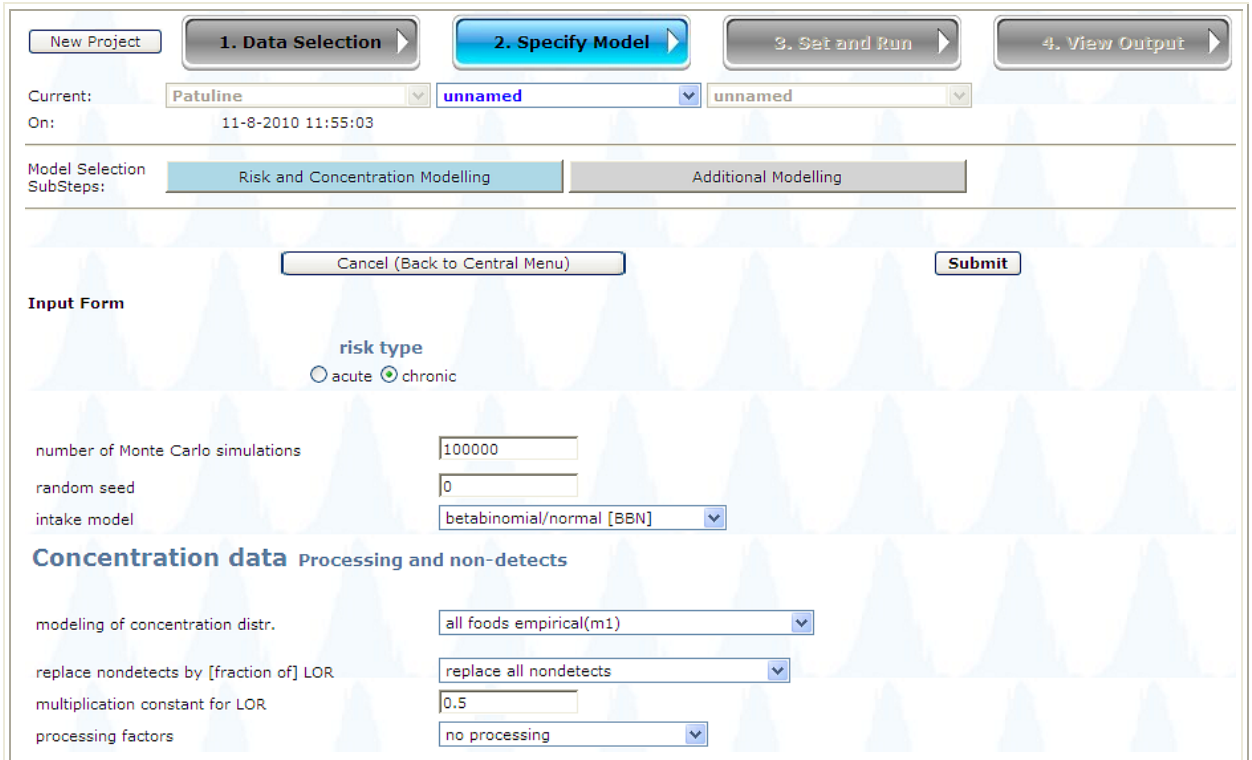
Specification of input options

In the second step the model with which the data will be analysed must be specified. This consists of the following two steps:

2. Specify Model	
Risk and concentration modelling	Risk type (acute, chronic), number of Monte Carlo simulations, chronic intake model (BBN, ISUF, LNN, OIM), modelling of concentration data (empirical, parametric), processing and replacement of nondetects by the LOR
Additional modelling	Modelling of intake frequency and amounts

Risk and Concentration modelling

- check chronic
- select intake model betabinomialnormal
- select all foods empirical
- select replace all nondetects
- submit



