SCIENTIFIC OPINION

EFSA Journal

ADOPTED: dd mmmm yyyy PUBLISHED: dd mmmm yyyy AMENDED: dd mmmm yyyy

doi:10.2903/j.efsa.20YY.NNNN

Guidance on Uncertainty in EFSA Scientific Assessment

EFSA Scientific Committee[1, 2]

European Food Safety Authority (EFSA), Parma, Italy

Abstract

To meet the general requirement for transparency, all EFSA scientific assessments must include consideration of uncertainties. Assessments must say clearly and unambiguously what uncertainties have been identified and what is their impact on the overall assessment outcome. The Guidance is applicable to all areas of EFSA and all types of scientific assessment. It does not prescribe specific methods for uncertainty analysis but rather provides a harmonised and flexible framework within which different methods may be selected, according to the needs of each assessment. Worked examples are provided to illustrate different methods. Assessors should be systematic in identifying uncertainties, checking each part of their assessment to minimise the risk of overlooking important uncertainties. Uncertainty may be expressed qualitatively or quantitatively. It is not necessary or possible to quantify separately every individual source of uncertainty affecting an assessment. However, assessors should always aim to express overall uncertainty in quantitative terms to the extent that is scientifically achievable. Uncertainty analysis should be conducted in a flexible, iterative manner, starting with simple approaches and then refining the analysis as far as is needed or possible within the time available. Some steps may be reduced or omitted in emergency situations and in routine assessments with standardised provision for uncertainty. Sensitivity analysis is used to target refinement on those sources of uncertainty where it will contribute most. The methods and results of all steps of the uncertainty analysis should be reported fully and transparently. Every EFSA Panel and EFSA Units that produce scientific outputs should apply the draft Guidance to at least one assessment during an initial trial period, involving relevant decision-makers and supported by specialists in uncertainty analysis where needed. When the trial period is completed and any resulting improvements to the Guidance Document have been agreed, uncertainty analysis will be unconditional for EFSA Panels and staff and must be embedded into scientific assessment in all areas of EFSA's work.

1 Scientific Committee members: Jan Alexander, Diane Benford, Qasim Chaudhry, John Griffin, Antony Hardy, Michael Jeger, Robert Luttik, Ambroise Martin, Simon More, Alicja Mortensen, Birgit Nørrung, Bernadette Ossendorp, Joe Perry, Josef Schlatter Kristen Sejrsen and Vittorio Silano. Correspondence: scer@efsa.europa.eu

2 Acknowledgement: The Panel wishes to thank the members of the Working Group on Uncertainty in Risk Assessment: Jan Alexander, Diane Benford, Peter Craig, Alessandro Di Domenico, Antony Hardy, Andrew Hart, Arie Havelaar, Michael Jeger, Robert Luttik, Ambroise Martin, Alicja Mortensen, Kaare Nielsen, Birgit Nørrung, Bernadette Ossendorp, Derek Renshaw, Josef Schlatter, Hans-Hermann Thulke, Natalie Von Goetz and Detlef Wölfle, for the preparatory work on this scientific opinion and the EFSA staff members Andrea Germini, Laura Martino, Caroline Merten, Olaf Mosbach-Schulz and Antony Smith for the support provided to this scientific opinion.

ADOPTED: dd mmmm yyyy PUBLISHED: dd mmmm yyyy AMENDED: dd mmmm yyyy

doi:10.2903/j.efsa.20YY.NNNN

Keywords: (max. seven keywords)

Requestor: add requesting party

Question number: EFSA-Q-2013-00738

Correspondence: xxx@efsa.europa.eu

Suggested citation: EFSA (European Food Safety Authority) [or EFSA ACRONYM Panel (EFSA Panel name)] [or EFSA Scientific Committee], 20YY. [Full title, including output category]. EFSA Journal 20YY;volume(issue):NNNN, 219 pp. doi:10.2903/j.efsa.20YY.NNNN

ISSN: 1831-4732

Summary

EFSA's role is to provide scientific advice on risks and other issues relating to food safety, to inform decision-making by the relevant authorities. A fundamental principle of EFSA's work is the requirement for transparency in the scientific basis for its advice, including scientific uncertainty. The Scientific Committee considers that *all EFSA scientific assessments must include consideration of uncertainties* and that application of this Guidance on uncertainty analysis should be unconditional for EFSA. Assessments must say clearly and unambiguously what uncertainties have been identified and what is their impact on the overall assessment outcome.

This document provides Guidance on how to characterise, document and explain all types of uncertainty arising in EFSA's scientific assessments. Uncertainty is defined as referring to *all types of limitations in the knowledge available to assessors at the time an assessment is conducted and within the time and resources available for the assessment*. The Guidance is applicable to all areas of EFSA and all types of scientific assessment, including risk assessment and all its constituent parts (hazard identification and characterisation, exposure assessment and risk characterisation). 'Assessor' is used as a general term for those providing scientific advice, including risk assessment, and 'decision-maker' for the recipients of the scientific advice, including risk managers.

The Guidance does not prescribe specific methods for uncertainty analysis but rather provides a harmonised and flexible framework within which different methods may be selected, according to the needs of each assessment. Worked examples are provided to illustrate different methods. For simplicity the examples are all based on a single case, an EFSA Statement on melamine that was published in 2008. [Section 1]

As a general principle, *assessors are responsible for characterising uncertainty, while decision-makers are responsible for resolving the impact of uncertainty on decisions*. Resolving the impact on decisions means deciding whether and in what way decision-making should take account of the uncertainty. Therefore, assessors need to inform decision-makers about scientific uncertainty when providing their advice.

In all types of assessment, the primary information on uncertainty needed by decision-makers is: *what is the range of possible answers, and how likely are they?* Assessors should also describe the nature and causes of the main sources of uncertainty, for use in communication with stakeholders and the public and to inform targeting of further work to reduce uncertainty, when needed.

The time and resources available for scientific assessment vary from hours in emergency situations to months or years for complex opinions. Therefore, this guidance provides a flexible framework for uncertainty analysis, so that assessors can select methods that are fit for purpose in each case.

Assessors and decision-makers should agree a *well-defined question for assessment*, such that a precise answer could be given if sufficient information were available. If that is not possible, or if the decision-makers' question is an open one, assessors should specify in a precise way what their conclusions refer to, as this is required for characterising the associated uncertainty. [Section 3]

Uncertainty may be expressed qualitatively (descriptive expression or ordinal scales) or quantitatively (individual values, bounds, ranges, or distributions). It is not necessary or possible to quantify *separately* every individual source of uncertainty affecting an assessment. However, *assessors should always aim to express overall uncertainty in quantitative terms to the extent that is scientifically achievable,* as is also stated in EFSA Guidance on Transparency and the Codex Working Principles for Risk Analysis. The principal reasons for this are the ambiguity of qualitative expressions, their tendency to imply value judgements outside the remit of assessors, and the fact that many decisions inherently imply quantitative comparisons (e.g. between exposure and hazard) and therefore require quantitative information on uncertainty. [Section 4]

When it is not possible to quantify uncertainty, assessors should avoid expressing their conclusions using words that could be interpreted as implying a probability statement (e.g. 'likely'). They should also avoid words with risk management connotations, such as 'negligible' or 'concern', unless scientific criteria have been agreed for their use. These restrictions apply only to language used in expressing scientific conclusions. [Section 3]

Key concepts for uncertainty analysis are introduced [Section 6]:

