Mathematical modelling and quantitative methods


*Deutsches Krebsforschungszentrum, German Cancer Research Center, Abteilung Biostatistik, R 0700, Postfach 10 19 49, D-69009 Heidelberg, Germany
bTERA, Toxicology Excellence for Risk Assessment, 1757 Chase Avenue, Cincinnati, OH 45223, USA
cILSI Europe, avenue E. Mounier B3, Box 6, B-1200 Brussels, Belgium
dTechnology Sciences Group, Inc., Toxicology, Ecotoxicology and Risk Assessment Division, 1101 17th Street, NW, Washington, DC 20036, USA
eAjinomoto Switzerland AG, Innere Güterstrasse 2-4, PO Box 4559, CH-6304 Zug, Switzerland
fUniversity of Southampton, Clinical Pharmacology Group, Biomedical Sciences Building, Bassett Crescent East, Southampton SO16 7PX, UK
gRIVM, National Institute of Public Health and the Environment, PO Box 1, NL-3720 BA Bilthoven, The Netherlands
hCoca-Cola Nordic & Baltics, Strandrejen, 60-5th, DK-2900 Hellerup, Denmark

Summary

The present review reports on the mathematical methods and statistical techniques presently available for hazard characterisation. The state of the art of mathematical modelling and quantitative methods used currently for regulatory decision-making in Europe and additional potential methods for risk assessment of chemicals in food and diet are described. Existing practices of JECFA, FDA, EPA, etc., are examined for their similarities and differences. A framework is established for the development of new and improved quantitative methodologies. Areas for refinement, improvement and increase of efficiency of each method are identified in a gap analysis. Based on this critical evaluation, needs for future research are defined. It is concluded from our work that mathematical modelling of the dose–response relationship would improve the risk assessment process. An adequate characterisation of the dose–response relationship by mathematical modelling clearly requires the use of a sufficient number of dose groups to achieve a range of different response levels. This need not necessarily lead to an increase in the total number of animals in the study if an appropriate design is used. Chemical-specific data relating to the mode or mechanism of action and/or the toxicokinetics of the chemical should be used for dose–response characterisation whenever possible. It is concluded that a single method of hazard characterisation would not be suitable for all kinds of risk assessments, and that a range of different approaches is necessary so that the method used is the most appropriate for the data available and for the risk characterisation issue. Future refinements to dose–response characterisation should incorporate more clearly the extent of uncertainty and variability in the resulting output. © 2002 ILSI. Published by Elsevier Science Ltd. All rights reserved.

Keywords: Hazard characterisation; Risk assessment; Mathematical modelling; Benchmark; Probabilistic methods; Toxicokinetic models; Categorical regression

Abbreviations: AEL, adverse effect level; ADI, acceptable daily intake; ALARA, as low as reasonably achievable; ALARP, as low as reasonably practicable; AUC, area under the plasma concentration–time curve; BMD, benchmark dose; BMDL, benchmark dose lower confidence limit; BMDP, benchmark dose point estimate; BMR, benchmark response level; CED, critical effective dose; CES, critical effect size; CSAFE, chemical-specific adjustment factor; CSTEE, EU Scientific Committee for Toxicity, Ecotoxicity and the Environment; ECETO, European Centre for Ecotoxicological and Toxicological Chemicals; ED50, effective dose 50%; EF, extrapolation factor; EPA, US Environmental Protection Agency; EU, European Union; FAO, Food and Agriculture Organisation of the United Nations; FDA, US Food and Drug Administration; FEL, frank-effect level; FEMA, US Federal Food and Drug Administration; FOSIE, Food Safety in the European Union; GLP, good laboratory practice; IOFI, International Organisation of Flavour Industry; IPCS, International Programme on Chemical Safety (WHO); ITR, International Toxicity Estimate for Risk; JECFA, Joint FAO/WHO Expert Committee on Food Additives; LD50, lethal dose 50%; LMS, linearised-multistage model; LOAEL, lowest-observed-adverse-effect level; MLE, maximum likelihood estimate; NAEL, no-adverse-effect level; NOAEL, no-observed-adverse-effect level; NOEL, no-observed-effect level; OECD, Organisation for Economic Co-operation and Development; OSHA, US Occupational Safety and Health Administration; PAFA, Priority-based Assessment of Food Additives of the FDA; PB-PK/PD, physiologically-based pharmacokinetic model/pharmacodynamic; PBTK, physiologically-based toxicokinetic; PBTK/TD, physiologically-based toxicokinetic/toxicodynamic; PTWI, provisional tolerable weekly intake; QSRP, qualitative structure-activity relationship; RA, risk assessment; RAC, reference concentration; RFD, reference dose; RIF, Registry of Toxic Effects of Toxic Substances; SAR, structure-activity relationships; SCF, EU Scientific Committee on Food; TDI, tolerable daily intake; TSH, thyroid stimulating hormone; TTC, Threshold of Toxicological Concern; UF, standard uncertainty factor; VSD, virtually safe dose.

* Corresponding author. Tel.: +32-2-771-00-14; fax: +32-2-762-00-44.
E-mail address: jkleiner@ilsi.europe.be (J. Kleiner).

0278-6915/02/$ - see front matter © 2002 ILSI. Published by Elsevier Science Ltd. All rights reserved.
PII: S0278-6915(01)00116-8
**Contents**

Introduction ........................................................................................................... 285
  1.1. Methods used in the European Union (EU) .................................................. 286
  1.2. Outline of this document ............................................................................ 287

2. Structure-activity relationships (SAR) and the Threshold of Toxicological Concern... 288
  2.1. Models and methods .................................................................................. 289
  2.2. Subpopulations ........................................................................................ 289
  2.3. Data requirements ..................................................................................... 290
  2.4. Strengths, limitations and weaknesses ....................................................... 290
  2.5. Applicability ............................................................................................ 290
  2.6. Acceptability to regulatory agencies and authorities ................................... 290
  2.7. Gaps and research needs .......................................................................... 290

3. Threshold methods ............................................................................................. 291
  3.1. Models and methods .................................................................................. 291
  3.2. Subpopulations ........................................................................................ 292
  3.3. Data requirements ..................................................................................... 292
  3.4. Strengths, limitations and weaknesses ....................................................... 293
  3.5. Applicability ............................................................................................ 293
  3.6. Acceptability to regulatory agencies and authorities ................................... 293
  3.7. Gaps and research needs .......................................................................... 294

4. Categorical regression ......................................................................................... 294
  4.1. Models and methods .................................................................................. 294
  4.2. Subpopulations ........................................................................................ 296
  4.3. Data requirements ..................................................................................... 296
  4.4. Strengths, limitations and weaknesses ....................................................... 297
  4.5. Applicability ............................................................................................ 297
  4.6. Acceptability to regulatory agencies and authorities ................................... 297
  4.7. Gaps and research needs .......................................................................... 297

5. Chemical-specific adjustment factors ................................................................ 298
  5.1. Models and methods .................................................................................. 298
  5.2. Subpopulations ........................................................................................ 299
    5.2.1. Infants and children ............................................................................ 300
    5.2.2. Ethnic differences .............................................................................. 300
    5.2.3. Polymorphic metabolism .................................................................... 300
  5.3. Data requirements ..................................................................................... 300
    5.3.1. Toxicokinetics ................................................................................... 300
    5.3.2. Toxicodynamics ............................................................................... 300
  5.4. Strengths, limitations and weaknesses ....................................................... 301
  5.5. Applicability ............................................................................................ 301
  5.6. Acceptability to regulatory agencies and authorities ................................... 301
  5.7. Gaps and research needs .......................................................................... 301

6. Non-threshold methods ....................................................................................... 302
  6.1. Models and methods .................................................................................. 302
  6.2. Subpopulations ........................................................................................ 303
  6.3. Data requirements ..................................................................................... 303
  6.4. Strengths, limitations and weaknesses ....................................................... 303
  6.5. Applicability ............................................................................................ 304
  6.6. Acceptability to regulatory agencies and authorities ................................... 304
  6.7. Gaps and research needs .......................................................................... 304
Low molecular weight chemicals, micronutrients and nutritional supplements, macronutrients, whole foods, novel foods and food processing may in principle pose a hazard to human health. Therefore, human food safety has been of continuous interest for risk assessors, public health officials, and the food industry. The presence of small amounts of ingredients in human food that are known to be harmful in high doses has emphasised the need for toxicological and methodological research suitable for risk assessment. The fundamental aim has been the establishment of a safe level for human exposure, which is achieved by the processes of hazard identification, hazard characterisation and exposure assessment.

The methods described in the following sections have been compiled to define the state of the art of mathematical modelling and quantitative methods that can be used for regulatory decision-making in Europe. Existing practices of the Joint FAO/WHO Expert Committee on Food Additives (JECFA), the US Food and Drug Administration (FDA), the US Environmental Protection Agency (EPA), etc., are compared. This should aid risk assessors, public health officials and the food
industry in applying appropriate testing, estimation and assessment procedures, and in developing objective quantitative risk characterization.

The task of this working group of the concerted action on risk assessment of chemicals in food and diet in Europe (FOSIE) has been to carry out a thorough and comprehensive investigation of quantitative methods of hazard characterization used in food safety assessment. This was achieved by conducting an evaluation of the presently available state of the art of mathematical methods and statistical techniques for hazard characterization, and by proposing potential future developments for the use of quantitative methods. It was the working group's intention to establish a framework for the development of new and improved quantitative methods for risk assessment of chemicals in food and diet. Areas for refinement, improvement and increase in efficiency of the methods used at present were identified in a gap analysis. Based on this assessment, needs for future research were defined. We should remark that the quantitative methods discussed below primarily address chemicals in food such as food contaminants or naturally occurring toxicants, but they are in principle also applicable to other categories such as macroingredients and whole food. The work of the theme group of mathematical modelling was aimed at major aspects of mathematical and statistical modelling for food safety assessment. Exposure models and risk models for data obtained from human epidemiological studies were not covered. The models described below have been designed and developed for toxicity data that can use experimental human data, but are not considered as models applicable in the field of epidemiology. Epidemiological risk modelling, based on standard methods of analytical epidemiology, will be addressed by van den Brandt et al. (2002) and exposure modelling is covered in Kroes et al. (2002), both in this issue. It is strongly recommended that modelling and quantitative analysis are based on data the quality of which has been assured, for example, by principles of good laboratory practice (GLP).

Before the results of this concerted work are presented, we summarise briefly the present status of quantitative risk assessment methods for food safety in the European Union.

1.1. Methods used in the European Union (EU)

Scientists working for JECF A developed the concept of acceptable daily intakes (ADIs) for food additives in the early 1960s. This was a further development of work by the FDA in the early 1950s. This approach assumes that a maximum safe dose (a dose without significant adverse health effects) can be determined from studies in experimental animals or humans and that division of this dose by appropriate safety factor(s) defines the 'safe' intake dose for the human population. At present, the EU applies this traditional approach for the evaluation of chemicals in food. The test methods applied are those of the Organisation for Economic Cooperation and Development (OECD) guidelines for the testing of chemicals, which the EU has adopted under the directives on chemical substances and preparations. A description of the minimum requirements for data for the assessment of food additives was published in 1989 in a guideline for industry. Otherwise there are no generally established quantitative methods at this time. The EU Scientific Committee on Food (SCF) was established in 1974 to advise the Commission on any problems relating to the protection of the health and safety of persons arising from the consumption of food and in particular to the composition of food, processes which are liable to modify food, the use of food additives and other processing aids as well as the presence of contaminants. Evaluations are undertaken on novel food ingredients, such as additives, and when contaminants are identified. Evaluations may be reconsidered when new data, which are expected to change the existing evaluation, are presented to the Committee. The EU Scientific Committee for Toxicity, Ecotoxicity and the Environment (CSTEE) issues reports on risks to the health of humans and the environment to industrial chemicals and other exposure factors.

In Europe there is a difference in the approach adopted for characterising carcinogens between the official bodies evaluating environmental problems and those that deal only with foodstuffs. The approach used by the EPA is determined by how well the mechanism/mode of action of the carcinogenic response is understood. If the mode of action is not sufficiently well characterised or the dose response is shown to be linear in nature, mathematical modelling for low-dose extrapolation, assuming linearity, is employed. If the mode of action is understood to have a dose response indicative of a threshold, then a margin of exposure/safety factor approach, similar to that used for non-cancer effects, is employed. The SCF and the JECFA consider that quantitative low-dose risk extrapolation does not have an adequate science basis, and their advice for carcinogens, is that their presence in the diet should be reduced "As Low As Reasonably Achievable (ALARA)" or to "Irreducible Levels".

The SCF has evaluated food additives and contaminants and published their evaluations since 1975. Safety evaluation of flavourings has been under consideration for many years within the EU, by the Council of Europe and by other international organizations such as the International Organization of Flavour Industry (IOFI) and the US Flavour and Essence Manufacturers Association (FEMA). The Council of Europe has established lists of flavouring materials used either as flavouring substances per se or as a source material from which flavourings are prepared (first edition in 1970). The SCF has, based on the Council of Europe
work, established a short list of the most “active principles with limit values in final food”. This has been included in the Directive on Flavours for Foodstuffs (EC, 1988) and this list excludes their direct use as flavouring substances. Guidelines for evaluation of chemically defined flavourings were established (EC, 1992) in which the use of structure-activity relationships (SAR) should be considered for cases where the chemical structure is similar to that of compounds with known toxicological and biological properties. In 1995, the JECFA (44th meeting) introduced an assessment scheme based on chemical structure, predicted metabolism and the experience from the FDA evaluations of packaging migrants using the concept of “Threshold of regulation”. This was based on the work of Hattan and Rulis (1986), who made an analysis of the FDA’s Priority-based Assessment of Food Additives (PAFA) database on substances with subchronic and chronic toxicity data, contained in the Registry of Toxic Effects of Toxic Substances (RTECS) and the carcinogen potency database of Gold et al. (1984). This method has been developed further by Munro and Kroes (1998) and Munro et al. (1999). The SCF has agreed with the JECFA evaluations on flavours in cases where they have considered the same substances.

With regard to packaging materials in contact with food, the SCF has made toxicological evaluations of monomers and other starting substances to be used in the manufacture of plastic materials which come into contact with foodstuffs since 1976. Also the migration of heavy metals from ceramics and degradation products from cellophane have been considered and regulated through directives. A guideline for data presentation for evaluation was published in 1990. The SCF is not using the US FDA “Threshold of Regulation” principle in evaluating substances and materials in contact with food. The SCF has published reports on issues for consideration in the evaluation for safety of novel foods and on requirements for information to be submitted for review before marketing. A decision tree has been established for their evaluation in which toxicological information is evaluated traditionally in accordance with the ADI concept, and where nutritional aspects are included.

1.2. Outline of this document

In the following sections we describe eight approaches for food safety/risk assessment that differ in their data requirements, degree of complexity, their applicability in different situations and the type and quality of resulting risk estimates. Each method is examined according to a standardised scheme. First, the mathematical models and the quantitative methods relevant to the specific approach are presented. This is followed by consideration of subpopulations and the data requirements for application of the model. The method is then discussed with respect to its strengths, limitations and weaknesses. The applicability of the methods is clarified and, finally, the acceptance by regulatory agencies is reported as far as this information was obtainable. Gaps and research needs are formulated at the end of each section.

Section 2 addresses the use of structure-activity information, which can be applied when adequate toxicity data do not exist for the chemical. If the extent of exposure to humans is extremely low, a pragmatic approach to define a reference dose for regulation is the determination of the so-called Threshold of Toxicological Concern (TTC). If human intake of a food component is below the TTC, a safety assessment can be undertaken without chemical specific toxicity data. In section 3, threshold methods are presented under the assumption that there is a level of exposure below which a biologically significant effect is not produced following exposure to the food chemical. The use of no-observed-adverse-effect levels (NOAELs), or related exposure levels as points of departure is explained and the application of standard uncertainty factors (UFs) to estimate an ADI or its equivalent is presented. This approach, which uses default uncertainty factors, can be undertaken despite limited knowledge of the biological processes involved in the generation of the adverse effect. Categorical regression is addressed in section 4. This method has received limited use to date, but allows the integration of all observed adverse effects in one regression analysis. It yields an estimate of a point of departure for the application of uncertainty factors, but also provides information about toxicity at higher doses. A refinement of the UF approach used with threshold methods employs chemical-specific adjustment factors (CSAFs). This method, described in section 5, makes direct use of toxicological and biological knowledge specific to the compound of interest and is seen as a substantial refinement over earlier default methods. The improvement that results when replacing the standard UFs with CSAFs is recognised in subsequent sections when performing low-dose extrapolation from a benchmark dose (section 7), or deriving a probabilistic distribution of an ADI (section 8). Standard mathematical models for low-dose risk estimation are discussed in section 6. These can be applied if it is assumed that no level of exposure is without a biologically significant risk. Historically, the application of simple mathematical models such as the one-hit or linearised multistage model allowed useful comparisons between different compounds, but the actual numerical values were determined more by the model than the data. We show how limitations of this approach argue for the development of methods, which use all the available dose-response information and are based more on biological principles. The next three sections (7–9) present further
methods and models that make more extensive use of all data available, but which are useful only if sufficient data exist. The benchmark dose, defined as a statistical lower confidence limit for a dose that produces a predefined change in response over background, denoted BMDL, is discussed in section 7. This method is able to provide a model-defined point of departure for low-dose extrapolation. Probabilistic risk assessment, described in section 8, uses probability distributions of the uncertainty factors in order to estimate the probabilistic distribution of a true underlying biological no-adverse-effect level (NAEL). If mechanistically-based toxicokinetic and toxicodynamic data are obtainable, risk characterisation can be improved considerably. This is illustrated in section 9 by physiologically-based toxicokinetic modelling (PBTK models), which can be used at various stages of risk assessment (definition of a target dose, extrapolation across species and routes of exposure, integration into a comprehensive toxicokinetic and toxicodynamic model, refinement of CSAs). For a more detailed discussion of the role of kinetics for in vitro and in vivo toxicology and for sources for information to build PBTK models we refer to Eisenbrand et al. (2002), and especially to Dybing et al. (2002). In section 10 we present the results of a general evaluation and comparison of the methods introduced in the earlier sections 2–9 through the use of summary tables. We examine the approaches for their main characteristics, data requirements, applicability and acceptance by regulatory bodies. We discuss reliability, precision, validity and predictive power. Research needs identified earlier in each section are formulated comprehensively in section 11. Finally, the conclusions of the review are drawn and formulated in section 12.