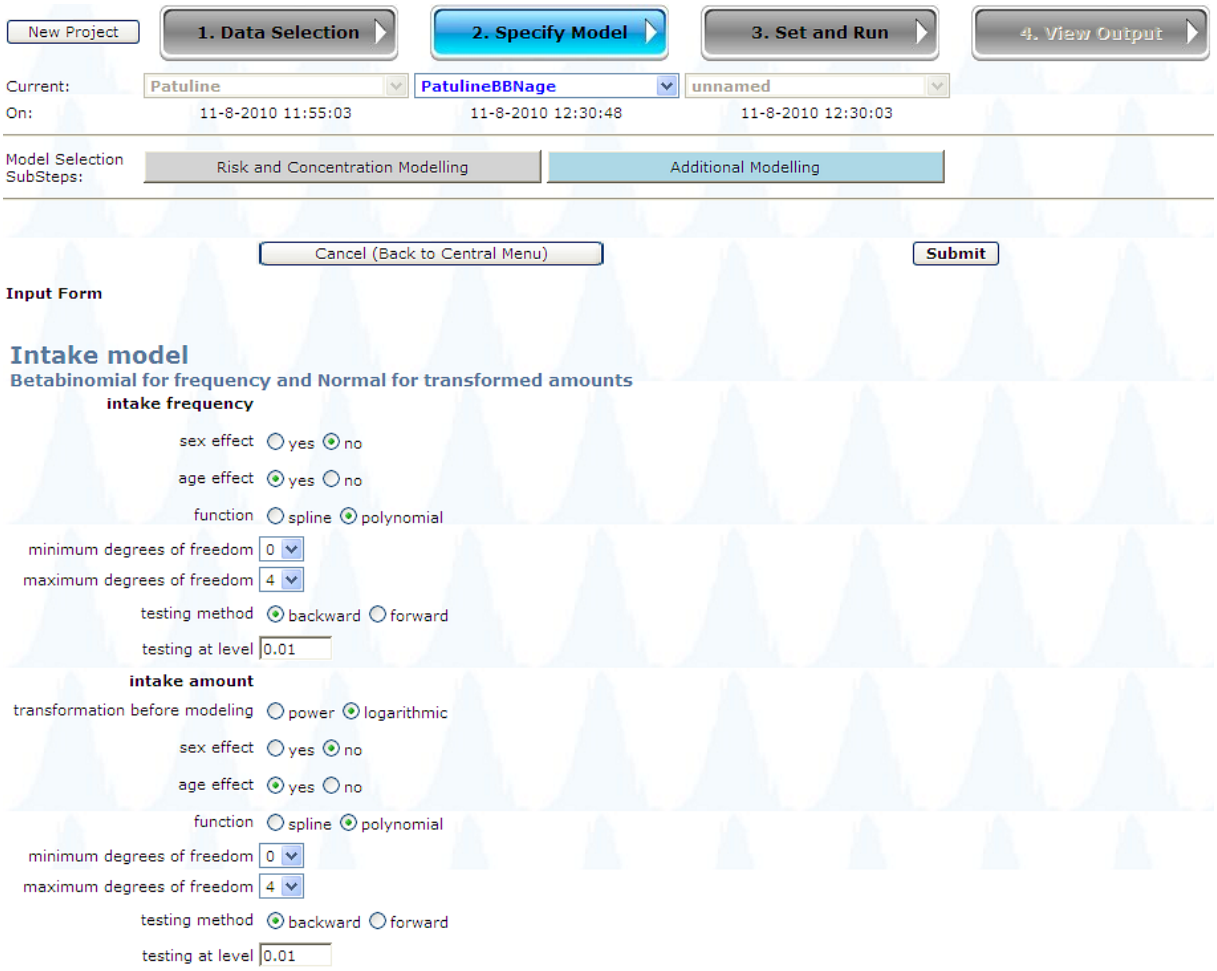
In this example we consider a chronic exposure assessment, using the betabinomialnormal (BBN) model and a lognormal transformation to normality, where both the daily intake frequency and the transformed intake amount of patulin are modelled as polynomial functions of age. The intake is calculated as the consumption on each day of each consumer multiplied by the average value of the compound concentrations levels (empirical) divided by the body weight. No processing is applied and a worstcase scenario is investigated (all nondetects were replaced by 0.5 x LOR). After modelling the daily intake frequency and amount distribution, the usual intake distribution is derived by numerical integration. This is done by Monte Carlo sampling, multiplying both distributions. When covariates are included in the model, a usual daily intake distribution is derived for each combination of the levels of the covariates. Note, the concept of unit variability is only relevant for acute risks.

