APPENDIX 4

REPORT ON PROBABILISTIC RISK ASSESSMENT

INTRODUCTION

The essence of probabilistic risk assessment is that it aims at ranges of plausible values, rather than single values or point estimates, as is usual in current risk assessments. A drawback of point estimates is, that they can be best estimates, or worst case values, or anything in between, and such can only be made known in a narrative way. In doing so, the level of conservatism of a worst case value remains vague. Further, a series of worst case assumptions may be multiplied, and thereby result in an unnecessarily conservative and possibly even unrealistic risk assessment.

In probabilistic risk assessment one tries to quantify uncertainties associated with any of the steps involved in the risk assessment process, be it data or assumptions. These uncertainties are then combined using statistical techniques, in order to quantify the uncertainty in the end result of interest. A common technique for combining uncertainties is the The Monte Carlo method. As a simple example, consider the assessment of exposure to a compound in a food item by multiplying the concentration in the food by the intake rate of the food. In a probabilistic approach we do not multiply the two single numbers, but two distributions, reflecting the uncertainty in the concentration and the intake rate, respectively. The Monte Carlo method randomly draws a large number of pairs of values from the two distributions, multiplies the two values in each pair, and pools the resulting list of products together in a histogram; this histogram gives the required uncertainty distribution for the exposure to the compound.

For more complicated assessments, the Monte Carlo approach remains easy to understand and implement. Making a probabilistic assessment meaningful is mostly a matter of a proper conceptual understanding of the various uncertainties involved, and how they relate to each other. It is important to distinguish between uncertainties that reflect imperfect scientific knowledge from uncertainties that reflect variability in a population (sometimes denoted as type I and type II uncertainties, respectively). The imprecision in any point estimate as reflected by its standard error or confidence interval is an example of type I uncertainty. The uncertainty that the human being could be more, or maybe less sensitive than a factor of ten compared to the test animal, is another example. The variation in sensitivity among individuals exemplifies type II uncertainty, but the question how large this variation exactly is constitutes type I uncertainty.
It is not meaningful to combine these two types of uncertainty into a single uncertainty distribution. It should always be clear how a derived uncertainty distribution must be interpreted. For example, a 5^{th} percentile may reflect either a 5% probability of being wrong (type I uncertainty), or a fraction of 5% of the population considered (type II uncertainty). By maintaining the distinction between these two types of uncertainty in the Monte Carlo analysis, one may end up with statements such as: There is a 95% probability (level of confidence) that at most 10% of the population exceeds the acceptable daily intake (ADI). Uncertainty distributions that result from mixing type I and

type II uncertainties are difficult to interpret, and can only be used as a sort of worste case approach to see if there might be a potential problem.

A second crucial aspect obviously is the quantification of the magnitude of the uncertainties involved. In the case of uncertainties of estimates resulting from (experimental) data, common statistical techniques can be used. It is more difficult to quantify uncertainties in assumptions that have to be made in situations of lacking data, such as in extrapolating from no adverse effect levels observed in animals to humans. In those situations one may base an estimate of the uncertainty associated with the extrapolation factor on relevant data for other compounds, for which both human and animal data are available, and consider the statistical characteristics of such data.

The approach of probabilistic risk assessment will be briefly discussed for assessing health-based exposure limits such as reference dose (RfD), ADI, and tolerable daily intake (TDI), as well as for assessing exposure levels in the population of interest.

HEALTH-BASED EXPOSURE LIMITS

In the default method, acceptable intake or exposure limits are obtained by dividing the no observed adverse effect level (NOAEL) by a number of uncertainty factors (UF):

$$ADI, TDI, RfD \;=\; \frac{NOAEL}{UF_1 \;x\; UF_2 \;x\; UF_3 .....} \qquad (1)$$

As opposed to the operational definition of the ADI by equation (1), the probabilistic approach starts from the notion that we are interested in a certain unknown dose level that we consider as sufficiently protective, e.g. the dose level that does not lead to adverse effects in the majority of people. The aim is to estimate this unknown value from any relevant information that we may have, and to assess the precision (uncertainty) of that estimate, depending on the quality of the data available. To that end equation (1) is re-formulated as:

$$NAEL_{\,sens.human} \;=\; \frac{NAEL_{animal}}{EF_{interspec}\, EF_{intraspec}} \qquad (2)$$

where the $NAEL_{sens.\ human}$ denotes the true, but unknown dose level at the borderline of adverse effects for a sensitive human being. Similarly, the $NAEL_{animal}$ is defined as the unknown dose level at the borderline of adverse effects for animals. The extrapolation factors (EFs) in the numerator are simply the ratios of the relevant NAELs, i.e.