2. Structure–activity relationships (SAR) and the threshold of toxicological concern

Structure–activity relationships (SAR) may be used in the absence of adequate toxicity data on the chemical under certain circumstances, such as when there is an urgent need for an assessment, for example after an accident, or when the extent of exposure of humans is extremely low. In the area of food safety there may be negligible risk, irrespective of the toxicological properties of the chemical, providing that the exposure is sufficiently low. This approach has been applied mainly to the assessment of packaging migrants and flavouring substances.

Structure–activity relationship analysis is based on the premise that the biological activities of a compound are related to its structure, and the functional groups present. It is a powerful method, especially when used to predict the activity of a chemical based on data for a structurally-related series of compounds, or for a specific and well-defined target (see Barlow et al., 2002 for a more complete discussion) for a more complete discussion. SAR has been used extensively by the pharmaceutical industry to predict receptor binding, potential therapeutic actions and kinetics. For risk assessment, SAR has been applied most widely for hazard identification, for example the assessment of potential genotoxicity, and also in the field of ecotoxicology. SAR can be used for predicting the potential toxicity of a new chemical, based on the analysis of the potencies of a wide range of existing chemicals. SAR, based on data on the oral LD50 values for over 2000 compounds, could predict the toxicity within a factor of 8-fold of the true potency for 95% of compounds, although this approach was less successful when applied to eye irritancy (DiPasquale and Hayes, 2001). A similar predictability was derived in an analysis of a smaller database (234 diverse chemicals) using the lowest observed adverse effect level from chronic bioassays (Mumtaz et al., 1995). A number of computer-based expert systems are available for the application of SAR to different data series. These methods are not used in modelling dose response. A combination of SAR and PBTK modelling is being explored as a way of providing information without the need for in vivo animal testing (see Eisenbrand et al., 2002).

The term “threshold” is used with two different meanings. For most forms of toxicity the dose or exposure has to reach a certain level before an adverse response will be produced, and hence it is considered that there is a “threshold” in the dose–response curve (the validity of this is discussed in section 3). In addition, the term threshold is used for an exposure that is so low that it is not of toxicological concern, in other words there is a pragmatic “threshold for concern”, or a “threshold for regulation”.

2.1. Models and methods

The establishment of safety on the basis of structure–activity considerations, in the absence of chemical-specific toxicity data, is a pragmatic approach taken to avoid the unnecessary use of human resources and test animals to investigate trivial risks. The approach is based on an assessment that the human exposure is sufficiently below the toxic doses of structural analogues, that there is an adequate safety margin. The safety margin would have to allow for the usual aspects of extrapolating from animal toxicity data (see below), and also allow for possible differences in potency compared with structural analogues.

In 1993, the FDA in the USA proposed that a dietary concentration of 0.5 ppb could be used as a threshold for regulation, in relation to the migration of food contact materials (Federal Register, 1993). Concentrations in the diet below this limit would give exposures that would not be of concern. This value was chosen because it was 2000 times less than doses of non-pesticides that cause toxicity in animals, and 200 times lower than the...
toxic doses of pesticides. The Federal Register (1993) states that "known carcinogens and substances whose chemical structures provide reason to suspect that they may be carcinogens will be excluded from review under the proposed regulation". However, because it was possible that an untested compound could be a carcinogen, the threshold for regulation was compared with "safety" factors of 100. Thus, the value of 0.5 ppb was supported by data on both threshold and non-threshold considerations.

The scheme used by JECFA to evaluate groups of flavouring substances represents a refinement of the FDA approach. The scheme involves a series of exposure thresholds that are based on the structural characteristics of the chemical. Structural characteristics and the potential for toxicity are taken into account by application of the Cramer decision tree (Cramer et al., 1978), which divides chemicals into three different classes determined by the chemical structure and the functional groups present.

Class I. Substances that have simple chemical structures and efficient modes of metabolism, and that would have low oral toxicity.

Class II. Substances that have less innocuous structural features than those of substances in Class I, but are not suggestive of toxicity. Substances in this class may contain reactive functional groups.

Class III. Substances that have structural features that permit no strong initial presumption of safety, or that may even suggest significant toxicity.

The thresholds for chemicals in structural classes I, II, III, were derived from the distribution of NOAEL values (see later) for all chemicals that have been tested in chronic, or subchronic studies (Munro et al., 1996). The distributions of NOAEL values for chemicals in each class were analysed and the 5th percentile of the NOAEL distribution determined. The threshold for each class was calculated by dividing the 5th percentile NOAEL value from the class data by the usual "safety" factor of 100 (see section 5 below). The 5th percentile NOAEL values for structural classes I, II and III (based on 137, 28 and 448 compounds, respectively) were 3.0, 0.91 and 0.15 mg per kg body weight per day, resulting in human exposure thresholds of 1800, 540 and 90 μg per person per day assuming 60 kg body weight (Munro et al., 1996). Reservations have been expressed about the validity of the class I–III thresholds for compounds that are neurotoxins, developmental toxins or immunotoxins. A recent review (Kroes et al., 2000) analysed the NOAELs for compounds showing these different types of toxicity, and also evaluated the effects produced by the 19 compounds in the Munro et al. (1996) analysis with NOAELs below the class III threshold. The review supported the thresholds derived by Munro et al. (1996), and concluded that neurotoxicity was the most sensitive endpoint considered, and that the distribution of NOAELs for neurotoxic chemicals occurred at lower doses than the distribution for class III compounds.

The scheme used for flavours by the JECFA incorporates an additional human exposure threshold of 1.5 μg per day. This would be applied to compounds that are not metabolised to innocuous products, and for which there are no toxicity data on the chemical itself or on structural analogues. This threshold was derived by analysis of the dose response data for known carcinogens, with extrapolation using a highly conservative model to the intakes estimated to be associated with a 1 in a million risk. The threshold was selected based on the distribution of 10−6 risk estimates and the probability of an untested compound proving to be a carcinogen. Despite the conservative approach used to derive the threshold of 1.5 μg per day, JECFA decided that the procedure and the threshold would not be applied to compounds with structural alerts for genotoxicity.

Flavouring substances are evaluated using a series of questions related to the potential for toxicity, based on structural and metabolic considerations, and the exposure (Fig. 1).

Examples of flavouring substances evaluated by the JECFA include a series of 47 primary alcohols, aldehydes, carboxylic acids, acetals and esters, all of which were considered to be metabolised to innocuous products. Forty-one of the substances were considered not to represent a safety concern because the estimated intakes were below the corresponding thresholds for the structural class. Of the remaining six substances, five were endogenous compounds and the other substance was evaluated based on a NOAEL in a 2-year animal feeding study (WHO, 1994).

2.2. Subpopulations

Subpopulations are not taken into account specifically other than in the threshold and non-threshold approaches used to derive the values of the thresholds of concern for different structural classes (see sections 3 and 6).

2. Data requirements

The use of these structure-based thresholds depends on both the structure and the exposure. The chemical structure has to be known, so that the structural class can be determined using the Cramer decision tree. Thresholds are usually applied to compounds for which few data are available and therefore exposure estimates are usually based on models and predictions.
Fig. 1. Decision tree used by the JECFA for the evaluation of flavouring substances (adapted from the Fifty-first Report of the JECFA (WHO, 2000)).

In the flavour evaluation scheme used by the JECFA, the intake that is compared with the threshold is a per capita estimate, which is based on annual production volume and the numbers of consumers, using data for Europe and the USA. Built-in conservative assumptions are that only 60% of the total amount that is used by food and beverage manufacturers is reported, and that only 10% of the population consume the flavour.

2.4. Strengths, limitations and weaknesses

The strength of the approach is that it allows a simple assessment of whether the current exposures represent a safety concern. The main weaknesses are the assumptions about the structure–activity relationship for the chemical under assessment compared with those in the databases that were used to derive the thresholds. Another difficulty is that the assessment relates to the exposure estimates that were used in the safety evaluation, and an increase in exposure would require a re-evaluation.

2.5. Applicability

Theoretically, the procedure (Fig. 1) is applicable to all compounds, although in practice it is used by JECFA either for compounds without structural alerts for genotoxicity and with low exposures, or for compounds that form normal endogenous metabolites.

2.6. Acceptability to regulatory agencies and authorities

In the EU, the JECFA scheme has been accepted by the SCF (1999) in general terms and its use for the evaluation of chemically defined flavouring substances is included in EU legislation (EC, 2000). However, the SCF has, for the time being, excluded the application of the TTC concept. Neither JECFA nor the SCF has accepted the 1.5 μg per day threshold for possible genotoxic carcinogens, despite the fact that it was derived using the highly conservative linear low-dose extrapolation model (see below). This is probably because of the undesirability of intentionally adding any level of genotoxic compound to the human diet. The value of the 1.5 μg per day threshold for the assessment of natural genotoxicants in the diet has not been considered, but could provide a useful threshold for risk management in the absence of adequate dose–response data on the compound.

2.7. Gaps and research needs

The Cramer decision tree is an integral part of the JECFA decision scheme in relation to establishing the structural class of the flavouring substance, and the determination of thresholds. This is an old publication and should be re-validated in the light of recent toxicological and metabolic data.
3. Threshold methods

For some hazards, such as non-genotoxic effects, it is assumed that there is a threshold of exposure below which no biologically significant effect will be produced. In such cases it is current practice within the EU to estimate the level of exposure without significant adverse effects. However, an absolute threshold in a scientific sense is not required to use the following methods. Instead, a threshold may be considered as a matter of biological or regulatory acceptability.

There is considerable debate about whether there are true biological thresholds in the dose response curves for the majority of toxic effects, i.e., excluding cancer involving genotoxicants and other genotoxic outcomes. Effect or response data that are determined from studies in animals may represent either a quantal or a continuous variable. An example of quantal data would be the incidence of a specific lesion, such as a histopathological change, while examples of a continuous variable would be a change in organ weight or body weight. Continuous variables can be converted to quantal data provided that a range of normality can be defined; in other words, upper and lower level limits can be set, below and above which any observation would be considered abnormal (outside the usual range), and hence the magnitude of the response in an individual would be converted to quantal effect.

The presence of a threshold cannot be demonstrated readily from experimental data, because any experimental dose–response relationship (whether or not it has a threshold) may include doses without a measurable (or statistically significant) biological effect in the test system. In consequence, the existence of biological thresholds cannot be proven or disproven, and the possibility of a level of exposure that does not produce any effect (rather than any measurable effect) has to be based on experience and expert judgement of the underlying biology of the test system. Although it can be argued that thresholds cannot exist in absolute terms, for instance a very low concentration will still interact with the biological system, in reality the presence of homeostatic and cytoprotective processes means that the interaction between the chemical and the biological system has to exceed the homeostatic or other protective processes in order to elicit a toxicologically relevant response (see Dybing et al., 2002). Although a biological threshold may occur in a single cell, due to homeostatic and cytoprotective processes, the same threshold would not apply to all cells in a population; consequently a true biological threshold cannot be defined for a population of cells. A similar situation applies to individuals, since thresholds can vary between individuals, and also in the same individual with time. Dose response data always relate to populations and not to single cells or single individuals. The issue of quantitative thresholds has been further elaborated by Slob (1999), who has concluded that a practical solution is to define thresholds in terms of the magnitude of changes in toxicological endpoints that may be considered as non-adverse for an individual organism.

Although the proof of the presence or absence of a threshold remains a matter for debate in risk assessment, subdivision of toxic effects into threshold and non-threshold has been the basis for risk assessment for the past 30–40 years (WHO, 1999b).

3.1. Models and methods

Low-dose extrapolation of the dose–response relationship (see section 6.1) is not used for threshold effects, and the approach that is adopted is the application of safety factors or uncertainty factors, (which allow for inter-species differences and for human variability) to a dose that is considered to be a surrogate for the threshold. The result is considered to be an exposure for humans, which would be without a biologically significant adverse health effects (safety assurance). Safety assurance is an example of quantitative risk assessment with the level of risk being considered to be "biologically insignificant" rather than a specific quantitative estimate. The term "safety assurance" gives the implication of complete absence of risk, but all outputs such as the acceptable daily intake (ADI), tolerable daily intake (TDI), reference dose (RfD), etc., are defined as "without significant" or "without appreciable" adverse health effects (WHO, 1987, 1994, 1999b).

Safety assurance is based on using the dose–response relationship to define an approximation of the threshold for toxicity in the animal study. The endpoint normally used as a surrogate for the threshold is the NOAEL (WHO, 1999b). The term no-observed-effect level (NOEL) is also used, but not all observed effects are adverse and would be the basis for quantitative risk assessment. In consequence, the NOAEL has tended to be used more in recent years, although in the context of JECFA evaluations the NOAEL and NOEL are interchangeable, because the term "adverse" is included in the definition of NOEL (WHO, 1987). The NOAEL is a level of exposure in which the treated animals do not differ significantly, in a biological and/or statistical sense, from untreated control animals in measurements related to the adverse effect(s) recognised at higher doses. The NOAEL is a dose without measurable activity, and therefore in practice it is considered to be close to the threshold in animals. Although the dose–response relationship above the NOAEL is not used quantitatively, it contributes to an assessment of whether the effects detected at different dose levels are likely to be treatment related.

The lowest-observed-adverse-effect level (LOAEL) is used instead of the NOAEL for contaminants when all
test groups produce a significant effect compared to controls. In consequence, this estimate is clearly above the threshold, and in risk assessment this is usually taken into account by the use of additional uncertainty factors (see below). The LOAEL, like the NOAEL, is an experimental observation and dependent on the design of the study. The benchmark dose (see section 7) would represent a more scientifically credible way of dealing with risk assessment for databases that do not allow for determination of a NOAEL.

The uncertainty factor that is most commonly applied to a surrogate threshold dose (such as the NOAEL or benchmark dose) to derive a safe level for human exposure (such as the ADI) for non-genotoxic chemicals is 100, which was introduced some 40 years ago by Lehman and Fitzhugh (Lehman and Fitzhugh, 1954). The 100-fold factor is considered to be the product of two 10-fold factors; one to allow for the possibility that the average human is 10-fold more sensitive to a toxicant than the average test species (interspecies factor) and the second (interindividual, intraspecies or human variability factor) to allow the average human to differ from the most sensitive human by a factor of 10 (WHO, 1987, and section 5). The use of default uncertainty factors to allow for interspecies differences and human variability are designed to move the dose–response curve from the response for a group of experimental animals down to the curve for sensitive humans. The 10-fold factors are multiplied on the assumption that these factors are independent and that for some compounds the test species and humans will vary by a factor of 10, and a factor of 10-fold will separate average and sensitive humans. The interspecies 10-fold factor is considered to move the dose-response curve from the exposure from a NOAEL for the typical (average) individual to a NOAEL for the sensitive individual. In dose–response terms it is to allow for differences in the position of the dose–response curve for the sensitive individual, compared with the population mean.

Only a few controlled studies have been conducted in humans for risk assessment purposes in which a response has been produced deliberately, for obvious ethical reasons; clinical trials of this nature would only be given ethical approval when the biological processes underlying the toxicology in the sensitive test species are fully understood and the measured endpoints both mild and reversible. An example of using human response data to derive an ADI is the iodine-containing food colour erythrosine (FD&C Red No. 3). The critical toxic effect after chronic feeding in the rat was thyroid tumours and several experiments have shown that this is due to a non-genotoxic, secondary event on the feedback mechanism involved in the secretion of thyroid stimulating hormone (TSH). A controlled study conducted in humans showed that the most sensitive biomarker of an adverse effect on the thyroid was changes in serum TSH. The NOAEL for this response was 60 mg/day (about 1 mg/kg body weight/day), and this was divided by a factor of 10 to allow for human variability, generating an ADI of 0.1 mg/kg/day (WHO, 1991).

3.2. Subpopulations

The factor of 10 for human variability can be considered to allow for differences within the human population, and includes "at risk" groups such as children and the elderly. The analysis of human variability by Renwick and Lazarus (1998) demonstrated that the 10-fold factor was an adequate default, but that situations could be envisaged in which the compound might show metabolic characteristics that would greatly increase human variability (for example polymorphisms in xenobiotic metabolism, such as CYP2D6, or polymorphisms in cytoprotective pathways, such as G6PD deficiency).

