EXTERNAL SCIENTIFIC REPORT

A European tool for usual intake distribution estimation in relation to data collection by EFSA[1]

Jacob D. van Klaveren[a],

Paul W. Goedhart[b],

Dagmar Wapperom[a]

Hilko van der Voet[b] [2]

[a] National Institute for Public Health and the Environment (RIVM)
P.O.Box 1, 3720 BA Bilthoven, The Netherlands
Jacob.van.Klaveren@rivm.nl

[b]Biometris
Wageningen University and Research Centre (WUR)
P.O. Box 100, 6700 AC Wageningen, The Netherlands
Paul.Goedhart@wur.nl and Hilko.vanderVoet@wur.nl

DISCLAIMER

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

[1] Question No EFSA-Q-2010-01342

[2] This report was prepared by Jacob D. van Klaveren, Paul W. Goedhart, Dagmar Wapperom and Hilko van der Voet. We wish to thank Davide Arcella, Andy Hart, Heiner Boeing, Polly Boon, Amélie Crepet, Waldo de Boer, Arnold Dekkers, Kevin Dodd, Pietro Ferrari, Fanny Heraud, Marc Kennedy, Victor Kipnis, Sven Knüppel, Marga Ocké, Wout Slob, Gerie van der Heijden and Janneke Verkaik for fruitful discussions during the ETUI workshops and contributions to the project.
Supporting Publications 2012:EN-300. [42 pp.]. Available online: www.efsa.europa.eu/publications

© European Food Safety Authority, 2012

SUMMARY

This report is a deliverable of the project 'European Tool Usual Intake' (ETUI) commissioned by the European Food Safety Authority (EFSA). The aim of the ETUI project is to review available usual intake models and to make a choice of the optimal usual intake model based on statistical consensus. The usual intake model chosen should be operable in the context of EFSA's data collection and data handling.

Two workshops were organised. In the first workshop an inventory was made of the models, issues and questions relevant to the EFSA work. The workshop proceedings can be found on the EFSA website. The second workshop was to confirm the approach, the models and the computational tool and to find consensus among the selected statistical experts with regard to how and when usual intake models should be applied.

We describe the computational tool as MCRA version 7.1, which was implemented as part of the web-based system Monte Carlo Risk Assessment (MCRA), enabling an optimal link with the EFSA Comprehensive Database to assess usual intake models. We describe MCRA's extensive options for data handling, detailed modelling, uncertainty analysis, diagnostics and output. The usual intake models available in MCRA are the Observed Individual Means (OIM) approach, the BetaBinomial-Normal (BBN) model, two variants of the LogisticNormal-Normal (LNN) model (also known as the NCI model), and, for historical reasons, an implementation of the Iowa State University Foods (ISUF) model. The majority of the models were already part of MCRA version 7.0 but the LNN model has been adjusted for international uses. Furthermore statistical aspects have been added as a follow up of the discussions during the workshop.

The OIM, BBN and LNN model, as well as models available in other software were tested using a simulation study approach. A summary is given of the characteristics of these usual intake models and the main results of this simulation study. Statistical justification and an overview of detailed results for each simulation scenario are reported in a separate document (Goedhart et al. 2012).

Along with the computational tool, statistical guidance is provided for choosing the most optimal statistical model in practical cases. The statistical guidance is largely based on the experience of the simulation study. According to the statistical guidance, first the objective of the assessment should be defined: interest can be in consumers with a relatively high or a relatively low intake corresponding with the right or left tail of the distribution, and either realism or conservatism of the exposure assessment compared to the expected real exposure may be of prime importance. Second, the available input data should be investigated and described. Then a tentative optimal statistical model can be chosen. The two LNN models appeared to perform best for realistic assessments of the right tail of the distribution (e.g. relevant to risk assessors and managers in case of exceedance of toxicological reference values). The choice between both LNN models depends on the presence of a correlation between intake frequency and intake amount.

The OIM model is suitable for conservative assessment of the right tail of the distribution in most cases. A known exception is that of multi-modal distributions. Diagnostic plots are needed to judge the appropriateness of the chosen tentative optimal usual intake model. These plots are also part of the computational tool. In some cases the plots point to complex situations and statistical expertise is needed to use the tools in a scientific sound manner.

Two case studies were performed using the computational tool. One case study on lead served as example to show that the usual intake tool was able to estimate usual intake using consumption data available in the EFSA comprehensive database and concentration data such as present at EFSA. In this example, interest was in a realistic assessment of the right tail of the intake distribution. Intake of lead was daily for most individuals, so BNN, LNN0 and LNN models gave essentially the same results. For

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

most countries transformed lead intake was normally distributed, so the model-based estimates were preferred. For one country this was not the case and model-assisted estimates might be better. The estimated percentage of the population with lead exposures above the benchmark dose lower confidence limit (BMDL) in the upper 10% of the exposure distribution was at least 50% lower than the percentage estimated using the OIM method currently in use by EFSA.

The second case study on a pesticide was conducted because EFSA referred to the ETUI project in a recent published draft guidance on the use of probabilistic methodology for modelling dietary exposure to pesticide residues. The document describes among other things the need of an exposure scenario in which a bimodal distribution is expected.

Both case studies illustrated the usefulness of diagnostics. First, insight in sources of high levels of intake learned that there were probably outliers in the input databases which might be judged as unrealistic. A risk assessor might decide to rerun the simulations without these outliers providing that it can be justified that the reported values are unrealistic. Second, the intake distribution plot showed that the distribution was bimodal. In this case the OIM estimates of the high percentiles were not conservative because the underlying assumption of LNN and BBN models were not met . Usual intake of this example was then modelled separately for separate sources of intake and then combined in total usual intake (‘model-than-add’, rather than ‘add-than-model’) which improved the applicability of the LNN and BBN models significantly.

In conclusion, the ETUI project delivered a computational tool with several models to estimate usual intake distributions. Connected to this, guidelines to choose the most optimal model for a given situation were prepared. Because of on-going developments in this research field, it is recommended to review and update the tool and guidelines after a period of use in practice.

KEY WORDS

Usual intake model, exposure assessment, probabilistic

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

TABLE OF CONTENTS

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

BACKGROUND AS PROVIDED BY EFSA

Dietary measurements collected with 24-HR and dietary records cover food consumption over a limited timeframe , i.e. one day when one single measurement per subject is collected. For chronic risk assessment calculations, there is a need for data on "usual" daily intake of a dietary component, commonly identified as the average daily intake over the past year . In this context, estimates of exposure to a given hazardous compound are customarily compared to values of Acceptable Daily Intake (ADI) for deliberately added compounds or Tolerable Daily Intake (TDI) for contaminants, which represent estimates of the amount of the compound that can be ingested daily through food over a lifetime without appreciable health risk.

Measuring "usual" intake of foods from surveys with a limited number of short-term dietary measurements per subject is challenging. An important limitation for the use of short-term food consumption instruments to estimate "usual" intake is that individual diets can vary greatly from day to day. In addition, short-term dietary assessments are affected by measurement errors. These factors contribute to considerable within-person variability, thus making measured intakes over a limited number of days a poor estimate of "usual" intake. In practice, within-person variability tends to inflate the observed distribution of the intake of a given food, thus leading to overestimation of extreme percentiles in the observed distribution of intake, e.g. percentiles 90th, 95th, or 99th.

To translate short-term measurements of intake into estimates of "usual" consumption, the use of the Nusser method was first advocated. Since then, a number of statistical procedures have been developed to correct consumption distributions, and thus obtain individual unbiased estimates of "usual" intake. The various statistical methods share a common framework, while differences arise from the set of assumptions about the measurement characteristics, and from the adopted computational solutions.

In addition, rarely consumed foods might not be captured at all in short-term measurements, making it very difficult to distinguish between zero-values and real non-consumers. This is particularly true if the number of replicates is small, thus presenting unique challenges for statistical modelling. Ideally, 24-HR and dietary record measurements should be supplemented by additional information about frequency of consumption, usually available in dietary assessment methods for long-term dietary exposure, such as Food Frequency Questionnaires (FFQ), Food Propensity Questionnaires (FPQ) or Dietary History Questionnaires (DHQ). In this way, in a probabilistic framework, it is possible to identify real non-consumers of a given food in a population.

TERMS OF REFERENCE AS PROVIDED BY EFSA

The main objectives of the project are to:

1. Review and discuss the existing methodologies that are used to correct short-term dietary assessment and provide individual estimates of "usual" intake for data in the EFSA Comprehensive Database.
2. Develop procedures for the correction of short-term assessments of food intake to estimate individual "usual" consumption in the EFSA Comprehensive Database, thus addressing MS data with one short-term measurement only, and data where more than one 24-HRs or dietary records per subject are available.
3. Implement the above mentioned procedures to regularly and rarely consumed foods in the EFSA Comprehensive Database, possibly with limited or no information on the frequency of consumption of specific foods.
4. Analyse the results from point 3 to evaluate the procedures developed in point 2, providing recommendations for potential improvements and/or revisions of the procedures and in view of the future design of the EFSA pan-European survey.

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

This grant was awarded by EFSA to:

Beneficiary: RIVM, on behalf of the Dutch Ministry of Public Health, Welfare and Sport

Grant title: Probabilistic tool to estimate “usual” intake distribution in the Comprehensive Food Consumption Database

Grant number: CFP/EFSA/DATEX/2009/03

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

1. INTRODUCTION AND OBJECTIVES

The project 'European Tool Usual Intake' (ETUI) was commissioned by the European Food Safety Authority (EFSA) (CFP/EFSA/DATEX/2009/03)[3]. The project started in December 2009 and ended in March 2012. The

Different terms are in use to describe the exposure assessment phase regarding food. For the purpose of this document dietary intake is considered synonymous with dietary exposure, that is we do not consider bioavailability issues and focus on the food as consumed. Usual intake (or habitual intake) is the long-term average intake of an individual. Typically usual intake distributions in a specified population are estimated from repeated food consumption survey data and associated concentration data of the chemicals or nutrients of interest. Statistical models are needed to obtain usual intake distribution estimates from data at person-day level (see e.g. Dodd et al. 2006). The most important aims of the ETUI project were threefold:

1. to review and to optimize already existing usual intake models;
2. to find consensus among statisticians regarding the optimal model;
3. to produce a computational tool workable at the European level.

The ETUI project aimed to produce four deliverables:

1. two workshops to find consensus among the statistical experts;
2. an interim report discussing statistical modelling of usual intake and two final reports;
3. a computational tool for usual intake assessment which can be used in combination with EFSA's data collection, including several models to estimate usual intake distributions from repeated survey data, and a procedure to generate usual intake results from food consumption database with only one day of food recording based on extrapolating information from other sources;
4. statistical guidance for choosing a model for usual intake distribution estimation.

This report describes the aim of the ETUI project and its deliverables; the computational tool, a short overview of the usual intake models addressed in the simulation study and the general conclusion of the simulation study. Furthermore it describes the statistical guidance on how to select the most optimal usual intake model based on data characteristics and risk management questions, the case study to test the statistical guidance and the and a summary of the discussions and presentations from the second ETUI workshop. Next to this report, a separate report was written on the statistical details and full results of the simulation study on estimating usual intake distributions for episodically consumed foods (Goedhart et al. 2012). Both reports should be seen as deliverables. It was decided to make two reports because the level of detail and the scope of both reports varied too much.

A first workshop was held on April 27th and 28th 2010 to find consensus on the statistical issues relevant for usual intake modelling. The usual intake models that were in use at that time were discussed and an inventory was made of the statistical issues relevant for the development of the computational tool. EFSA provided an overview of their data collections and statistical requirements as far as relevant for EFSA panels. A report has been published on the EFSA website (van der Voet and van Klaveren 2010).

---

[3] The project was co-funded by the Netherlands Food Safety and Consumer Product Authority and the Dutch Ministry of Economic Affairs, Agriculture and Innovation.