$$EF_{\,interspec} = \frac{NAEL_{animal}}{NAEL_{average\,human}} \quad , \quad EF_{\,intraspec} = \frac{NAEL_{average\,human}}{NAEL_{sens.\,human}} \qquad (3)$$

Note that equation (2) can be expanded when necessary to allow for other areas of uncertainty, for example with a factor $EF_{subchronic}$ when only subchronic studies are available.
In the probabilistic approach each of the entities at the right hand side of (2) is given by an uncertainty distribution. The uncertainty in the $NAEL_{sens.\ human}$ can be evaluated using the Monte Carlo approach, as illustrated in Fig. 1.

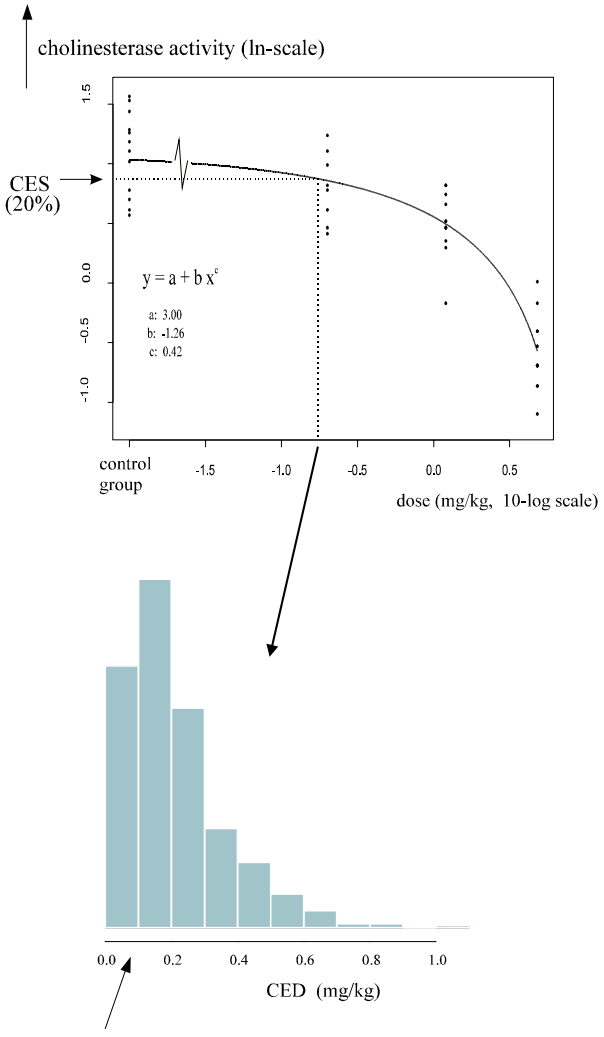

Fig. 1. Upper panel: cholinesterase activity (μmole/ml, ln-scale) in erythrocytes as a function of ^{10}log-dose (dots refer to individual animals), with fitted regression function, and the estimated CED (value: 0.17 mg/kg) at a CES of 20% cholinesterase inhibition. Lower panel: the associated uncertainty distribution (obtained from 500 Monte Carlo runs from the fitted regression model) for the CED. The lower 5th percentile of this distribution (0.04 mg/kg) is comparable to the benchmark dose.

Uncertainty in $NAEL_{animal}$

The usual way of deriving a dose level from the animal study that can be used as the starting point for risk assessment is by (statistically) comparing each dose level with the controls, resulting in the NOAEL. WOUT – I have put this change in to try to keep NAEL separate in reader's mind from NOAEL The NOAEL suffers from various drawbacks, one of them being that it is not possible to assess the uncertainty in the NOAEL as an estimator of the (true) $NAEL_{animal}$. This makes the NOAEL unsuitable for a probabilistic risk assessment. The benchmark approach does allow for deriving a point estimate of the $NAEL_{animal}$, together with an uncertainty distribution. The general idea of the benchmark approach is to estimate the dose associated with a particular, small effect, based on a dose-response function that has been fitted to the data. When this small effect is regarded as nonadverse, the associated dose can be regared as a(n estimate of the) $NAEL_{animal}$. Figure 1 illustrates this for a set of continuous dose-response data. In the case of continuous data, the severity of an effect can best be quantified in terms of a percent change relative to the level of the endpoint observed in the controls. In the example of Figure 1, where the response is acetylcholinesterase activity, it is assumed that a 20% inhibition is adverse, and the Critical Effect Size (CES) is postulated to be 20%. After fitting a dose-response model to the data, the associated dose, the Critical Effect Dose (CED), is derived from the fitted model. The latter is in fact a point estimate. The complete uncertainty distribution for the CED may be derived by various statistical techniques, for example by bootstrapping (Slob and Pieters 1998) or by maximum likelihood based methods. Note that the benchmark dose as originally defined by Crump (1984) is a lower percentile of the uncertainty distribution of the CED (i.e., a lower confidence bound). Thus, the uncertainty distribution of the CED can be seen as simply an extension of Crump's benchmark dose.