The concept that infants and children may be a sensitive subgroup relates to their relative immaturity compared with adults (Renwick et al., 2000). Such immaturity could affect principally their ability to eliminate the compound (toxicokinetics) or their target organ response (toxicodynamics). For obvious reasons there are few in vivo response data available in humans that have investigated this. An analysis of published literature (kinetic and dynamic) showed that the use of an additional 10-fold factor for infants and children (as required under the Food Quality Protection Act in the USA) does not appear to be justifiable based on considerations of the usual 100-fold factor used to allow for species differences and human variability (Renwick et al., 2000). However, an additional factor could be appropriate if reproductive and developmental toxicity studies were not available, or if the current testing methods were considered inadequate, or if the effects produced in the neonate and young were essentially irreversible so that a "severity" factor is introduced for risk management.

It is important that the rationale underlying an extra factor for potentially sensitive subgroups is based clearly on scientific principles and understanding.

3.3. Data requirements

There are established guidelines for the extent and design of toxicity studies necessary for the approval of a
chemical, such as a food additive or a pesticide (EC, 1980, 1989; FDA, 1982, 1993; WHO, 1987; OECD, 1993). The adequacy of the database can affect the selection of appropriate uncertainty factors (WHO, 1994). For contaminants where a risk assessment may have to be undertaken on a non-ideal database, additional uncertainty factors may be used, in order to allow for deficiencies in the database, (Dourson et al., 1992; Beck et al., 1993; WHO, 1994; Vermeire et al., 1999) such as the absence of a NOAEL, or of a chronic (long-term) study in animals (see section 8 for a probabilistic analysis of these factors).

3.4. Strengths, limitations and weaknesses

The main strength of this approach is that it is simple and pragmatic, and can be applied readily to a wide range of databases.

However, there are limitations in each of the numerical values that are the basis of the calculation of an ADI, TDI or RfD, that is the surrogate for the threshold (usually the NOAEL), and the uncertainty or safety factor that is applied to the surrogate for the threshold.

The main problem with the use of the NOAEL is that it may be below, at, or above the threshold, depending on the following characteristics of the study design.

Group size — the larger the group size, the greater will be the sensitivity. The group sizes currently recommended in testing guidelines represent the best compromise between sensitivity and practicability.

Sensitivity of measurement of the adverse effect — the more sensitive the method of detection, the lower will be the NOAEL. Because poor or inadequate methods would result in a higher NOAEL, studies which are submitted for regulatory purposes should comply with GLP or be reported in sufficient detail to provide quality assurance to those undertaking the risk assessment.

Dose spacing — a determinant of the NOAEL in real databases is the selection of the dose levels given to the animals, and the magnitude of effect at the LOAEL. If doses are widely spaced in relation to the slope of the dose–response curve, and the dose above the NOAEL produces only a slight effect, then the NOAEL will probably underdose the threshold. However, if the doses are more closely spaced then the NOAEL may be at or above the threshold.

A problem arises when the top dose studied is the only dose showing an adverse effect, and it is not clear whether this is a true response or a random "false positive" finding. Under such circumstances using the dose below the top dose as a NOAEL will result in a conservative assessment.

The NOAEL approach does not provide any information about risk above the ADI. Such information can be provided by dose–response modelling. The risk assessment process would have a more scientific basis if the simple use of the NOAEL were to be replaced, when possible, by an approach that utilised more of the dose–response data and gave a more robust measure, such as the BMD. The use of a benchmark dose means that the position or existence of a threshold is irrelevant, because the benchmark dose can be set at a level of response that is considered to be non-adverse.

The rationale for the value of 100 is that it represents a 10-fold factor for interspecies differences and a 10-fold factor for human variability (WHO, 1987). A number of recent reviews have assessed the validity of this 100-fold value as a general default, which is applied to widely different compounds that may show a range of toxicokinetic or toxicodynamic properties (Dourson and Stara, 1983; Calabrese, 1985; Hattis et al., 1987; Lewis et al., 1990; Sheehan and Gaylor, 1990; Renwick, 1991; Calabrese et al., 1992; Naumann and Weideman, 1995; Dourson et al., 1996; Renwick and Lazarus, 1998). These reviews have been post hoc analyses of the validity of the "uncertainty factors" which were selected in the 1950s, before recent advances in the fields of toxicology and risk assessment. The reviews have identified various situations where they may be either inadequate (for example when the metabolism of the compound shows a genetic polymorphism), or excessive (for example when the compound undergoes negligible absorption from the gut). In consequence, the risk assessment process would have a more scientific basis if the simple default factor were replaced, when possible, by a value that was related more specifically to the database under evaluation, such as the use of a chemical-specific adjustment factor (see section 5).

5. Applicability

The use of 100-fold and 10-fold uncertainty factors applied to the NOAEL is the standard approach adopted by all agencies undertaking risk assessment (safety assurance) on chemicals considered to show threshold effects. The approach is readily applicable and there are established guidelines and procedures for its implementation (WHO, 1987).

3.6. Acceptability to regulatory agencies and authorities

The NOAEL expressed on a body weight basis (e.g. mg/kg body weight/day) is divided by an uncertainty factor (or safety factor) to derive the level of human exposure that will be without significant adverse effects. Although the terminology differs between regulatory bodies (NOEL vs NOAEL; acceptable daily intake (ADI) vs tolerable daily intake (TDI) vs reference dose (RfD — used in the USA); safety factor vs uncertainty factor), there is a common underlying approach. The
The 100-fold factor has been used internationally for 40 years to calculate an ADI or TDI, which is applied to the NOAEL from animal studies to estimate the acceptable daily intake (ADI) or provisional tolerable weekly intake (PTWI) for chemicals that accumulate. The usual default uncertainty factor of 100, which is introduced in the late 1950s by the FDA and subsequently adopted by the JECFA, is related to the potential for accumulation, for example, acceptable daily intake (ADI) or provisional tolerable weekly intake (PTWI) for chemicals that accumulate. The usual default uncertainty factor of 100, which is applied to the NOAEL from animal studies to calculate an ADI or TDI, was introduced in the late 1950s by the FDA and subsequently adopted by the JECFA. The 100-fold factor has been used internationally for 40 years, and is accepted as a reasonable default value.

7. Gaps and research needs

Available data on toxicokinetics, toxicodynamics and in vivo response in animals and humans should continue to be studied in order to investigate the adequacy and limitations of the inter- and intraspecies default uncertainty factors. The use of default uncertainty factors can be improved by a probabilistic application of uncertainty factors (see section 8), and/or the development of pathway- and process-related categorical default factors (see section 5).

The percentage of a population that would be covered by the use of a default factor is an important risk management issue. Estimates may be made using the standard deviation, or the geometric standard deviation, combined with a population distribution model (Renwick and Lazarus, 1998). The development of this approach using chemical-specific data or categorical defaults will allow a quantitative assessment of the adequacy of the uncertainty factors on a case-by-case basis.

4. Categorical regression

4.1. Models and methods

A method that has been proposed for quantitative dose–response analysis for non-cancer toxicity data is that of categorical regression. Categorical regression is a statistical tool that can be used to estimate potential health risk from chemical exposures. Using regression of ordered categories of toxic severity, the method can estimate the likelihood that a given category of severity may occur at a given dose level. Of course one may not be able to directly observe the effect, but one may estimate the probability that it could occur.

Categorical, or ordinal, regression involves statistical regression on the experimental doses associated with various severity categories of overall toxicity (Hertzberg and Miller, 1985; Hertzberg, 1991; Hertzberg and Wymer, 1991). Typically, these severity categories may be assigned to a scheme qualitatively defining different levels of effect (e.g. no-effect level, no-adverse-effect level, low-adverse-effect level and frank-effect level as the most severe level, see Table 1). Using this type of severity scheme, doses and their associated effects from different studies may be combined to predict effect severity. In this way, all categorised adverse effects may be taken into account rather than focusing on the critical effect only. This method allows for dichotomous, continuous and descriptive data within a single system of severity scoring. Furthermore, in some cases, this mathematical approach can utilise both concentration (dose) and duration of exposure as independent variables that determine response. Thus, toxicity data

<table>
<thead>
<tr>
<th>Ranking</th>
<th>Toxicological effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Enzyme induction or other biochemical change with no pathologic changes and no change in organ weight</td>
</tr>
<tr>
<td>2</td>
<td>Enzyme induction and subcellular proliferation or other changes in organelles, but no other apparent effects</td>
</tr>
<tr>
<td>3</td>
<td>Hyperplasia, hypertrophy or atrophy, but no change in organ weights</td>
</tr>
<tr>
<td>4</td>
<td>Hyperplasia, hypertrophy or atrophy with changes in organ weights</td>
</tr>
<tr>
<td>5</td>
<td>Reversible cellular changes: cloudy swelling, hydropic change, or fatty changes</td>
</tr>
<tr>
<td>6</td>
<td>Necrosis, or metaplasia with no apparent decrement of organ function. Any neuropathy without apparent behavioral, sensory or physiologic changes</td>
</tr>
<tr>
<td>7</td>
<td>Necrosis, atrophy, hypertrophy or metaplasia with a detectable decrement of organ functions. Any neuropathy with a measurable change in behavioral, sensory or physiologic activity</td>
</tr>
<tr>
<td>8</td>
<td>Necrosis, atrophy, hypertrophy or metaplasia with definitive organ dysfunction. Any neuropathy with gross changes in behavior, sensory or motor performance. Any decrease in reproductive capacity. Any evidence of fetotoxicity</td>
</tr>
<tr>
<td>9</td>
<td>Pronounced pathologic changes with severe organ dysfunction. Any neuropathy with loss of behavioral or motor control or loss of sensory ability. Reproductive dysfunction. Any teratogenic effect with maternal toxicity</td>
</tr>
<tr>
<td>10</td>
<td>Death or pronounced life shortening. Any teratogenic effect without signs of maternal toxicity</td>
</tr>
</tbody>
</table>

Table 1

<table>
<thead>
<tr>
<th>Toxicological severity scoring of NOAELs, LOAELs and FELs for use in categorical regression analysisa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category</td>
</tr>
<tr>
<td>NOE</td>
</tr>
<tr>
<td>NOAE</td>
</tr>
<tr>
<td>LOAE</td>
</tr>
<tr>
<td>AE</td>
</tr>
<tr>
<td>FE</td>
</tr>
</tbody>
</table>

* Adapted from DeRosa et al. (1985).
* NOE = no-observed-effect; NOAE = no-observed-adverse-effect; LOAE = low-observed-adverse-effect; AE = adverse effect; FE = frank effect. 
* These risk categories are approximate and may vary depending on the actual data used and experience of the toxicologist ranking the effects.
Dourson et al. (1997) constructed a case study using the carbamate insecticide aldicarb. The case study illustrated how categorical regression can be used to assess risks above the RID. The experimental animal data supported the observations seen in the human studies. Final analysis was conducted on the two data sets for two different human exposure studies that measured blood cholinesterase inhibition. These data suggested that a maximum likelihood risk estimate of adverse effects is 0.008% at a 10-fold higher dose than the RID when blood cholinesterase inhibition is not considered as an adverse effect. When blood cholinesterase inhibition of 20% or more is considered as an adverse effect, a maximum likelihood risk estimate of adverse effects is 0.1% at a dose 10-fold higher than the RID. Although the application of this categorical regression was noted as model dependent when estimating risk above the RID, such a mathematical approach may have value as an adjunct to accepted methods of risk assessment.

One other promising application of categorical regression is to describe the exposure-response relationship for a chemical and to help the risk assessor to determine the most appropriate data set and model for...
risk assessment. Guth et al. (1997) demonstrated that by analysing the exposure response to inhaled tetrachloroethylene, using both concentration and duration as determinants of exposures, stratified categorical regression could be used to assess risk. A generalized linear model for ordinal data was used to estimate the probability of response for effects of various severities and exposures. This is useful for accidental exposure situations, including concentration and duration that correspond to severity of response ranging from mild or transient effects through adverse or disabling effects to severe or life-threatening effects. Using the results from 12 different studies in humans and rodents, the authors were able to assign severity categories to the data (endpoints ranged from general central nervous system (CNS) impairment through anaesthesia to lethality in rodents and functional disturbances, dizziness and sleepiness in humans) and constructed a categorical regression for the exposure response function for acute inhalation exposure. The model of choice was a stratified common-slope categorical regression model that can be used to set human exposure limits for acute exposure to tetrachloroethylene for durations up to 12 h. A stratified regression model is a form of meta-analysis that generalizes the scope of the categorical regression model by allowing for subgroup differences. Thus, a risk assessor now has a method of estimating risk by severity of effect for both concentration and time along a continuum over this acute exposure period. Similar types of analyses could be performed for commonly ingested food contaminants (e.g. pesticides).

4.2. Subpopulations

One advantage of the categorical regression approach is that data can be combined from multiple studies, including studies in different species. Although the model will allow estimation of risk, it does not necessarily address the issue of sensitive subpopulations, although categorical regression could be used to address sensitive subpopulations if data are available from those subpopulations.

4.3. Data requirements

A homogenous categorical regression model assumes that all experimental groups were treated under the same experimental conditions aside from the variables (dose and duration). When parameters are different, the heterogeneous nature of the data can be accounted for the modelling software by stratification and will average over the random study effects. Simpson et al. (1996) modified the approach to allow the combination of continuous and dichotomous endpoints, by assigning group-level severity scores, and making some assumptions about the correlation of the individual responses within such groups.

Both continuous and quantal data may be used in a categorical regression analysis; however, they require different considerations for determination and assignment to the severity categories. Quantal data require a decision by the toxicologist to assignment of the category based on the nature of the endpoint. When considering continuous data, the response magnitude must be determined for each severity category and exceeding that response results in classification into the corresponding severity category. Categorical effect measures are often found in the reported toxicological literature, such as descriptions of histopathology findings or hematochemical or enzyme function changes. Other data that do not necessarily fit this definition may still be used if they can be assigned a severity category and analysed in combination with other data that has been appropriately assorted by severity categorisation.

4.4. Strengths, limitations and weaknesses

One strength of the method is that it incorporates all relevant studies and toxicological judgement into the estimation of a starting point for risk assessment, for example a NOAEL. In addition, it allows for quantifying the uncertainty associated with that estimate. By combining multiple studies of limited individual use for risk assessment purposes, the risk assessor can dramatically increase the statistical power of the analysis. This method allows the risk assessor to use multiple independent variables to explain a given response, to predict exposures related to various levels of effect severity and to combine multiple studies. Thus, categorical regression allows the user to predict the effect category in the subject for any particular dose (or, when time is included in the model, for any combination of dose and exposure duration), by interpolation or extrapolation of the fitted model.

An example of this might be predicting the effect for a bracketed time interval or concentration for which there are no measurements. The model also allows for the possibility to predict exposures at related responses for different severities, allowing the risk assessor to make judgements to protect a population from increasingly severe effects.

Categorical regression can also provide information about potential toxicity at doses exceeding the risk assessment dose such as the ADI, TDI or RfD. The NOAEL approach is limited in that it is focused on the determination of "safe" or "acceptable" intake levels. In situations where exposures may exceed these levels, however, information is needed to help determine the urgency of a situation (e.g. in the case of contaminated drinking water, at what point should public utilities shut off supplies?). Risk assessment numbers such as the
Note that these assessments employed animal studies in benchmark calculated using categorical regression. Ods) for deriving a safe acute exposure level from a uncertainty factor of 30 (10 for human variability and regulatory context. These assessments typically use an although these assessments have not yet been used in a particularly across temporal intervals. The assignment of an experimental dose and the resultant effects to severity categories is not standardised and needs some type of harmonisation of criteria. A major problem occurs when the regression includes different types of adverse effects. A ranking of different effects would be highly subjective, for instance when combining effects of reproductive toxicity nature with neurological or renal effects, and would involve subjective and non-scientific judgements.

Other limitations in using this method is that a goodness-of-fit approach can be misleading; for example, accuracy is not confirmed in the extrapolation and may not have sufficient power to differentiate or reject models. Also, confidence bands can hide key assumptions especially when combining data across species and across temporal intervals or both.

4.5. Applicability

The application of this method provides an estimate on the continuum of severity for effects of the same type or in the same organ system. However, combining severity rankings for different effects will diminish the utility of this model.

Use of categorical regression requires expertise in toxicological risk assessment and statistics. Few risk assessments to date have employed this method. Categorical regression has been used to assess data from both acute inhalation of tetrachloroethylene (Guth et al., 1997) and oral exposure to aldicarb (Dourson et al., 1997). A number of other case studies have been conducted for acute inhalation regulatory exposure, although these assessments have not yet been used in a regulatory context. These assessments typically use an uncertainty factor of 30 (10 for human variability and 3 for interspecies extrapolation using dosimetric methods) for deriving a safe acute exposure level from a benchmark calculated using categorical regression. Note that these assessments employed animal studies in the risk assessment and thus needed to consider the animal-to-human extrapolation (interspecies) uncertainty factor.

The use of this mathematical model requires specific software designed for categorical regression. Such a software package has been developed by the EPA as a pilot program to assess and investigate the feasibility of using categorical regression to assess acute inhalation toxicity. This has resulted in the publication of newly developed software to assess acute inhalation toxicity in a program called CatReg (EPA, 2000b,c). Application of this software is being used by risk assessors and presented along with other accepted methods by the EPA to demonstrate the feasibility of using this software to develop risk assessments (EPA, 2000b,c).