Additional modelling

- intake frequency: check
  - sex effect no
  - age effect yes
  - function polynomial
  - use default for minimum and maximum degrees of freedom for polynomial fit
  - use default backward selection for testing the degrees of freedom of polynomial fit
  - use default significance level $\alpha = 0.01$ for backward testing
- intake amount: check
  - transformation before modelling logarithmic
  - sex effect no
  - age effect yes
  - function polynomial
  - use default for minimum and maximum degrees of freedom for polynomial fit
  - use default backward selection for testing the degrees of freedom of

polynomial fit
- use default significance level α = 0.01 for backward testing

- select scrolldown
- save as…PatulineBBNage
- submit



A smoothing spline is a complicated function, constructed from segments of cubic polynomials with constraints to ensure smoothness. A polynomial function is based on orthogonal linear, quadratic, cubic or quartic curves. The degree of smoothness of the spline or polynomial function is controlled by increasing or decreasing the degrees of freedom. A spline or polynomial with the maximum degrees of freedom is less smooth than a spline or polynomial with the minimal degrees of freedom. To determine automatically the degrees of freedom of the spline or polynomial two testing methods are available. Backward selection means that testing starts with a spline or polynomial of the highest degree. Then, in each elimination round the number of degrees of freedom is decreased, one at a time, and the process is stopped when the resulting decrease in fit is significant at the specified significance level as judged on the basis of a deviance test. Forward selection means that the evaluation of the degree of the spline or polynomial is started with a function of the lowest degree. In all evaluations the testing level is 0.01.
Before modeling the positive amounts, a logarithmic or power transformation is applied to approximate normality. The analysis provides mean intakes of the transformed intake distribution dependent on explanatory variables. The total variance of the non-zero transformed intake amounts is divided into a between individuals and a between days within

individuals variance component. The between-individuals component is the basis for the estimation of the distribution of the usual intake. If both cofactor and covariable are included in the analysis, decide on modeling the interaction (not shown). For a polynomial, the interaction means that curves are no longer parallel and intercepts may differ.