Probabilistic Extrapolation Factors

The denominator of expression (2) consists of a number of extrapolation factors that are typically unknown for any particular chemical. What can be done, however, is try to find indirect information, e.g. historical data on other chemicals, that may give an indication of what are plausible values for each of these factors. This information can be summarized in the form of a distribution for each EF. For example, one may imagine that the ratio $NAEL_{average\ human}$ / $NAEL_{animal}$ (i.e. the interspecies EF) varies from chemical to chemical. If for a number of chemicals this ratio could be estimated (using those compounds for which human data are available) the resulting distribution of these ratios represents the variation between chemicals. Examination of ratios of NOAELs related to two animal species (e.g. rat vs. mouse, dog vs. rat) shows that metabolic dose-scaling (dose per $BW^{0.75}$) results in distributions with medians close to unity (see, e.g. Baird et al. 1996, Vermeire et al. 1999). This indicates that dose per $BW^{0.75}$ is a better dose-scale to correct for size differences between species than dose per BW. Given this assumption, the distribution of the interspecies EF should have a species-specific median (given by the allometric BW ratio for the specific test animal and the human being), and a spread that

may be derived from the empirical distributions of NOAEL ratios between test animal species.
Similarly, one may postulate a distribution for the $EF_{intraspecies}$. Here two approaches may be followed. On the one hand, it may be assumed that variation in sensitivity in the human population is more or less homogeneous, i.e. the magnitude of that variation does not depend on the chemical. This variation might be estimated from the variation in observed characteristics in the human population, such as absorption rates and metabolic clearances (e.g. Hattis et al, 1999). On the other hand, one may focus on differences between chemicals, i.e. for one chemical the difference between a sensitive and an average individual is larger than for another chemical (e.g. Slob and Pieters, 1998). In the first approach, the uncertainty is of type II, in the second of type I. The illustration in Fig. 2 uses the second approach, and results in a single overall (type I) uncertainty distribution for the $NAEL_{sens.\ human}$. Therefore, this distribution reflects scientific uncertainty in the $NAEL_{sens\ human,}$ where the sensitive human is not specified. When the $EF_{intraspecies}$ is embodied with a type II distribution, it should be evaluated separately from the type I uncertainties that are associated with the other entities (e.g. $NAEL_{animal}$ and the $EF_{interspec}$). In this second interpretation the outcome of the risk assessment can, at least theoretically, be associated with a particular (small) proportion of the population at risk (e.g. Evans et al. 2001). The results of such an analysis could be a lower confidence limit (resulting from the type I uncertainties) for a no-adverse-effect dose for a particular (low) percentage of the population (resulting form the type II uncertainty in $EF_{intraspec}$).

In general it may be assumed that each EF (i.e. ratio of true NAELs) is approximately lognormally distributed. The plausibility of this assumption is confirmed by observed (ratios of) NOAELs, which are well described by lognormal distributions (e.g. Kramer et al. 1996).

When chemical specific information is available, the EF distributions may be (partly) based on that specific information. When no chemical specific information on any of the EFs is available, default EF distributions have to be used. Default distributions for the various EFs have been proposed by Vermeire et al (1999), after reviewing the relevant literature.