- *Uncertainty is personal and temporal*. The task of uncertainty analysis is to express the uncertainty of the assessors, at the time they conduct the assessment: there is no single 'true' uncertainty.
- It is important to *distinguish uncertainty and variability* and analyse them appropriately, because they have differing implications for decisions about options for managing risk and reducing uncertainty.
- *Dependencies* between different sources of uncertainty can greatly affect the overall uncertainty of the assessment outcome, so it is important to identify them and take them into account.
- *Evidence, agreement, confidence and conservatism* are related but distinct concepts. Measures of evidence and agreement may be useful in assessing uncertainty but are not sufficient alone. Confidence and conservatism are partial measures of uncertainty, and useful if adequately defined.
- *Probability* is the preferred measure for expressing uncertainty, as it quantifies the relative likelihood of alternative outcomes, which is what decision-makers need to know. All well-defined uncertainties can be quantified using subjective probability, which enables rigorous calculation of their combined impact.
- *Subjective judgment* of uncertainty is inherent and unavoidable in scientific assessment, but vulnerable to various psychological biases. These may be countered using formal methods for eliciting expert judgments, and combining uncertainties by calculation where possible.
- When assessors are unable to quantify some uncertainties individually, then those uncertainties cannot be included in quantitative characterisation of overall uncertainty. The quantitative assessment is then *conditional* on assumptions made for those uncertainties that could not be quantified, and it should be made clear that the likelihood of other conditions and outcomes is unknown.
- *Assessment questions* may be *quantitative* (estimation of a quantity) or *categorical* (e.g. yes/no questions). Many questions may usefully be divided into sub-questions for assessment. The structure of an assessment is subject to uncertainty, as well as its inputs, and both contribute to the uncertainty of the assessment output.

Assessors should be systematic in identifying uncertainties, checking each part of their assessment for different types of uncertainty, to minimise the risk of overlooking important uncertainties. All identified uncertainties should be documented, in an annex if desired, together with any initial assessment that is made to prioritise them for further analysis. [Section 7]

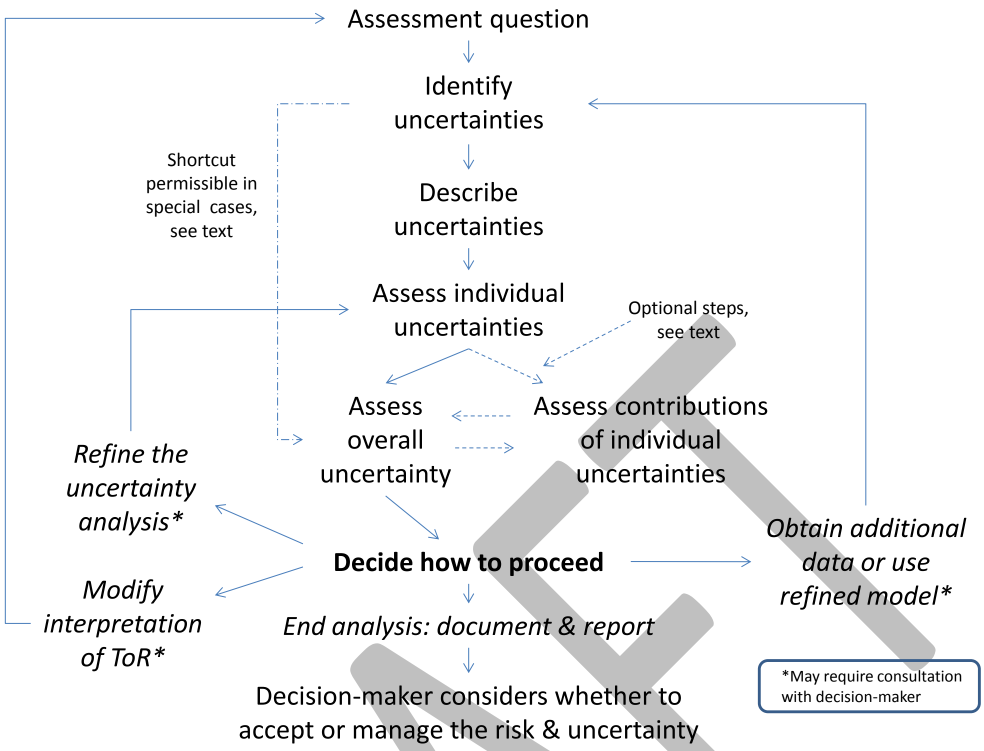
Six main steps in uncertainty analysis are distinguished: identifying uncertainties, describing uncertainties, assessing individual sources of uncertainty, assessing the overall impact of all identified uncertainties on the assessment output, assessing the relative contribution of individual uncertainties to overall uncertainty, and documentation and reporting. [Section 5]

Uncertainty analysis should be conducted in a flexible, iterative manner, as illustrated in Figure S.1, rather than a fixed set of tiers. Analysis starts with simple approaches and is then refined as far as is needed or possible within the time available. Some steps may be reduced or omitted in *emergency situations* and in *routine assessments* with standardised provision for uncertainty (e.g. default assessment factors), when suitably calibrated.

Sensitivity analysis should be used to help target refinement on those sources of uncertainty where it will contribute most. Consequently, in many assessments, different uncertainties will be analysed at different levels of refinement, which must be integrated in the overall characterisation of uncertainty.

Uncertainty analysis plays an important role in decisions about whether and how far to refine the overall assessment, and in what way (Figure S.1). Therefore, *uncertainty analysis should begin early in the assessment process*, and not be left to the end. [Section 8]

Figure S.1: Iterative approach for uncertainty analysis. ToR = Terms of Reference for the assessment.



Within the framework provided by Figure S.1, assessors should select methods that meet the needs of their assessment. The Guidance describes a selection of qualitative and quantitative methods and illustrates their application to the melamine example. The qualitative methods are [Section 9]:

- *Descriptive approaches*, using narrative phrases or text to describe uncertainties.
- *Ordinal scales*, characterising uncertainties using an ordered scale of categories with qualitative definitions (e.g. high, medium or low uncertainty).
- *Uncertainty matrices*, providing standardised rules for combining two or more ordinal scales describing different aspects or dimensions of uncertainty.
- *NUSAP method*, using a set of ordinal scales to characterise different dimensions of each source of uncertainty, and its influence on the assessment outcome, and plotting these together to indicate which uncertainties contribute most to the uncertainty of the assessment outcome.
- *Uncertainty tables for quantitative questions*, listing sources of uncertainty affecting a quantitative question and assessing their individual and combined impacts on the uncertainty of the assessment outcome on an ordinal scale.
- *Uncertainty tables for categorical questions*, listing lines of evidence contributing to answering a categorical question, identifying their strengths and weaknesses, and expressing the uncertainty of the answer to the question.

The quantitative methods reviewed are:

- *Quantitative uncertainty tables*, similar to the qualitative versions but expressing uncertainty on scales with quantitative definitions.
- *Interval analysis*, computing a range of values for the output of a risk calculation based on specified ranges for the individual inputs.