4.6. Acceptability to regulatory agencies and authorities

As with all new approaches to risk assessment, there needs to be an appropriate time period to work the model and see where the pitfalls are before applying guidelines for use. This mathematical approach has great promise but still requires that the issues (as outlined above) be sufficiently answered before regulatory agencies are comfortable with its use.

4.7. Gaps and research needs

Use of this mathematical model has been limited to date, and although promising, there are still a number of issues that need to be addressed either by setting guidelines for use or by conducting research to further investigate and validate the model.

(i) Can the method address sensitive populations?
(ii) What guidelines should be used for assigning severity and assigning severity categories?
(iii) What guidelines should be used for combining studies? For excluding studies? For accepting or rejecting specific endpoints?
(iv) What guidelines should be applied for using this model in lieu of other accepted risk assessment methodologies?
(v) Can a policy or guideline be instituted to interpret and accept the conclusions from the categorical regression model?
(vi) How should the slope of the fitted model be interpreted? Is it noise, or variation between studies, or anything else?

Additionally, given a data set with fewer studies, rather than a larger array of studies, may not guarantee that a categorical regression will be feasible given biological variation, different criteria for measuring endpoints, and potentially insufficient power to construct the model.
5. Chemical-specific adjustment factors

5.1. Models and methods

The specific quantitative model which is adopted for the risk assessment of a chemical present in food is dependent on the extent of scientific data which describes both the toxicity and the biological processes which underlie the final response. For chemicals which are rich in data this allows the relatively simplistic default approaches using the 100-fold factor as a product of two 10-fold factors (see section 3) to be replaced by more scientifically-credible, quantitative, dose-response analyses, using chemical-specific data.

Although a dose–response analysis from animal toxicity data is more common, a framework has been developed which allows the incorporation of quantitative data describing the different biological processes that contribute to a response (Renwick, 1993; WHO, 1994). In this model, each interspecies and inter-individual 10-fold factor is subdivided to allow for differences in toxicokinetics (which determines the delivery of the chemical to the target site) and toxicodynamics (which determines the reaction of the target site to the presence of the chemical) (Fig. 3). Each of the separate subfactors is to allow for one specific source of uncertainty (Fig. 3; Renwick, 1993). The International Programme on Chemical Safety (IPCS) (WHO, 1994) supported the subdivision of the 100-fold factors with the human variability factor given equal weighting for toxicokinetics and toxicodynamics (Fig. 3), which is in agreement with the recent analysis of Renwick and Lazarus (Renwick and Lazarus, 1998). The framework (Fig. 3) allows the replacement of any of the four default subfactors if chemical-specific data are available.

The product of the CSAFs (See the summary of the IPCS meeting in Ottawa, 2000 in http://www.ipcharmonize.org/CSAFsummary.htm) with the remaining default factors would give a composite chemical-specific adjustment/uncertainty (previously called a data-derived uncertainty factor).

For interspecies differences in toxicodynamics or toxicokinetics, the difference in the mean value for the appropriate parameter (see section 5.3) in the test species compared to the mean of the same parameter in humans can be used to estimate the magnitude of difference, and replace the interspecies defaults.

Human variability in a toxicokinetic or toxicodynamic parameter is best viewed as a distribution, with the proportion of the population to be covered by the CSAF determined by the variability for a particular parameter measured. A CSAF for human variability would be calculated as the ratio between the mean or median for the population and the parameter value corresponding to a pre-defined and accepted proportion of the population covered, such as the 95th, 97.5th or 99th percentile of the population (Fig. 4). The nature of the population distribution of the relevant kinetic or dynamic parameter could be unimodal, bimodal, normal, log-normal, skewed, etc. In most cases the nature of the population distribution will not be known, and a pragmatic default nature of the distribution, such as unimodal and log-normal, will usually be necessary. An analysis by Renwick and Lazarus (1998) used this approach to determine the proportion of the population not covered by the default factors. Using pharmacokinetic data (e.g. area under the plasma concentration-

![Fig. 3. The subdivision of the usual 100-fold uncertainty factor into toxicokinetics and toxicodynamic aspects (based on WHO, 1999b).](http://www.ipcharmonize.org/CSAFsummary.htm)

---

**AD** - Animal to human dynamic uncertainty factor
**AK** - Animal to human kinetic uncertainty factor
**HD** - Human variability dynamic uncertainty factor
**HK** - Human variability kinetic uncertainty factor

Chemical specific data can be used to replace a default uncertainty factor (UF) by an adjustment factor (AF)

![Fig. 4. Development of CSAFs using chemical-specific data. The CSAF would be determined by the ratio of the mean and the parameter value for a predefined proportion of the population e.g. 95th percentile (adapted from Silverman et al., 1999).](http://www.ipcharmonize.org/CSAFsummary.htm)
time curve) for 60 compounds and pharmacodynamic data (e.g. in vitro response data) for 49 compound-related effects they showed that the default factor for kinetics covered 99.9% of a healthy adult population assuming a normal distribution and 99% assuming a log-normal distribution, with 99.7 and 98% covered for dynamics assuming normal and log-normal distributions. An advantage of using population distributions is that data for subgroups can be analysed either separately (Renwick and Lazarus, 1998), or by modification of the population distribution of the appropriate parameter estimate.

The use and development of CSAFs using chemical-specific kinetic or dynamic data has been discussed in relation to a number of cases at the time the concept was developed (see Kroes et al., 1993). It has also been discussed for boron, an element found naturally in both water and food (Dourson et al., 1998). An analysis of the published toxicokinetic and toxicodynamic parameters in humans and the rat by Dourson et al. (1998) showed that an interspecies default of 10-fold (4-fold for kinetics and 2.5-fold for dynamics) was recommended since the systemic clearance of boron is 3-4-fold higher in rats than in humans and data on dynamic differences (for modifying the dynamic default subfactor) between species were unavailable. For refining the inter-individual 10-fold default, a factor of 6-fold (1.8-fold for kinetics and 3.1 for dynamics) rather than the default was estimated based on the variability in glomerular filtration rate in pregnant women (the route by which boron is cleared from the systemic circulation) and the lack of data on dynamics, for which the default value was used (Dourson et al., 1998). By combining the chemical-specific toxicokinetic values with the remaining toxicodynamic default subfactors the composite chemical-specific adjustment or uncertainty factor (or data-derived uncertainty factor) for boron was 60 \( [(4 \times 2.5) \times (1.8 \times 3.1)] \).

5.2. Subpopulations

In terms of human variability in the total population, the magnitude of a chemical-specific subfactor for toxicokinetics or toxicodynamics is dependent on the proportion of the population that needs to be covered (a risk management decision), the presence of potentially "at risk" subgroup(s) and the variability within the main population and the subgroup, for the parameter considered. There are two approaches by which an adjustment factor for a subgroup(s) could be calculated based on population distribution analysis. Silverman et al. (1999) proposed that the adjustment factor for a sensitive subgroup(s) should be based on a comparison of the 95th percentile for a subgroup with the 50th percentile of the general population (Fig. 5). The percentage of the total population covered would depend on the incidence of the sensitive subpopulation. An alternative approach would be to develop an adjustment factor that covers a predefined proportion of the whole population [the healthy adult and subgroup(s) distributions]. Using computer software capable of randomly selecting data points (e.g. Monte Carlo), the distribution for each subgroup could be combined to give one model representative of the likely mean and variability for a parameter in the whole population. This analysis would take into account differences between the main population and the subgroup(s) for the mean parameter estimates and the coefficient of variation (Renwick and Lazarus, 1998) and could also include the incidence of the subgroup(s) in the total population.

The questions that would have to be addressed by a risk manager when considering models which could incorporate subgroups for which chemical-specific data is available are:

(i) Should a constant proportion of the subgroup be covered (e.g. 90, 95, 99%, etc.)?
(ii) Should the incidence of the subgroup be taken into account using the demographic data for a population?
(iii) Would it be logical to have a factor, which covered 90% of a rare subgroup (e.g. 1 in 10,000) because this would cover 99.9999% of the whole population?
(iv) Would it be logical to have a factor, which covered 99% of a whole population including subgroups?

In reality, it is unlikely that there will be toxicokinetic or toxicodynamic data available for a susceptible sub-group for a chemical present in food, and a risk assessment for a particular "at risk" group may have to be based on subgroup data for a different compound for which the route of elimination (toxicokinetics) or the mechanism of action (toxicodynamics) is the same as the chemical under assessment (see gaps and research needs at the end of this section).
5.2.1 Infants and children

A recent review of the differences in kinetics between adults and children concluded that children, but not infants, frequently eliminate drugs and foreign compounds more rapidly by metabolism and excretion, compared with adults (Renwick, 1998). In consequence, a smaller proportion of a population of children may be at risk compared with adults, although this would have to be assessed on a case-by-case basis. However, the immaturity of hepatic metabolism and low clearance in neonates suggests that they represent a potentially vulnerable subgroup (Dorne et al., 2001).

5.2.2 Ethnic differences

Ethnic differences can arise from genetic, dietary or environmental factors and could result in differences in kinetics or response or both. In many cases, differences in mean kinetic parameter estimates between different ethnic groups are small and ethnicity would not influence the validity of the default factor of 3.16 (Renwick and Lazarus, 1998). However, in cases where a different ethnic group showed a decrease in clearance, combined with an increase in variability, for example for the therapeutic drugs desipramine, diazepam, methylprednisolone and nifedipine, it is possible that small groups of the healthy adult population will have a distribution different to that of the majority. Ethnicity should be considered for some P450-mediated oxidation reactions, although this would need to be on a chemical-specific basis.

5.2.3 Polymorphic metabolism

Genetically determined differences are of greatest relevance to risk assessment when the polymorphic pathway is the major route of elimination for a specific chemical. The knowledge that a specific chemical is a substrate for a metabolic pathway that shows a polymorphic phenotype may raise questions about the validity of the 3.16-fold default uncertainty factor for kinetics. If chemical-specific data showed a polymorphic distribution, then chemical-specific adjustment still applies.

5.3 Data requirements

The types and quality of data that would be appropriate for the generation of a CSAF have been the subject of review. In May 2000, a major workshop organised by the IPCS on Human Variability and Uncertainty in Risk Assessment (Berlin, Germany) used case studies to examine the different types, and quality, of data needed to replace the defaults for interspecies and interindividual differences in toxicokinetics and toxicodynamics. This workshop produced a guidance document to assist risk assessors in the use of experimental data in deriving CSAFs (http://www.ipcsharmonize.org/CSAFsummary.htm).

5.3.1 Toxicokinetics

The absorption, distribution, metabolism and excretion of a chemical in humans and animals are processes that contribute to the levels of a chemical at the active site, although in isolation such data could not be used to replace the toxicokinetic defaults. Data that could be used to refine the interspecies or interindividual toxicokinetic default should be derived from in vivo toxicokinetics studies (using the most relevant route of exposure), or in vitro measurement of the process involved in the elimination (e.g. hepatic metabolism to give an estimate of the intrinsic clearance) combined with a PBTK model, or in some cases from in vivo measurements after environmental or occupational exposures.

For toxicokinetics, it is necessary to know whether the critical effect observed is related to the maximum concentration (Cmax), or the overall total exposure (area under the plasma concentration–time curve, AUC or clearance), of the toxicant at the target site. Using in vivo data the relevant chemical-specific data for acute toxicity could be based on either AUC (1/clearance) or Cmax. The AUC is a predictor of toxicity during subchronic or chronic treatment (e.g. during steady-state kinetics). The main advantage of using in vivo physiologically-based parameters (such as clearance) is that all in vivo processes are included in the estimate, including the contribution of extrahepatic tissues. Such studies are most appropriate when the parent compound is active. If the toxicity is due to a metabolite formed in one tissue and acting in another, then the plasma toxicokinetics should relate to the active metabolite.

Alternatively, PBTK models can be used to develop measurements of the target organ dose based on animal data and appropriate interspecies scaling (see section 9). PBTK models are also appropriate when they include parameters for the conversion of the parent compound into an active metabolite, and for route-to-route extrapolation of the internal dose. PBTK analysis has been applied largely to take interspecies differences into account for low-dose extrapolation of animal data, and could be used to replace the kinetic component of the interspecies uncertainty factor.

5.3.2 Toxicodynamics

The interspecies and interindividual toxicodynamic default subfactors could theoretically be replaced using data from any step ranging from the interaction of a toxicant with its molecular target (e.g. using an in vitro receptor binding assay) up to the final toxic response. For interspecies and interindividual variability in response these data would have to come from an in vivo or in vitro (using animal and human tissue) study for the final toxic effect itself or for an intermediate in the chain of events leading to the final endpoint, where variability due to toxicokinetics has been excluded.
use data from a surrogate endpoint it is important that the surrogate is related to the mode of action and critical to the dose-response relationship of the final response, and this would have to be characterised in the critical test species. To replace the interspecies dynamic default with a CSAF, the data would have to be calculated using equipotent doses (such as the ED₁₀ for a continuous variable or the ED₅₀ for a quantal response), in animals and humans for the same endpoint, and not calculated from differences in response to the same dose. Information from in vivo response studies in humans would describe variability in both toxicodynamics and toxicokinetics, and a safe exposure value could then be derived from a NOAEL in humans using the 10-fold uncertainty factor for human variability, without a need to calculate CSAFs.

5.4. Strengths, limitations and weaknesses

By dividing both the interspecies and interindividual 10-fold factors into toxicokinetic and toxicodynamic subfactors, a general framework has been provided that allows for the use of chemical-specific data related to certain aspects of the risk assessment process. When such data are available to replace the default adjustment subfactors, this will obviously result in a more scientific risk estimate for that chemical. In the absence of suitable data the product of the two subfactors gives the usual default of 10.

However, the chemical-specific adjustment factor approach also has some weaknesses; it is now recognised that only a few databases are currently available which contain the appropriate information to allow the replacement of any of the four default subfactors with chemical-specific data related to the interspecies toxicokinetic aspect of uncertainty. Any data available on response or kinetics in humans (to replace the interindividual kinetics and/or dynamic defaults) will normally use a relatively small number of subjects and therefore the nature of the population distribution (e.g. normal, log-normal, bimodal or skewed) will not be known.

5.5. Applicability

The use of CSAFs requires supplementary studies relating to toxicokinetics and/or toxicodynamics in the test species and in humans, at toxicologically relevant doses. This method is a refinement of the standard 100-fold default (10 x 10) that would be used in the absence of chemical-specific data, so that it is compatible with the current default approach (section 3). Hence, the ability of a risk assessor to utilise this approach depends on the availability and quality of the data for a chemical.

5.6. Acceptability to regulatory agencies and authorities

Guidance on CSAFs for use by regulatory agencies is being developed in the context of one of the projects of the initiative of the International Programme Chemical Safety (IPCS) on Harmonisation of Approaches to the Assessment of Risk from Exposure to Chemicals. The principal objective of this project is to define the criteria for the adequacy of chemical-specific data to quantitative interspecies differences and human variability in toxicokinetics and toxicodynamics. As a result of the IPCS guidance documents there has been a gradual acceptance of the CSAF approach; the concept of the subdivision of uncertainty factors into kinetics and dynamics was adopted by the SCF for the risk assessment of cyclamate and dioxins.

5.7. Gaps and research needs

For the majority of chemicals, there is an absence of toxicokinetic or toxicodynamic data that can be used to make an informed decision about the magnitude of CSAFs. However, in many cases the metabolic fate of the chemical maybe known, or could be determined easily, in animals and/or humans, from in vivo or in vitro studies or both. Although such metabolic data currently contribute qualitatively to the risk assessment process, it has not been used in a quantitative manner. The development of different default uncertainty factors for different metabolic fates in the different test species and humans would allow the usual default subfactors to be refined to “species- and/or pathway-related defaults” (Fig. 6). These defaults could then be applied, as appropriate, to compounds for which metabolic but not
kinetic data are available. This refinement would bridge the gap between the use of chemical-specific data and the kinetic and dynamic defaults currently available (Renwick and Lazarus, 1998; Renwick et al., 2001; Silverman et al., 1999; Dorne et al., 2001; Walton et al., 2001a,b).

It is envisaged that species- and pathway-related defaults for the different routes of elimination (toxicokinetics) or mechanisms of action (toxicodynamics), in animals and humans, will provide an additional level in the regulatory framework (Fig. 6), allowing supplementary data (such as in vitro metabolism studies) on a chemical to inform the regulatory process.

6. Non-threshold methods

For some hazards, such as those involving genotoxicity, it is considered that there may be no threshold for the mode of action, and therefore a level of exposure without significant adverse effects cannot be determined. In such cases, estimates are made of the possible magnitude of the risk (usually incidence) at human exposures (quantitative risk assessment; dose–response extrapolation).

In order to provide a quantitative risk estimate for non-threshold effects, the incidence data in the experimental range has to be extrapolated to low-dose and low-risk levels. Quantitative risk assessment can produce either the exposure associated with a particular level of risk, or the risk associated with a particular level of exposure. Attempts to provide estimates of exposure associated with risks in the region of 1 in 10^6 have been made for the past 40 years and a number of mathematical models of increasing sophistication have been developed.