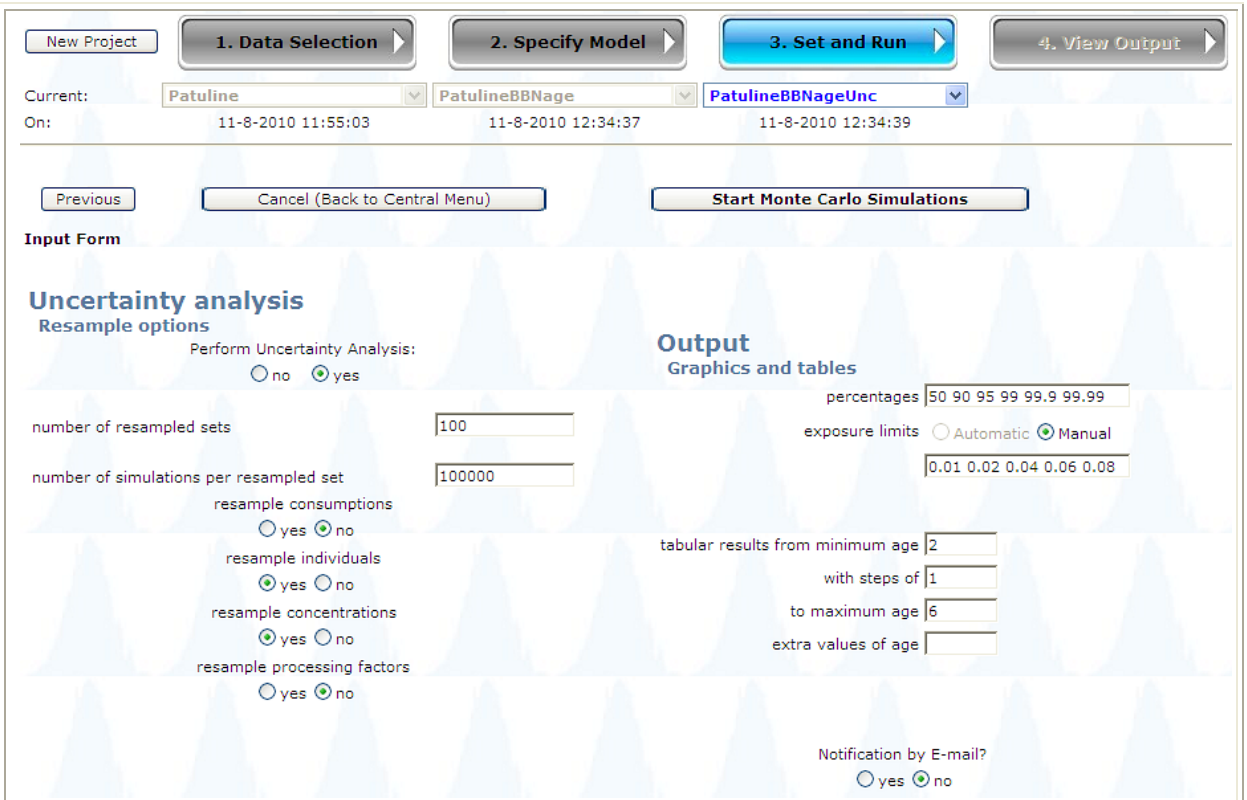
Specification of output options

In the third step options for graphical and tabular output must be specified. This consists of the following steps:

	
Output options	Uncertainty analysis (yes, no), resample consumptions, individuals, concentrations and processing factors (yes, no), options concerning graphical and tabular output.
Start Monte Carlo Simulations	Start a MCRA analysis

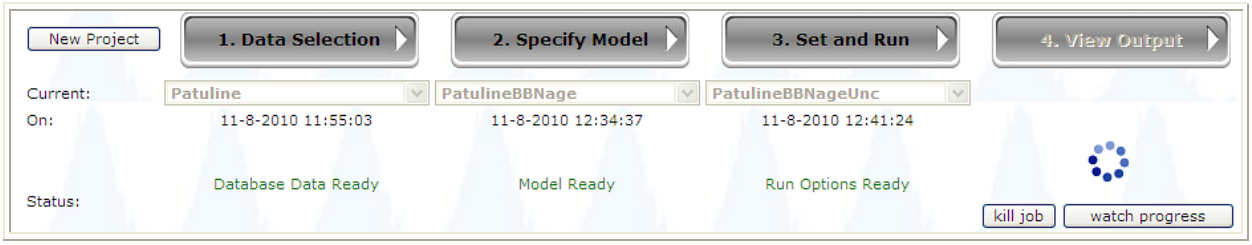
Start Monte Carlo Simulations

- check Perform Uncertainty Analysis is yes
- use default number of resampled sets is 100
- use default number of simulations per resampled set is 10.000
- check resample individuals is yes
- check resample concentrations is yes
- select scrolldown
- save as…PatulineBBNageUnc
- Start MCRA Analysis



The uncertainty of output statistics (*e.g.* mean or percentiles of the intake distribution) is assessed by resampling datasets. Resampling can be applied on the level of fresh Monte Carlo-samples, on the level of consumptions (portion size), on the level of individuals, on the level of the concentration values and on the level of processing factors from a parametric uncertainty distribution. To examine the uncertainty due to MC-variability in each analysis only, set all resample options to no (not shown). Then data are resampled from the original data. Here, the uncertainty due to resampling individuals and concentrations is established. From each dataset, data are resampled (with replacement) to construct a so-called bootstrap sample. From the resampled data sets and parameters an intake distribution is simulated and each resampled set provides a mean, maximum and percentiles according to the specified percentages. All replicates together contain the information to make inferences from the data, *e.g.* to establish the uncertainty of mean, maximum and percentiles.
In this example, 100 resampled sets are specified and on each set 100,000 Monte Carlo iterations are made. Note that the number of values within a set restricts which percentiles are displayed. Here, the highest possible percentage for which uncertainty information can be calculated is the 99.999th percentile, for a set containing 1000 simulations this is the 99.9th percentile.