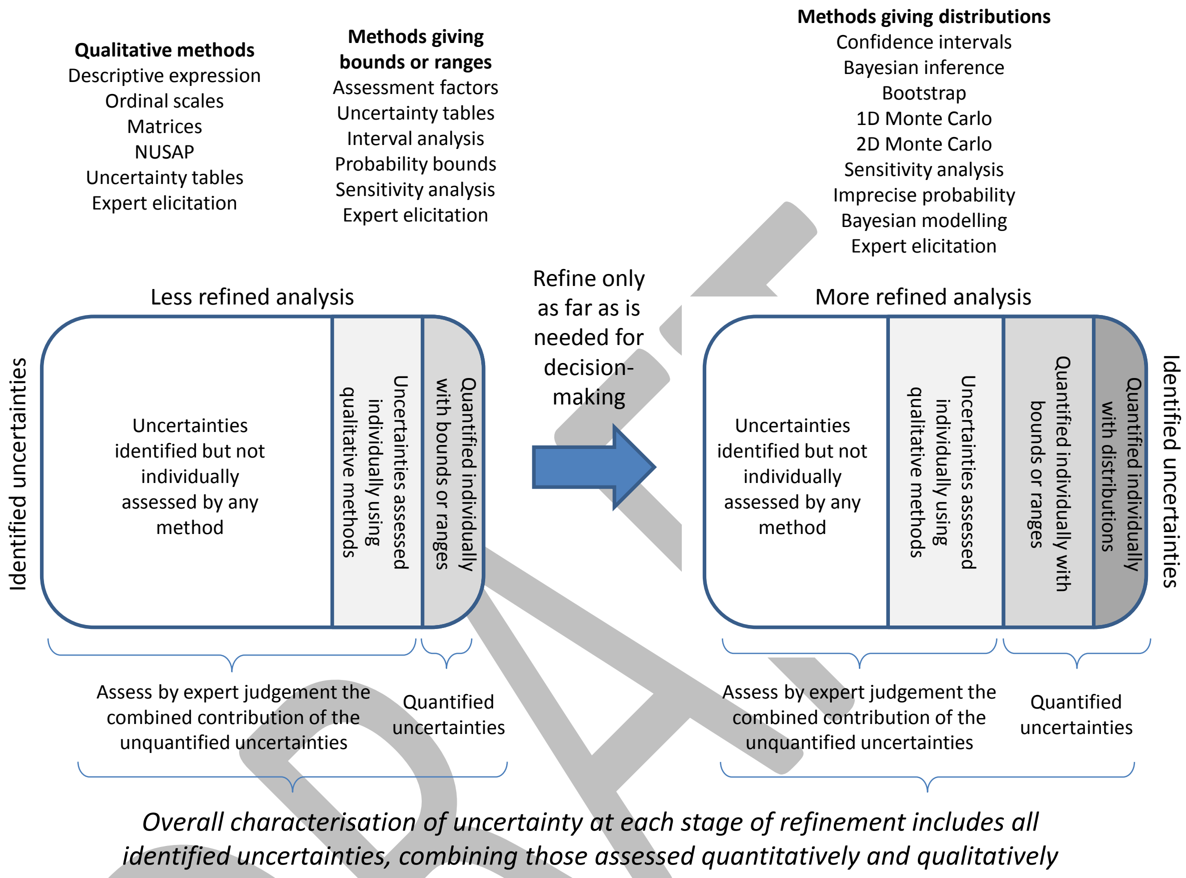
- *Expert knowledge elicitation* (EKE), a collection of *formal* and *informal* methods for quantification of expert judgements of uncertainty, about an assessment input or output, using subjective probability.
- *Confidence intervals* quantifying uncertainty about parameters in a statistical model of variability on the basis of data.
- *The bootstrap*, quantifying uncertainty about parameters in a statistical model of variability on the basis of data.
- *Bayesian inference*, quantifying uncertainty about parameters in a statistical model of variability on the basis of data and expert judgements about the values of the parameters.
- *Probability bounds analysis*, a general method for combining limited probability specifications about inputs in order to make a limited probability specification about the output of a risk calculation.
- *Monte Carlo simulation*, taking random samples from probability distributions representing uncertainty and/or variability to: (i) combine uncertainty about several inputs in the risk calculation by numerical simulation when analytical solutions are not available; (ii) carry out certain kinds of sensitivity analysis.
- *Deterministic calculations with conservative assumptions* are a common approach to uncertainty and variability in EFSA assessments. They include default values, assessment factors and decision criteria ('trigger values') which are generic and applicable to many assessments, as well as conservative assumptions and adjustments that are specific to particular cases.
- *Sensitivity Analysis*, a suite of methods for assessing the sensitivity of the output of the risk calculation (or an intermediate value) to the inputs and to choices made expressing uncertainty about inputs. It has multiple objectives: (i) to help prioritise uncertainties for quantification: (ii) to help prioritise uncertainties for collecting additional data; (iii) to investigate sensitivity of final output to assumptions made; (iv) to investigate sensitivity of final uncertainty to assumptions made.
- *Other quantitative methods* described more briefly: uncertainty expressed in terms of possibilities, imprecise probabilities, and Bayesian modelling.

All of the methods reviewed have stronger and weaker aspects. Qualitative methods score better on criteria related to simplicity and ease of use but less well on criteria related to technical rigour and meaning of the output, while the reverse tends to apply to quantitative methods. It would be premature to give prescriptive guidance on the choice of methods, apart from the general need to be quantitative where possible, as most methods have not yet been tried in sufficient EFSA assessments to form conclusions on their usefulness. More specific guidance may be given when more experience is gained. Until then, the following *strategy for method selection* is suggested [Section 9.3]:

1. Identify the uncertainties affecting the assessment.
2. Decide which classes of methods will be used in the initial assessment: usually this will include qualitative expression and bounds or ranges, but sometimes also distributions.
3. Within each class of methods to be used, consider which of the available methods are best suited to the assessment in hand.
4. Check which *steps* of uncertainty analysis are addressed by the chosen method in each class. Choose additional methods to address the remaining steps.
5. Carry out the uncertainty analysis and review the results. Refine the analysis iteratively until it is sufficient to support decision-making.
6. Document in a concise and clear way all of the uncertainties identified and how they have been addressed in the assessment.

The final output of uncertainty analysis should be an *overall characterisation of uncertainty* that takes all identified uncertainties into account. In this final step the contribution of those uncertainties that have been quantified individually with those that have been assessed qualitatively and those that have not been individually assessed by any method. This concept is illustrated in Figure S.2. [Section 10]

Figure S.2: Illustration of the methods options available for uncertainty analysis at lower and higher levels of refinement, and the process for overall characterisation of uncertainty.



Overall uncertainty should be characterised in terms of how different the assessment outcome might be and how likely that is, and quantified to the extent that is scientifically achievable. This should include those uncertainties that have been quantified individually, and also the additional uncertainties that have been assessed qualitatively or not individually assessed by any method. There are several ways in which the contribution of the additional uncertainties can be quantified and incorporated into the assessment [Section 10]:

1. If the some of the additional uncertainties could not be quantified individually, then they cannot be included in the overall quantitative assessment. In such cases, the assessor should still quantify those that they can and combine them with the uncertainties that have been quantified individually, using the methods described in the following steps. They should make clear to the decision-maker that this is an incomplete picture of the identified uncertainties, and conditional on whatever assumptions have been made about those uncertainties that remain unquantified.

2. If the assessors judge that the additional uncertainties are so unimportant that, collectively, they would make no difference to the bound, range or distribution obtained for the uncertainties that have been quantified individually, then the latter can be taken as representing the overall uncertainty.

3. Estimate by informal expert judgement what size of adjustment to the outcome of the assessment would be needed to allow for the effect of the additional uncertainties, expressed as a distribution or range. This is equivalent to the well-established practice of using case-specific assessment

factors to allow for extra sources of uncertainty. If the additional uncertainties are large enough to influence decision-making, consider using formal rather than informal elicitation to quantify them.

4. Combine the estimated contribution of the additional uncertainties with that of those uncertainties that have been quantified individually. Do this by calculation if possible, taking account of potential dependencies between them.