6.1. Models and methods

Quantitative risk assessment for non-threshold effects (e.g. cancer) usually uses the dose response for the incidence data from the animal study to estimate the risk at levels of exposure more relevant to humans. The incidence of the risk detected in an in vivo animal study would normally be greater than 1 in 20, but a "virtually safe dose" for a genotoxic compound is usually considered as an incidence of 1 in 100,000 or 1 in 1,000,000 (this is a risk-management decision). Therefore, this approach normally requires extrapolation of the dose–response relationship over at least four orders of magnitude. The model used is the major influence on the estimate derived by quantitative risk assessment, when extrapolation has to be made over three or four orders of magnitude (ECETOC, 1996). The animal dose–response curve in the observed range is not normally extrapolated down to low risk estimates, because of imprecision in the curve-estimate. Consequently, a choice about the appropriate model(s) has to be made. A number of different models have been proposed (Fig. 7), although in practice only a restricted number have been used (ECETOC, 1996).

The most commonly used mathematical model for extrapolation to very low incidences, such as a 10^-6 risk, is the linearised multi-stage (LMS) model. Alternative approaches that are commonly used are to undertake simple linear extrapolation from either (i) a selected incidence within the experimental range, or (ii) an incidence derived by fitting a model to the response data. Simple linear extrapolation requires the selection of a point on the dose–response curve, either within or close to the experimental dose range, that is the starting value for the linear low-dose extrapolation. The starting point on the dose–response curve selected for extrapolation to low doses is either the LOAEL, or a fixed value such as a 10 or 25% response (ED10 or ED25), which may be derived by fitting a mathematical model to the data in the experimental range, or the lower 95th percentile confidence limit on the dose giving such a response. Alternative mathematical models of extrapolating the dose–response data, such as the Weibull model, are not normally used.

Species differences can be taken into account by correcting the dose in the animal studies to a human equivalent dose by interspecies scaling, or by the incorporation of a PBTK model giving the target organ dose of the active chemical species (see section 9). However, despite such refinements, the final value is still determined largely by the mathematical model selected for extrapolation. When species differences in the target organ response in relation to the concentration of the compound have also been investigated, this can also be included to modify the dose–response relationship so that a full biologically-based dose–response model, which includes all toxicokinetic and toxicodynamic
variability into account is the use of the one-sided upper approach that has been claimed to take inter-individual differences into account. An example of this approach is the use of the dose-response relationship for a carcinogen in humans. The low-dose risk estimate derived from this approach is the value that is the "starting point" for extrapolation of the dose-response curve for three or four orders of magnitude outside the range of the experimental observations.

Extrapolation over several orders of magnitude is not always necessary when the risk estimate is based on epidemiology data related to dietary intakes by the general population, because large numbers of subjects are usually studied, and the risk estimate useful for risk managers relates to the risk associated with human exposures close to those that were present in subjects who took part in the epidemiology study. A recent example of the use of the dose–response relationship for a carcinogen in humans was the evaluation by JECFA of aflatoxin (WHO, 1999a,b) (see also Van den Brandt et al., 2002). Aflatoxin is a known animal and human carcinogen affecting primarily the liver in humans. It arises from contamination of the food supply by mycotoxins produced by species of Aspergillus, and there are wide geographical variations in potential intake related to agricultural practices. However, the geographical variations in exposure and liver cancer incidence are confounded by additional variations in hepatitis B and possibly hepatitis C virus, which are also risk factors for liver cancer. There is an extensive database available on aflatoxins, including in vivo animal data and mechanistic studies, epidemiological data and a large number of studies defining the extent of contamination. JECFA used the available epidemiological data to estimate the risk for subjects, with or without hepatitis B surface antigen, that would be associated with two different proposed maximum levels of contamination. JECFA highlighted some of the uncertainties in the data, including (i) only studies showing a positive association between aflatoxin exposure and liver cancer were included, (ii) the shape of the dose–response curve is unknown, (iii) the reliability and precision of the estimates of exposure to aflatoxins are unknown, (iv) current levels of intake were related to current levels of liver cancer, which may be inappropriate for a carcinogen, (v) the earlier studies underestimated the prevalence of hepatitis B in the patients, and (vi) histological confirmation for the cases of primary liver cancer was limited in most of the studies. This analysis of aflatoxin illustrates the types of problems that may be inherent in the interpretation of epidemiological data.

6.2. Subpopulations

Human variability is rarely taken into account in these dose–response extrapolation procedures. An approach that has been claimed to take inter-individual variability into account is the use of the one-sided upper 95th% confidence limit of the dose–response relationship, because the variability in response will be reflected in the variability in the experimental data. However, much of this will relate to variability arising from the small size of the experimental groups, and will relate to variability within the test species and not within the humans.

6.3. Data requirements

The data requirements for extrapolation are directly related to the sophistication of the model to be used. Application of the usual default approach of linear low-dose extrapolation requires only a single dose group with an increased incidence of tumours. It is the general applicability of this simple method that has resulted in its wide adoption. Sufficient groups to define the dose response within the experimental range are necessary for fitting dose–response models, but the usual study design of one control plus three dose groups cannot differentiate between the different mathematical models. In addition an assumption about the shape of the dose response between the bottom of the experimental range and a human risk of 1 in 1,000,000 is always necessary.

6.4. Strengths, limitations and weaknesses

The principal strength of simple linear extrapolation of animal data to low doses, or low risks, is that it is readily applicable, and can even be used with data sets that have only a single dose that shows a positive tumorigenic response. In addition, linear extrapolation represents a conservative method, which is unlikely to underestimate the risk.

Extrapolation of the dose–response curve for three or four orders of magnitude outside the range of the experimental observations is the subject of both assumptions and errors (Lovell and Thomas, 1996). Low-dose extrapolation is based on the assumption that there is a theoretical possibility of an effect with exposure to a single molecule of the substance. Although this proposition was defensible when first introduced about 20 years ago, our increasing understanding of DNA repair mechanisms, and other cytoprotective and homeostatic processes, makes this less plausible, and the dose response may actually be sublinear. However, since there are no methods currently available to assess the dose response at levels of expected human exposure, the linear extrapolation method has been retained as a default approach.

The major determinant of the low-dose risk estimate [e.g. the virtually safe dose (VSD)] using simple linear extrapolation is the value that is the "starting point" for linear extrapolation. The low-dose risk estimate derived by simple extrapolation is not influenced greatly by the shape of the curve within the observed experimental range.
Selection of the mathematical model cannot entirely be based on statistical criteria of the "goodness of fit" to the experimental data, because different models can usually be found to fit the data equally well in the experimental range. Additional criteria used for model selection include biological plausibility, applicability, simplicity, or matters of preserving consistency with prevailing public health policy.

6.5. **Applicability**

Low-dose linearity is the simplest model and can be readily applied to a wide range of experimental data. It simply requires the selection of a point on the dose-response curve as representative of the dose-response relationship, and which can be used as the starting point for linear extrapolation. The starting point may represent the only dose group in a study that showed a positive response. Linear extrapolation from a fixed incidence, such as a \( \text{TD}_{10} \), can be performed using a simple calculator, whereas fitting a model to the experimental data requires the use of an appropriate computer program.

6.6. **Acceptability to regulatory agencies and authorities**

Various approaches have been used internationally, but without a clear consensus and harmonisation. Approaches include the linearised-multistage model, and simple linear extrapolation from a fixed point on the dose-response curve such as the \( \text{TD}_{10} \), \( \text{TD}_{25} \), \( \text{TD}_{10} \), \( \text{TD}_{05} \) or \( \text{LOAEL} \). The main criterion for acceptance is that the model should be applicable to a wide variety of dose-response data; that is, it is a simplistic but pragmatic approach. The LMS model and methods using a simple linear relationship are the default approaches adopted by those agencies that undertake low-dose risk estimation. A number of agencies do not accept the use of low-dose extrapolation and a risk-specific dose is not calculated; in such circumstances the advice to risk managers is to reduce the exposure to as low as reasonably practicable (ALARP), or reasonably achievable (ALARA).

6.7. **Gaps and research needs**

The shape of the dose-response curve at very low doses, and low estimated risks, is not known. The assumption of a linear dose-response relationship between zero dosage and the dose with an observed increase in incidence of adverse effects appears increasingly questionable, based on biological principles, unless the mechanism of action is identical to a process with a background (Crump, 1995). Research is needed on the following aspects in order to provide a better understanding of the dose-response relationship at very low doses:

1. Improvement of study design to provide greater accuracy within the experimental dose-response relationship, for example by optimising the numbers of dose groups and the numbers of animals in different dose groups.
2. Analysis of the dose-response data for tumor precursors (e.g. liver foci, biomarkers, DNA adducts and repair).
3. Toxicokinetic studies as part of, or adjuncts to, chronic carcinogenicity studies. Knowledge of dose-dependency, for example by the development of PBTK models of the target organ dose of the active chemical moiety, can increase understanding of the possible biological effects of low doses. Such studies should take into account factors such as food consumption, growth, aging, etc.
4. Studies using genetically modified animals can provide information on the roles of different biological processes, such as DNA repair or apoptosis. Such studies should define any changes in the dose-response relationship compared with normal animals, and investigate their biological bases.
5. Knowledge of the mechanism of action of particular compounds will be critical to the selection of the appropriate model for extrapolation to very low risk estimates such as the VSD.

7. **Benchmark dose**

7.1. **Models and methods**

An alternative approach to the assessment of non-cancer endpoints is the benchmark dose (BMD) method. The EPA (EPA, 1995) has introduced a benchmark dose level as "a statistical lower confidence limit for a dose that produces a predetermined change in response rate of an adverse effect ... compared to background," denoted hereafter as the benchmark dose lower confidence limit (BMDL) in contrast to a point estimate of the dose that produces the predetermined change denoted as benchmark dose point estimate (BMDP) which is sometimes also used.

Note that the BMD approach, per se, is not low-dose extrapolation. Estimation of exposure levels that are considered acceptable, such as ADIs, is completed, just as in the case of NOAELs, by the use of uncertainty factors applied to the BMDL. The same uncertainty factors apply to NOAELs and to BMDLs. Although the BMDL offers several advantages over the NOAEL because more of the dose response information is used, it can only be used in cases where available data are suitable for modelling. It is not, therefore, a replacement...
for the NOAEL, but should be considered as an additional tool which may offer advantages for some risk assessments. This method, which was first described by Crump (1984) and Dourson et al. (1985), was developed in an attempt to remedy some notable shortcomings of the use of a NOAEL in the default approach described above. For example, the NOAEL is limited by the experimental doses chosen by the investigators in the toxicity studies. The larger the dose spacing, the less accurate the experimental NOAEL (or LOAEL) is likely to be. Also, the shape of the dose-response curve provides valuable information that is not explicitly used in the NOAEL approach (although this shape could possibly influence the choice of uncertainty factors). The main problem of the default approach is that it does not recognise that relevant effects actually may occur at the NOAEL: the test may not have been sensitive enough to detect them.

The BMD method attempts to use more of the available dose-response information by fitting a mathematical model to the data, and then determining the dose associated with a specified response level. This can be a specified incidence of an adverse effect over background in the case of quantal data, or a specified percent change in the level of the endpoint in the case of continuous data. In this way, the benchmark dose is not limited to the experimental doses chosen by the investigators. In addition, one can be fairly confident that the effect at the BMDL does not exceed a certain level (e.g. the benchmark response level, BMR). A common fitting approach is to use maximum likelihood methods, and the resulting point estimate of the dose corresponding to the BMR (the BMDP) is often referred to as the maximum likelihood estimate (MLE) (see Fig. 8). A statistical lower bound (often the 95% lower bound on the dose) is often used instead of the MLE, for added health protective characteristics, and to account for statistical uncertainties. This bound is referred to as the benchmark dose lower confidence limit (BMDL), or sometimes simply as the BMD. Figs. 9 and 10 show a specific application for deoxinivalerol found in cereal crops.

Calculated in this fashion, the BMD is derived based on data from the entire dose-response curve for the critical effect, rather than only from the single dose (e.g. the NOAEL). In this way, the BMD reflects the characteristics of the dose-response curve. Thus, for example, two studies with identical NOAELs and LOAELs could have different BMDs, due to differences in the slope of the dose-response curve above the LOAEL (Fig. II). In this example, the data with the more gradual slope has the lower BMDL, because the response in the range of the NOAEL is higher.

The BMD may be used a point of departure for the development of an intake limit (ADI, TDI or RfD). As with other dose-response assessment approaches, the fitted model will also allow estimation of risk at doses both above and below the BMDL. But low-dose extrapolation including the estimation of a toxicological threshold is not recommended because we generally do...
not understand the mechanisms of toxic action for individual chemicals and how such extrapolations should be performed. Moreover, in estimating risks near the BMD, the issue of using an experimental animal dose-response relationship as the surrogate for a human response needs to be well justified by the risk assessor. In the case of a BMD based on human data, the risk assessor must take care to judge whether the population from which the BMD is developed is an adequate sample from the population being protected, or must justify the use of the surrogate human data.

There are a number of decisions to be made in applying the BMD method and determining a BMDL; for example, which mathematical model to use; what degree of confidence to use in calculating confidence limits; what response level to predetermine as the benchmark response (e.g. BMR = 1, 5 or 10% incidence of an effect, or a 5 or 10% change in a continuous endpoint, such as body weight or red blood cell counts). For more information, the reader is referred to a guidance document on the use of the benchmark dose approach in risk assessment that was issued by the US EPA's Risk Assessment Forum (EPA, 1995). Special software has been developed e.g. ToxRisk Software by EPA (USA) or PROAST by RIVM (The Netherlands). The EPA has also published software in order to make it easier to develop BMDs (EPA, 2000a,d). This software offers several different decisions based on model choice, degree of confidence and incidence rate.

The EPA is also in the process of revising its methods and subjecting these revisions to external peer review (EPA, 2000d). These revisions are expected to be published sometime in 2001.

7.2. Subpopulations

Applying the BMD approach instead of the NOAEL approach leaves the problem of subpopulations basically unchanged, although the use of the BMD will generally lead to more precise determinations of hazard and sensitivity. Thus, the BMD approach may contribute in producing a more accurate estimate of the difference in sensitivity, if data on subpopulations are available. For example, one may have dose-response data in healthy rats as well as in rats suffering from diabetes, which are both exposed at various doses of the same compound. By fitting dose-response models to these data, the difference in sensitivity between animal models can be accurately estimated, and this in turn may have applicability in extrapolation to humans.

Using models based on human data to extrapolate below the BMD (e.g. to sensitive populations) is inherently easier since interspecies extrapolation is avoided. However, the risk assessor must take care to judge whether the BMD adequately represents the population being protected. If it appears to be adequate, extrapolation to doses slightly below the BMD may be appropriate (but as noted above such extrapolation is not considered appropriate for very low doses or threshold determination). If the BMD is based on an observed population that might exclude extremes of sensitivity, uncertainty factors may be needed to address
sensitive individuals (e.g. a BMD developed from a worker population may not be directly applicable to a population that includes infirm people without the use of an uncertainty factor).

Baird et al. (2001) proposed several methods of dealing with heterogeneity in the human population by applying probabilistic assessment factors to the experimental BMD. These are discussed further in section 8.

7.3. Data requirements

The BMD can only be used in cases where data are available that are suitable for modelling. This implies that at least three doses (including the controls) showing different response levels are required, but preferably more. It has been shown by computer simulations (Kavlock et al., 1996; Slob and Pieters, 1997) and illustrated by toxicity studies (Piersma et al., 2000; Woutersen et al., 2001) that study designs with more than the usual four dose groups are better suited for assessing the BMD. It should be noted that the total number of animals in the study can be kept the same, and that the implied lower number of animals per dose group in multiple dose studies does not impair the precision in the BMD estimate.

Because most toxicity studies include descriptions of several effects, care must be exercised in interpretation of BMD modelling. For example, it is usually found that minimally adverse effects occur first as dose rate increases, some of which disappear at higher doses where more severe effects become evident. Modelling of such minimally severe effects might lead to spurious results, or results that are difficult to interpret because the response does not increase with dose.

7.4. Strengths, limitations and weaknesses

The main strength of the BMD method is that an explicit response level can be associated with it. Thus, the dose serving as a starting point for human risk assessment is based more on toxicological judgement (choice of benchmark response level), and somewhat less on the statistical characteristics of the data (which is sometimes more the case in assessing a NOAEL). At the same time this strength reveals an inherent difficulty in risk assessment: it is often not clear what response level BMR can be considered as non-adverse. For example, should a 5% decrease in red blood cell counts be considered as adverse, or should a smaller (or larger) change be chosen? Or, should a 5% increased incidence in hepatocellular hypertrophy be considered as acceptable in an animal study, or is a 10% increase adequate? These and other choices need additional discussion among toxicologists and medical practitioners. Although an explicit statement on the BMR is an improvement compared to the generally unknown response level associated with a NOAEL, choices of a BMR need consensus building. An important contribution of the BMD approach is that it helps the risk assessor to further consider the toxicological and statistical aspects of the data, thereby allowing any discussion on the appropriateness of the BMR to be more influenced by toxicological issues.

Other strengths of the BMD approach are the following. It uses the whole range of experimental dose-response data in its determination as compared to the NOAEL, and is not limited by the doses selected by the investigators. Further, it allows for assessing confidence limits, which is not possible for the NOAEL. The BMD can also be estimated from a study which does not define a NOAEL, thus obviating the necessity of conducting another, and sometime expensive, study. These and other strengths of using this approach in risk assessment have been noted by other authors (Crump, 1984; Dourson et al., 1985; Kimmel and Gaylor, 1988; Barnes et al., 1993; Slob and Pieters, 1998).