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Based on the review and the first workshop, criteria were defined for model selection. The initial perception of EFSA was that preferably one model should be selected for use in all cases. It was recognised during the workshop and thereafter that different questions or different data sets might require different usual intake modelling approaches. The outcome of the workshop was discussed with EFSA resulting in the definition of the computational tool as a toolbox of various models, rather than a single model.

This toolbox will allow the use of usual intake models on a case-by-case approach depending on the type of data and refinement needed in the exposure assessment.

For practical reasons the ETUI project was restricted to the selection of methods which were already available in the public domain and directly accessible to the project partners so that relevant adjustment in the source code could readily be made. For the implementation the existing software MCRA Release 7.0 was considered as a starting point because it was already able to handle many relevant usual intake models before the start date of the ETUI project. It was recognized that usual intake models are or might become available also in different software environments like the macros provided by the National Cancer Institute (NCI), the Multiple Source Model (MSM) developed by the German Institute for Nutrition (DIfE), and a program developed at RIVM (SPADE).

In chapter 2 we describe the computational tool developed in this project and available as part of the MCRA release 7.1 software. In the ETUI project only already existing knowledge and models were included, but statistical issues raised at the workshop were translated into additional computer programmes and were added to the already existing models. Addressing and implementing also new developments in the field of usual intake models were beyond the scope of the EFSA request and budget.

In this simulation study shortly summarised in chapter 3 we investigated the role of food frequency questionnaire (FFQ) data as additional information when estimating usual intake distributions, as well as the effect of correlation between intake amounts and frequency of consumption. For a detailed description of the simulation study and a full overview of the results and conclusions we refer to Goedhart et al. 2012. A conclusion from the simulation study is that when the right model is used, inclusion of FFQ information is not beneficial when interest is in the upper percentiles only.
Statistical guidance on how and when to use usual intake models has been made based on the simulation study and based on statistical discussions during project meetings and during the first workshop. The statistical guidance is given in chapter 4 of this report.

How the computational tool can be used in connection to the EFSA comprehensive database is demonstrated in the case studies reported in chapter 5. The case studies were also used to test the statistical guidance. The case studies were not part of the original contract but it was agreed with EFSA that a few case studies could be very relevant to demonstrate the potential of usual intake models in relation to the work that EFSA's Unit for Dietary and Chemical Monitoring has to perform for the EFSA Panels or to be seen as examples if Member States want to use the usual intake models in the nearby future. Lead was chosen because of the ongoing work of the EFSA Contam Panel on this issue as well as the reported exceedances of lead exposure compared to benchmark dose lower confidence limits (BMDL)[4] to evaluate the risk to lead exposure (EFSA 2010).

[4] A benchmark dose lower confidence limit (BMDL) was used as point of departure for comparing the toxicological findings with the calculated exposure because the EFSA Contam Panel could not derive a Tolerable Daily Intake. The BMDL is based on human data.

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

The simulation study, the computational tool, the statistical guidance and the case studies were discussed during a second ETUI workshop, which was held on February 8th 2012, In chapter 6, we summarize the discussions and presentations given during this final workshop. Overall conclusions and recommendations are given in chapter 7.

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

2. COMPUTATIONAL TOOL FOR USUAL INTAKE ESTIMATION

A main deliverable of the ETUI project is a web-based computational tool for estimating usual intake distributions, allowing to connect selected statistical models to the data from the EFSA Comprehensive database. The implementation of this computational tool has been made as a further development of the existing Monte Carlo Risk Assessment (MCRA) system, and is now available as part of MCRA Release 7.1 (**mcra.rivm.nl**). This document describes the main features of the program relevant for usual intake estimation, and provides guidelines for choosing the optimal model available in MCRA 7.1. For usage of MCRA 7.1 itself we refer to the manual on the website (de Boer and van der Voet 2011).

MCRA 7.1 implements the following statistical models for usual intake:

- OIM – Observed Individual Means. Naive method just averaging the individual intake over survey days. May be used as first tier or as replacement when BBN or LNN approaches fail because they are not fit for the purpose of use, depending on data characteristics.

- BBN – BetaBinomial-Normal model. Two part model, applicable for daily and/or episodical intakes. Based on a BetaBinomial distribution for intake frequencies and a Normal distribution for log or Box-Cox transformed intake amounts (see Slob 2006, de Boer et al. 2009). Provides model-based [5] and model-assisted [6] distribution estimates (Goedhart et al 2012).

- LNN0 – LogisticNormal-Normal-nocorrelation model. Two part model, applicable for daily and/or episodical intakes. Based on a LogisticNormal distribution for intake frequencies and a Normal distribution for log or Box-Cox transformed intake amounts (see Tooze et al. 2006, Goedhart et al 2012). Provides model-based and model-assisted distribution estimates.

- LNN = LogisticNormal-Normal model. Model also known as NCI model. Similar to LNN0, but accounting for a correlation between logistically transformed frequencies and log or Box-Cox transformed intake amounts (see Tooze et al. 2006, Goedhart et al 2012). Currently only provides model-based distribution estimates.

- ISUF – Iowa State University Foods model. Two part model, applicable for daily and/or episodical intakes. Implements the main ideas of a semiparametric transformation approach to estimating usual intake distributions (see Nusser et al. 1996, 1997, Dodd et al. 2006, de Boer et al. 2009). The ISUF model is mainly present in the computational tool for historical reasons.

Data handling and options for detailed modelling

MCRA 7.1 provides extensive possibilities for linking consumption data and nutrient or chemical concentration data such as are needed to obtain daily intakes as input for the models above:

- Data can be entered separately for the amounts of consumed foods obtained (usually from consumption surveys) and nutrient or chemical concentration in a given quantity of foods (usually from occurrence monitoring programs). The computational tool produced in the ETUI

[5] A model-based estimate of the usual intake amounts distribution is back-transformed from the normal distribution assumed for transformed amounts.
[6] A model-assisted estimate of the usual intake distribution is back-transformed from a shrunken version of the transformed OIM distribution, where the shrinkage factor is based on the variance components estimated using the linear mixed model for amounts at the transformed scale.

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

project links the consumption and concentration data and sum the intake of the nutrient or chemical over the foods that contain it to obtain person-day intake values. Advantages of entering the data on consumption and concentrations as part of the computational tool are:

- To diagnose bi- or multimodal distributions. The type of distribution after transformation is mainly related to the input data. If the data are causing the bi- or multimodality food items can be separated and the 'model-then add' approach can be applied (see also chapter 5).
- The possibility for a detailed drill-down from a specific percentile of the estimated usual intake distribution. The drill-down is an important part of the computational tool and enables the risk assessor to check the realism of usual intake modelling for example by plotting unrealistic outliers in the input data.

- Alternatively, pre-calculated daily intake data can be entered, as commonly used for most other software programs for usual intake estimation. In this way the ETUI computational tool can be used by risk assessors not having access to detailed input data.

- Food coding used for the monitoring programs (foods-as-measured) are often different from food codes used in food consumption measurements (foods-as-consumed), and a code conversion can be made based on food code translation tables (see Boon et al. 2009). Food conversion is an essential part of the computational tool because it's use can effect the precision of the calculation. A food conversion model is used in the case study addressing pesticides (see also chapter 5).

- The food code conversion algorithm allows for hierarchical codes, such that it is easy to link consumption data on a food level to concentration data on a food group level. In the EFSA comprehensive database the food is coded with the FoodEx1 code, which is a hierarchical code. The food code conversion algorithm is an essential requirement of the computational tool in the data environment of EFSA because concentration data is not always available at the food level, but more often at the level of subgroup or food group level (one or two levels higher in the hierarchical coding system).

- Subset selection can be used for e.g. characteristics of individuals in the population (e.g. age, gender), foods-as-consumed and foods-as-measured. The risk assessor can use this functionality of the computational tool to correctly link the toxicological relevant effect to the right age or gender group (e.g. in the case of a neurotoxic effect of lead the exposure should be calculated for babies and infants instead of adults because these are the relevant age group for the toxic effect). As another example, for chemicals having an effect on the development of the unborn child the subpopulation of females in the age range of 15-45 years is considered to be the appropriate exposure population.

- Processing factors can be specified to correct for processing effects on concentration (e.g. vitamin C in boiled potatoes vs. raw potatoes). It is even possible to specify that these processing effects themselves are variable by providing a high percentile of a distribution of processing factors. Brand loyalty can be modelled when concentrations have been measured for separate brands of a food, but consumption data are generic. Brand loyalty might cause bimodal distribution.

- Left-censored data points or non-detects may be included in the modelling of chemical concentrations, or they can be imputed by either 0 or a specified fraction of the Limit of Reporting (LoR). The use of the limit of reporting as a replacement for non-detects is a requirement in basic probabilistic assessment prescribed in the EFSA (draft) guidance on the use of probabilistic methodology for modelling dietary to pesticide residues (EFSA 2012).

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

The replacement of non-detects might effect the shape of the usual intake distribution and the applicability of usual intake models. This functionality is seen as part of the computational tool.

- For the case where only one day per individual is available in the consumption survey, MCRA allows to use externally generated statistics (e.g. from other MCRA runs) to quantify between-day variation as input for the statistical models BBN, LNN0 and LNN. This was an additional requirement of the computational tool set by EFSA.

- From the consumption survey data a covariable (quantitative characteristic, e.g. age) and/or a cofactor (qualitative characteristic, e.g. gender) can be selected to model the usual intake as a function of these. This is useful when it is known a priori that usual intake will depend on these covariates. For example, children are often experiencing a higher exposure per kg body weight in comparison to adults.

- Uncertainties were identified as an important issue during the first ETUI workshop (van der Voet and van Klaveren 2010). Input data for the usual intake models are invariably uncertain. MCRA 7.1 is based on the idea of embedding the statistical analysis in a loop for uncertainty analysis. For datasets with replication (consumption surveys, nutrient or chemical concentration datasets) the non-parametric bootstrap can be used to quantify the effects of sampling uncertainty. Uncertainty analyses can also be performed for other relevant exposure parameters such as processing factors and portion sizes. Distributions fitted to concentration data can be corrected for measurement uncertainty.

Diagnostics

'All models are wrong, some are useful'. In order to evaluate whether the use of a specific statistical model for usual intake is reasonable in a given case, it is essential to be able to have some means for diagnosing data, model fits and outcomes. MCRA 7.1 has, among other things, the following options:

- Assessment of normality of the estimated residuals for individuals in the models. QQ plots and statistical tests are given, so that assumptions such as normality or presence or absence of correlation between amounts and frequency, can be formally tested and (often more usefully) plots might give guidance to the optimal decision.

- Graphs showing the daily intake distribution on the original and the transformed scale, and a zoom-in on the right tail.

- Pie charts showing the contribution of individual foods to the overall intake, and to the highest levels of intake.

- Drill-down, showing the detailed data (list of relevant consumptions, concentrations, processing factors) for nine individuals around a chosen percentile (e.g. a high tail percentile) of the usual intake distribution.

- Intake levels as different percentiles (both in tables and in plots).

- Percentage of population exceeding nutrient intake reference values as Average Nutrient Requirement (ANR), Estimated Average Requirement (EAR), Upper Nutrient Level (UNL) or toxicological reference values (ADI, TDI) or other Point of Departures to compare exposure levels with toxicological findings like the BMDL.

- Overview of values of all input variables (e.g. how much was eaten, what were the concentration values from multiple foods used in the exposure assessment).

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

- Additional output to check input and output used in the exposure assessment and the setting of various variables such as left-censored data used in the models.

For more details of the MCRA software not directly linked to the ETUI project we refer to the MCRA reference manual (de Boer and van der Voet 2011).