5. If the additional uncertainties cannot be combined with the rest of the analysis by calculation, then this must be done by expert judgement. This is much less rigorous than calculation, but still much better than ignoring the additional uncertainties. In this case one option is to quantify overall uncertainty using a *standard scale of probability ranges* [Section 10.3], if these provide sufficient information for decision-making.

6. When assessors cannot provide even a conditional bound or range for overall uncertainty, one option may be to present quantitative estimates for one or more possible scenarios, provided their limitations are made clear to decision-makers. Another option is to characterise overall uncertainty qualitatively, using descriptive expression or ordinal scales. However, as above, the assessor should avoid any language that implies a probability judgement.

The basis for the assessment of overall uncertainty must be *documented and justified*. The nature and cause of any uncertainties that remain unquantified must be described, so that decision-makers can consider what strategies to adopt. [Section 10]

The methods and results of all steps of the uncertainty analysis should be *reported fully and transparently*, in keeping with EFSA's (2012) Guidance on Transparency, and placed in a separate section within the main document of the assessment it relates to. Wherever statistical methods have been used, reporting of these should follow EFSA's (2014) Guidance on Statistical Reporting. A layered approach to reporting is recommended, to address the needs of different audiences and enable readers to access easily the different levels of information they require. [Section 11]

Various arguments have been made both for and against *communicating uncertainty* to the general public, but there is little empirical evidence to support either view or to define best practice. From EFSA's perspective, communicating scientific uncertainties is of fundamental importance to its core mandate, reaffirming EFSA's role in the Risk Analysis process. Therefore further work is recommended to test approaches for handling uncertainty in public communications and incorporate them in EFSA's Handbook on Risk Communication. [Section 12]

In conclusion, this draft Guidance provides a framework and principles for uncertainty analysis, with the flexibility for assessors to select different methods to suit the needs of each assessment. It is proposed that *every EFSA Panel and EFSA Units that produce scientific outputs should apply the draft Guidance to at least one assessment during an initial trial period, involving relevant decision-makers and supported by specialists in uncertainty analysis where needed*. When the trial period is completed and any resulting improvements to the Guidance Document have been agreed, uncertainty analysis will be *unconditional* for EFSA Panels and staff and must be embedded into scientific assessment in all areas of EFSA's work.

The final Guidance should be implemented in a *staged process*, starting by focussing on uncertainties specific to individual assessments. The implications for standardised assessment procedures should be considered over a longer period, as part of the normal process for evolving EFSA approaches. Where appropriate, this should be done in *consultation* with international partners and the wider scientific community. [Section 13]

Table of contents

1. Introduction

1.1. Background and Terms of Reference as provided by EFSA

Background

The EFSA Science Strategy for the period 2012-2016 identifies four strategic objectives: i) further develop excellence of EFSA's scientific advice, ii) optimise the use of risk assessment capacity in the EU, iii) develop and harmonise methodologies and approaches to assess risks associated with the food chain, and iv) strengthen the scientific evidence for risk assessment and risk monitoring. The first and third of these objectives underline the importance of characterising in a harmonised way the uncertainties underlying in EFSA risk assessments, and communicating these uncertainties and their potential impact on the decisions to be made in a transparent manner.

In December 2006, the EFSA Scientific Committee adopted its opinion related to uncertainties in dietary exposure assessment, recommending a tiered approach to analysing uncertainties (1/ qualitative, 2/ deterministic, 3/ probabilistic) and proposing a tabular format to facilitate qualitative evaluation and communication of uncertainties. At that time, the Scientific Committee "strongly encouraged" EFSA Panels to incorporate the systematic evaluation of uncertainties in their risk assessment and to communicate it clearly in their opinions.

During its inaugural Plenary meeting 23-24 July 2012, the Scientific Committee set as one of its priorities to continue working on uncertainty and expand the scope of the previously published guidance to cover the whole risk assessment process.

Terms of reference

The European Food Safety Authority requests the Scientific Committee to establish an overarching working group to develop guidance on how to characterise, document and explain uncertainties in risk assessment. The guidance should cover uncertainties related to the various steps of the risk assessment, i.e. hazard identification and characterisation, exposure assessment and risk characterisation. The working group will aim as far as possible at developing a harmonised framework applicable to all relevant working areas of EFSA. The Scientific Committee is requested to demonstrate the applicability of the proposed framework with case studies.

When preparing its guidance, the Scientific Committee is requested to consider the work already done by the EFSA Panels and other organisations, e.g. WHO, OIE.

1.2. Interpretation of Terms of Reference

The Terms of Reference (ToR) require a framework applicable to all relevant working areas of EFSA. As some areas of EFSA conduct types of assessment other than risk assessment, e.g. benefit and efficacy assessments, the Scientific Committee decided to develop guidance applicable to all types of scientific assessment in EFSA.

Therefore, wherever this document refers to scientific assessment, risk assessment is included, and 'assessors' is used as a general term including risk assessors. Similarly, wherever this document refers to 'decision-making', risk management is included, and 'decision-makers' should be understood as including risk managers and others making policy decisions.

1.3. Definition of uncertainty

Uncertainty is a familiar concept in everyday language, and may be used as a noun to refer to the state of being uncertain, or to something that makes one feel uncertain. The adjective 'uncertain' may be used to indicate that something is unknown, not definite or not able to be relied on or, when applied to a person, that they are not completely sure or confident of something (Oxford Dictionaries, 2015). Its meaning in everyday language is generally understood: for example, the weather tomorrow is uncertain, because we are not sure how it will turn out. In science and statistics, we are familiar with concepts such as measurement uncertainty and sampling uncertainty, and that weaknesses in methodological quality are a source of uncertainty. General types of uncertainty that are common in EFSA assessments are outlined in Section 7.

In the context of risk assessment, various formal definitions have been offered for the word 'uncertainty'. For chemical risk assessment, IPCS (2004) defined uncertainty as 'imperfect knowledge concerning the present or future state of an organism, system, or (sub) population under consideration'. Similarly, EFSA's (2011) guidance on environmental risk assessment of plant pests defines uncertainty as 'inability to determine the true state of affairs of a system'. In EFSA's previous guidance on uncertainties in chemical exposure assessment, uncertainty was described as resulting from limitations in scientific knowledge (EFSA, 2006a) while EFSA's BIOHAZ Panel has defined uncertainty as 'the expression of lack of knowledge that can be reduced by additional data or information.' (EFSA, 2012a). The US National Research Council's Committee on Improving Risk Analysis Approaches defines uncertainty as 'lack or incompleteness of information' (NRC, 2009). Recently, the EU non-food scientific committees SCHER, SCENIHR and SCCS (2013) described uncertainty as 'the expression of inadequate knowledge'. The common theme emerging from these and other definitions is that uncertainty refers to limitations of knowledge. It is also implicit in these definitions that uncertainty relates to the state of knowledge for a particular assessment, conducted at a particular time (the personal and temporal nature of uncertainty is discussed further in Section 7).

> In this document, uncertainty is used as a general term referring to *all types of limitations in the knowledge available to assessors at the time an assessment is conducted and within the time and resources agreed for the assessment*.

There are many sources and types of uncertainty in scientific assessment. Cataloguing these can be helpful when identifying the uncertainties affecting a particular assessment, and is discussed further in Section 7.

1.4. Scope, audience and degree of obligation

The mandate for this document is to provide guidance on how to characterise, document and explain all types of uncertainty arising in EFSA's scientific assessments. The Guidance is aimed at all those contributing to EFSA assessments and provides a harmonised, but flexible framework that is applicable to all areas of EFSA and all types of scientific assessment, including risk assessment. It should be used alongside other cross-cutting guidance on EFSA's approaches to scientific assessment including, but not limited to, existing guidance on transparency, systematic review, expert knowledge elicitation and statistical reporting (EFSA, 2009, 2010, 2014a, 2014b) and forthcoming guidance on weight-of-evidence assessment[3], biological relevance[4] and EFSA's Prometheus project[5].