A practical difficulty in applying the BMD approach may arise if the study design of toxicity tests is based on the NOAEL approach, which is suboptimal for dose-response modelling. As already discussed, the BMD approach requires various dose groups with different response levels rather than many replicates (animals) within dose groups. In a re-evaluation of five occupational chemicals (Appel et al., 2001) the critical studies were re-analysed by the BMD approach. Although these studies all employed three dose groups and a control group, the dose–response data were not in all cases suitable for the benchmark approach, in particular in those cases where an effect was only observed in the top dose group. To prevent such situations, more dose groups should be used in toxicity studies. Another problem encountered in these re-evaluations was that in some cases a single dose group deviated from the general response pattern formed by the other treatment groups. Whether such a deviation was due to an experimental artefact rather than an effect from the dose, cannot be decided, since dose groups in toxicity studies are not replicated. Again, this problem would be less important if more dose groups are used (but with fewer numbers of animals per group). In multiple study designs similar dose groups may serve as replicates for the other dose groups, by showing a general pattern in the dose–response relationship. An area that would require additional study is that properly designed toxicity studies for the determination of a BMDL, rather than a NOAEL, might be able to reduce the number of animals needed.

While the BMD approach fails when the number of dose groups is too small, both the determination of the BMDL and that of the NOAEL may fail at different points when the number of animals per dose group is
too small. For example, when the critical effect is seen in a larger experimental animal, such as the dog, with few animals per dose group, the NOAEL may be too high due to the insensitivity of the test (i.e. it fails). The BMD approach, however, is able to fully utilize all of these sparse dose–response data, and quantify the resulting uncertainty. However, since an apparent dose–response relationship in the data remains a requirement, the BMD may also fail. Therefore, a typical four-dose study with a few animals per dose may in practice be unreliable whatever method, NOAEL or BMD, is applied.

7.5. Applicability

The applications of the BMD method with data sets of three or more doses are promising, and several assessments to date have employed this method. For example, the EPA and Health Canada has several BMDL-based risk values available (e.g. EPA’s methylmercury RfD, or EPA’s carbon disulfide, antimony trioxide and 1,1,1,2-tetrafluoroethane RfCs, or Health Canada’s 1,3-butadiene TDI). These can be viewed at the International Toxicity Estimates for Risk (ITER, available at www.tera.org/iter). Independent evaluations using the BMD method are also available including those for 1,2-dibromo-3-chloropropane (Pease et al., 1991), hydrogen fluoride (Alexeeff et al., 1993), chromium (Malsch et al., 1994) and soluble nickel (Haber et al., 1998).

7.6. Acceptability to regulatory agencies and authorities

The BMD method has not been used with food additives in either the US or Europe, and its acceptance and proper use may conflict somewhat with the protocols of guideline toxicity studies where a well-characterised NOAEL is the desired outcome. However, in North America, both the EPA and Health Canada now routinely estimate BMDs for chemicals of high interest for environmental regulation. Other states and independent organisations have followed these leads and multiple publications are available that emphasise BMDs for other chemicals. Interest in consistent BMD application has resulted in EPA’s publication of newly developed software (EPA, 2000a). International agencies, such as the International Programme on Chemical Safety, also describe the BMD method in their background methods documents (Dourson et al., 1985; WHO, 1994).

In situations regarding the evaluation of food and food components where risk assessors need a tool other than the NOAEL for proper evaluation, the BMD may be useful and will allow the introduction of this method into a new area. It should be noted here that from current practice, implementation of the BMD approach leads to doses that are quite similar to NOAELs for the studies in question. The adoption of the BMD method does not lead to a wholesale change in the notion of what constitutes acceptable doses, and the continued use of NOAELs (when BMDLs cannot be calculated) is eased because they are considered to be approximately comparable.

7.7. Gaps and research needs

Use of the BMD method has been somewhat limited to date for environmental chemicals and chemicals in the workplace, but has not been used for food or food components. Some publications have compared the use of the BMD to the NOAEL (see previous citations), but for food additives this needs to be done. This limited use has allowed the determination of a number of issues, which may be addressed either by setting guidelines for use or by conducting research to further investigate the model.

The potential application of the BMD approach to simultaneously describe dose–response relationships in different groups (e.g. rat and mouse, chronic and sub-chronic exposure, healthy and infirm rats) should be further investigated by analysing suitable data sets. In this way a better quantitative estimate of differences in sensitivity between groups can be obtained. Such information is highly useful in giving a better quantitative basis to the use of CSAFs.

Other research areas include:

(i) Require rules for minimal dose–response data for deriving a BMDL.
(ii) Distinguish whether different types of toxicity tests, such as a gavage vs dietary exposure, affect the choice of either BMD or NOAEL method.
(iii) Develop rules for combining studies. It should be possible to simultaneously describe the dose–response on the same endpoint but observed in different studies, and might be possible to integrate the observed differences between studies into a cohesive picture of the chemical’s toxicity.
(iv) Apply rules for using a BMDL when acceptable NOAELs are available. Develop a science policy or guideline to interpret and accept the conclusions from the BMDL.
(v) Formalise the choice of the dose–response model, such that the choice of the model depends on the data, and not on the person doing the analysis.
(vi) Select a specific benchmark response (BMR) to model. Various definitions (measures) of BMR have been proposed in the literature, e.g. a particular additional risk for quantal data, or a certain percent change in a continuous endpoint. We must show how these relate to each other, state the various pros and cons, and describe what measures should be preferred for different toxicity endpoints.
Optimise the study design for deriving a BMD, including number of dose groups and animals per group, in order to achieve a required precision of the BMD estimate.

Investigate whether the differences between animal and humans are similar at the low-end of the response curve.

8. Probabilistic risk assessment

8.1. Models and methods

8.1.1. General framework

In the default method, discussed in section 3, acceptable intake or exposure limits are obtained by dividing the NOAEL by a number of uncertainty factors:

\[
\text{ADI, TDI, RID} = \frac{\text{NOAEL}}{\text{UF}_1 \times \text{UF}_2 \times \text{UF}_3}
\] (1)

In other words, the ADI, TDI, RID is defined in a purely operational way; that is, in terms of how to assess its value. This operational definition implies that it results in a single value, the quality (uncertainty) of which cannot be quantified, using this procedure. Thus the level of conservatism for each ADI derived is unknown, and, as a matter of fact, may substantially differ between compounds, depending on the number of uncertainty factors used, which relates to the nature and adequacy of the database.

The probabilistic approach to be discussed in this section suffers less from this drawback. Instead of deriving a single value the aim is to derive a range of values that are plausible, given the uncertainties in our general scientific knowledge, as well as in the available data. In this way, an ADI can be based on a given level of conservatism, so that ADIs assessed for different compounds are more comparable.

As opposed to the operational definition of the ADI by Eq. (1), the probabilistic approach starts from the notion that we are interested in a certain unknown dose level that we consider as sufficiently protective; for instance, the dose level that does not lead to adverse effects in the majority of people. The aim is to estimate this unknown level 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.

At the level of an individual organism we may define the NAEL as the (unknown) dose not resulting in adverse effects in the particular individual. Thus, prior to considering any experimental data we may define the NAEL\textsubscript{animal} as the true NAEL in the average individual of the species (or rather strain) that was observed in a toxicity study. (Notice that the NAEL is a fundamentally different concept from the NOAEL, which only has meaning for a particular data set; the NAEL is the true, but never exactly known value of the dose level at the borderline of adverse effects.) Similarly, we may define the NAEL\textsubscript{average human} and the NAEL\textsubscript{sens. human} as the true NAELs for the average and the sensitive human being, respectively. For any particular compound we define the ratio of two NAELs as the relevant extrapolation factor, for example:

\[
\text{EF\textsubscript{inter}spec} = \frac{\text{NAEL\textsubscript{sens. human}}}{\text{NAEL\textsubscript{average human}}}
\] (2)
\[
\text{EF\textsubscript{intraspec}} = \frac{\text{NAEL\textsubscript{average human}}}{\text{NAEL\textsubscript{sens. human}}}
\]

From these two relations it follows directly that

\[
\text{NAEL\textsubscript{sens. human}} = \frac{\text{NAEL\textsubscript{animal}}}{\text{EF\textsubscript{inter}spec} \times \text{EF\textsubscript{intraspec}}}
\] (3)

Expression (3) can be extended based on the same idea if other extrapolations are to be made, for example from subchronic to chronic exposure.

The question now is how to estimate the unknown NAEL\textsubscript{sens. human} and to quantify the uncertainty in the estimate. The probabilistic approach allows for that by estimating each of the entities on the right hand side of Eq. (3) by a range of plausible values, with various degrees of plausibility. Thus, each entity on the right hand side of Eq. (3) is estimated by a distribution rather than by a single value. These distributions are then combined resulting in a distribution for the NAEL\textsubscript{sens. human}, representing the degree of uncertainty in that estimate. This will be illustrated below, but first the estimation of the entities in Eq. (3) will be addressed.

8.1.2. Uncertainty in NAEL\textsubscript{animal}

The usual way of estimating the NAEL\textsubscript{animal} is by (statistically) comparing each dose level with the controls, resulting in the NOAEL. The NOAEL suffers from various drawbacks (see section 3), one of them being that it is not possible to assess the uncertainty in the NOAEL as an estimator of the (true) NAEL\textsubscript{animal}. This makes the NOAEL unsuitable for a probabilistic risk assessment. The benchmark approach (section 7) allows derivation of a point estimate of the NAEL\textsubscript{animal}, together with an uncertainty distribution, as illustrated in Fig. 12. The benchmark approach is based on fitting a dose–response model to the data. After postulating a certain critical effect size (CES), the associated critical effect dose (CED) is derived from the fitted dose–response 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 an extension of Crump’s benchmark dose.

8.1.3. Probabilistic extrapolation factors (EF)

The denominator of expression (3) consists of a number of extrapolation factors that are typically unknown for any particular chemical. What can be done, however, is to try to find indirect information, for example historical data on other chemicals, that may give an indication of what are plausible values for each of these factors. This information can be summarised in the form of a distribution for each EF. For example, one may imagine that the ratio \( \text{NAEL}_{\text{animal}}/\text{NAEL}_{\text{average human spec}} \) (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 such distributions shows that metabolic dose-scaling (dose per BW\(^{0.75}\)) results in distributions with medians close to unity (see Baird et al., 1996; Vermeire et al., 1999). This indicates that, using this dose-scale, either of the two species is equally likely to be more sensitive to a given compound than the other species. This species-specific dose-scaling factor can be used for a more accurate interspecies extrapolation than a default factor of 10 as currently applied for all test species.

Similarly, one may postulate a distribution for the \( \text{EF}_{\text{intrasp}} \). The same approach can be adopted for other extrapolation factors, for example the \( \text{EF}_{\text{subchronic}} \) (to be used for extrapolating to chronic exposure when only a subchronic study is available).

All these distributions reflect the across-chemical variation (uncertainty). The \( \text{EF}_{\text{intrasp}} \) distribution, however, may be interpreted in either of two ways: as reflecting the across-chemical variation in sensitivity differences for a particular, but unspecified human subpopulation (i.e. as scientific uncertainty), or as the variation in sensitivities in the human population as a whole, where the across-chemical variation is ignored. Theoretically, it is possible to extend the methodology to cover both these aspects, but the data that would be required appear a limitation to do that.

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 (Kramer et al., 1996).

When chemical specific information is available, the EF distributions may be (partly) based on that specific information (see also section 5). When no chemical specific information on any of the EFs is available, default EF distributions have to be used. These default distributions may be based on data that are available for other compounds. An obvious way of estimating an EF distribution is by considering a compilation of relevant NOAEL ratios from historical data. For example, one may consider the ratios of chronic to subchronic NOAELs for a number of compounds to estimate the distribution of the \( \text{EF}_{\text{subchronic}} \). To inform the distribution of the \( \text{EF}_{\text{intrasp}} \) human data would be necessary, e.g., using drug files. Another approach is to examine NOAEL ratios between animal species, and consider the distributions of these ratios as a surrogate for the animal-to-human \( \text{EF}_{\text{intrasp}} \) distribution (see e.g. Vermeire et al., 1999).

Information on the distribution for \( \text{EF}_{\text{intrasp}} \) may be obtained from studies examining the variability in

---

**Fig. 12.** Upper panel: cholinesterase activity (nmol/mg in scale) in erythrocytes as a function of 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.
human kinetics, or relevant physiological or biochemical parameters, and, furthermore, by examining the variation in the interindividual variability across chemicals (Hattis et al., 1999b). In that case, differences in sensitivity are considered to result from general variability in the population as a whole, and sensitive subgroups as the tails of the overall distribution. In this interpretation the outcome of the risk assessment can, at least theoretically, be associated with a particular (small) fraction of the population at risk (Evans et al., 2001). Another way of informing the $\text{EF}_{\text{intraspec}}$ distribution is by searching for chemicals (including drugs), or other agents for which dose–response information is available for specific sensitive subgroups, as well as for the average human being. In that case, the resulting distribution for $\text{EF}_{\text{intraspec}}$ reflects across chemical variation for all sorts of sensitive subgroups that vary among agents and endpoints.

Although most available dose–response data in the literature report NOAELs, it should be noted that NOAELs, and therefore ratios of NOAELs even more so, 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, impairing a correction for this additional nuisance variation so as to obtain more realistic EF distributions. A better way to estimate EF distributions would be to consider ratios of CEDs, obtained by the benchmark approach (see research needs stated in section 8.1.3).

Default distributions for the various EFs have been proposed by Vermeire et al. (1999), after reviewing the relevant literature.

### 8.1.4. Probabilistic assessment of ADI, TDI, RfD

The procedure for deriving a probabilistic ADI is illustrated in Fig. 13. 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 postulated; that is, a certain percent change relative to the level of the 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 postulated; that is, 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. Then the associated critical effect dose ($\text{CED}_{\text{animal}}$) 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, such as for subchronic to chronic extrapolation.

In the Fig. 13 the resulting distribution for the $\text{CED}_{\text{sens. human}}$ is interpreted as the scientific uncertainty for the CED in a particular, but unspecified, human subpopulation. So in this case, the distribution cannot be used to read the fraction in the human population having a lower than pre-specified CED.

Finally, a lower percentile of the final distribution for the $\text{NAEL}_{\text{sens. human}}$ may be chosen, and this value may serve as the ADI (or TDI, RfD). An obvious choice for this lower percentile is 5%, since this is generally considered in science as an acceptable (type I) error (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.

It is important to realise that the final distribution should be 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 $\text{EF}_{\text{intraspec}}$ distribution (see section 8.1.3).

### 8.1.5. 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 focused 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 shown in Fig. 14. 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. Baird et al., 1996; Vermeire et al., 1999), 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, it should be noted that the final distribution reflects scientific uncertainty only: the interpretation of “sensitive human” depends on the interpretation of the EF_{intrasp} 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. (2002), who estimated, for various endpoints, the possible effect sizes in the sensitive human subpopulation resulting from the estimated current intake of deoxynivalenol in cereal crops. For example, based on the 95th 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 1-year-old girls was 0.2-24.6%, based on the 95th 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 1-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.

8.2. Subpopulations

When no chemical-specific information related to sensitive subgroups is available, a probabilistic risk assessment applies a default distribution for intraspecies extrapolation, analogous to the default factor applied in the NOAEL approach (as described in section 3). When

![Diagram](https://example.com/diagram.png)

**Fig. 13.** 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.
such specific information is available, the default distribution for $E_{\text{intraspec}}$ can be adjusted accordingly. This may result, for example, in a distribution which has a higher median than the default distribution, but being less wide (lower geometric S.D.). In this way, one may express the fact that additional information on sensitive subgroups has led to less uncertainty.

8.3. Data requirements

The data requirements concerning the numerator of expression (3) are the same as for the benchmark dose approach: dose-response data are required that allow for dose-response modelling; that is, at least three dose groups (including controls) with different response levels (see section 6.3), but preferably more. If such data are not available, and a reliable benchmark dose cannot be derived, one may do a partial probabilistic assessment based on a NOAEL instead of a CED distribution by replacing the uncertainty distribution of the NEL$_{\text{animal}}$ by the NOAEL.

For the probabilistic extrapolation factors, default distributions have to be used in the usual situation that no chemical-specific information is available. Default distributions have been proposed (Vermeire et al., 1999). If chemical specific information is available on any of the assessment factors, that information could obviously be used to better specify the relevant default distribution. Thus, the probabilistic approach adds to the CSAF approach (section 5) by using a distribution around the CSAF which reflects the remaining uncertainty.

8.4. Strengths, limitations, weaknesses

The strong point of the probabilistic approach is that it quantifies the uncertainty associated with any particular assessment, apart from the risk level itself: the former determines the width of the distribution, the latter the location of it. In the default uncertainty factor approach a low ADI (or TDI, RID) 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 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

---

Fig. 14. 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.
information, is not (yet) firmly based on data, and that the default distributions appear to some extent arbitrarily chosen. However, it should be noted that this similarly holds for the default factor of 10 that is often applied in current practice (section 3).

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 (see section 8.1.5). 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. None the less, the latter situation is informative and helpful in making decisions. 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 results in an improved guidance to the decision-making process.