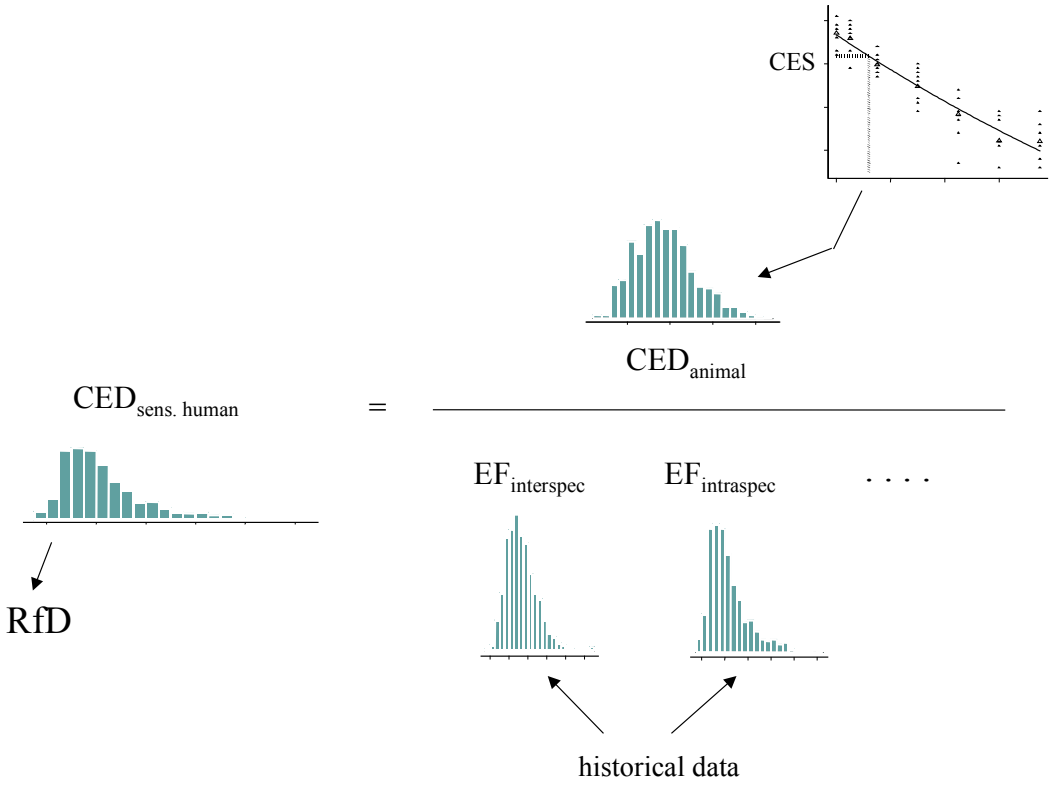

Fig. 2. Illustration of probabilistic assessment of the ADI, TDI, or RfD. The upper right corner shows dose-response data for a continuous endpoint from an (animal) study, with the fitted model (decreasing curve) used for deriving the CED distribution (numerator). The denominator comprises the distributions for the extrapolations to be made. The resulting distribution (left hand side of the equation) denotes the uncertainty in the CED for the sensitive human subpopulation, a lower percentile of which may be taken as the ADI, TDI, or RfD.

Although most available dose-response data in the literature report NOAELs, it should be noted that NOAELs are critically dependent on study design and dose-selection. In consequence the ratios of NOAELs do not accurately reflect the ratios of NAELs and are subject to large estimation errors, resulting in ratio distributions overestimating the variation of the relevant EF. Unfortunately it is not possible to quantify the estimation error of a NOAEL, and therefore correction cannot be made to obtain more realistic EF distributions. A better way to estimate EF distributions would be to base them on the ratios of CEDs, obtained by the benchmark approach. This is an important research need that may improve general risk assessment methodology.

Probabilistic assessment of ADI, TDI, RfD

The procedure for deriving a probabilistic ADI is illustrated in Fig. 2. In the upper right corner the data are presented related to the (continuous) endpoint considered as critical from the available studies on the particular compound. In this case the data points relate to the observations in individual animals, the larger marks indicating the group means. First a certain Critical Effect Size (CES) is determined, i.e., a certain percent change relative to the level of the endpoint observed in the controls, assuming that this particular percent change is non-adverse for the endpoint considered. Note that the CES is not a measure of incidence, but of (acceptable) severity. Then the associated Critical Effect Dose (CED_{animal}) is derived from the fitted dose-response model, together with its uncertainty distribution. This distribution is then "divided" by the distributions for the relevant EFs, usually inter- and intraspecies, and, if necessary, for other EFs, e.g. for subchronic to chronic extrapolation. Finally, the resulting distribution for the $NAEL_{sens.\ human}$ has to be analyzed to derive an ADI (or TDI, RfD), for example by selecting a low percentile of the uncertainty distribution. An obvious choice for this lower percentile is 5%, since this is generally considered in science as an acceptable error in significance testing (including significance testing in the classical approach aimed at deriving NOAELs). Thus, the interpretation of a probabilistically derived ADI (or TDI, RfD) is that it is unlikely (with quantitative information on how unlikely) that the true NAEL in the sensitive human is lower than the derived value.

In this illustration the final distribution is strictly interpreted as reflecting scientific uncertainty concerning the NAEL in the sensitive human. The meaning of the term "sensitive human" cannot be read from the final distribution: it is determined by the interpretation of the $EF_{intraspec}$ distribution.