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

3. MODELS ADDRESSED IN THE SIMULATION STUDY

The purpose of the ETUI simulation study is to compare different methods to estimate the usual intake distribution for episodically consumed foods. An important criterion is the bias and precision with which the upper tail of the usual intake distribution is estimated. An extensive overview of the purpose of the simulation study, the way the simulation study has been set up and how it has been performed, and a detailed overview of all results generated are described in a separate report (Goedhart et al. 2012). In this chapter we shortly repeat the models and software that has been addressed in the study and the main conclusions from the simulation study. It should be noted that simulation studies can be very useful but are always limited with respect to the models used for simulation and the scenarios that can be investigated. Specifically, in this simulation study the attention was restricted to one or two foods, whereas in reality substances in a risk assessment may occur on a much larger number of foods, where perhaps one of the rarely consumed foods is still responsible for very high intakes. Further research would be needed to investigate the validity of the conclusions from the current study for such cases.

3.1. Observed Individual Means (OIM)

The Observed Individual Means (OIM) method uses individual mean intakes to characterize the usual intake distribution, see e.g. Dodd et al (2006). This method does not separate the between-individual variation from the within-individual variation. As a consequence the distribution of the individual means contains considerable within-individual variation and the OIM estimate of the usual intake distribution is thus too wide. For episodically consumed foods there can be a lot of non-consumers which implies that lower percentiles of the usual intake distribution are estimated by zero and are thus under-estimated. OIM cannot handle covariates. Despite these obvious drawbacks, OIM is still popular because it gives conservative, i.e. too large, estimates of the upper percentiles of the usual intake distribution, which might reduce the risk of setting too low food safety limits. The simulation study uses the results of the MCRA 7.1 implementation of OIM.

3.2. Iowa State University Foods Model (ISUF)

The second model was developed at Iowa State University (Nusser et al 1996, Nusser et al 1997). Dodd et al (2006) refer to this model as the Iowa State University Foods (ISUF) model, and we will follow that notation. The original implementation of ISUF is in the C-SIDE program (Dodd 1996). This method is the first attempt to model both the frequency and the amount of consumption followed by an integration step to estimate the usual intake distribution. The ISUF model is rather complex and involves many steps especially with respect to finding an appropriate transformation of the positive intakes and with respect to modelling the frequency of consumption on a discrete grid. The model cannot address covariates, although in the C-SIDE implementation *a priori* data adjustments can be made for non-individual specific biases such as season or day of the week.

The simulation study employs the MCRA 7.1 implementation of ISUF which implements the main features of the C-SIDE program (de Boer et al 2009).

3.3. BetaBinomial-Normal model (BBN)

The BetaBinomial-Normal model uses the betabinomial distribution to model the frequency of consumption and a one-way normal random effects model for transformed positive intakes. The betabinomial model assumes that the probability of consumption varies among individuals according to the beta distribution. The intake is assumed to have an approximate normal distribution after a natural logarithmic or Box-Cox transformation. The random effects model for the transformed positive intakes separates between-individual variation from within-individual variation. Both models can be

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

extended with covariates such as age and gender. The results of both models are combined to estimate the usual intake distribution. This model has been part of MCRA from 2005 onwards and is fully described in de Boer et al (2009). A similar model is employed by Slob (2006), although Slob relates the α parameter of the beta distribution to covariates rather than the mean probability $\frac{\alpha}{\alpha+\beta}$ as is more usually done.
The simulation study employs the MCRA 7.1 implementation of the BBN model. The frequency model is fitted by means of a C# implementation of GenStat procedure betabinomial (Goedhart, 2011), while the amounts model employs LME in the R statistical package (Pinheiro and Bates, 1999).

3.4. LogisticNormal-Normal model without and with correlation (LNN0, LNN, NCI)

The LogisticNormal-Normal model employs the same random effects model for the transformed positive intakes as BBN, but uses a different model for the frequency of consumption. The frequency model in LNN0 is an ordinary logistic regression model with an added random between-individual effect on the scale of the linear predictor (Williams, 1982). The resulting distribution is termed Logistic-Normal. The combination of this frequency model with the random effects model was brought to usual intake estimation by Tooze et al (2006). Their seminal paper also included an extension by incorporating a correlation between frequency and amount, see the LNN model below. However the LNN0 method described here ignores this correlation, or, equivalently, fixes the correlation to equal zero.

The simulation study employs the MCRA 7.1 implementation of LNN0 and also the most recent implementation using SAS macros obtained from the National Cancer Institute in the United States (Dodd, 2011). The latter implementation will be termed NCI0 in this report, where NCI is the original name for the model. There is no analytical expression for the LogisticNormal distribution and therefore both implementation use Gauss-Hermite quadrate to obtain a generally accurate approximation of the distribution. LNN0 uses 32-point integration while NCI0 uses an adaptive form of Gauss-Hermite integration. Both implementations estimate the parameters by means of maximum likelihood employing a general optimization method. NCI0 uses SAS PROC NLMIXED while LNN0 in MCRA employs the Simplex optimization method.

The LogisticNormal-Normal (LNN) model is the model as proposed by Tooze et al (2006), also see Tooze et al (2009). It extends the LNN0 model by including a correlation between the individual random effect in the frequency model and the between-individual random effect in the amounts model. More specifically, instead of two separate normal distributions for these two random effects, a bivariate normal distribution is used to model the joint random effect. The resulting joint distribution of frequency and intake can be approximated by two-dimensional Gauss-Hermite integration.

The simulation study employs the MCRA 7.1 implementation of LNN and also the most recent implementation using SAS macros obtained from the National Cancer Institute in the United States (Dodd, 2011). The latter implementations will be termed NCI in this report. NCI employs SAS PROC NLMIXED while LNN in MCRA uses the Simplex optimization method. The default number of integration points used for LNN is 32. Since two dimensional integration is used, the total number of integration points is the square of this number, i.e. 32^2=1024. This is computationally very expensive and therefor LNN is also employed with 4^2, 8^2 and 16^2 integration points. These are termed LNN(4), LNN(8), LNN(16).

3.5. Statistical Program to Assess Dietary Exposure (SPADE)

The Statistical Program to Assess Dietary Exposure (SPADE) was developed in the context of the Dutch National Food Consumption Survey 2007-2010 (van Rossum et al., 2011). A preliminary

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

version is used in the European Food Consumption Validation project (EFCOVAL, Souverein et al., 2011). SPADE is based on ideas of the program AGEMODE for daily intakes (Waijers et al., 2006) which models the usual intake distributions as a function of age. The SPADE model is in principle the same as the BNN and the LNN0 model and includes both methods. It does not include the LNN model. SPADE uses a Box-Cox transformation in order to obtain normally distributed observations. SPADE is implemented in R (R Development Core Team, 2011) and uses the R packages *mfp* (Ambler and Benner, 2010) for modelling the fractional polynomials, *amer* (Scheipl, 2011) for the truncated polynomials, *lme4* (Bates et al., 2011) for fitting the logistic normal model and for fitting the random effects model for the transformed positive intakes, and finally *gamlss* (Rigby and Stasinopoulos, 2005) for the BNN model.

In this simulation study the BBN variant of SPADE is used, which employs the packages *lme4* and *gamlss*.

3.6. Multiple Source Method (MSM)

The Multiple Source Method (Haubrock et al, 2010) is specifically designed to estimate individual usual intakes which are then employed to characterize the usual intake distribution. MSM is also developed to include FFQ information, both as a covariate as well as by using the FFQ to identify never-consumers. The model employs ordinary logistic regression for the frequency of consumption and ordinary linear regression for the positive amounts. The residuals of both models are then transformed and subsequently modelled by means of a one way random effects model without any covariates. The shrunken residuals are then back-transformed to their original scale and the individual usual intake is obtained by multiplication of the frequency and amounts result.

The implementation employed in this simulation study uses an offline version of the program available at https://msm.dife.de/ (Harttig et al, 2011). Note that the website also contains a user guide.

3.7. Model-based and model-assisted approaches

The model-based usual intake distribution employs estimated model parameters to simulate the usual intake of many hypothetical individuals which results in a usual intake distribution across individuals. This can only be applied for full parametric models. When there are covariates in the frequency and/or amounts model a usual intake distribution can be derived for various values of the covariates. When the dataset is considered to be representative for the population, the individual covariate patterns are also representative. A usual intake distribution for the population can then be simulated by simulating a fixed number of hypothetical individuals for every individual covariate pattern. The usual intake of all hypothetical individuals then form an estimate of the usual intake distribution of the population. The model based approach is used for ISUF, BBN, LNN0, LNN, NCI0, NCI and SPADE. All models, except ISUF, use a form of Gauss-Hermite integration for the back-transformation of the amounts model (Dekkers et al, 2009).

Model-assisted usual intake uses the fitted model to predict the usual intake of every individual in the dataset. These individual predictions are then employed to form a usual intake distribution. OIM is an example of this approach: the individual predictions are then set to the individual means. MSM uses its own form of model- assisted usual intake. Model-assisted usual intakes are not included in SPADE, NCI0 and NCI.

3.8. Main conclusions simulation study

The purpose of the ETUI simulation study was to compare different methods to estimate the usual intake distribution for episodically consumed foods. An important criterion is the bias and precision with which the upper tail of the usual intake distribution is estimated. It is further investigated whether

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

it is beneficial to include food frequency questionnaire (FFQ) information in the models as a covariate. The data are simulated with the logistic-normal model for the frequencies and a two-way random effects model for the log-transformed amounts.

Four different frequency models are employed with low, moderate and high consumption frequencies, as well as a so-called bathtub model. This is combined with two ratios (1 and 4) for the variance components in the amount model as well as three values (0, -0.5 and 0.5) for the correlation between frequency and amount. This gives a total number of 24 different scenarios. The answer to a hypothetical FFQ with seven response categories is derived by discretizing the simulated consumption. Three datasets were simulated for each scenario with 6250 individuals and two recall days.

The models and software described in section 3.1 to 3.7 were used in the simulation study. In an additional simulation, with 50% never-consumers and four scenarios, it was tested whether it is advantageous to include information on which individuals are never-consumers and which are episodic consumers which happen to have zero consumption on the recall days. Also, for three practically relevant scenarios with two foods, the ‘add-then-model’ approach is compared with the ‘model-then-add’ approach.

The main conclusions are that a practical approach for single foods would be to fit the LNN model and to revert to the LNN0 model when the estimated correlation is low, or when the frequency of consumption is large. For data similar to those in this simulation study, when the right model is used, inclusion of FFQ information is not beneficial when interest is in the upper percentiles only. For multiple foods the ‘model-then-add’ approach seems to be quite promising. Correlations between foods can then be accommodated by using the model-assisted approach.

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

4. STATISTICAL GUIDANCE FOR USUAL INTAKE CALCULATIONS

The choice for a usual intake model should be made case by case. Below we focus on standard situations, which are situations resembling the practical datasets as we know them from EFSA practice (see also chapter 5 case studies), and which formed the basis for our simulated datasets. It should be clearly stated that in non-standard situations other methods and approaches may be preferable.

As the outcome of the discussion during the first ETUI workshop it was agreed with EFSA to study the models OIM (representing much of current practice), BBN (used before in EFSA opinions) and LNN0/LNN (closely resembling the NCI methods) in the simulation study. The above mentioned models were made available in the MCRA 7.1 software. Because the computational tool is a contractual deliverable as part of the ETUI project the focus in the statistical guidance is on the models available in this MCRA software. Other tools studied were: SPADE (which implements a model very close to BBN), NCI (implementing the same models as LNN0 and LNN) and MSM. Experience with food frequency questionnaires (FFQ) has been tested with MSM and LNN0.

Based on the experience of the simulation study and the statistical discussions we propose a three step approach to guide the user on how and when to use usual intake models. Different types of assessment can lead to selection of different usual intake models. The three steps are:

A. Define the assessment type
B. Investigate the available data
C. Choose a tentative model

Step A. Define the assessment type:

1. Right tail of usual intake distribution is of primary interest: realism is more important than conservatism. This might occur when a toxicological reference value is exceeded;
2. Right tail of usual intake distribution is of primary interest: conservatism is more important than realism.
   This might occur when legislation requires a conservative approach or when exposure levels are far from exceeding the toxicological reference values;
3. Left tail of usual intake distribution is of primary interest. This might occur when a deficient intake of nutrient is addressed in the exposure assessment.