8.5. Applicability

For assessing the distribution of the CED the applicability of the probabilistic approach is given by the benchmark approach, for example suitable software for dose–response modelling and derivation of the benchmark dose from the fitted model. The only addition is that the software should be able to derive the uncertainty distribution associated with each CED, for example by bootstrapping. In order to combine the various distributions, software is needed that provides for Monte Carlo analysis. To that end, commercial software is available (e.g. Crystal Ball, @Risk).

8.6. Acceptability to regulatory agencies and authorities

Although the probabilistic approach has not yet received an official status in the regulatory agencies, the value of the probabilistic approach is being increasingly recognised by risk assessors and regulators. For example, RIVM (The Netherlands) has recently produced a fact sheet of distributional assessment factors (Vermeire et al., 2001). A probabilistic assessment has been performed (Pieters et al., 2002) on deoxynivalenol occurring in cereal crops at levels exceeding the ADI derived by the default method, which has been seriously taken into account by the regulators addressing this issue.

8.7. Gaps and research needs

The quality of a (probabilistic) risk assessment hinges on the realism of the distributions of the extrapolation factors. Therefore, the main research need here is to find quantitative support for each of these distributions. This may be done by literature studies gathering historical dose–response data related to the relevant assessment factor. The original data should be re-analysed by the benchmark approach, since NOAEL ratios are less suitable for that purpose. NOAELs are subject to large errors, thereby contaminating the distributions of NOAEL ratios considerably. When various distributions are used in the denominator of expression (3), this extra nuisance variation in each single distribution is magnified in the process of multiplying the various distributions with each other. By applying the benchmark approach, the resulting ratios will have smaller estimation errors. Besides that, the estimation errors can actually be quantified, offering a potential tool to correct the resulting ratio distribution for that extra variation.

Chemical-specific data or pathway-related data could be used to refine the distribution of the relevant EF (see section 5). Therefore, the research needs as mentioned in section 5 are of interest in a probabilistic framework as well.

It may be necessary to perform specific experimental studies to fill certain gaps in the literature. For example, it may occur that good human dose–response data are available, that are not mirrored by a similar animal study. Human data that are complemented with animal data, are highly useful in informing the default EF interspec distribution to be used for other chemicals.

9. Physiologically-based modelling

9.1. Models and methods

Accurate characterisation of dose–response relationships and extrapolation of results from animal models to humans are both of primary importance in toxicology and risk assessment. One approach that has the potential to improve how both of these critical issues are addressed is PBTK modelling (see also Eisenbrand et al., 2002; Dybing et al., 2002). A PBTK model is a series of mathematical equations based on organism-specific and chemical-specific information that describe the absorption, distribution, metabolism and elimination of a chemical within an organism (Gerlowski and Jain, 1983; Andersen, 1991). Incorporated into PBTK models are organism-specific physiological information such as cardiac output, tissue blood flows, and tissue volumes, and chemical-specific information on parameters including tissue/blood partition coefficients and kinetic constants that describe elimination and metabolism of the compound under study. Solutions of the PBTK models equations provide estimates of the concentrations of a chemical and its metabolites in any tissues or organs over time. The power of PBTK modelling lies
in its ability to calculate the amount of the active form of a chemical at its target site within the body over time, given virtually any dose, exposure route and exposure scenario. In addition, extrapolation of a valid animal PBTK model to humans can be attempted through substitution of animal physiological data with human physiological descriptors and, when available, human chemical-specific information.

The construction of a PBTK model begins with the description of the body as a series of compartments representing individual tissues or tissue groups. Those tissues represented by compartments are selected on the basis of their relevance to the disposition and/or action of the chemical under consideration. For example, a central nervous system depressant would require the brain as a compartment, while hepatic metabolism of a drug would entail the use of the liver as a compartment. Following selection of appropriate compartments, differential equations are written that describe the fate of a chemical as it passes through each tissue. Many examples of such equations have been published, and even the novice will find that a specific model can be used easily. Values for organism-specific parameters that must be supplied to the differential equations include: tissue blood flows, organ volumes, cardiac output, and in certain instances, ventilation rate. Many of these values that describe physiological parameters can be obtained from the literature (Brown et al., 1997). Values for chemical-specific parameters that often must be determined experimentally are, at a minimum: elimination rate constants and tissue/blood partition coefficients. The specific methodology employed to determine these values depends to a large extent on the physico-chemical and toxicokinetic characteristics of the chemical under study. If metabolite disposition is to be included in the model, chemical-specific parameters for the parent compound and the metabolite(s) must be determined. Solution of the model requires a personal computer equipped with software that can solve simultaneous differential equations. One software package commonly used for this purpose is Advanced Continuous Systems Language (ACSL) (Pharsight Corporation, Mountain View, CA, USA), although many others can be used as well.

A number of PBTK models have been constructed, validated and the results published. Some interesting examples include a PBTK model to assess combined dermal and inhalation exposure of humans to the gasoline additive methyl tert-butyl ether and its metabolite tert-butyl alcohol (Rao and Ginsberg, 1997). In this assessment, the authors evaluated the inhalation and dermal exposures that one would experience through bathing that might cause acute effects on the central nervous system. They determined that while concentrations of methyl tert-butyl ether in the groundwater might present a groundwater resource concern, levels would have to increase to a range where organoleptic concerns would preclude water use before acute neurotoxicity would occur.

Trichloroethylene, a common contaminant of ground and surface water, is a compound that has been the focus of a number of PBTK modelling exercises (e.g. Fisher et al., 1991). One model in particular includes a component to simulate the formation and enterohepatic recirculation of the metabolites of trichloroethylene, including trichloroethanol and the rodent carcinogen, trichloroacetic acid (Stenner et al., 1998). Trichloroethylene is rapidly absorbed from the gastrointestinal tract of rodents, and is virtually eliminated from the blood within 2 hours, though the trichloroacetic acid blood level reaches its peak after 12 hours. Incorporation of enterohepatic recirculation and metabolism of trichloroethylene provides a more precise description of the variables important in determining the effective dose of the active metabolites at the target tissues than was available in earlier models.

Arsenic is an element widely distributed in nature that is particularly difficult to characterise. Arsenic chemistry is complicated because it may be trivalent or pentavalent and forms many different compounds of concern in toxicology. A PBTK model of inorganic arsenic exposure in humans has been developed that is an extrapolation of a model developed for hamsters and rabbits, with adjustments for body weight, metabolic rates and absorption rates (Mann et al., 1996). The model describes the absorption, distribution, metabolism and excretion of the four major metabolites of inorganic arsenic: arsenate, arsenite, methyl arsinate and dimethyl arsinate. The intake pathways evaluated were inhalation of arsenic dust and fumes and oral intake of arsenic via drinking water and food. The PBTK model simulates the effects on the kinetics of exposure via different routes, and allows simulations of various realistic exposure scenarios.

9.2. Consideration of sensitive subpopulations

The effects of chemicals on sensitive subpopulations can be evaluated using PBTK models in a number of ways. The impact of genetic variability on toxicity of a chemical can be evaluated by incorporating the kinetic parameters associated with genetic polymorphisms in the PBTK model (Hattis et al., 1999a). For example, it is known that alcohol abusers are at an increased risk of developing cirrhosis, though only 12–13% of them actually develop the disease. The cause of this differential sensitivity is not known. Pastino and co-workers (Pastino et al., 2000) noted that genetic polymorphisms in alcohol dehydrogenase impact alcohol elimination rates among various populations, though this does not fully account for differential susceptibility to cirrhosis. To determine whether kinetic parameters and differ-
ential expression of the alcohol dehydrogenase isoforms can explain susceptibility to cirrhosis, additional kinetic data on each isoform could be collected and incorporated in a PBTK model. Other parameters needed for such a model include blood alcohol levels from populations with known alcohol dehydrogenase phenotypes. A PBTK model that incorporates kinetic constants for each isoform of alcohol dehydrogenase in the liver would help in assessing the relative contribution of each isoform to cirrhosis susceptibility.

Variability in human physiology may account for differences in sensitivity to a given toxicant that may give rise to a sensitive subpopulation. These differences in sensitivity due to variability in human physiology may be accounted for in the PBTK model by using a distributional approach for input of parameter values (Hattis et al., 1999b).

### 9.3. Data requirements

Even the simplest PBTK models require an extensive set of data. In turn, the more complicated the model the more data are required. The data required for development of PBTK models include: physiological and anatomical descriptors, partition coefficients of the compounds between various media, descriptors of metabolic transformation pathways and, when the assumption of diffusion limited uptake is made, transport parameters. These data may be measured directly or estimated based on expert judgment or statistical methods. Key species-specific physiological descriptors, such as organ volume, cardiac output, regional blood flow renal clearance rates and alveolar ventilation rates are obtained experimentally and may be available in the open literature (Brown et al., 1997). Chemical-specific partition coefficients describe chemical transport across membranes, based on the affinity of the chemical for different types of tissues. Knowledge of the partition coefficient allows estimation of the amount of unbound chemical in a given tissue. Partition coefficients needed for a given chemical and model may include blood/air coefficient, muscle/blood and fat/blood coefficients.

Descriptors of metabolic transformation pathways are needed if the metabolites of the parent compound are of interest as in models of methyl tert-butyl ether and trichloroethylene mentioned above in section 9.1. Chemical-specific information from humans should be incorporated in the model if available, though use of surrogate data for metabolically related compounds has been described (Barton et al., 2000). As with any model, there are areas of uncertainty in PBTK modelling, both in the models themselves and the data that are input. There are methods to account for uncertainty in PBTK models and uncertainty in the model parameters (Nestorov et al., 1997). For example, distributions may be used to represent physiological parameter values in place of single values. Sensitivity analysis can be performed to characterise the impact of variability in certain parameters on the model result.

### 9.4. Strengths and limitations of the method

The greatest strength of PBTK modelling lies in the ability to calculate the amount of the active form of a chemical at its target site within the body over time, given virtually any exposure scenario, and extrapolation to humans based on human physiological and chemical-specific information. The model's capacity to describe the tissue dosimetry of a toxicant in different animal species, sexes and conditions through selection of appropriate model parameter values, reduces the uncertainty associated with extrapolation of toxicity data from laboratory animal studies to human scenarios using other methods. This reduction in uncertainty is effected by incorporation of mechanistic data in equations that represent physiological processes. Mechanistic data combined with a PBTK model can be helpful in understanding dose-response data derived from toxicity studies, and thereby reduce or eliminate the need for default uncertainty factors (see section 3). These models allow the use of data in place of interspecies scaling factors in risk or safety assessment (section 5), and provide a means to evaluate safety or risk based on all the available scientific data.

Although PBTK modelling is a powerful toxicokinetic tool, it is not without drawbacks. There are certain limitations intrinsic to this approach that may render PBTK modelling impractical for routine use. Building PBTK models can require considerable time and resources, if risk assessment is difficult. At a minimum, moderate programming skills are required, and tissue/blood partition coefficients and elimination rate constants can be difficult and time-consuming to determine accurately for many types of chemicals. Moreover, although validation of animal PBTK models can be accomplished by comparing model predictions with empirical toxicokinetic data, such data are often missing and unobtainable in humans. Often there is a lack of transparency in models that should be overcome in order for the models to gain acceptance. Validation PBTK model can be problematic, particularly if unknown parameters are estimated indirectly by fitting the model to the data. The toxicokinetic estimates and statistical methods should be clearly described and documented. In this case, it is possible that various sets of parameter values in the model could result in an equally good fit to the data, but produce different estimations of internal dose in extrapolation from animals to humans. An important challenge to modellers is to address these complicating issues so that in the future PBTK modelling can become routine and widespread in all areas of toxicology.
9.5. Applicability

PBTK models are applicable to evaluate target organ dose following exposure to chemicals by any route. It follows that PBTK models can be useful tools in risk assessment of chemicals in food.

9.6. Acceptability to regulatory agencies and authorities

Regulatory agencies in the United States have begun to accept toxicity evaluations and risk assessments based on PBTK modelling. The US Occupational Safety and Health Administration (OSHA) has relied on PBTK modelling to develop a risk estimate in the specific case of methylene chloride, a chemical with more extensive information on metabolism than exists for most other substances. To that end, OSHA adopted a Bayesian approach in which all of the physiological and methylene chloride-specific data could be used to generate a distribution of estimates of the carcinogenic risks of methylene chloride. OSHA used the mean and the upper 95th percentile estimator of the distribution of human PBTK parameters, coupled with the maximum likelihood estimator of cancer potency, to generate its final estimates of risks.

9.7. Gaps and research needs

PBTK models can be extended and improved by adding components to describe the effects on the target tissue. The action of the chemical on the target tissue, known as the toxicodynamics of a chemical, can be described mathematically once the mechanism of action of a chemical is known. Physiologically-based-pharmacokinetic-pharmacodynamic (PBPK/PD) models have been used in pharmacology research to link the dose-concentration relationships with concentration-effect relationships in the target tissue. PBTK/TD modelling can be used as a tool in human health risk and safety assessment (Becking, 1995; Medinsky, 1995; Della Paschoa et al., 1998; Yang et al., 1998; Derendorf and Meibohm, 1999). PBTK/TD models allow refined characterisation of dose-response relationships that are critical in quantitative risk assessment. In these models, tissue dosimetry from the PBTK model is linked quantitatively to the action of the effective dose at the target site. These models can predict the time course of toxic effects resulting from a given exposure. Incorporation of mechanistic data in these models can provide a rational means for extrapolating from animal data to human risk, and may address whether or not the mechanisms that mediate toxicity in animals are similar or different than in humans.

PBTK/TD models have been developed to evaluate the impacts of multiple chemicals on biological systems and can be used to evaluate interactions such as inhibition or potentiation (Mumtaz et al., 1993; Krishnan et al., 1997). Because there are many chemicals in food that may interact with each other and the body in both beneficial and undesirable ways, this type of model may be important in characterising both the toxic and beneficial effects of food components. Specialised software that is capable of solving coupled differential equations numerically, including optimisation of algorithms, is required.

The expansion of PBTK/TD models from use largely in pharmacological research to use in toxicology and risk assessment will bring these models more into mainstream use by risk assessors. These models can be used to refine risk estimates for potentially sensitive subpopulations, and in some cases identify the basis for the increased sensitivity.

10. General evaluation and comparison of the methods

Quantitative methods and the application of mathematical modelling have been described in the previous sections. They all are considered as valuable tools useful for application in food risk assessment and are summarised in this section. Table 2 provides a compilation of the main characteristics of the eight methods including their ability to address subpopulations and their practical requirements in relation to the type, amount and quality of data necessary.

The first methods we considered in this document were those based on SAR used together with the concept of the TTC. They are used when no adequate toxicity data exist, but human exposure is estimated to be extremely low. These methods do not rely directly on dose–response data, although they may use dose–response information for structural analogues. Therefore one might consider them separately from the other methods. Nevertheless, we found it useful to compare this approach with the others based on our general criteria addressing subpopulations, data requirements, applicability and acceptance. SAR and TTC methods have been applied to a larger extent to food packaging migrants and flavours, because of their very low estimated exposures. Threshold methods using default assumptions focus on the determination of a dose as a point of departure for defining exposures with negligible risk through the use of UFs. The basic approach can be refined by the use of CSAs in place of default UFs. Obviously, there can be a strong interrelation between the characterisation of CSAFs and the use of PBTK models. PBTK/TD modelling is an efficient tool to derive CSAFs given sufficient biological information and data for their construction. If the mode of action of a compound suggests that there may exist no dose level without significant adverse effect, a model can be used which extrapolates to low doses assuming no threshold.
Table 2
Summary of the main characteristics, the ability to address subpopulations and the type and extent of data required for the eight methods presented in sections 2-9

<table>
<thead>
<tr>
<th>Method</th>
<th>Main characteristics</th>
<th>Subpopulations</th>
<th>Data requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAR and TTC</td>
<td>• applied in the case of no adequate toxicity data but extremely low human exposure</td>
<td>• not specifically considered</td>
<td>• chemical structure, intake estimates</td>
</tr>
<tr>
<td></td>
<td>• thresholds are derived based on toxicity for other compounds including carcinogens</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Threshold</td>
<td>• assumes the existence of a (biological) threshold below which “no significant biological effect exists”</td>
<td>• UF = 10 for differences within the human population but subpopulations are not specifically considered</td>
<td>• for NOAEL determination at least one dose with positive and one without statistical/biological significant positive response is needed for the same endpoint</td>
</tr>
<tr>
<td></td>
<td>• uses the NOAEL as a surrogate for the (biological) threshold</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• default, CSAF or probabilistic UFs can be used for interspecies and intrahuman extrapolation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSAF modelling</td>
<td>• allows chemical specific information to replace default UFs</td>
<td>• adjustment factors for sensitive groups can be determined from percentile differences of the general population and the sensitive group based on chemical and population specific data</td>
<td>• in vivo toxicokinetic data and/or in vitro measurement of elimination combined with a PBTK model and/or in vivo or in vitro toxicodynamic data using equipotent doses or concentrations</td>
</tr>
<tr>
<td></td>
<td>• is based on the comparison of mean parameter values (interspecies) and on the difference between means and percentiles (human variability)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-threshold</td>
<td>• estimates the magnitude of risk at human exposure levels</td>
<td>not specifically considered</td>
<td>• for linear extrapolation from a point of departure single observed incidence can be used for fitting</td>
</tr>
<tr>
<td></td>
<td>• extrapolation to low-dose effects over a range of approximately five orders of magnitude</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• methods usually applied are</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• LMS: linearised multistage model to estimate doses as low as 10^{-6}</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• linear extrapolation from a point of departure chosen from the data, e.g. ED_{10} or ED_{25} or from an ED_{10} estimated by LMS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMD</td>
<td>• makes use of all dose response information for getting a point estimate of the dose associated with, e.g., 5%, 10% response</td>
<td>• UF = 10 for differences within the human population but subpopulations are not specifically considered</td>
<td>• dose–response data (at least three dose groups and a control, showing different response levels)</td>
</tr>
<tr>
<td></td>
<td>• fits empirical models</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• default, CSAF or probabilistic UFs can be used for interspecies and intrahuman extrapolation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probabilistic RA</td>
<td>• provides an ADI, TDI or RID and the associated uncertainty</td>
<td>• specific distributions can be used for subpopulations based on chemical and population specific data</td>
<td>• no data needed when using default distributions for EFs applied to NOAELs and BMDs</td>
</tr>
<tr>
<td></td>
<td>• uses the concept of the true NAEL for average and for sensitive humans</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• probabilistic extrapolation factors (EFs) replace single values for UFs</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• the method is applicable also for estimating effect size (non-cancer) at actual exposure</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
These models frequently apply simple linear extrapolation from some point of departure on the dose–response relationship within the experimental range. Whereas both the threshold and the non-threshold default methods make only limited use of the full dose–response data, the more recently developed benchmark dose methodology aims to use all dose–response information. The three other methods — probabilistic risk assessment, categorical regression and PBTK modelling — are considered in this report to be relevant for food risk assessment as supportive measures for the evaluation of dose–response data and the quantitative analysis of a dose–response relationship. The data requirements of these methods are quite different (Table 2). Importantly, CSAFs, probabilistic risk assessment and PBTK modelling are methods that allow improved consideration and evaluation of subpopulations; in contrast the other methods either rely on default methods or do not allow specific consideration of subpopulations. Non-threshold methods and the BMD approach require almost the same type and amount of data. The complexity of a model determines both the quantity and quality of data required to obtain a dose–response relationship. In the vast majority of cases, risk assessment for compounds faces the situation that only limited data are available, which may not allow a full characterisation of the dose–response relationship. Then, either the risk assessment has to be suspended until sufficient data become available or simple default methods have to be adopted.