Estimating risk at actual exposure levels

The probabilistic approach may also be used to estimate possible health effects at any given exposure level in the human population, be it in the general population or in a particular exposure group. The actual exposure level may be below or above the ADI (or TDI, RfD), but in practice one will mostly be interested in situations where these exposure limits are exceeded.

When exposure levels increase, the response rate (fraction of the population) and the response size (magnitude of effects in individuals) are expected to increase simultaneously. However, applications of the probabilistic approach thus far have focussed on either of these two. Baird et al (1996) and Evans et al. (2001) base their analyses on a fixed effect size considered as adverse, and aim to estimate the uncertainty distribution of the dose where a specified fraction of the human population may suffer from that effect. Slob and Pieters (1998) take the other approach, and aim to estimate the size of the effect in the individuals of a (unspecified) sensitive subpopulation. The latter approach is illustrated in Fig. 3. After scaling the actual human exposure level (possibly

related to a specific exposure group) to the animal dose level (depending on the animal species, see e.g. Vermeire et al. 1999, Baird et al. 1996), the size of the effect in the animal is estimated from the dose-response data obtained from a relevant toxicological study. The critical dose-response data being of a continuous nature (e.g. red blood cell counts, organ weights, enzyme levels), the size of the effect is defined in terms of a percent change in the level of the endpoint compared to the normal level in the controls. The uncertainty associated with the estimated effect size, due to experimental error, is quantified, resulting in a distribution for the expected effect size in the animal. Then this distribution is combined with the distributions for the extrapolation factors relevant for the particular assessment. The resulting distribution reflects the scientific uncertainty in the estimate of the effect-size in the sensitive human at the actual exposure level. Thus, a higher percentile of this distribution may be chosen to assess the upper bound of the size of the effect in the (sensitive) human population. When this upper percentile of the effect size is very small, or toxicologically insignificant, one may decide that human health risks can be disregarded. Or one might report both the 5th and the 95th percentiles as a 90%-confidence interval for the expected effect size in the sensitive human being. Again, in the illustration of Fig. 2., the final distribution reflects scientific uncertainty only: the interpretation of "sensitive human" depends on the interpretation of the $EF_{intraspec}$ distribution that was used in the analysis. An important assumption in this approach is that the dose-response relationship for the (continuous) endpoint used in the analysis is similar in animal and human, except for a dose factor that reflects the possible difference in sensitivity between the species.

An example of a probabilistic risk assessment at actual exposure levels can be found in Pieters et al. (2001), who estimated, for various endpoints, the possible effect sizes in the sensitive human subpopulation resulting from the estimated current intake of deoxynivalenol (DON) in cereal crops. For example, based on the 95^{th} percentile of the intake distribution in 20 year old women they estimated the additional risk of anomalous sternebrae in embryos between 0.0 and 0.6% (90%-confidence interval). For body weight reduction, the 90%-confidence interval for the estimated effect size in one-year old girls was 0.2 – 24.6%, based on the 95^{th} percentile of the intake distribution in this subpopulation. Thus, it may be concluded that the risk of anomalous sternebrae is minimal, since even the upper limit of the confidence interval is small. However, in one-year old girls the reduction in body weight could both be very small (around 0.2%), or quite substantial (around 25%), not allowing a positive or negative conclusive answer for this endpoint.