In risk assessment the percentage of the population exceeding toxicological reference values like the Acceptable Daily Intake (ADI) or Tolerable Daily Intake (TDI) or benchmark dose lower confidence limit (BMDL) is the most important output. Risk managers want to avoid unneeded investment in risk mitigation measures, and unnecessary public concern and therefore an over-estimation of the percentage exceeding these limits is of particular concern for risk management. The percentage of the population exceeding the ADI or TDI will vary among chemicals and countries.

Sometimes legislation defines a certain cut-off value of the exposure distribution above which value a chemical can not be authorised. The risk assessor then should report the exposure results at and around the percentile corresponding with the cut-off value. Furthermore precision can be an issue in decision making e.g. how sure does the risk manager want to take the decision? Depending on the toxicity and the relevance the risk assessor might need to report the usual intake as a function of covariates? Questions can be raised about whether these should be done quantitative or qualitative?

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Step B. Investigate the available data:

- How many days per person are available in the consumption survey data?
- Is FFQ information available additional to the survey? In case of episodical intake, does this include information about the proportion of never-consumers?
- Is the intake daily or episodical? For episodical: Is the percentage of positive intakes high or low?
- Is the intake mainly from a single food or from multiple foods?
- Is the logarithmically transformed positive intake distribution approximately normal? Note that judgment of normality of both between-person and within-person distributions is in principle impossible based on the data alone, so the judgment will have to involve subjective aspects based on expert knowledge.

Step C. Choose a tentative model:

General:

- For daily intakes LNN0, BBN, LNN implemented as models in MCRA 7.1, NCI and SPADE as other software packages including the NCI (=LNN) model or BBN model respectively are identical.
- OIM and ISUF cannot handle covariates other than using these methods on subsets of the data.
- There is currently little to be said about a preference between BBN, SPADE and LNN0 in standard situations. Results are typically very similar (Goedhart et al. 2012). A reason to prefer LNN0 is the possible extension to LNN (including correlation between intake frequency and intake amount).
- LNN can be used to test for a correlation between intake frequency and intake amount. In case this correlation is significant and leads to a relevant change in outcomes, LNN is to be preferred.
- BBN and LNN0 produce both model-based and model-assisted estimates of the usual intake distribution (chapter 3). If the normality after transformation is judged to be OK, then the model-based output is preferred, otherwise the model-assisted output might be preferable. In a 'model-than-add' approach these preferences may be different (see below).
- ISUF predictions deviated from the true simulated values and the method can not include covariables and therefore ISUF should be considered as not suitable for use.
- There are no reasons to prefer MSM over other available model-assisted methods in standard situations. Results from the simulation study demonstrated deviation from the true intake in several simulation scenarios (Goedhart et al. 2012).
- OIM does not make model assumptions and can therefore be used if model assumptions for the other models are blatantly wrong. However, it can lead to strongly biased results, notably underestimating in the left tail, but also under- or overestimating in the right tail: in the simulations (Goedhart et al. 2012.) we found under-estimating of high percentiles in some scenarios, notably with small frequency of consumption and a large between-day variance.
- Most parametric models are based on normality on a transformed scale. The 'model-than-add' approach may alleviate non-normality problems for intake based on multiple foods. In the current computational tool (MCRA release 7.1) the 'model-than-add' approach requires the performance of multiple analyses and combining the results afterwards in a spreadsheet program (plans are to allow this within MCRA in a future version).
- The adding involved in the 'model-than-add' approach can be applied to model-based or model-assisted estimates. The former may have better precision under normality assumptions and when there are no correlations between foods, the latter retains the correlations between

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

foods. Case-by-case judgment should be made to decide for model-based or model-assisted addition.

- For estimating percentiles of the usual intake distribution or estimating the percentages exceeding an intake limit, the use of additional FFQ information is not needed if sufficient survey data are available. Note that this may be different if the amount of information in the survey is less that in the data as simulated in this study, and it is also different if the purpose is to estimate a relationship between intake and diseases (regression calibration) as occurring in epidemiological research.
- Information on the percentage of never-consumers can slightly improve results when the modelling is restricted to the complementary part of the population, but just ignoring this information is not expected to alter results much when upper percentiles are of primary interest.
- When only one day per person is available, preferably an external estimate of the variance components should be used in the parametric models. Otherwise OIM can be employed as the only alternative left.

Depending on the assessment type determined in step A the following guidelines apply:

Assessment type A1: realistic right-tail assessments

- Start with the LNN0 or LNN model. Use LNN if normality is OK (pragmatically) and the intakes are episodical. Test if the additional correlation between intake frequency and amount is significant. If no correlation are diagnosed use the LNN0 approach. The BBN is for daily intakes without correlation comparable with LNN0.
- If the diagnostics indicate non-normality of the transformed intake distribution based on the qq-plot, consider to apply the ‘model-than-add’ approach, applied to sensibly chosen groupings of the underlying foods.
- If there is a mild non-normality then model-assisted estimates from LNN0 are thought to be preferable to model-based estimates. However, in the extreme tails there is no empirical basis for model-assisted estimates, and only model-based estimates make sense. An advantage of using both model-based and model assisted in parallel is that a drill down can be made to check for potential errors in input data (see diagnostics described in chapter 2 and the example given in chapter 5).
- If non-normality remains a problem, the ultimate fall-back option is OIM, but it should be realised that bias (mostly over-estimation in the extreme right tail) can be severe (see Goedhart et al. 2012).

Assessment type A2: conservative right-tail assessments

- If interest is in extreme high percentiles (>p95) the OIM method provides a simple approach.
- If also lower percentiles are of interest and/or if some degree of realism is also needed, fit the models as given under A1, using a bootstrap procedure to generate uncertainty intervals for the percentile(s) of interest. The upper uncertainty limits can then be used as conservative estimates.

Assessment type A3: left-tail assessments

- In principle use the models as described under A1.
- In general for episodic data the left tail is more difficult to estimate than the right tail, because it is more dependent on the non-consumptions.

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

- Model-based methods are preferable to model-assisted approaches.
- If there is a correlation between intake frequencies and intake amounts it is more important than for the right-tail to include it in the modelling, so model-based LNN is a method of choice then.
- Do not use OIM for left-tail assessments of episodical intakes.

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

5. CASE STUDIES

5.1. Introduction to the case studies

Lead was selected as a case study because in the EFSA Scientific Opinion on Lead in Food it is stated that extreme consumers can exceed the BMD. for chronic kidney disease and the BMDL for systolic blood pressure (EFSA 2010a). Therefore a type A1: a realistic right tail assessment is required (chapter 4).

A **pesticide** was chosen because of issues raised in the PPR panel addressing cumulative dietary pesticide risk assessment. In January 2012 EFSA published a draft guidance on how probabilistic methodologies can be used as part of the pesticide risk assessment. Two different scenarios for chronic pesticide exposure assessment are described in the draft guidance: 1) actual exposure assessments as part of annual reporting of monitoring results, and 2) authorization of plant protection products in which field trial data is used for the focal commodity for which an authorization is requested and background exposure using monitoring data for all other food items (EFSA, 2012). A focal commodity is the commodity for which an Maximum Residue Limit has to be set. Both scenarios should be performed for acute and chronic effects. For chronic exposure assessment the draft guidance makes reference to the ETUI project. For the pesticide case the normality of the exposure levels after being transformed to the lognormal scale might be questionable especially in the authorization scenario in which scenario higher residue concentrations levels in the focal commodity, for which commodity a Maximum Residue Limit (MRL) has to be set, are combined with much lower residue concentrations levels in background commodities. Consumers eating the focal commodity are expected to have significant higher exposures levels compared to consumers without consumption of the focal commodity. The result might be a bimodal distribution. In the case study we focus on the chronic scenario for the purpose of authorization

EFSA uses the raw individual food consumption data of the Comprehensive database to conduct risk assessments and other scientific analyses within the activities related to EFSA's mandate. For any other use of the data, formal authorisation from the data providers is needed. For this the ETUI project was presented during the 4th meeting of Expert Group on Food Consumption Data in which all European Member States are presented. The ETUI project was received very well by the members of this group, and access was granted on a confidential basis by signing a confidentiality agreement by the RIVM and WUR/Biometris co-workers. Also the concentration data were provided to the project by EFSA on a confidential basis. For the reasons of confidentiality not all results are reported in full detail in this chapter:

The aim of the case studies was threefold:

1. to gain practical experience with the selected usual intake models also in terms of impact for risk assessment and risk management;
2. to test applicability of the statistical guidance as described in chapter 4;
3. to test the computational tool in EFSA's data environment.

5.2. Methods and data used in case studies

Consumption data

An overview of the consumption data used in the ETUI project is given in table 5.2.1. The foods in all databases were coded using one standardized coding system, the FoodEx1 system (EFSA, 2011b). The

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 5.2.1: overview of consumption data used in the ETUI case study. A full description of the data characteristics of the comprehensive database is given on the EFSA website (EFSA, 2011a). Data of 18 EU countries were provided to the ETUI project of which 11 countries reported food consumption data on two or more days. Six countries reported consumption data on one day.

Country	Name survey	Survey period	Number of subjects	Age range (yrs)	Method (replicates)
Austria	ASNS	2005-06	2,123	19-65	24hr recall (1)
Bulgaria	NSFIN	2004	1,204	> 16	24-hour recall (1)
Czech Republic	SISP04	2003 – 04	1,751	> 4	24-hour recall (2)
Estonia	NDS 1997	1997	1,866	19-64	24-hour recall (1)
Finland	FINDIET 2007	2007	2,038	25-74	48-hour recall (1)
France	INCA2	2005 – 07	4,079	3-79	Food record (7)
Hungary	National Repr Survey	2003	1,360	> 18	Food record (3)
Ireland	NSIFCS	1997 – 99	958	18-64	Food record (7)
Italy	INRAN-SCAI 2005–06	2005 – 06	3,323	> 0.1	Food record (3)
Latvia	EFSA_TEST	2008	2,070	7-66	24-hour recall (2)
Netherlands	DNFCS-2003	2003	750	19-30	24-hour recall (2)
Poland	IZZ-FAO-2000	2000	4,134	1-96	24-hour recall (1)
Slovakia	SK MON 2008	2008	2,761	19-59	24-hour recall (1)
Slovenia	CRP-2008	2007 – 08	410	18-65	24-hour recall (1)
Spain	AESAN-FIAB	1999 – 2001	1,068	17-60	Food record (3)
	AESAN	2009	418	18-60	24-hour recall (2)
Sweden	RIKSMATEN 1997-98	1997 – 98	1,210	18-74	Food record (7)
United Kingdom	NDNS	2000 – 01	1,724	19-64	Food record (7)

preliminary version of the hierarchical food classification system 'FoodEx1', developed by EFSA, was used to codify all foods and beverages present in the Comprehensive Database. FoodEx1 is a hierarchical system based on 20 main food categories that are further divided into subgroups up to a maximum of 4 levels. It was demonstrated that all data providers were able to classify the large majority of their food correctly at least at the 2nd level of the FoodEx1.

Concentration data

Concentration data on lead was provided by EFSA on a confidential basis, for which reason we are not allowed to report the concentration data in full detail. Lead concentration data were submitted to EFSA by 14 Member States and Norway. Approximately 94,126 results of lead concentrations in food commodities and tap water monitored between 2003 and 2009 were used in the case study. Analysed

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

commodities were also classified according to FoodEx 1. For a description of the concentration data we refer to the Scientific Opinion on Lead in Food (EFSA 2010a).

The majority of analysed commodities could be linked directly to food consumption data. An exception was the concentration data of coffee, tea and cocoa. These were measured as purchased (undiluted / unprepared), whereas the consumption data refer to the amounts as consumed (diluted / prepared). To make a correct link between those data, dilution factors 18, 60 and 10 were used for coffee, tea and cocoa, respectively. Unrealistic high concentrations analysed in uncertain food items like wakame, milkshake and special nutrition were not used in the exposure assessment.