The strengths, limitations and weaknesses of the quantitative methods described in sections 2–9 are summarised in Table 3. There is an obvious gradient of decreasing simplicity and, therefore, increasing complexity in progressing from the default methods (e.g. non-threshold and NOAEL) to full dose–response modelling (e.g. BMD or categorical regression) and from the usage of default UF's to the use of CSFAs based on toxicokinetic and toxicodynamic information. The strength of simple usage and ease of understanding of a method is offset by its dependency on assumptions. Conversely, the greater complexity and increased usage of all toxicological information in a mathematical modelling approach is offset to some extent by the efforts and costs required to obtain sufficient data. Fewer animals may be required if the dose–response experiments are designed appropriately and if the information extracted from the animal studies is used more fully.

Risk assessment of chemicals usually involves the extrapolation of information obtained from animal studies to the human situation. High doses are given to animals in order to maximise hazard identification. The results from these high-dose animal studies often must be extrapolated to predict effects at lower levels of expected human exposure. Risk assessment of chemicals in food is guided by the demand that the human food supply should represent a negligible or no risk, and therefore the adverse effect detected in animal studies would not be acceptable if they occurred in humans. Thus, acceptable levels of exposure of humans generally are set two or more orders of magnitude lower than those producing no adverse effects in animal studies. In consequence, from a methodological point, the approaches considered above are appropriate for the assessment of minor and non-essential components of the human diet, such that there can be a substantial difference between the animal and human exposures. While this approach is suitable for determining acceptable intakes for food additives and tolerable intakes for contaminants, it cannot be applied to all food components. For example, a large difference between doses in animal studies and expected human exposures may not be possible for macronutrients, because it is not possible to exaggerate the exposure in animal feeding studies. For micronutrients the establishment of a wide margin between
doses that do not produce adverse effects in animals and expected human intakes could result in intakes in humans that would lead to nutritional deficiency.

The applicability of the quantitative methods described in sections 2–9 is shown in Table 4, and ranges from limited usage to broad applicability depending on the expected exposure level, the availability and quality of the data, the ease of extrapolation, and software availability. Resources for the programming of model parameters or the means to get good estimates (especially for populations) are further criteria for applicability. Mathematical and numerical complexity is not an intrinsic limiting factor for most of the methods described above. However, complexity could require sub-

Table 3
Strengths, limitations and weakness of the eight methods presented

<table>
<thead>
<tr>
<th>Method</th>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAR and TTC</td>
<td>• is simple to apply and readily understood</td>
<td>• assumes that structure predicts toxicity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• depends on current exposure estimates for the population</td>
</tr>
<tr>
<td>Threshold</td>
<td></td>
<td>• assumes the existence of a threshold</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• the NOAEL does not exclude biologically significant effects below the sensitivity of the test</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• the value of the NOAEL depends on experimental conditions such as group size, sensitivity of measurement of the adverse effect, and dose spacing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• does not make full use of the dose–response information</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• uses default UFs</td>
</tr>
<tr>
<td>CSAF modelling</td>
<td>• chemical specific data can be incorporated to reduce uncertainty</td>
<td>• depends on the validity of the subdivision of the 10-fold factors</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• is a data intensive method</td>
</tr>
<tr>
<td>Non-threshold</td>
<td>• linear extrapolation is simple to apply</td>
<td>• linear extrapolation is thought to be highly conservative</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• LMS cannot be validated as a model for low doses and extrapolation model dependent</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• differing balances between reactivity and repair between low and high doses are not accommodated</td>
</tr>
<tr>
<td>BMD</td>
<td>• makes full use of the dose–response data</td>
<td>• obtaining consensus defining a benchmark response level for the adverse effect (e.g. 5 or 10%) is difficult</td>
</tr>
<tr>
<td></td>
<td>• allows confidence limits for point estimates</td>
<td>• is not applicable to studies with few dose groups</td>
</tr>
<tr>
<td></td>
<td>• an optimal experimental design may allow reduction of the number of animals tested (does not require a large number of animals per group)</td>
<td></td>
</tr>
<tr>
<td>Probabilistic RA</td>
<td>• uncertainties associated with all aspects of the quantitative methods of the RA process can be taken into account</td>
<td>• requires use of default distributions in most cases</td>
</tr>
<tr>
<td></td>
<td>• appropriate chemical specific information can be incorporated to reduce uncertainty</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• provides effect estimates at actual exposure levels</td>
<td></td>
</tr>
<tr>
<td>Categorical regression</td>
<td>• takes all studies into account and not only the most sensitive one</td>
<td>• requires toxicological judgement for the categorisation</td>
</tr>
<tr>
<td></td>
<td>• allows the prediction of a severity effect category at a particular dose (e.g. above ADI)</td>
<td>• the interpretation of fitted model (different endpoints, observer variation etc.) is difficult</td>
</tr>
<tr>
<td>PBTK</td>
<td>• is able to model the time course of the amount of the active compound at the target site</td>
<td>• is a data intensive method</td>
</tr>
<tr>
<td></td>
<td>• it is possible for any species and for different exposure (e.g. route to route extrapolations) and lifetime conditions</td>
<td>• does not address the dynamics</td>
</tr>
<tr>
<td></td>
<td>• allows extrapolation from animal to human without having to have human exposure data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• allows target organ dose–response relationships to be used for low-dose extrapolation</td>
<td></td>
</tr>
</tbody>
</table>
stantial effort for some applications of the BMD, probabilistic risk assessment, categorical regression and, especially, PBTK modelling. There is a rather mixed degree of acceptance and application of the various methods between the EU, its member countries, and other countries and agencies involved in food risk assessment. Table 4 provides a summary of acceptability by EU, EPA, Health Canada and WHO and the policies of JECFA and SCF.

The applicability of the different methods to different categories of chemicals in food and diet (such as additives, flavourings, substances used in production, contaminants, non-threshold methods for population distribution analysis are needed.

Table 4
Applicability, computational complexity and availability of software, and acceptance by regulatory agencies or their advisory committees of the eight methods presented

<table>
<thead>
<tr>
<th>SAR and TTC</th>
<th>Threshold</th>
<th>CSAF modelling</th>
<th>Non-threshold</th>
<th>BMD</th>
<th>Probabilistic RA</th>
<th>Categorical regression</th>
<th>PBTK</th>
</tr>
</thead>
<tbody>
<tr>
<td>• to extremely low exposures only</td>
<td>• software is not essential, except the use of SAR software and databases</td>
<td>• limited because it depends on the availability of appropriate data for the specific chemical</td>
<td>• linear extrapolation is widely applicable because of its limited data requirementsthough LMS model is less widely applicable than other models using linear extrapolation</td>
<td>• less widely applicable for current studies than threshold methods</td>
<td>• less widely applicable for current studies than simple deterministic methods</td>
<td>• limited because it needs special methodology</td>
<td>• limited because it depends on the availability of appropriate data for the specific chemical</td>
</tr>
<tr>
<td>• FDA accepted 0.5 ppb as threshold</td>
<td>• all agencies accept the NOAEL and the use of UF or margins of safety/exposure</td>
<td>• IPCS uses this approach</td>
<td>• linear extrapolation is accepted in various EU countries</td>
<td>• special software is developed, e.g. ToxRisk software or software by EPA (see 6.1), PROAST (RIVM)</td>
<td>• is not currently accepted in EU</td>
<td>• not accepted in EU</td>
<td>• OSHA uses and relies on PBTK models for some chemicals</td>
</tr>
<tr>
<td>• JECFA uses this approach for flavouring substances</td>
<td>• guidelines have been developed</td>
<td>• LMS model is available through several sources including ToxRisk</td>
<td>• LMS is not widely accepted in EU</td>
<td>• no software is needed when using default distributions otherwise software allowing for Monte Carlo methods (Crystal Ball. @Risk) is needed</td>
<td>• is used in some case studies (e.g. by EPA) in conjunction with other methods</td>
<td>• OSHA uses and relies on PBTK models for some chemicals</td>
<td>• is accepted by EPA</td>
</tr>
<tr>
<td>• 1.5 μg TTC is not accepted by JECFA and SCF for possibly genotoxic carcinogens</td>
<td>• Health Canada has used the approach</td>
<td>• programming with software systems as, e.g. SAS, SPSS, Splus is possible</td>
<td>• LMS is accepted by EPA</td>
<td>• no software is needed when using default distributions otherwise software allowing for Monte Carlo methods (Crystal Ball. @Risk) is needed</td>
<td>• is used in some case studies (e.g. by EPA) in conjunction with other methods</td>
<td>• is accepted by EPA</td>
<td></td>
</tr>
<tr>
<td>• genotoxic carcinogens</td>
<td>• SCF have used the subdivision of the 10-fold factor (cyclamate, dioxins)</td>
<td>• is accepted by WHO</td>
<td>• is accepted by WHO</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Figures and tables are placeholders and should be replaced with actual data from the source.*
Further methodological development is to be encouraged, particularly in the areas of (i) the nature and application of uncertainty factors, (ii) distributional approaches to extrapolations, (iii) the full characterisation of the uncertainties inherent in ADI calculations, (iv) the separation of variability and uncertainty in risk estimations, (v) the issue of sensitive subpopulations, (vi) the definition of what represents an adverse effect, especially for continuous endpoints, and (vii) the question of how risks from exposures near or above the ADI are to be addressed.

11. Gaps and research needs

Gaps and research needs identified during the above evaluation are summarised in Table 5. Major gaps and research needs detected in our review include the expansion of existing databases, the validation of assumptions, additional experiments to fill data gaps in the toxicological and exposure literature, and guidelines for the application and validation of the models.

Major research needs include the development and validation of default UF's, development of improved dose–response models based on a better understanding of their biological basis, and by a better understanding of adversity, and assignment of severity. Obviously sensitive subpopulations and heterogeneity in the population are fields of prevailing uncertainty in modelling.

Table 5
Comprehensive summary of the gaps and research needs, identified for the eight methods presented in this review

<table>
<thead>
<tr>
<th>Gaps and research needs</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAR and ITTC</td>
</tr>
<tr>
<td>• expansion of databases of chemical structure and thresholds for different types of toxici ties</td>
</tr>
<tr>
<td>• updated validity of the Cramer decision tree</td>
</tr>
<tr>
<td>Threshold</td>
</tr>
<tr>
<td>• development and validation of appropriate UF's for different species and different categories of metabolic fate and mechanisms of action</td>
</tr>
<tr>
<td>• validity of the default UF by analysis of historical data concerning human vs animal or/and human subpopulations</td>
</tr>
<tr>
<td>CSAF modelling</td>
</tr>
<tr>
<td>• availability of toxicodynamic data for the subdivision of the 10-fold default inter- and intraspecies UF's</td>
</tr>
<tr>
<td>Non-threshold</td>
</tr>
<tr>
<td>• knowledge of the shape of the dose–response curve at low doses</td>
</tr>
<tr>
<td>• better understanding of the biological basis for the extrapolation, e.g. by using biomarkers, tumour precursors, genetically modified animals, and by including toxicokinetics for target dose estimation</td>
</tr>
<tr>
<td>• optimal designs with respect to number animals and number of dose groups</td>
</tr>
<tr>
<td>• understanding of adversity of the effect size for getting consensus on the benchmark response level</td>
</tr>
<tr>
<td>• combined analysis of different populations for more precisely estimating BMD ratios</td>
</tr>
<tr>
<td>• optimal designs with respect to number animals</td>
</tr>
<tr>
<td>• combining studies</td>
</tr>
<tr>
<td>Probabilistic RA</td>
</tr>
<tr>
<td>• analysis of historical data concerning human vs animal or/and human subpopulations for updating default uncertainty distributions</td>
</tr>
<tr>
<td>• specific experiments filling data gaps in the literature</td>
</tr>
<tr>
<td>• application of probabilistic RA in parallel with the traditional UF approach</td>
</tr>
<tr>
<td>Categorical regression</td>
</tr>
<tr>
<td>• guidelines for application and validation of models used</td>
</tr>
<tr>
<td>• criteria for the assignment of severity categories</td>
</tr>
<tr>
<td>• criteria for combining studies (e.g. weighting small and large studies)</td>
</tr>
<tr>
<td>• extension of the toxicokinetic model to a complete toxicokinetic/toxicodynamic model</td>
</tr>
<tr>
<td>• lack of basic information on human variability in physiological parameters</td>
</tr>
<tr>
<td>• knowledge to translate in vitro information to in vivo information</td>
</tr>
<tr>
<td>• validation of physiological parameters in humans using pharmaceutical research results</td>
</tr>
</tbody>
</table>
molecular weight chemicals, micronutrients and nutritional supplements, macronutrients, whole foods, novel foods and food processing may each require specific quantitative methods.

2. Most toxicity tests, as currently designed, allow hazard identification and the determination of a dose level (the NOAEL) that is expected not to produce adverse effects. Data produced using current testing guidelines are not always suitable for robust mathematical dose-response modeling.

3. Adequate characterization of the dose-response relationship by mathematical modeling requires a number of doses giving a range of different response levels.

4. Mathematical modeling of the dose-response relationship would improve the risk assessment process. Existing protocols should be redesigned to support more accurate modeling, for instance by increasing the number of treatment groups. An increase in the number of treatment (dose) groups would not necessarily require an increase in the number of animals in the study, and an appropriate design might even result in a reduction in the total number of animals studied.

5. Chemical-specific data relating to the mode or mechanism of action and/or the toxicokinetics of the chemical, for example by a PBTK model, should be introduced into dose-response characterization whenever possible. Toxicokinetic data would allow correction of the external dose to an internal dose (or target-organ dose). Mode of action data could be used to extend a PBTK model into a full biologically-based dose-response model (PBTK-TD model).

6. Chemical-specific data relating to the mode or mechanism of action and/or the toxicokinetics of the chemical can be used to modify the default (10-fold) uncertainty factors by calculating chemical-specific adjustment factor(s). Such information can also be used to refine the distribution of the extrapolation factors used in a probabilistic assessment.

7. Future refinements to dose-response characterization should incorporate more clearly the extent of uncertainty and variability in the resulting output. Probabilistic analysis allows uncertainty to be characterized, and can be used to provide an output, such as an ADI, which relates to a predetermined degree of conservatism, for example the 5th or 10th percentile of the uncertainty distribution of the output. Probabilistic analysis can model the impact of different degrees of conservatism on the output.

8. Simple default approaches, such as the use of the NOAEL and uncertainty factors or linear low-dose extrapolation, are widely applicable, but are likely to give imprecise outputs, without an estimate of uncertainty. More advanced methods, for both cancer and non-cancer effects, are capable of giving greater precision or an estimate of uncertainty or both, but may not be applicable given the paucity of suitable data. Such methods may be particularly valuable for situations where quantification of risk is required, for example when the intake exceeds the ADI, TDI or RfD.

References


Woutersen, R.A., Jonker, D., Stevenson, H., Te Biesebeek, J.D., Slob, W., 2001. The benchmark approach applied to a 28-day toxicity study with Rhodorsil silane in rats: the impact of increasing the number of dose groups. Food and Chemical Toxicology 39, 697–707.