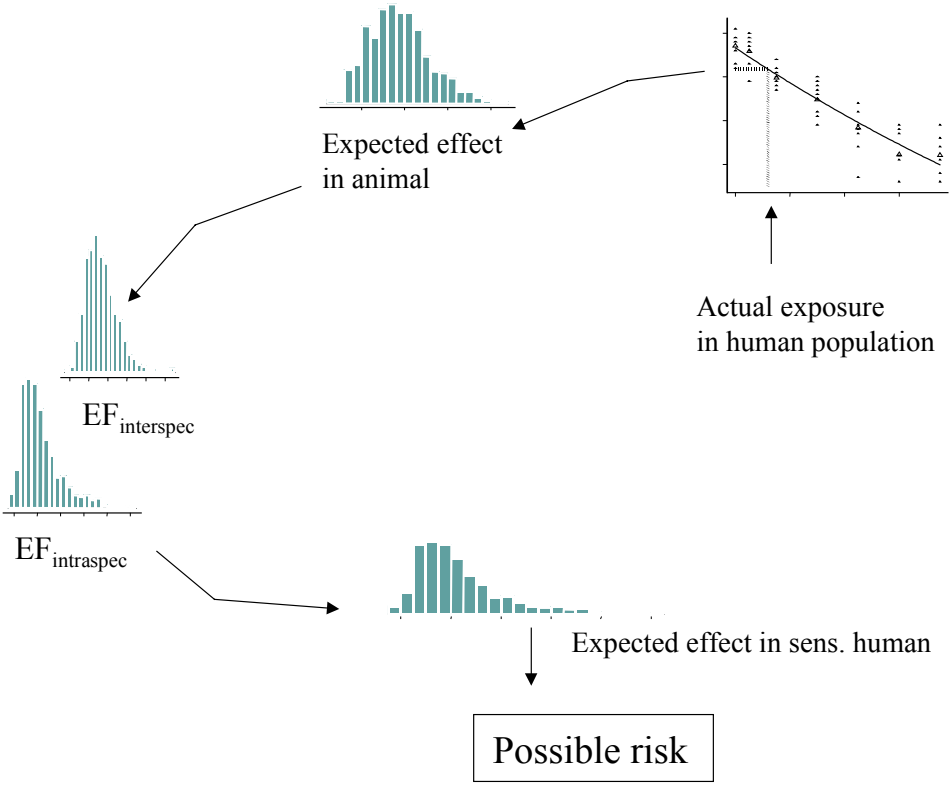

Fig. 3. Illustration of a probabilistic risk assessment aimed at the possible human health effects given a particular exposure level. The upper right corner shows dose-response data for a continuous endpoint from an (animal) study, together with the fitted model (decreasing curve) used for estimating, at that dose, the magnitude of the effect in the animal with the associated uncertainty distribution (top, middle). Combining this distribution with the EF distributions results in a distribution of the expected effect size in the sensitive human subpopulation.
Note that in this example the observations in the animal study refer to a continuous endpoint, and therefore the final distribution of the expected effect in the sensitive human is defined in terms of a continuous effect size (e.g. percent change in red blood cell counts). The upper percentile, indicated as " Possible risk" can be seen as an upper confidence bound of the expected effect size in the sensitive human subpopulation.

PROBABILISTIC ASSESSMENT OF EXPOSURE

In exposure assessments data are often poor, resulting in type I uncertainties. Due to variations in human behaviour type II uncertainties are unavoidable. Therefore, probabilistic exposure assessments are generally faced with both types of uncertainties, which need to be evaluated separately. To illustrate how this may be done, consider the example from the introduction, where exposure to a compound from a food item was estimated by multiplying the concentration in the food by the consumption rate of the food:

intake = concentration x consumption.

Both concentrations and consumption rates usually vary, and this may be expressed by a type II uncertainty distribution. Suppose now, that the data on concentrations in food are limited, so that the type II distribution for the concentrations itself is uncertain (i.e. type I uncertainty). For example, we may be quite sure that the distribution for the concentrations is lognormal, but we may have doubts on the exact value of the geometric mean (GM). This type I uncertainty may be expressed by yet another lognormal distribution. By randomly drawing values from the latter GM distribution we are in fact randomly drawing whole distributions for the concentrations. Then each distribution in the sample of distributions for the concentrations is multiplied with the consumption distribution, resulting in the sample of intake distributions.
This idea of hierarchical Monte Carlo analysis is further illustrated in Fig. 4. Here, not only the GM but also the GSD (geometric standard deviation) is assumed to be uncertain due to lack of data. First, a pair of values for the GM and the GSD is drawn from their type I distributions. Since a lognormal distribution (describing the variation in the concentrations) is completely determined by the GM together with the GSD, these two values form a randomly drawn type II uncertainty distribution of the concentrations. Then, this concentration distribution is multiplied with the consumption distribution, by a full Monte Carlo analysis, resulting in a (type II) distribution for the product, i.e., the intake. This whole process is repeated many (N) times, starting with two new random drawings from the type I distributions for the GM and GSD underlying the concentration distribution. The final result of this hierarchical Monte Carlo scheme is a whole bunch of type II distributions for the intake, which can all be plotted as cumulative distributions in one plot (see e.g. fig. 7.1 in Risk Assessment of Food Borne Bacterial Pathogens). Or, one may calculate, say, the 95th percentile of each (type II) intake distribution, which taken together compose a type I uncertainty distribution for that percentile. The latter distribution can be used to derive a confidence interval for the 95^{th} percentile of the intake in the population. For example, suppose the 90%-confidence interval were (12, 25), then the verbal conclusion from this analysis would be that 95% of the population should have an intake lower then 25, a statement that is made with 90% confidence.