> The Scientific Committee considers that all EFSA scientific assessments must include consideration of uncertainties. Therefore the application of this guidance document is unconditional for EFSA. For reasons of transparency and in line with EFSA 2006, the assessments must say what uncertainties have been identified and what their impact on the overall assessment outcome is. This must be reported clearly and unambiguously.

This document provides guidance on overall principles and a menu (toolbox) of different approaches and methods which can be used to help assessors to systematically identify, characterise, explain and account for uncertainties at different stages of the assessment process. For brevity, we refer to these processes collectively as 'uncertainty analysis'. This also describes how methods and steps can be

[3] Guidance on the use of the Weight of Evidence Approach in Scientific Assessments, EFSA-Q-2015-00007.
[4] Self-tasking mandate proposed to EFSA by the Scientific Committee for developing guidance for the identification of biological relevance of adverse positive health effects from experimental & human studies, EFSA-Q-2014-00746.
[5] PRO-METH-EU-S: Promoting Methods for Evidence Use in Science, EFSA-Q-2015-00106.

combined in an efficient and integrated assessment. The reader is referred to other sources for technical details on the implementation and use of each method.

The Scientific Committee emphasises that assessors do not have to use every method but the guidance is intended to help the selection of a suitable method to use at an appropriate point in the scientific assessment. This document aims at reviewing the general applicability of principles and approaches to EFSA's work. It does not critically assess specific applications of those methods by EFSA or other bodies, such as existing or new approaches to uncertainty in chemical hazard characterisation, as this would require in-depth assessment by experts from the subject area concerned.

Uncertainties in decision-making, and specifically in risk management, are outside the scope of EFSA and of this Guidance, as are uncertainties in the framing of the question for scientific assessment. When uncertainties about the meaning of an assessment question are detected, they should be referred to the decision-makers for clarification, which is likely to be an iterative process requiring discussion between assessors and decision-makers.

The primary audience for the document comprises all those contributing to EFSA's scientific assessments. Some sections will also be of particular interest to other groups, for example Chapters 3 and 12 are especially relevant for decision-makers and Chapter 12 for communications specialists.

2. Approach taken to develop this Guidance

The approach taken to developing this Guidance was as follows. A Working Group was established, comprising members of EFSA's Scientific Committee and its supporting staff, a Panel member or staff member nominated by each area of EFSA's work, some additional experts with experience in uncertainty analysis (identified and invited in accordance with EFSA procedures), and an EFSA communications specialist. Activities carried out by the Working Group included: a survey of uncertainties encountered by different EFSA Panels and Units and their approaches for dealing with them (which were taken into account when reviewing applicable methods); consideration of approaches that deal with uncertainty described in existing guidance documents of EFSA, of other bodies and in the scientific literature; meetings with selected risk managers in the European Commission and communications specialists from EFSA's Advisory Forum; and a public consultation on a Draft of the Guidance Document. These activities informed three main strands of work by the Working Group: development of the harmonised framework and guidance contained in the main chapters of this Guidance; development of annex sections focussed on different methods that can be used in uncertainty analysis; and development of illustrative examples using a common case study.

When evaluating the potential of different methods of uncertainty analysis for use in EFSA's work, the Working Group considered two primary aspects. First, the Working Group identified which of the main steps of uncertainty analysis (introduced in Section 5) each method can contribute to. Second, the Working Group assessed each method against a set of criteria which it established for describing the nature of each method and evaluating the contribution it could make. The criteria used to evaluate the methods were as follows:

- Evidence of current acceptance
- Expertise needed to conduct
- Time needed
- Theoretical basis
- Degree/extent of subjectivity
- Method of propagation
- Treatment of uncertainty and variability
- Meaning of output
- Transparency and reproducibility
- Ease of understanding for non-specialist

Definitions for these criteria are shown in Section 9.3 where the different methods are reviewed.

2.1. Case study

Worked examples are provided in Annexes to the Guidance to illustrate different steps in uncertainty analysis and different methods for addressing them. To increase the coherence of the document a single case study was selected enabling people to compare the different methods, based on an EFSA Statement on melamine that was published in 2008 (EFSA, 2008). While this is an example from chemical risk assessment for human health, the principles and methodologies illustrated by the examples are general and could in principle be applied to any other area of EFSA's work, although the details of implementation would vary.

The EFSA (2008) statement was selected for the case study in this guidance because it is short, which facilitates extraction of the key information and identification of the uncertainties and makes it accessible for readers of this guidance who would like more details, and also because it incorporates a range of uncertainties.

An introduction to the melamine case study is provided in Annex A, together with examples of output from different methods used in uncertainty analysis. Details of how the example outputs were generated are presented in Annex B, together with short descriptions of each method. It is emphasised that the case study is provided for the purpose of illustration only, is limited to the information that was available in 2008, and should not be interpreted as contradicting the subsequent full risk assessment of melamine in food and feed (EFSA, 2010).

3. Roles of assessors and decision-makers in addressing uncertainty

Some of the literature that is cited in this section refers to risk assessment, risk assessors and risk managers, but the principles apply equally to other types of scientific assessment and to the more general roles of assessor and decision-maker.

Risk analysis is the general framework for most of EFSA's work including food safety, import risk analysis and pest risk analysis, all of which consider risk analysis as comprising three distinct but closely linked and interacting parts: risk assessment, risk management and risk communication (EFSA, 2012b). Basic principles for addressing uncertainty in risk analysis are stated in the Codex Working Principles for Risk Analysis:

- 'Constraints, uncertainties and assumptions having an impact on the risk assessment should be explicitly considered at each step in the risk assessment and documented in a transparent manner'
- 'Responsibility for resolving the impact of uncertainty on the risk management decision lies with the *risk manager*, not the risk assessors' (Codex, 2015).

These principles apply equally to the treatment of uncertainty in other areas of science and decision-making. Thus, in general, assessors are responsible for characterising uncertainty and decision-makers are responsible for resolving the impact of uncertainty on decisions. Resolving the impact on decisions means deciding whether and in what way decision-making should be altered to take account of the uncertainty.

This division of roles is rational: assessing scientific uncertainty requires scientific expertise, while resolving the impact of uncertainty on decision-making involves weighing the scientific assessment against other considerations, such as economics, law and societal values, which require different expertise. The weighing of these different considerations is defined in Article 3 of the EU Food Regulation 178/2002 as risk management. The Food Regulation establishes EFSA with responsibility for scientific assessment on food safety, and for communication on risks, while the Commission and Member States are responsible for risk management and for communicating on risk management measures. In more general terms, assessing and communicating about scientific uncertainty is the responsibility of EFSA, while decision-making and communicating on management measures is the responsibility of others.

Although risk assessment and risk management are conceptually distinct activities (NRC, 1983, p. 7), they should not be isolated – interaction between them is essential (NRC, 1996, p. 6) and needs to be conducted efficiently. Discussions with risk managers during the preparation of this Guidance identified opportunities for improving this interaction, particularly with regard to specification of the question for assessment and expression of uncertainty in conclusions (see below).

