MCRA 7

a web-based program for Monte Carlo Risk Assessment

Overview 2011-12-19

documenting MCRA Release 7.1

Waldo J. de Boer
Hilko van der Voet

Contributions to MCRA were made by: Paul W. Goedhart, Gerie W.A.M. van der Heijden, Sanne Korzec, Helen Owen (Food and Environment Research Agency, UK), Jac T.N.M. Thissen

Overview

The MCRA program (Monte Carlo Risk Assessment) is a computational tool for assessment of risks due to the intake of substances found on foods. MCRA builds on databases of food consumption and measured concentrations of substances in food. Tables to specify the conversion between food-as-eaten and food-as-measured can be provided.
MCRA provides a number of options to model the data.

Acute versus chronic risk assessment

Health effects may be acute reactions to a short term exposure or may be of a chronic nature as a result of a long term exposure. Both risk assessments are covered by MCRA. In an acute risk assessment, daily consumptions are simulated by sampling a food consumption database and, combined with a random sample from a compound database, an acute intake distribution is derived. In a chronic risk assessment, daily intakes are calculated based on a mean of the concentration values multiplied by consumption. The usual daily intake distribution is estimated by removing the effects of day to day variation within a person from the dietary intake.

Acute risk assessment

MCRA calculates the intake distribution (in mg, microgram, nanogram, picogram, femtogram substance per kg body weight per day) from input data on consumption and concentration levels of substances in food. Through MC-sampling, food consumption patterns are combined with concentration levels and the derived intake distributions are used to make inference about the percentage of the population at risk. The intake distribution is characterised by percentiles, *i.e.* substance concentration levels exceeded with only a small specified probability (for example the 99^{th} percentile p99 is exceeded only in 1% of the cases).
Although MCRA is able to do the calculations using aggregated intake values per person as input data, the use of primary data as stored in a relational databases offers the possibility to pre-process the data and to investigate the results of the analysis in greater detail.

Chronic risk assessment

Some substances on food do not have instantaneous health effects but are of a chronic nature as a result of long-term exposure to relatively low concentrations. The daily intakes, here the aggregate of food consumption times a mean value of the concentrations, is modelled using four different methods. The first two models, the betabinomialnormal (BBN) and the logisticnormalnormal (LNN) model, estimate the usual daily intake distribution in a two step procedure. The LNN model is also offered in a quicker version that does not estimate the correlation between intake frequency and intake amount (LNN0). In the first step, the intake frequencies are modelled, in the second step the parameters for the usual intake amounts are estimated. Through numerical integration the usual intake distribution is derived. In an extended version of both models, frequencies and amounts are related to e.g. covariables like sex or age. The Iowa State University Foods (ISUF) model follows a discrete/semi-parametric approach to counteract nonnormality, but does not have the possibility to estimate age or sex related percentiles for usual daily intake. The last model, observed individual means (OIM), calculates the average of single-day intakes for each individual and the empirical distribution of these within-individual means is used to estimate the usual intake distribution for the entire population.

Empirical versus parametric

A special feature of concentration data is that the large majority of measured values is recorded as non-detect. These values may correspond to true zero's or to censored values, *i.e.* concentrations below a Limit of Reporting (LOR). In the empirical approach, concentration data are sampled at random from the available data. The empirical approach requires enough data to obtain a satisfying representation of the full distribution.

When positive data are scarce, parametric modelling becomes important. MCRA offers five methods to model non-detects and positive concentration values based on a lognormal distribution with parameters μ and σ . When all positive concentrations are above the LOR or at least the fraction of positive values < LOR is very small, a mixture of a non-detect spike and the lognormal distribution may be satisfying. A mixture of a non-detect spike and a truncated lognormal may be used when the fraction of censored observations is smaller than the fraction of non-detects. When the assumption is that true zero's are absent, which is reasonable in case of contaminants, a censored lognormal or a censored lognormal with estimated LOR may be used. The last model, a mixture of a zero spike and a censored lognormal, fits a distribution where the non-detects are divided over a spike of true zero's and the censored tail of the lognormal distribution.