After starting the MCRA analysis, the spinning wheel indicates that the job is running. Note that all main tasks are available except the View Output task. The View Output window automatically opens after a successful run (output not shown).

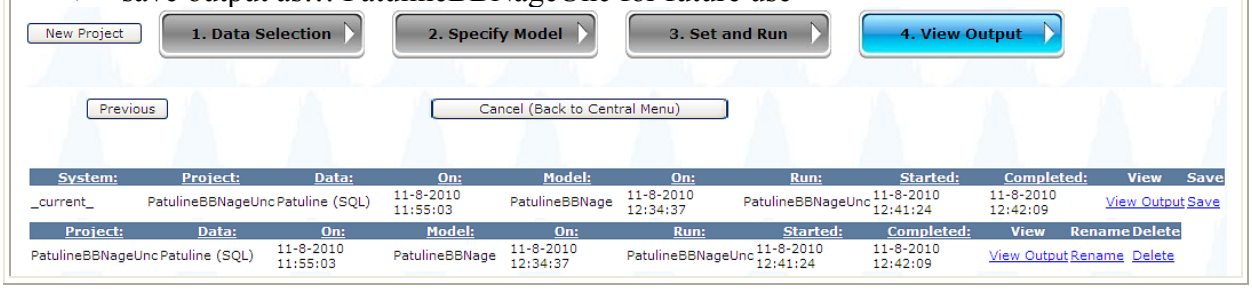


Viewing and saving output

4. View Output	
Manage output	Save, Rename, Delete and View output

View Output

- check save
- save output as… PatulineBBNageUnc for future use



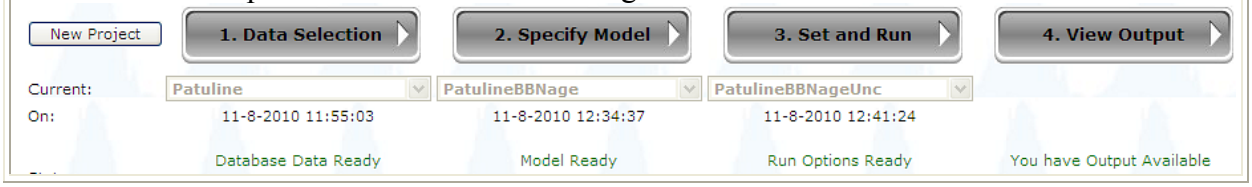
User profiles

The scrolldown boxes below the main task button enables the user to manage the data, input models, output options and output of MCRA. At any moment these user profiles can be modified. Available options are:

- save a unnamed selection or model (save as…)
- rename a selection or model (rename)
- delete a selection or model (delete)
- retrieve a former selection or model

Scrolldown

- Data Selection: save as… Patuline
- Specify model: save as… PatulineBBNage
- Set and Run: save as… PatulineBBNageUnc
- View Output: save as… PatulineBBNageUnc