3.1. Information required for decision-making

Given the division of responsibilities between assessors and decision-makers, it is important to consider what information decision-makers need about uncertainty. Scientific assessment is aimed at answering questions from managers about risks and other issues, to inform managers' decisions on how to manage them. Uncertainty refers to limitations in knowledge, which are always present to some degree. This means scientific knowledge about the answer to the manager's question will be limited, so in general a range of answers will be possible. Therefore the decision-maker needs to know the range of possible answers, so they can consider whether any of them would imply risk of undesirable management outcomes (e.g. adverse effects). Decision-maker's questions relate to real-world problems that they have responsibility for managing. Therefore, when the range of possible answers includes undesirable outcomes, the decision-maker needs information on how likely they are, so they can weigh options for management action against other relevant considerations (economic, legal, etc.). This includes the option of provisional measures when adverse outcomes are possible but uncertain (the precautionary principle, as described in Article 7 of the Food Regulation). Therefore, decision-makers need assessors to provide information on the range and likelihood of possible answers to questions submitted for scientific assessment.

Some EFSA work comprises forms of scientific assessment that do not directly address specific risks or outcomes. For example, EFSA is sometimes asked to review the state of scientific knowledge in a particular area. Conclusions from such a review may influence the subsequent actions of decision makers. Scientific knowledge is never complete, so the conclusions are always uncertain to some degree and other conclusions might be possible. Therefore, again, managers need information about how different the alternative conclusions might be, and how likely they are, as this may have implications for decision-making.

In summary, in all types of assessment, the primary information on uncertainty needed by decision-makers is: what is the range of possible answers, and how likely are they? In addition, decision-makers need to decide whether to commission further investigation or analysis to reduce uncertainty, and may need to communicate with other stakeholders and the public about the reasons for uncertainty (especially if it affects their decisions). Therefore, decision-makers also need information on the main sources of uncertainty affecting the outcomes of assessment, scientific options for reducing those uncertainties, and the time and resources required by those options.

3.2. Time and resource constraints

Decision-makers generally need information within specified limits of resources and time, including the extreme case of emergency situations where advice might be required within weeks, days or even hours. To be fit for purpose, therefore, EFSA's approaches to assessing uncertainty must include options for different levels of resource and different timescales, and/or methods that can be implemented at different levels of detail/refinement, to fit different timescales and levels of resource. Consideration of uncertainty is always required, even in emergency situations, because reduced time and resource for scientific assessment increases uncertainty and its potential implications for decision-making.

3.3. Defining questions for assessment

Questions for assessment must be specified in precise terms. Imprecise questions make it hard for assessors to focus their efforts efficiently, and may result in the answer not being useful to managers, or even being misleading. If the meaning of the question is imprecise or ambiguous (could be interpreted in different ways by different people), more answers become possible, hence adding to the overall uncertainty of the response. Assessors and decision-makers should therefore aim to agree on a formulation of the question such that a precise answer could be given if sufficient information were available. For example, 'what will the exchange rate of euros and dollars be in 2016' is an imprecise question: it is necessary to specify which type of dollars, whether the rate is from euros to dollars or dollars to euros, what date in 2016, and on which exchange (e.g. the European Central Bank). Similarly, terms such as 'typical', 'worst case' or 'high consumer' must be clearly defined. If the question relates to a quantity, then that quantity and the population and time period of interest must be specified. . If the question refers to the occurrence of a state, condition or process (e.g. is chemical X genotoxic) then that state, condition or process must be unambiguously specified. When there is uncertainty about the meaning of an assessment question, assessors should consult with the decision-maker to clarify it. If that is not possible, assessors must specify their interpretation of the question in precise terms both at the start of the assessment and when reporting conclusions.

Occasionally, decision-makers pose open questions to EFSA, for example a request to review the state of scientific knowledge on a particular subject (e.g. chicken welfare). In such cases, the assessors and decision-makers should identify the principal conclusions of the assessment (those that may have implications for decision-making) and the assessor should specify in precise terms what each conclusion refers to, such that its uncertainty can be assessed and communicated.

3.4. Acceptable level of uncertainty

The Food Regulation and other EU law relating to risks of different types frequently refer to the need to 'ensure' protection from adverse outcomes. The word 'ensure' implies a societal requirement for some degree of certainty that adverse outcomes will not occur, or be managed within acceptable limits. Complete certainty is never possible, however. Deciding how much certainty is required or, equivalently, what level of uncertainty would warrant precautionary action, is the responsibility of decision-makers, not assessors. It may be helpful if the decision-maker can specify in advance how much uncertainty is acceptable for a particular question, e.g. about whether an outcome of interest will exceed a given level. This is because the required level of certainty has implications for what outputs should be produced from uncertainty analysis, e.g. what probability levels should be used for confidence intervals. Also, it may reduce the need for the assessor to consult with the decision-maker during the assessment, when considering how far to refine the assessment (see Section 8). Often, however, the decision-maker will not be able to specify in advance the level of certainty that is sought or the level of uncertainty that is acceptable. In general, therefore, assessors will need to provide more information to decision-makers, e.g. confidence intervals with a range of probabilities, so that decision-makers can consider at a later stage what level of uncertainty to accept.

3.5. Expression of uncertainty in assessment conclusions

In its Opinion on risk terminology, the EFSA Scientific Committee (SC) recommended that 'Scientific Panels should work towards more quantitative expressions of risk and uncertainty whenever possible, i.e. quantitative expression of the probability of the adverse effect and of any quantitative descriptors of that effect (e.g. duration), or the use of verbal terms with quantitative definitions. The associated uncertainties should always be made clear, to reduce the risk of over-precise interpretation' (EFSA, 2012b). The reasons for quantifying uncertainty are discussed in Section 4, together with an overview of different forms of qualitative and quantitative expression. This section considers the implications for interaction between assessor and decision-maker in relation to the assessment conclusions.

Probability is the natural metric for quantifying uncertainty and can be applied to any well-defined uncertainty. This means that both the question for assessment and the eventual conclusion also need to be well-defined, in order for its uncertainty to be assessed. For example, in order to say whether an estimate might be an over- or under-estimate, and to what degree, it is necessary to specify what the assessment is required to estimate. Therefore, if this has not been specified precisely in the terms of reference (see Section 3.4), assessors should provide a series of alternative estimates (e.g. for different percentiles of the population), each with a characterisation of uncertainty, so that the decision-maker can choose which to act on.

Sometimes it may not be possible to quantify uncertainty (Section 6.7). In such cases, assessors must avoid using any language that could be interpreted as implying a probability statement (e.g. "likely", "unlikely", etc.), as this would be misleading. In addition, as stated previously by the Scientific Committee (EFSA, 2012b), the assessor should avoid any verbal expressions that have risk management connotations in everyday language, such us "negligible" and "concern". When used in EFSA opinions, such expressions should be clearly defined with objective scientific criteria so as to avoid the impression that assessors are making risk management judgments (EFSA, 2012b). Some time may be required to develop explicit criteria in some parts of EFSA's work, where such terms are currently part of standard assessment procedure (see also Section 8.3). The Scientific Committee notes that these restrictions on the use of verbal expressions apply only to *scientific conclusions*, and not to the everyday use of such words in other parts of EFSA outputs.

The remainder of this Guidance Document sets out a framework and principles for assessing uncertainty using methods and procedures that address the needs identified above, including the need to distinguish appropriately between risk assessment and risk management, and the requirement for flexibility to operate within varying limitations on timescale and resource so that each individual assessment can be fit for purpose.

4. Qualitative and quantitative approaches to expressing uncertainty

4.1. Types of qualitative and quantitative expression

Expression of uncertainty requires two components: expression of alternative outcomes or states, and some expression of their relative likelihoods. Quantitative approaches express the alternative outcomes on a numerical scale, if they refer to a quantity, and express likelihood on a numerical scale. Qualitative approaches express range of outcomes and relative likelihoods using words, categories or labels, and do not provide a numerical scale.