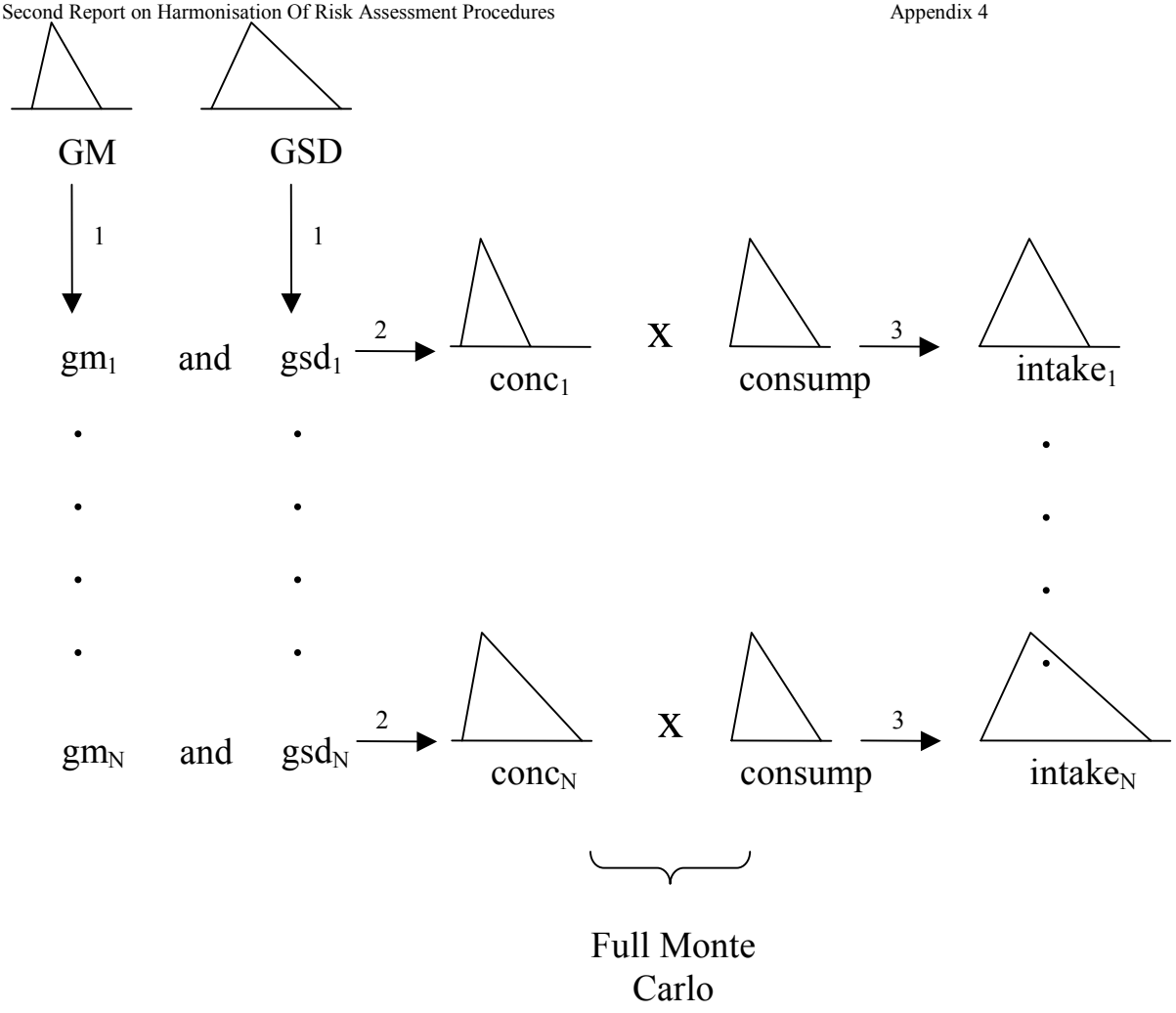

Fig. 4. Illustration of a hierarchical sampling scheme evaluating type I and type II uncertainties. A set of N pairs of gms and gsds is randomly drawn from their (type I) uncertainty distributions, indicated by arrows #1. Each pair determines (arrows #2) a type II distribution, expressing the variation in concentrations. Each concentration distribution is combined with the type II distribution for consumption by a full Monte Carlo analysis, resulting (arrows #3) in N type II distributions for the intake. Thus, the variation in the intake distributions reflect type I uncertainty.