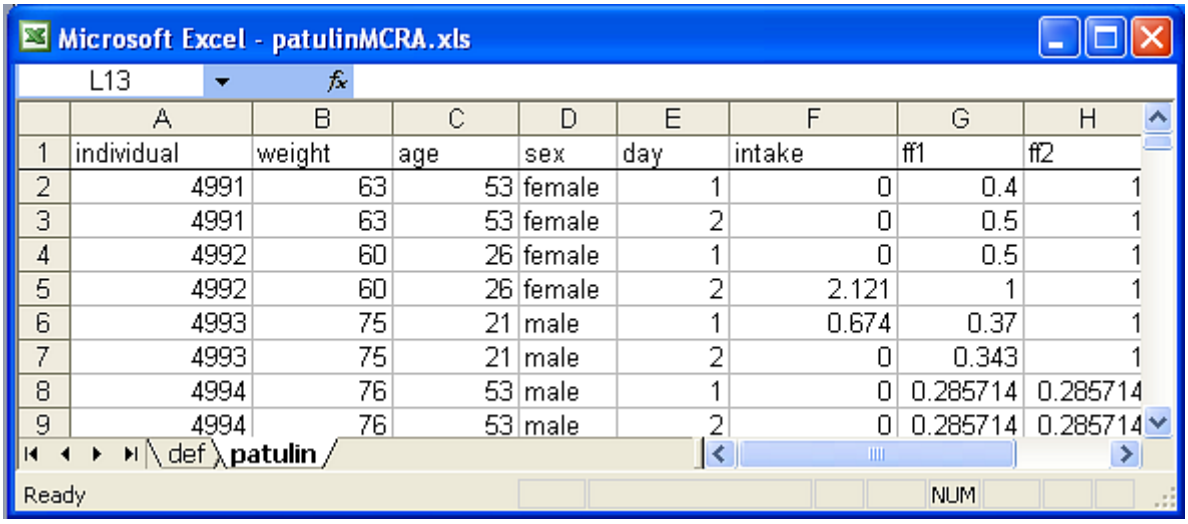
MCRA 7 Example: patulin (BBN, Excel data)

After login, all tasks with corresponding actions are started from the MCRA central menu.

Excel spreadsheet

We consider the example of the patulin data (see also de Boer *et al*. 2009).

Here is an example how patulin intake per personday, if calculated outside MCRA, can be entered as an Excel table, part of which may look as follows:



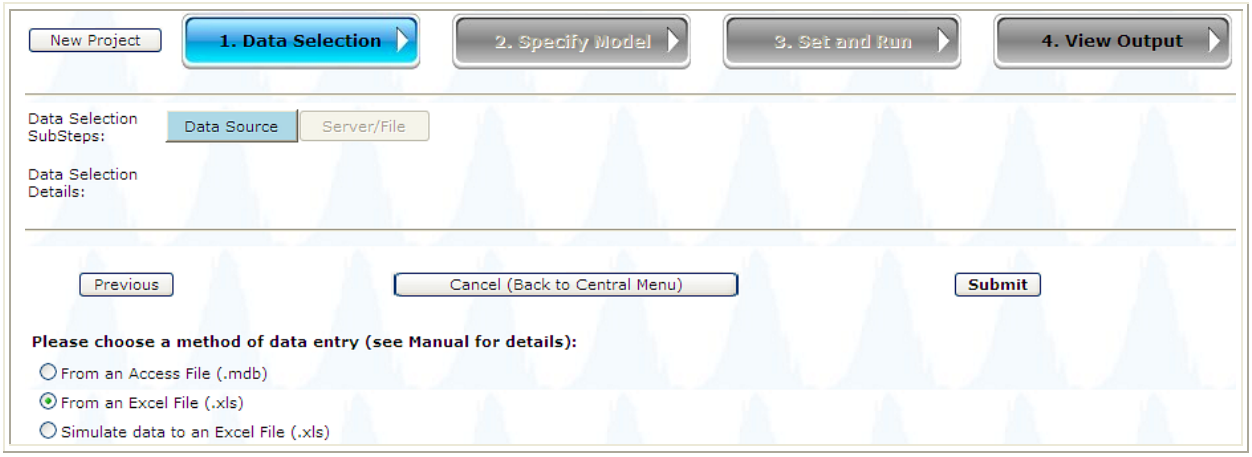
	A	B	C	D	E	F	G	H
1	individual	weight	age	sex	day	intake	ff1	ff2
2	4991	63	53	female	1	0	0.4	1
3	4991	63	53	female	2	0	0.5	1
4	4992	60	26	female	1	0	0.5	1
5	4992	60	26	female	2	2.121	1	1
6	4993	75	21	male	1	0.674	0.37	1
7	4993	75	21	male	2	0	0.343	1
8	4994	76	53	male	1	0	0.285714	0.285714
9	4994	76	53	male	2	0	0.285714	0.285714

Note, that there should be a record for all persondays, also when the intake was zero. The spreadsheet contains some additional columns (ff1, ff2) which will not be used. Use worksheet **def** to specify which data are to be analysed.