Concentration data for the selected pesticide were sent by EFSA on a confidential basis and concentrations are therefore not reported here. Pesticide concentration data were obtained from national and European monitoring programs reported to EFSA by Member States on an annual basis. Field Trial data were obtained from the FAO report Pesticide residues in food (FAO, 2003).

The EFSA concentration database for lead and for the pesticide contains many non-detects. In the exposure assessments reported here all concentrations of non-detects were set to zero. The assumption for the non-detects were the same for the various exposure assessment using different usual intake models in order to make a good comparison between the models.

Link between food consumption data and concentration data

Lead was mainly analysed in foods suspected to contain lead. There were many foods in which the contaminant was not measured. These data gaps were partly filled with averaged values from comparable food items and partly with an average of all concentration data in the same FoodEx1 (sub)group. In this way the software dealt with the hierarchal structure of the FoodEx1 codes (see for more details the paragraph on linking consumption and concentration data).

The lead concentration data as well as the consumption data were both coded according to FoodEx1. The computational tool matches the concentration and food items at the lowest hierarchical level of the FoodEx1 (level 4) provided that both concentration and consumption data were available at this level. If no match was available at this level, the tool moves up to a higher hierarchical level (level 3) to establish a link between concentration and consumption. It stops when a match is found.

The concentration data of the selected pesticide were analysed in raw agricultural commodities (RACs) using a coding system prescribed in pesticide legislation. To model the dietary exposure to pesticides using concentrations analysed at RAC level a link between concentration and food consumption data needs to be established. The Food conversion database developed in the project "Conversion of foods coded according to FoodEx1 into raw agricultural commodities (RACs)" was used for this purpose (Boon et al, 2011). In this database foods classified according to FoodEx1 were converted to their RAC-ingredients for fruit and vegetables only, including weight adjustments.

Reference values

For lead, EFSA could not derive a TDI as a toxicological reference value. However, EFSA reported a bench mark dose lower confidence limit BMDL value of 0.63 μg/kg body weight for nephrotoxic effects and 1.50 μg/kg body weight for an increase in blood pressure based on an evaluation of epidemiological data using the benchmark dose approach (EFSA 2010a). The BMDLs are related to human data.
The results of the exposure assessments using different usual intake models were compared to the respective BMDLs reference values. The ADI for the pesticide addressed in the case study equalled 30 μg/kg bw/day.

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Exposure calculations

The dietary exposure was calculated using the Monte Carlo Risk Assessment programme (MCRA) release 7.1 which is the computational tool developed in the ETUI project (see chapter 2). EFSA and RIVM agreed to focus on the OIM, BBN, LNN0 and LNN models. Model-based as well as model assisted approaches were run in parallel for the BBN and LNN0 models. For a description of the models we refer to chapter 3, and to Goedhart et al. (2012).

5.3. Results case study lead

Data characteristics and diagnostic plots

When specific examples are given we have used either the Dutch food consumption data as owned by the Dutch government and available for use by RIVM or we have made the data not traceable to a specific country.
As consumers are exposed to lead on almost all days (see figure 5.3.1), there is no frequency modelling, which means that BBN, LNN0 and LNN will give essentially the same results. According to the statistical guidance LNN0 would be a preferable starting point.

```
--------------------------------------------------------------------------------
Summary of intake frequencies (before modelling)
Substance: Lead (Pb)
--------------------------------------------------------------------------------

Number and percentage of population with positive intakes

  nPosDays    %PosDays    nindividuals    %individuals
         0         0.0               0             0.0
         1        50.0               0             0.0
         2       100.0             750           100.0


                                   All    Positive    %Positive
                               intakes     intakes      intakes
  Number of observations          1500        1500        100.0
     Number of individuals         750         750        100.0
```

Figure 5.3.1:. diagnostic table containing information about the frequencies of positive intake on survey days (Dutch data).

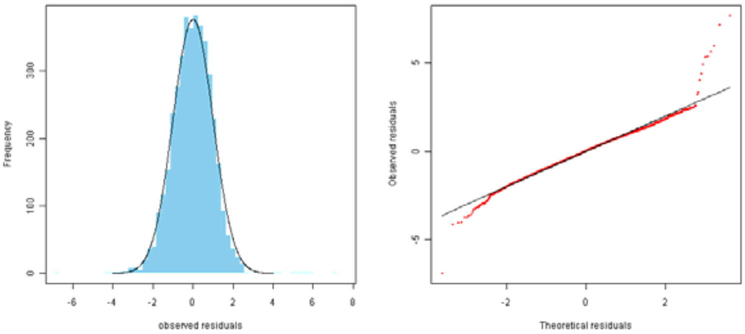
The transformed intake distribution of lead was approximately normal for 10 countries, and mildly skewed for one country. This was observed from diagnostic plots as shown in figure 5.3.2 a and b. The results reported in this paragraph refer to 11 countries and are based on the LNN0 model based. For the country with a mildly skewed transformed intake distribution, it may be better not to rely too much on the normality assumption, and the model based approach might therefore be better replaced by a model-assisted approach.

Results of countries with 2 or more reporting days of food consumption

When using the OIM model for the 11 countries, with two or more days of food consumption recording, a range of 4-34% exceedances of the BMDL for nephrotoxic effects was reported. A range of 0-3% exceedances of the BMDL for blood pressure was observed. The mean percentage of the population which exceeded the BMDL for nephrotoxic effects was 13.4% according to the OIM model and 0.7% for blood pressure. When comparing LNN0 and OIM at the 90th, 97.5th and 99th percentiles,

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

a) Example of an approximately normal log-intake distribution, with some deviation from normality at extreme high percentiles.



b) Example of a mildly skewed distribution

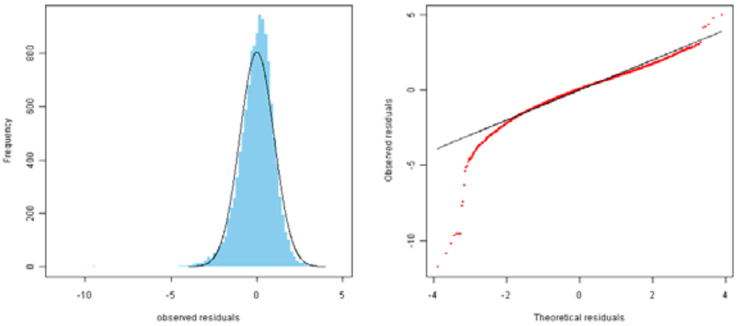


Figure 5.3.2: diagnostic plot for a typical normal distribution and a mildly skewed distribution after lognormal transformation of daily intakes.

An example of specific percentiles using Dutch data is presented in figure 5.3.3. In this case the model-assisted distribution resembled reasonably well the normal distribution generated in the model-based approach, which was found for nearly all countries.

The results of the 11 countries showed that the percentage of the population exceeding the BMDL for nephrotoxic effects , decreased on average from 13.4% using the OIM model to 12.4% using the LNN0 model based approach. The percentage of the population exceeding the BMDL for blood pressure decreased on average from 0.7% (OIM) to 0.1% (LNN0) as an average results for the 11 countries. A large variation between countries was observed . The use of BBN or LNN models provided similar results compared to the LNN0 model.

Based on average exposure levels no difference was found between the results generated with the OIM and LNN0 models. However at the upper part of the distribution the results, either expressed as absolute exposure levels on a body weight basis or as percentage exceeding BMDL levels, generated with the LNN0 model were significantly lower compared to those generated with the OIM model. Table 5.3.1 provides an overview of the impact of using LNN0 on the percentage exceeding the BMDL. The results are split in three percentile classes. In the 51-80th percentile class, the LNN0 model resulted in a 2% increase in the percentage of intake exceeding the BMDL compared to the OIM model. In the class of 81-90th the percentage decreased by 12% and in the highest class (91-100th) even with 50%.

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

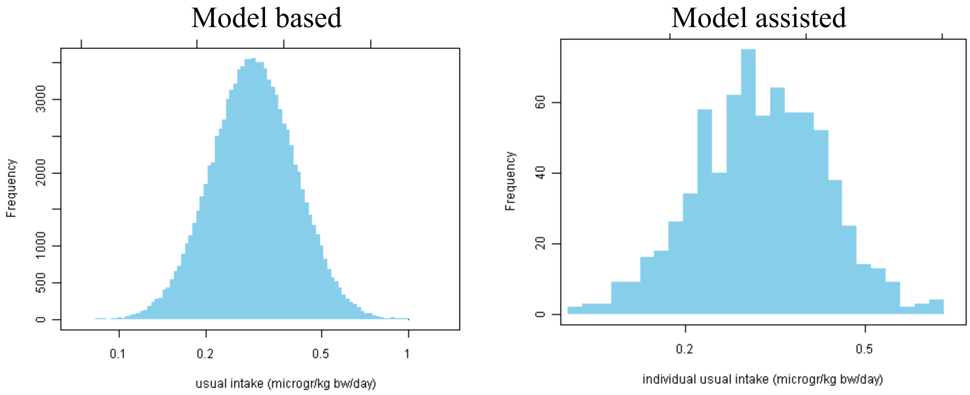


Figure 5.3.3: examples of exposure distributions using Dutch data.

Table 5.3.1: percentage exceedance of BMDL, for either nephrotoxic effects or blood pressure, using OIM or LNN0 at different parts of the exposure distribution.

Percentile classes	Number of scenarios exceedance BMDL	Percentage increase (+) or decrease (-) using either OIM or LNN0
51-80%	2	+2
81-90%	3	-12
91-100%	9	-50

Results of models for countries with one day of food recording

For the six databases with food consumption data for one day per person, the between person variation was calculated for each country and the within person variation was calculated using the variance ratio (between/within) as derived from the countries that reported two or more days of food consumption. This variance ratio was on average 0.74 and entered into the computational tool as external data using a screen as shown in figure 5.3.4.

In the higher percentiles a decrease in exposure was observed when using the LNN0 models compared to the OIM model. The 90th, 97.5th and 99th percentiles calculated with the LNN0 were on average 2, 52, 59% lower compared to the results generated with the OIM model, respectively for the six countries having only one day of food reporting.

The percentage of the population exceeding the BMDL for nephrotoxic effects using the OIM model was on average 19% for the six countries . When using the LNN0 model based approach the percentage exceeding this BMDL was on average 22.4% and 19% when using the LNN0 model approach. The example using these consumption databases for six countries showed that the computational tool can handle food consumption databases with only one day of food reporting providing that the variance component of the within-person variation is borrowed from other calculations using food consumption databases with two or more days of food reporting.

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

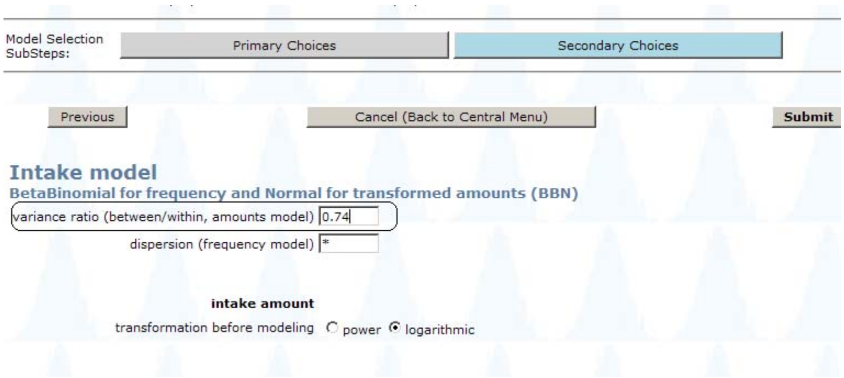


Figure 5.3.4: option to enter variance of the between/within person variation into the computational tool.