As already noted in the introduction, an appropriate conceptual analysis of the problem at hand is crucial in an uncertainty analysis. A proper interpretation and use of any data should be part of that. A particularly important example in relation to risk assessment and the ADI (or TDI, RfD) is the use of food consumption survey data as a means to estimate interindividual variation in long-term intake rates of food. Food surveys usually consist of (consecutive) records of daily food consumption. The observed variation in daily consumption is the "sum" of the long-term variation in consumption habits between individuals, and the variation between (short-term) differences between days (within individuals). Usually the aim is to compare the exposure in the population with a health-

based exposure limit (RfD, ADI, TDI), which is typically based on chronic exposure. Therefore, the long-term variation in consumption habits between individuals is of interest. The observed variation in the raw data overestimates that variation, and needs to be corrected for the daily fluctuations. Here, Monte Carlo analysis does not apply. What can be done is an analysis based on a statistical model (Slob, 1993). This exemplifies the phenomenon that observed variation in the available data may not match the variation that is needed, due to the fact that other, not relevant sources of variation are incorporated in the observed variation.

DISCUSSION

The strong point of the probabilistic approach is that it quantifies the uncertainty associated with any particular assessment, in addition to the magnitude of the risk level: the former is given by the width of the distribution, the latter by the location of it. In the default uncertainty factor approach a low ADI (or TDI, RfD) may result from high toxicity as well as from poor data (large uncertainty).

Being an extension of the default approach, the probabilistic approach solves a few of the existing weaknesses, but importantly it does not introduce any new ones. For example, one might argue that a weakness of the probabilistic approach is that the default interspecies distribution to be used in the absence of chemical-specific information, is not firmly based on data, and that the default distributions appear to some extent arbitrarily chosen. However, it should be noted that this criticism similarly holds for the default factor of ten that is usually applied in current practice.

An important advantage of the probabilistic approach is that it allows for the estimation of possible health effects given the actual exposure in the population (Fig. 3). The outcome of such an analysis may lead to the conclusion that health effects in the human population are likely (so that measures are required) or unlikely (so that measures are not required), but the outcome may also be inconclusive. Nonetheless, the latter situation is informative and helpful in that the probabilistic approach can provide useful information on the consequences of proposed risk management options that may either increase or decrease exposure. One may compare the costs involved in reducing exposure with costs of reducing the uncertainties in the risk assessment, taking the severity of the possible effects in the human population into account. Thus, an important strength of the probabilistic approach is that it enhances the decision making process.

REFERENCES

S.J.S. Baird, J.T. Cohen, J.D. Graham, A.I. Shlyakter, and J.S. Evans, " Noncancer Risk Assessment: A Probabilistic Alternative to Current Practice," *Human Ecolog. Risk Assess*. 2:79-102 (1996).

K.S. Crump, "A New Method for Determining Allowable Daily Intakes," *Fund. Appl. Toxicol*. 4:854-871 (1984).

Evans, J.S., Rhomberg, L.R., Williams, P.L. and Wilson, A.M. and Baird, S.J.S. 2001. Reproductive and Developmental Risks from Ethylene Oxide: A Probabilistic Characterization of Possible Regulatory Thresholds, Risk Analysis, Accepted.

Hattis, D., Banati, P., and Goble, R. (1999) Distributions of individual susceptibility among humans for toxic effects. How much protection does the traditional tenfold factor provide for what fraction of which kinds of chemicals and effects? Annals of the New York Academy of Sciences, 895, 286-316.

Kramer, H.J., W.A. Van den Ham, W. Slob, M.N. Pieters, "Conversion Factors Estimating Indicative Chronic NOAELs from Short-Term Toxicity Data," *Regul. Toxicol. Pharmacol*. 23:249-255 (1996).

Pieters, M.N., Freijer, J., Baars, B, Fiolet, D.C.M., Van Klaveren, J, and Slob, W. (2001) Risk assessment of deoxynivalenol in food: Concentration limits, exposure and effects. In: Mycotoxins and Food Safety (Jackson, L.S., Trucksess, M.W. and DeVries, J.W., eds). In press.

Risk Assessment of Food Borne Bacterial Pathogens: Quantitative Methodology Relevant for Human Exposure Assessment. SSC report, draft.

Slob, W., and Pieters, M.N. (1998).
A probabilistic approach for deriving acceptable human intake limits and human health risks from toxicological studies: general framework.
Risk Analysis 18: 787-798.

W. Slob (1993).
Modeling long-term exposure of the whole population to chemicals in food.
Risk Analysis 13:525-530

Vermeire T, Stevenson H, Pieters MN, Rennen M, Slob W, Hakkert BC (1999)
Assessment Factors for Human Health Risk Assessment: A Discussion Paper
Crit. Rev. Toxicol. 29: 439-490.