It is useful to distinguish descriptive expression and ordinal scales as different categories of qualitative expression: descriptive expression allows free choice of language to characterise uncertainty, while ordinal scales provide a standardised and ordered scale of qualitative expressions facilitating comparison of different uncertainties. It is also useful to distinguish different categories of quantitative expression, which differ in the extent to which they quantify uncertainty: partial quantification requires less information or judgements but may be sufficient for decision-making in some assessments, whereas other cases may require fuller quantification.

Examples of important types of qualitative and quantitative expression of uncertainty are shown in the box below.

Differing approaches to expressing uncertainty

Qualitative expression

Descriptive expression: Uncertainty described in narrative text or characterised using verbal terms without any quantitative definition.

Ordinal scale: Uncertainty described by ordered categories, where the magnitude of the difference between categories is not quantified.

Quantitative expression

Individual values: Uncertainty partially quantified by specifying a number of possible values, without specifying what other values are possible or setting upper or lower limits.

Bound: Uncertainty partially quantified by specifying either an upper limit or a lower limit on a quantitative scale, but not both.

Range: Uncertainty partially quantified by specifying both a lower and upper limit on a quantitative scale, without expressing the relative likelihoods of values within the limits.

Bound/Range with Probability: Uncertainty partially quantified by specifying a bound or range with an accompanying probability.

Distribution: Uncertainty fully quantified by specifying the relative likelihood (probability) of alternative values on a quantitative scale.

When using bounds or ranges it is important to specify whether the limits are absolute, i.e. contain all possible values, or contain the 'true' value with a specified probability (e.g. 95%), or contain the true value with at least a specified probability (e.g. 95% or more). A 95% confidence interval is an example of a range with a specified probability. When an assessment factor (e.g. for species differences in toxicity) is said to be 'conservative', this implies that it is a bound that has sufficient probability of covering the uncertainty the factor is supposed to address, although the level of probability is often not specified. Sensitivity analysis is often conducted with alternative individual values for an assessment input, to explore their impact on the assessment output.

As well as differing in the amount of information or judgements they require, the different categories of quantitative expression differ in the information they provide to decision-makers. Individual values give only examples of possible values, although often accompanied by a qualitative expression of where they lie in the possible range. An upper bound provides a conservative assessment with specified degree of conservativism, while a range provides both a conservative assessment and an indication of the potential for less adverse outcomes and therefore the potential benefits of reducing uncertainty. A distribution provides information on the likelihood of all possible outcomes: this is useful when the decision-maker needs information on the relative likelihoods of multiple outcomes with differing levels of severity.

Assessments using probability distributions to characterise variability and/or uncertainty are often referred to as 'probabilistic'. Sometimes, the term 'deterministic' is applied to assessments using individual values without probabilities (e.g. EFSA 2006, IPCS 2008, ECHA 2008 but not IPCS 2014 which prefers 'non-probabilistic').

The term 'semi-quantitative' is not used in this Guidance. Elsewhere in the literature it is sometimes applied to methods that are, in some sense, intermediate between fully qualitative and fully quantitative approaches. This might be considered to include ordinal scales with qualitative definitions, since the categories have a defined order but the magnitude of differences between categories is undefined. Sometimes, 'semi-quantitative' is used to describe an assessment that comprises a mixture of qualitative and quantitative approaches or an ordinal assessment in which the numbers are not on a ratio scale.

4.2. Advantages of quantitative expression

The Codex Working Principles on Risk Analysis (Codex 2015) state that 'Expression of uncertainty or variability in risk estimates may be qualitative or quantitative, but should be quantified to the extent that is scientifically achievable'. A similar statement is included in EFSA's (2009) guidance on transparency. Advantages and disadvantages of qualitative and quantitative expression are discussed in the EFSA (2012b) Scientific Committee Opinion on risk terminology, which recommends that EFSA should work towards more quantitative expression of both risk and uncertainty.

It is not necessary, and indeed not possible, to quantify *separately* all the sources of uncertainty affecting an assessment. However, it is important that the *combined effect* of all *identified* sources of uncertainty is expressed in quantitative terms, to the extent that this is scientifically achievable. The principal reasons for this are as follows:

- Qualitative expressions are ambiguous: the same word or phrase means different things to different people. This has been demonstrated repeatedly (e.g. Theil 2002 and Morgan 2014). As a result, decision-makers may misinterpret the assessors' assessment of uncertainty, which will result in sub-optimal decisions. Stakeholders may also misinterpret qualitative expressions of uncertainty, which may result in overconfidence or unnecessary alarm.

- Decision-making often depends on quantitative comparisons, for example, whether a risk exceeds some acceptable level, or whether benefits outweigh costs. Therefore, decision-makers need to know whether the uncertainty affecting an assessment is large enough to alter the comparison in question, e.g. whether the uncertainties around an estimated exposure of 10 and an estimated safe dose of 20 are large enough that the exposure could in reality exceed the safe dose. This requires uncertainty to be expressed in terms of how different each estimate might be, and how likely that is.

- If assessors provide only best estimates and a qualitative expression of the uncertainty, decision-makers will have to make their own quantitative interpretation of how different the estimated values might be. Even if this is not conscious or explicit, such a judgement will be implied when the decision is made. Therefore a quantitative judgement is, in effect, unavoidable, and this is better made by assessors, since they are better placed to understand the uncertainties affecting the assessment and judge their effect on its outcome.

- Qualitative expressions often imply, or may be interpreted as implying, judgements about the implications of uncertainty for decision-making, which are outside the remit of EFSA. For example, 'low uncertainty' tends to imply that the uncertainty is too small to influence decision-making, and 'no concern' implies firmly that this is the case. Qualitative terms can be used if they are based on scientific criteria, so that assessors are not making risk management judgements (EFSA, 2012b). However, for transparency they need to be accompanied by quantitative expression of uncertainty, to make clear what likelihood of adverse outcomes is being accepted.

- When different assessors work on the same assessment, e.g. in a Working Group, they cannot reliably understand each other's assessment of uncertainty if it is expressed qualitatively. Assessors may assess uncertainty differently yet agree on a single qualitative expression, because they interpret it differently. Expressing uncertainties in terms of their quantitative impact on the assessment outcome will reveal such differences of opinion, enabling a more rigorous discussion and hence improving the quality of the final assessment.

For these reasons, assessors should always express overall uncertainty in quantitative terms to the extent that is scientifically achievable. This is in agreement with the requirement stated in the Codex Working Principles for Risk Analysis (Codex 2015) and in the EFSA Guidance on Transparency (EFSA, 2010). However, qualitative methods still have an important role to play, including in prioritising which uncertainties to quantify individually, and for informing judgements about overall uncertainty (see Section 10).

A range of methods for assessing and combining individual uncertainties are reviewed in Section 9. Overall characterisation of uncertainty combines the results of quantitative analysis with expert judgement of the contribution of other uncertainties that were identified but not quantified individually. This should include consideration of any uncertainties associated with assumptions or judgements made in the quantitative analysis (e.g. choice of distributions, treatment of dependencies). Overall characterisation of the identified uncertainties is discussed in detail in Section 10.

The limit to how much quantification is scientifically achievable, and the consequences of this for reporting to decision-makers, are discussed in Sections 6.7 and 6.8.

These recommendations refer to the immediate output of the assessment, and do not necessarily imply that all communications of that output should also be quantitative. It is recognised that quantitative information raises significant issues for communication with stakeholders and the public. These issues and options for addressing them are discussed in Section 12.