5.4. Results case study pesticide

Data characteristics and diagnostic plots

In the higher percentiles the exposure estimates calculated with the usual intake model LNN0 were higher than the OIM results. The 90th, 97.5th and 99th percentiles were on average 33, 81 and 110% higher. In none of the countries the ADI of 30 µg/ kg bw/ day was exceeded using the OIM model.
This result was due to a bimodal distribution which is explained by a combination of exposure levels from the focal commodity using field trial data and exposure levels from background commodities using monitoring data. An example of a diagnostic plot showing a bimodal distribution is presented in Figure 5.4.1. In those cases the transformation to normality is not satisfactory, and all usual intake models tested might result in unreliable results, such as higher percentiles compared to those obtained by the OIM method.

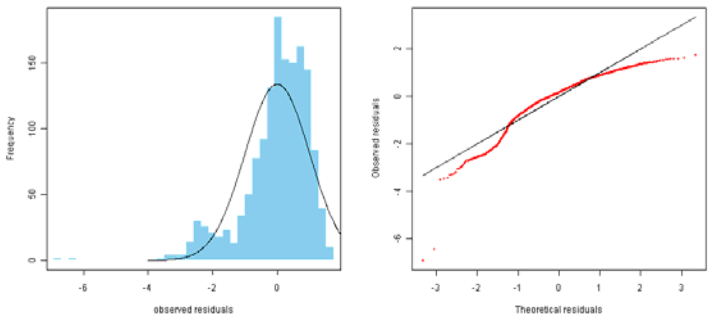


Figure 5.4.1: example of diagnostic plot combining exposure levels from a focal commodity with exposure levels from background commodities resulting in a not normal or multimodal distribution.

To solve this, we performed the calculations separately for the focal commodity and the background commodities. This resulted in two new exposure distributions, which were subsequently summed (‘model-than-add’). An example of such an approach is given for two different focal commodities in combination with monitoring concentrations in table 5.4.1.

In the pesticide case study the concentration database contained both monitoring and field trial data in which case we expected to diagnose a bimodal distribution. The usual intake models sum the observed intakes from all different foods per subject per day and subsequently adjust for within-person variation: the ‘add-than-model’ approach. The ‘model-than-add’ approach the focal commodity and

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Table 5.4.1: exposure assessment results based on exposure levels from focal commodities and from background commodities using different usual intake models.

Method	**P50**	**P90**	**P97.5**	**P99**	**% population exceeding ADI**
OIM	0.1	0.3	0.5	0.7	0
LNN Model based	0.3	1.1	2.1	2.9	0
LNN Model assisted	0.4	1.0	1.4	1.6	0
LNN 'model-than-add'	0.2	0.3	0.5	0.5	0

the background commodities are first modelled separately and then the results are added. The results generated with the 'model-than-add' approach are generated with the LNN0 model-assisted approach.

At the 50th percentile of the intake distribution the result of the 'model-than-add' approach was twice as high as the OIM results. The intakes at the 90th and 97.5th percentile were equal for both methods At the 99th percentile the 'model-than-add' approach results were 29% lower than results generated with the OIM model.

5.5. Discussion case studies

The results generated with the data and the models used in the case studies should not be seen as a new risk assessment for lead or pesticides, neither should it be regarded as a replacement of the conclusions in the Scientific Opinion on Lead in food (EFSA 2010a). In the case studies we aimed at comparing the usual intake models with the exposure assessment method currently in use by EFSA, which method is comparable with the Observed Individual Mean (OIM) model as implemented in MCRA. The OIM model calculates a distribution of mean intake over the survey days of each person as a proxy for usual intake. Since the mean intake still contains a considerable day-to-day variation, the model tends to overestimate the exposure at the high percentiles of the intake distribution (Goedhart et al. 2012).

In certain areas we deviated in these case studies from common practices in exposure assessment as ignoring the uncertainties, which should be included in a complete risk assessment. Uncertainties are part of the computational tool, but for reasons of comparison between the OIM model and the usual intake models we focussed on the comparison only. The experience gained with the case studies might be useful in future risk assessments to be performed by EFSA.

Current usual intake models are based on the assumption that the exposure is normally distributed after transformation at the log scale. If this condition is not met, the use of usual intake models might be debatable or not fit for purpose. As proposed by Boer et al. (2009), normality can be checked by visual inspection of q-q plots. In the lead case study both normally distributed and slightly right-skewed transformed intake distributions were observed using such plots. Four examples of this are given in Figure 5.5.1. Panel A of this figure is a typical normal distribution, because the dotted line of the q-q plot follows a straight line. Panel B shows a mildly skewed transformed distribution. When looking at the q-q plot at the range of between 1.3 and 2.3 (corresponding with the P90 – P99) the values are still on the straight line. Figure C is an example of a typical left-skewed distribution and figure D represents a bimodal distribution or a not normal distribution.

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

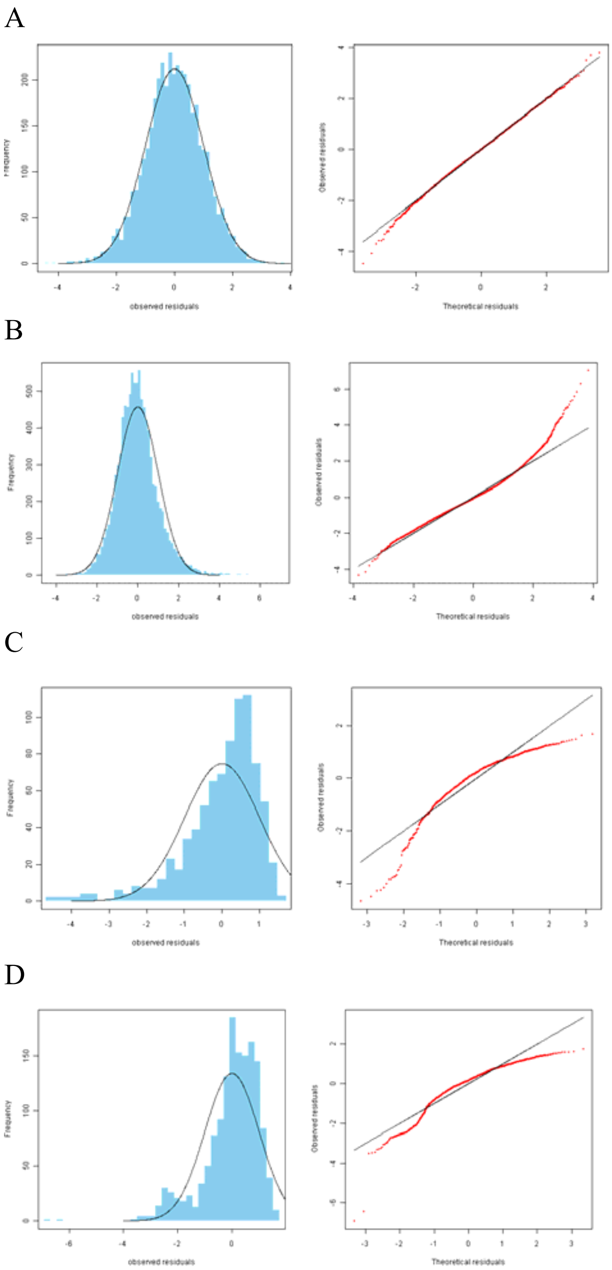


Figure 5.5.1: examples of different distribution types using diagnostic plots as part of the computational tool. A. normally distributed, B. (mildly) skewed distributed, skewed to the right, C. mildly skewed distributed, skewed to the left, D. multimodal or not normal distributed.

The drill down can be helpful to check the realism of all input data used in a simulation (see reasoning below), whereas the contribution of foods is a powerful option to identify possible risk drivers. Both functions are identified as requirements when performing probabilistic or usual intake assessments according to the EFSA Draft scientific opinion Guidance on the Use of Probabilistic Methodology for Modelling Dietary Exposure to Pesticide Residues (EFSA 2012).

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

The model-assisted approach is part of the computational tool and runs in parallel with the model-based approach. A major advantage of the model-assisted approach is that it includes a drill-down function of which an example is given in figure 5.5.2. The drill-down tabulates 10 persons around a certain percentile chosen by the risk assessor. The person respondent number, the person body weight and the level of exposure of that person is given calculated with the OIM and calculated with the selected usual intake model. Furthermore for each person the amounts of food eaten in combination with the concentration used as input for the usual intake calculation is tabulated. With the drill-down function the risk assessor can spot potential outliers. In this example we did not find extreme outliers although the beer consumption of respondent number 1806414023 is quite high with 2.6 litre. If unrealistic outliers were found the risk assessor can decide to exclude them from the exposure assessment.

Individual	180614414	180614158	180614124	180613695	180614134	180613880	180614227	180614406	180614023
Body Weight	70	57	66	68	83	63	75	65	70
individual usual intake	1.14	1.15	1.16	1.22	1.30	1.35	1.36	1.44	1.68
observed individual mean	1.54	1.64	1.57	2.01	1.92	2.16	2.12	2.48	3.11
Soft drinks	.	1253.80	833.80	.	.	.	.	.	.
Soft drink, fruit content	660.00	.	681.30	210.00	1087.50	761.30	.	.	.
Fruit soft drink, Orange	.	.	435.00	.	.	.	.	.	.
Fruit soft drink, mixed fruit	.	.	.	.	.	.	.	.	.
Carbonated drink, flavoured	.	.	870.00	330.00	330.00	.	.	.	.
Soft drink with bitter princip	.	.	.	.	.	.	.	.	.
Cola beverages, caffeinic	.	.	.	.	.	.	435.00	.	.
Tea (Infusion)	.	298.70	.	245.00	.	.	435.00	.	.
Black tea, infusion	186.70	.	.	386.70	165.00	722.00	.	298.60	560.10
Fruit tee, infusion	.	.	.	.	.	.	.	.	.
Coffee (Beverage)	288.80	.	.	.	.	.	.	678.10	324.00
Coffee drink, cappuccino	.	.	.	.	.	.	.	.	.
Cocoa beverage	.	.	.	.	.	.	.	.	.
Beer and beer-like beverage	.	.	.	.	.	.	.	.	.
Beer, strong	.	.	.	.	.	.	.	.	.
Beer, regular	.	.	.	1715.00	1470.00	.	.	300.00	2660.00

Figure 5.5.2: example of a drill-down using the Dutch Food Consumption to be used by the risk assessor to check realism for potential outliers and how they effect the shape of the distribution after transformation.

'Model-than-add' approach

The use of the LNN0 model in the pesticide case study resulted in a significantly higher exposure compared to the OIM approach due to bimodal transformed intake distributions for all countries (see figure 4.5.1.D). This was caused by the use of relatively high concentration levels for a focal commodity, combined with lower concentrations for 'background' commodities, using an 'add-then-model' approach. In this approach all daily exposures per food are summed and then the model is applied. Conversely, to improve the modelling, the 'model-than-add' approach was applied also in the pesticide case study. In this approach the exposure distribution of the focal commodity was separated from the exposure distribution of all background commodities, and then summed to calculate the total exposure.

Correlation frequency and amount

The two compounds used in this case study had both an intake frequency close to 100%. As a consequence there were no correlations between intake frequency and amount. If correlations are encountered the LNN model might be the preferable option above LNN0 because this model can address correlations. It would be interesting to perform a case study with a compound or food which is episodically consumed and shows a correlation between intake frequency and amount, e.g. a food colour with limited use in a food that is prone to episodically consumption.

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Conclusion case study

We were well able to follow the proposed statistical guidance described in chapter 4 using the data of the EFSA comprehensive database. For the exposure assessment of lead a realistic right tail assessment was required. The presentation of the results in this chapter were generalised to average conclusions because it was agreed in the confidentiality agreement that individual country data can not be presented in full detail in the report of the ETUI project.

However the general picture clearly shows that the upper ten percent of the exposure distribution for lead was significantly lower when a usual intake model was used. The percentage of the population that did exceed the benchmark dose lower confidence limit (BMDL) decreased for about 50% in this case study. It is emphasised that this should be taken as an example relevant for comparing usual intake models with the current approach. It should not be perceived as a full risk assessment for lead.