Agricultural use: percent crop treated

Occasionally, when the use of a substance on a crop or food is allowed and when data about the percentage of the crop or food that is treated is present, this information may be used to infer which part of the non-detects are true zero's and which part are censored ones. Based on the so called percent crop treated, part of the non-detects are replaced by a value *f* x LOR, with *f* a constant between 0 and LOR. Replacement of all non-detect values represents a worst case scenario.

Processing

Concentrations are measured on raw agricultural products, but the foods people eat are for the most part processed. Potatoes are cooked, banana's peeled, grapes are dried and consumers drink apple, orange or tomato juice. More often, processing of foods reduces the concentration of the substance. To account for this effect, MCRA offers the possibility to incorporate 16 types of processing in the model. The most frequently used processing types are: cooking, peeling, juicing, baking, frying and cleaning. The supplied processing factors are multiplied with the concentration values, offering a more realistic way of modeling concentration measurements. MCRA can incorporate processing factors as fixed effects or by sampling from a processing factor distribution. The latter possibility requires the specification of a nominal and an upper value which is used as a measure of variability.

Unit variability

Substances are often measured in large composite samples e.g. 10 apples mixed together, thus hiding part of the variability that exists between individual units. The mean of the measurements is represented well, but the variability is underestimated. MCRA has extensive possibilities to model the original variability between units, a concept known as unit variability. The beta, bernoulli or lognormal distribution are used to sample from, each distribution requiring special knowledge about the data. Depending on the available data, default variability factors may be used or variability factors based on expert opinion or trials are supplied. For the lognormal and beta distribution, a conservative approach applying censoring together with a realistic approach (no censoring) is implemented.

IESTI

MCRA is designed for probabilistic risk assessment. However, occasionally there is a need to compare probabilistic estimates with the deterministic ones. The International Estimate for Short Term Intake approach is implemented and the deterministic estimates together with the probabilistic ones are calculated.

Uncertainty

Parameters describing the risk are uncertain due to the limited size of the underlying datasets. MCRA uses the bootstrap methodology and parametric uncertainty estimates to make inference on the accuracy of the data. The result is a 95% uncertainty interval around the estimate. When all primary data are present, variability in a given population and variability in the available concentration levels may be explored. Occasionally, information about the uncertainty of the portion size consumed is present, or, when foods are processed, the variability of the processing factors is known. Then,

parametric inputs are sampled from parametric distributions. The MCRA output contains diagnostic plots on the amount of MC-variation and the amount of variability due to resampling individuals, concentrations and processing factors.

MS Access and Excel supported data entry

MCRA works with data stored in MS Access relational databases or Excel spreadsheets. The latter requires intakes aggregated at the level of individuals, but has less possibilities to pre-process data. In the database format, data on processing, unit variability, empirical of parametric modelling of concentration distributions, replacement of the LOR (based on percent crop treated) and all possibilities to explore variability originating from individuals, concentrations and uncertainty estimates are available, but data input may be laborious. The database format offers the possibility to decompose foods as consumed into ingredients (or foods as measured) using the composition of a food, *e.g.* convert pizza consumption to consumption of wheat, tomato, cheese, etc. Or, decomposition of foods as consumed into marketshares, *e.g.* for apple market shares are specified for Jonagold, Granny Smith and Golden Delicious. This option allows to incorporate a parameter for brand loyalty in the model, *e.g.* the tendency to repurchase the same brand.

Subset selection

Specific questions like the risk of males versus females, or children versus adults are easily answered using the subset facilities of MCRA (only database format). Based on variables like sex, age or weight, part of the population is selected and used in the exposure assessment. Also, concentration measurements may be limited to a specific year or *e.g.* country. It is also possible to restrict the risk assessment to specific foods or foodgroups or to consumptiondays only.