Selection of data from Excel spreadsheet (xls)

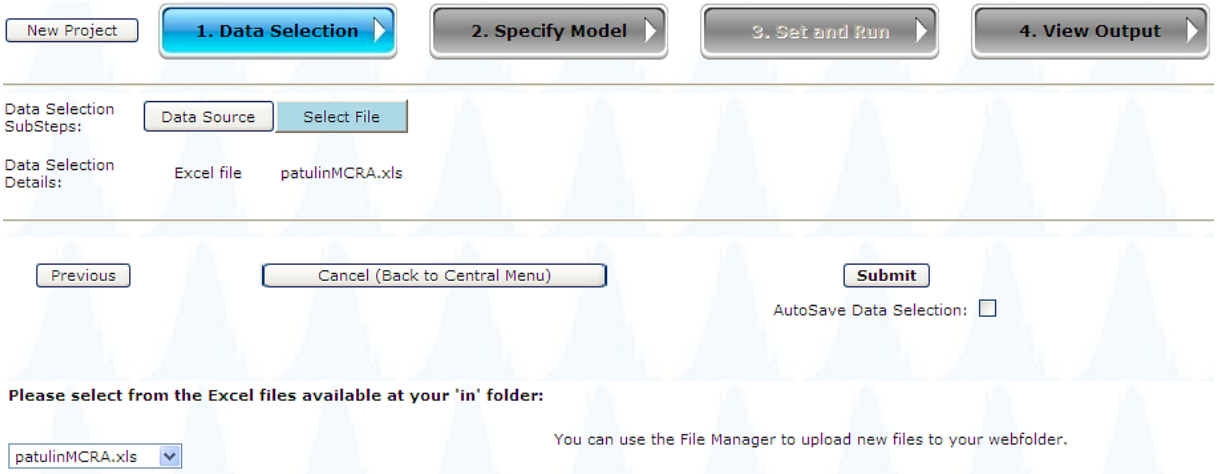
Clicking the Data Selection button presents the user with three different main data sources: (1) relational databases as stored in MS Access database files (MDB), (2) simple table data as stored in an Excel spreadsheet, (3) or one can simulate data. Since the example data are stored in MDB files the radio button “From an Excel File (.xls)” is checked followed by clicking the “Submit” button.

Data Source



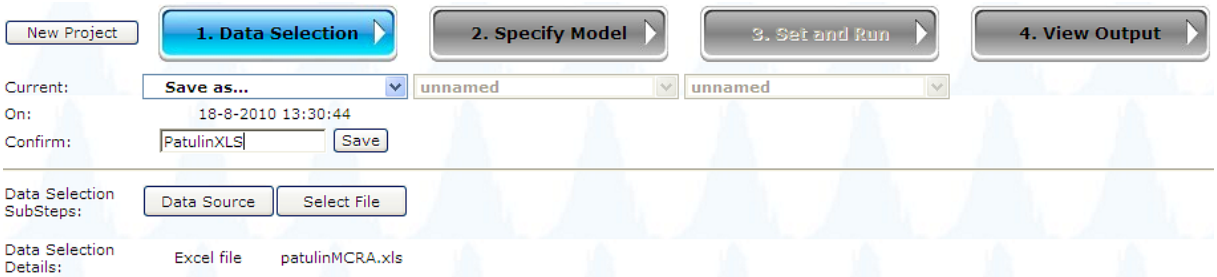
Select File

- Select patulinMCRA.xls
- submit



Check AutoSave Data Selection for saving data with a default name. Here, your selection is stored as patulinMCRA_xls (not shown).

After a successful Data Selection, the central menu indicates which steps were performed together with some short information. The selected data is displayed as **unnamed** and can be saved with a chosen name for future use.



- select scrolldown box under Data Selection
- save as…PatulinXLS

From here model options are specified using the Specify Model button.

References

Boer, de W.J., Voet van der, H., Bokkers B.G.H., Bakker, M.I., Boon, P.E. (2009). Comparison of two models for the estimation of usual intake adressing zero consumption and non-normality. Food Additives and Contaminants. Part A, 26:11,1433 - 1449