The diagnostic plots presented in this chapter are designed to support the risk assessor to decide whether the data are appropriate to allow the use of usual intake model. Data characteristics as plotted in the diagnostic tools also might influence the selection of the most optimal usual intake model.

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

6. FINAL WORKSHOP

The aim of the final workshop held on February 2012 8th in Utrecht was to report back about the work done in the ETUI project with respect to the statistical issues and the computational tool. Both the usual intake models implemented in the computational tool and the and work done with respect to integration of FFQ information were discussed. In addition the results of the simulation study and the case studies were presented and discussed. For the list of experts that attended the second workshop see the acknowledgment on page 1).

After a short introduction by **Jacob van Klaveren** (RIVM) the computational tool and guidelines for estimating usual intake distributions were presented by **Hilko van der Voet** (WUR). The first deliverable is the computational tool, which was implemented as part of the web-based system Monte Carlo Risk Assessment (MCRA), enabling an optimal link with the EFSA Comprehensive Database. The statistical models available are the Observed Individual Means (OIM) approach, the BetaBinomial-Normal (BBN) model, two variants of the LogisticNormal-Normal (LNN) model, and, for historical reasons, an implementation of the Iowa State University Foods (ISUF) model. A description was given of MCRA's extensive options for data handling, detailed modelling, uncertainty analysis, diagnostics and output. The second deliverable described consists of guidelines for choosing between the statistical models in practical cases. The type of assessment should be clearly defined: interest can be in the right or left tail of the distribution, and realism or conservatism may be more important. The available data should be investigated and described. Then a tentative model can be chosen based on general findings for the diverse statistical models, the properties of the data at hand, and issues that are specific for the assessment type. It was emphasized that the guidelines are an evolving set of best practices, due to ongoing research on new methods for the assessment of usual intake distributions and new comparison studies.

Paul Goedhart (WUR) presented the outcomes of the ETUI simulation study (see Goedhart et al. 2012). The purpose of the ETUI simulation study was to compare different methods to estimate the usual intake distribution for episodically consumed foods. An important criterion is the bias and precision with which the upper tail of the usual intake distribution is estimated. It is further investigated whether it is beneficial to include Food Frequency Questionnaire (FFQ) information in the models as a covariate. The data are simulated with the logistic-normal model for the frequencies and a two-way random effects model for the log-transformed amounts. Four different frequency models are employed with low, moderate and high consumption frequencies, as well as a so-called bathtub model. This is combined with two ratios (1 and 4) for the variance components in the amount model as well as three values (0, -0.5 and 0.5) for the correlation between frequency and amount. This gives a total number of 24 different scenarios. The answer to a hypothetical FFQ with seven response categories is derived by discretizing the simulated consumption probabilities for each individual and then applying the logistic transformation to the discretized probabilities. Three datasets were simulated for each scenario with 6250 individuals and two recall days. The following methods were used: OIM, ISUF, MSM, SPADE, BBN, LNN0 and LNN. The LNN model is also known as the NCI model, and the LNN0 variant is identical to the NCI model with zero correlation. In an additional simulation, with 50% never-consumers and four scenarios, it was tested whether it is advantageous to include information on which individuals are never-consumers and which are episodic consumers which happen to have zero consumption on the recall days. Also, for three practically relevant scenarios with two foods, the 'add-than-model' approach is compared with the 'model-than-add' approach. The main conclusions are as follows.

For the one food case ISUF produces very biased results even for the zero-correlation scenarios with high frequency of consumption. Also for the zero-correlation scenarios, MSM gives variable and sometimes biased results especially after inclusion of FFQ as a covariate. The remaining parametric

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

method is quite similar for the zero-correlation scenarios. However, as expected, LNN performs much better for the correlation scenarios although there is bias for scenarios with low frequencies and a large within day variance. It is also found that for high frequencies of consumption, a correlation between frequency and consumption is quite unimportant. A general finding is that the largest bias is in the lower tails of the usual intake distribution. OIM is not always that conservative for upper percentiles up to 95%. An alternative conservative estimate with better properties is offered by bootstrap confidence limits of LNN, or LNN0 when there is no correlation between frequency and amount. Including FFQ as a covariate is not beneficial for the zero-correlation scenarios. For the correlation scenarios it is advantageous to enter the FFQ covariate in both the frequency and the amounts model of LNN0 as this partly accounts for the correlation between frequency and amount. However, the LNN model without FFQ still produces better estimates of the percentiles. Information about never-consumers, even 50% of them, is not helpful when estimating upper percentiles. This is because, when omitting this information, the fitted frequency distribution preserves the mean and variances of the underlying distribution, although the two distributions are very different. The three simulation scenarios with two foods seem to suggest that the 'model-than-add' approach, which clearly violates the assumption of bi- or multimodality is not doing too bad. Model assisted percentiles are biased, but no more than 50%. The 'add-than-model' approach is doing better with small bias, if any, for the upper percentiles. For scenarios with a correlation between the two foods, model assisted percentiles perform somewhat better than model based percentiles. In conclusion: a practical approach for single foods would be to fit the LNN model and to revert to the LNN0 model when the estimated correlation is low, or when the frequency of consumption is large. When the right model is used, inclusion of FFQ information is not beneficial when interest is in the upper percentiles only. For multiple foods the 'model-than-add' approach seems to be quite promising. Correlations between foods can then be accommodated by using the model assisted approach.

The role of FFQ information in usual intake estimation was discussed by **Sven Knüppel** and **Heiner Boeing** (DIfE). It was recognized that the FFQ information can be helpful in modelling a population's usual intake distribution. One major advantage of using FFQ information is to distinct consumers from never consumers. The benefit of models with and without detailed FFQ information about habitual frequency is questionable. In the ETUI simulation study it was shown that some models benefit from the inclusion of FFQ, others get worse and some are still unchanged. The simulated FFQ information was assumed to be unbiased. This FFQ information provided only few additional information for discrimination of consumer (and non-consumers) because the simulated FFQ frequencies and the count of simulated consumption days are associated. Including FFQ information to model based models didn't improve the estimation of population's usual intake. To assess additionally the uncertainty of a biased FFQ information we simulated a FFQ with a correlation of r=0.2, 0.5, and 0.8 to the unbiased FFQ. Models with biased FFQ compared to the models with unbiased FFQ and low consumption probability showed more differences than models with a higher consumption probability. The variation of the strength of bias influenced only slightly the change of the results. Further investigations are needed to explore the weak effect of the FFQ and evaluate different correlation scenarios of FFQ and 'true' usual intake. Therefore, we propose to perform another simulation study to investigate the possibility to improve the estimation with FFQ information. The MSM method with FFQ failed. The MSM is based on transforming residuals after applying a regression model because it was assumed that the within and between variance components could be better identified by using the transformed residuals instead of using residuals from a regression model on transformed intake values. The MSM models with FFQ showed that the transformation of the residuals did not achieve normality. On the other hand the log-transformation of the positive amounts worked well because the data were simulated in this way. The validity of the estimation of variance components in the non-normal case is unclear and an appropriate back-transformation is missing. Further investigations are needed to evaluate the effect of non-normality while transforming the observed intake or the residuals.

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Some results from the case studies were presented by **Jacob van Klaveren** (RIVM). The intake of the contaminant Lead is an example where most intakes are daily. The intake of pesticides on the other hand represents the situation where the compound is more often absent than present from the daily diet, and where the two-part models as investigated in the EUI project are of great interest.

In the general discussion **Davide Arcella** (ETUI project leader at EFSA) concluded that very interesting work has been done in this project, and that EFSA would now take time to digest the report. A basis has been laid to start working, e.g. in collaboration with the Contaminants Panel. It would be good to take some time and allow space for further improvements to make results more focussed. Regarding the result that FFQ data did not seem to improve much the estimation of the right tail of the usual intake distribution, he noticed that FFQ data might have more value for episodically consumed nutrients in cases where the main interest is in the left tail of the distribution. **Pietro Ferrari** (formerly ETUI project leader at EFSA, now IARC) considered the work done a great achievement and a clear case of "mission accomplished".

In a discussion on the usefulness of the modelling of correlation between intake frequencies and intake amounts **Kevin Dodd** and **Victor Kipnis** (NCI) reported their experience with US data: correlations, if present, were found to be mostly positive, but mostly of moderate size, up to 0.3 or 0.4. Consequently, the correlation of 0.5 used in the simulation study already represents a sort of practical maximum.

In a discussion of model-based vs. model-assisted estimates of the usual intake distribution **Kevin Dodd** (NCI) suggested that a QQ-plot comparing the two distributions might be a useful diagnostic.

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

7. CONCLUSIONS AND RECOMMENDATIONS

The ETUI project produced four deliverables:

1. two workshops to find consensus among the statistical experts;
2. three reports: the interim report discussing statistical modelling of usual intake (van der Voet & van Klaveren 2010), and two final reports: next to the current general report, a separate report was written on the results of a simulation study on estimating usual intake distributions for episodically consumed foods (Goedhart et al. 2012);
3. a computational tool for usual intake assessment (implemented in the MCRA system available at mcra.rivm.nl) which can be used in combination with EFSA's data collection, including several models to estimate usual intake distributions from repeated survey data, and a procedure to generate usual intake results from food consumption database with only one day of food recording based on extrapolating information from other sources;
4. statistical guidance for choosing a model for usual intake distribution estimation.

The first workshop, held on 27-28 April 2010, identified the usual intake models used in practice at that time, and the statistical issues associated with the use of these models. EFSA and EFSA Panel representatives discussed the risk assessment questions that were most relevant to EFSA work. Improving the estimation of the percentage of the population exceeding risk limits was considered to be EFSA's current priority in this ETUI project. Future EFSA work might have other focuses e.g. more related to nutritional deficiency or epidemiological oriented issues. Research questions related to these issue might require further research on how to make optimal use of food frequency questionnaires for infrequently eaten food for estimating the percentage of inadequate nutrition using usual intake modelling.

Promising usual intake models have been tested by comparing simulated true usual intake with the estimated intake calculated with selected usual intake models. Depending on the simulation scenarios the performance between methods differed, but under the assumptions of the simulation model in general the logistic-normal normal models LNN0 and LNN were considered as models of first choice. For daily intakes, without correlation between intake amounts and intake frequency, the BBN model is also a possible choice.

The experience of the simulations study as well as the issues identified during the first ETUI workshop were summarised in statistical guidance aiming to select the usual intake model most fit for purpose in practical cases. We tested the statistical guidance and the computational tool in two case studies. Data was transferred from the EFSA database to the MCRA computational tool. The impact of using a usual intake model compared to the current methodology in use by EFSA was found to be significant. The case studies also demonstrated the complexity of when and how to use usual intake models. In general diagnostic plots were found to be essential to make the risk assessor aware whether or not the usual intake models can be applied in line with the underlying statistical assumptions.

The computational tool for usual intake modelling, the simulation study, the statistical guidance and the case studies were presented in the second ETUI workshop held on February 8th 2012. The results were well-received by the statistical experts. In general consensus on the major statistical issues as far as handled and as far as relevant to the ETUI project was reached among the experts, although details or issues related to other research areas are still open for future improvements.

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

In conclusion, the user-friendly computational web-based tool developed in this project can be used by everyone having access to the input data needed for the exposure assessment. The statistical guidance can be helpful to select an optimal usual intake model for the case at hand. We demonstrated the impact of a usual intake models with a case studies. Applying the usual intake models the percentage of the population exceeding risk limits decreased significantly compared to the currently used calculation method.

EFSA and the EFSA panels may want to practice with the different usual intake models in the computational tool, and evaluate the statistical guidance in real risk assessment questions. There are many on-going developments in the field of statistical modelling of usual intake. It is therefore recommended to consider an update of the computational tool and guidelines after a certain period of use.

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

REFERENCES

Ambler G, Benner A (2010). MFP: Multivariable Fractional Polynomials. http:/cran.r-project.org/web/packages/mfp/

Bates D, Maechler M, Bolker B (2011). LME4: Linear mixed-effects models using S4 classes. http://cran.r-project.org/web/packages/lme4/

Boon PE, Petersen A, Moussavian S, Debegnach F, Ruprich J, Van Klaveren JD (2009). Harmonisation of food consumption data format for dietary exposure assessments of chemicals analysed in raw agricultural commodities. Food and Chemical Toxicology, 47: 2883–2889.

Boon, PE, Wapperom, D, de Maeyer, M, Drijvers, J, Janssen-van der Vliet, M, van Donkersgoed, G. Conversion of foods coded according to FoodEx into raw agricultural commodities (RACs). Scientific report submitted to EFSA.

de Boer, WJ, van der Voet H (2011). MCRA, Release 7, a web-based program for Monte Carlo Risk Assessment. Reference Manual docmenting MCRA 7.1. Report. Biometris and RIVM, available at http://mcra.rivm.nl.

de Boer WJ, van der Voet H, Bokkers BGH, Bakker MI, Boon PE (2009). Comparison of two models for the estimation of usual intake addressing zero consumption and non-normality. Food Additives and Contaminants Part A. 26:1433-1449.

Dekkers ALM, Ocké MC, Verkaik-Kloosterman J (2009). An efficient method for back-transformation in modelling usual intakes, 7th International conference on Diet and Activity Methods (ICDAM), Washington, p. 145.

Dodd KW (1996). A technical guide to C-SIDE. Technical Report 96-TR 32, Department of Statistics and Center for Agricultural and Rural Development, Iowa State University, Ames, Iowa. Available at http://www.card.iastate.edu/publications/DBS/PDFFiles/96tr32.pdf

Dodd KW (2011). SAS macros for estimating the usual intake distribution by means of the NCI method. 2011. Personal communication. See also for older versions at http://riskfactor.cancer.gov/diet/usualintakes/macros.html

Dodd KW, Guenther PM, Freedman LS, Subar AF, Kipnis V, Midthune D, Tooze JA, Krebs-Smith SM (2006). Statistical methods for estimating usual intake of nutrients and foods: a review of the theory. J Am Diet Ass. 106:1640-1650.

EFSA (2006). Guidance of the Scientific Committee on a Request from EFSA Related to Uncertainties in Dietary Exposure Assessments. The EFSA Journal 438: 1-54.

EFSA (2010a) EFSA Panel on Contaminants in the food Chain (CONTAM) (2010). Scientific Opinion on Lead in Food. EFSA Journal 2010; 8(4):1570 [147 pp.]. doi:10.2903/j.efsa.2010.1570. Available online: www.efsa.europa.eu.

EFSA)2010b) Panel on Plant Protection Products and their Residues (PPR) (2010). Outcome of the first public consultation on draft guidance on the use of probabilistic methodology for modelling dietary exposure to pesticide residues. EFSA Journal 2010;8(11):1907. [8 pp.]

doi:10.2903/j.efsa.2010.1907. Available online: www.efsa.europa.eu/efsajournal.htm

EFSA (2012) Draft scientific opinion Guidance on the Use of Probabilistic Methodology for Modelling Dietary Exposure to Pesticide Residues. Available as a draft for public consultation online www.efsa.europa.eu

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

EFSA (2011a). Use of the EFSA Comprehensive European Food Consumption Database in Exposure Assessment. EFSA Journal 9(3):2097. [34 pp.] doi:10.2903/j.efsa.2011.2097. Available online: www.efsa.europa.eu.

EFSA (2011b). Evaluation of the FoodEx, the food classification system applied to the development of the EFSA Comprehensive European Food Consumption Database. EFSA Journal 9(3):1970. [27pp.] doi:10.2903/j.efsa.2011.1970. Available online: www.efsa.europa.eu.

EFSA (in prep.). Opinion of the Scientific Panel on Plant protection products and their residues on Guidance on the use of probabilistic methodology for modelling dietary exposure to pesticide residues. Part 1: Single active substances exposure assessment.

FAO (2003). Pesticide residues in food. Report of the Joint Meeting of the FAO Panel of Experts on Pesticide Residues in Food and the Environment and the WHO Core Assessment Group on Pesticide Residues. FAO Plant production and protection paper 176. Available online: www.fao.org

Goedhart PW, van der Voet H, Knüppel S, Dekkers ALM, Dodd, KW, Boeing H, van Klaveren JD (2012). A comparison by simulation of different methods to estimate the usual intake distribution for episodically consumed foods. report presented to EFSA.

Goedhart PW (2006). GenStat procedure Rbetabinomial. In: Goedhart PW and Thissen JTNM (editors), Biometris GenStat Procedure Library Manual 9th Edition. Biometris report 10.12.06. Biometris, Wageningen, The Netherlands. Available at http://www.biometris.wur.nl/UK/Software/Genstat+procedures/

Goedhart PW, van der Voet H & Slob W (2011). Validatie van blootstellingsberekeningen in geval van inname van een stof via diverse voedingsproducten. Briefrapport 12522. RIVM. Bilthoven.

Goedhart PW and van der Voet H (2012). Model assisted usual intake for episodically consumed foods. In preparation.

Harttig U, Haubrock J, Knüppel S, Boeing H (2011). The MSM program: Web-based statistics package for estimating usual dietary intake using the Multiple Source Method. Eur J Clin Nutr. 65:87-91.

Haubrock J, Nöthlings U, Volatier JL, Dekkers A, Ocke M, Harttig U, Illner AK, Knuppel S, Andersen LF, Boeing H (2011). Estimating Usual Food Intake Distributions by Using the Multiple Source Method in the EPIC-Potsdam Calibration Study. Journal of Nutrition. 141: 914-920.

Keogh RH and White IA (2011). Allowing for never and episodic consumers when correcting for error in food record measurements of dietary intake. Biostatistics, 12: 624-636.

Kennedy MC (2010). Bayesian modelling of long-term dietary intakes from multiple sources. Food and Chemical Toxicology, 48:250-263.

Kipnis V, Midthune D, Buckman DW, Dodd KW, Guenther PM, Krebs-Smith SM, Subar AF, Tooze JA, Carroll RJ, Freedman LS (2009). Modeling Data with Excess Zeros and Measurement Error: Application to Evaluating Relationships between Episodically Consumed Foods and Health Outcomes. Biometrics. 65:1003-1010.

Nusser SM, Carriquiry AL, Dodd KW, Fuller WA (1996). A semiparametric transformation approach to estimating usual daily intake distributions. J Am Stat Ass. 91: 1440-1449.

Nusser SM, Fuller WA, 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, Trewin D (editors). Survey Measurement and Process Quality. New York (NY), Wiley. p. 689-709.

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Pinheiro J and Bates D (1999). LME: linear mixed effects models. http://stat.ethz.ch/R-manual/R-patched/library/nlme/html/lme.html

R Development Core Team (2011). R: A Language and Environment for Statistical Computing. http://cran.r-project.org/

Rigby RA, Stasinopoulos DM (2005). Generalized additive models for location, scale and shape (with discussion). Applied Statistics 54, 507-554. Also see http://gamlss.org/.

Scheipl F (2011). AMER: Additive mixed models with lme4. http://cran.r-project.org/web/packages/amer/amer.pdf

Slob W (2006). Probabilistic dietary exposure assessment taking into account variability in both amount and frequency of consumption. Food Chem Toxicol. 44: 933-951. Corrigendum on in 44:1952.

Slob W, de Boer WJ, van der Voet H (2010). Can current dietary exposure models handle aggregated intake from different foods? A simulation study for the case of two foods. Food and Chemical Toxicology, 48:178-186.

Souverein, OW, de Boer WJ, Geelen A, van der Voet H, de Vries J, Feinberg M. van 't Veer P (2011). Uncertainty in intake due to portion size estimation in 24-hour recalls varies between food groups. Journal of Nutrition, 141: 1396-1401.

Tooze JA, Midthune D, Dodd KW, Freedman LS, Krebs-Smith SM, Subar AF, Guenther PM, Carroll RJ, Kipnis V (2006). A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution. J Am Diet Ass. 106: 1575-1587.

Tooze JA, Kipnis V, Buckmann DW, Carroll RJ, Freedman LS, Guenther PM, Krebs-Smith SM, Subar AF, Dodd KW (2010). A mixed-effects model approach for estimating the distribution of usual intake of nutrients: the NCI method. Stat Med. 29:2857–68.

van der Voet, H. and van Klaveren, J.D. (2010). Statistical modeling of usual intake. Scientific Report submitted to EFSA. http://www.efsa.europa.eu/en/scdocs/scdoc/86e.htm

van Rossum CTM, Fransen HP, Verkaik-Kloosterman J, Buurma-Rethans EJM, Ocké MC (2011). Dutch National Food Consumption Survey 2007-2010, in: RIVM (Ed.), RIVM. National Institute of Public HEalth and the Environment (RIVM), p. 143.

Waijers PMCM, Dekkers ALM, Boer JMA, Boshuizen HC, van Rossum CTM (2006). The Potential of AGE MODE, an Age-Dependent Model, to Estimate Usual Intakes and Prevalences of Inadequate Intakes in a Population. J. of Nutrition 136, 2916-2920.

Williams DA (1982). Extra-binomial variation in logistic linear models. Applied Statistics. 31:144-148.

Zhang S, Krebs-Smith SM, Midthune D, Perez A, Buckman DW, Kipnis V, Freedman LS, Dodd KW & Carroll RJ (2011a). Fitting a Bivariate Measurement Error Model for Episodically Consumed Dietary Components. International Journal of Biostatistics, 7, Article 1.

Zhang S, Midthune D, Guenther PM, Krebs-Smith SM, Kipnis V, Dodd KW, Buckman DW, Tooze JA, Freedman LS, & Carroll RJ (2011b). A new multivariate measurement error model with zero-inflated dietary data, and its application to dietary assessment. Annals of Applied Statistics, 5: 1456-1487.

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

GLOSSARY

ADI	Acceptable Daily Intake
ANR	Average Nutrient Requirement
BBN	BetaBinomial-Normal model. Two part model, applicable for daily and/or episodical intakes. Based on a BetaBinomial distribution for intake frequencies and a Normal distribution for log or Box-Cox transformed intake amounts. Provides model-based and model-assisted distribution estimates.
BMDL	benchmark dose model lower confidence limit
DIfE	German Institute for Nutrition
EAR	Estimated Average Requirement (EAR),
EFSA	European Food Safety Authority
ETUI	European Tool Usual Intake
FFQ	Food Frequency Questionnaire
ISUF	Iowa State University Foods model. Two part model, applicable for daily and/or episodical intakes. Implements the main ideas of a semi parametric transformation approach to estimating usual intake distributions.
MRL	Maximum Residue Limit
MSM	Multiple Source Model
NCI	National Cancer Institute
LNN0	LogisticNormal-Normal-nocorrelation model. Two part model, applicable for daily and/or episodical intakes. Based on a LogisticNormal distribution for intake frequencies and a Normal distribution for log or Box-Cox transformed intake amounts. Provides model-based and model-assisted distribution
LNN	LogisticNormal-Normal model. Model also known as NCI model. Similar to LNN0, but accounting for a correlation between logistically transformed frequencies and log or Box-Cox transformed intake amounts. Currently only provides model-based distribution estimates.
MCRA	Monte Carlo Risk Assessment
OIM	Observed Individual Means. Naive method just averaging the individual intake over survey days. May be used as first tier or as replacement when BBN or LNN approaches fail because they are not fit for the purpose of use, depending on data characteristics.
RAC	raw agricultural commodities

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

RIVM	National Institute for Public Health and the Environment, the Netherlands
TDI	Tolerable Daily Intake
UNL	Upper Nutrient Level
WUR	Wageningen University and Research centre

The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.