5. Main steps of uncertainty analysis

Conducting an uncertainty analysis generally requires a number of main steps: identifying the uncertainties that affect the assessment, describing and explaining them, characterising their effect on the assessment outcome, and documenting the analysis. For uncertainties affecting inputs to the assessment, an additional step is needed to characterise the uncertainty of the input, before determining the effect of that on the assessment output. It is often important to assess the relative contribution of different sources of uncertainty to overall uncertainty, either by sensitivity analysis or expert judgement, which adds another step. This results in a total of six main steps, as shown in the box below. These steps are often applied in an iterative manner, in which more detailed assessment is focussed on the most important sources of uncertainty. This is explained in Section 8, which also identifies some defined situations where some of the steps may be omitted.

Main steps in uncertainty analysis.

Identifying uncertainties. Systematic examination of all parts of the assessment to identify as many sources of uncertainty as possible (see Section 7).

Describing uncertainties. Qualitative description of source, cause and nature of identified uncertainties in terms comprehensible to non-specialists (see Section 9.1.1).

Assessing individual sources of uncertainty. Estimation of the magnitude of each source of uncertainty in terms of its impact on the part of the assessment it directly affects (see Section 9).

Assessing the overall impact of all identified uncertainties on the assessment output, taking account of dependencies. Calculation or expert judgement of the combined impact of multiple uncertainties on the assessment output, in terms of the alternative answers they might lead to and how likely they are (see Sections 9 and 10).

Assessing the relative contribution of individual uncertainties to overall uncertainty. Calculation (sensitivity analysis) or expert judgement of the relative contribution of different sources of uncertainty to uncertainty of the assessment outcome, based on the relation between the results of Steps 4 and 5 (for sensitivity analysis, see Section 9.2.3).

Documenting and reporting the uncertainty analysis, in a form that fully documents the analysis and its results and meets the general requirements for documentation and reporting of EFSA assessments (see Section 11).

6. Key concepts for uncertainty analysis

6.1. Personal and temporal nature of uncertainty

The uncertainty affecting a scientific assessment is a function of the knowledge available to those conducting the assessment, at the time that it is conducted. If additional relevant information exists elsewhere but is not accessible, or cannot be analysed within the time permitted for assessment, those limitations are part of the uncertainty of the assessment even though more information may be known to others. This is one of the reasons why uncertainty tends to be higher when a rapid assessment is required, e.g. in emergency situations.

Expressions of uncertainty are therefore *personal* and *temporal.* The task of uncertainty analysis is to express *the uncertainty of the assessors, at the time they conduct the assessment*: there is no single ‘true’ uncertainty.

Individuals within a group of assessors will have different expertise and experience. This is acknowledged in EFSA’s work by establishing Panels and WGs consisting of experts with complementary expertise. However, the personal nature of knowledge and uncertainty means it is legitimate, and to be expected, that different experts within a group may give differing judgements of uncertainty for the same assessment question. Structured approaches to eliciting judgements and characterising uncertainty should reveal the reasons for differing views and provide opportunities for convergence. Some degree of compromise may therefore be involved in reaching the consensus conclusion that is generally produced by an EFSA Panel or Working Group. Alternatively, expert elicitation methodology offers several different techniques to aggregate the judgements of multiple experts (see EFSA, 2014a). Where significant differences of view remain, EFSA procedures provide for the expression of Minority Opinions.

The personal, subjective nature of knowledge and uncertainty also contributes to cases where different groups of assessors reach diverging opinions on the same issue. Where this involves EFSA and other EU or Member State bodies, Article 30 of the Food Regulation includes provision for resolving or clarifying such differences and identifying the uncertainties involved.

6.2. Uncertainty and variability

The relation between uncertainty and variability is often discussed. Uncertainty refers to the state of knowledge, whereas variability refers to actual variation or heterogeneity in the real world. It follows that uncertainty may be altered (either reduced or increased) by further research, whereas variability cannot, because it refers to real differences that will not be altered by obtaining more knowledge. Distinguishing uncertainty and variability is therefore of practical importance, because it informs decisions about investing resources in research to gather more information. This applies both when the assessment is qualitative and when it is quantitative.

Variability is a property of the real world, but our knowledge of it is generally incomplete. Therefore there is generally uncertainty *about* variability. Some types of variability, for example the variation in human body weight, are much less uncertain than others, e.g. the nature and degree of genetic variation in different populations.

When there is interest in an individual instance within a population of individuals or outcomes, variability in the population causes uncertainty about the individual instance. For example, even if we were certain a coin is fair, i.e. that when tossed an infinite number of times it would land on heads precisely half the time, nevertheless at any point there is uncertainty about the outcome of the next toss. Uncertainty caused by variability is sometimes referred to as 'aleatory' uncertainty and distinguished from 'epistemic' uncertainty, which refers to other types of limitations in knowledge (e.g. Vose, 2008). How variability should be treated in an assessment therefore depends on whether the assessment question refers to the population or to a particular member of that population. Many assessment questions refer to populations, e.g. what proportion of a population will experience a given level of exposure. An important example of a risk assessment element relating to a particular instance of a variable quantity is provided by the default assessment factors used in chemical risk assessment, as discussed in Annex B15.

6.3. Dependencies

Variables are often inter-dependent. For example, body weight tends to be positively correlated with height and both are correlated with age. It is important to take account of dependencies between variables in assessment, so that different combinations of values are considered in proportion to their expected frequency and unrealistic or impossible combinations are excluded.

Uncertainties can also be inter-dependent. This happens when learning more about one aspect of an assessment would alter the assessor's uncertainty about another aspect. An example that may be surprising is that the uncertainties of the population mean and variance for a normal distribution are inter-dependent, when estimated from a measured sample. This is because, if one discovered that the true mean was a long way from the sample mean, this would change the uncertainty of the variance (because high variances would become more likely). Such dependencies can greatly affect the overall uncertainty of the assessment outcome, so it is important to identify them and take them into account. This is true not only when using distributions but also in qualitative assessment or when using bounds or ranges to take account of uncertainty. For example, it is important to avoid combining multiple conservative assumptions which, while individually plausible, are unlikely to occur together.

6.4. Evidence, agreement, confidence, conservatism & uncertainty

Evidence, agreement (e.g. between experts), confidence, conservatism and uncertainty are related but distinct concepts. Increasing the amount, quality, consistency and relevance of evidence or the degree of agreement between experts tends to increase confidence and decrease uncertainty.

However, the relationship between these concepts is complex and variable. For example, new evidence sometimes reveals new issues that were previously not considered, so confidence decreases and uncertainty increases. As another example, two experimental studies may provide the same amount and quality of evidence for the same measurement, but differing confidence intervals.

Because the amount, quality, consistency and relevance of evidence and the degree of agreement are related to the degree of uncertainty, measures of evidence and agreement may be useful in assessing uncertainty (e.g. Mastrandrea et al., 2010). However, such measures do not, on their own, provide sufficient information for decision-making. As discussed in earlier sections, what matters for decision-making is the range and likelihood of possible outcomes.

Levels of confidence are often used as an expression of the probability that a conclusion is correct. Sometimes they represent a subjective judgement (e.g. the confidence scale of IPCC (Mastrandrea et al, 2010)). In other cases it has a quantitative meaning, e.g. in frequentist statistics, a confidence interval is a region within which an estimated value would lie in a specified proportion of occasions (e.g. 95%) if the experiment and/or statistical analysis were repeated an infinite number of times. In Bayesian statistics, a credibility interval is the region within which the real value would lie with a specified probability. However, even a quantitative confidence or credibility interval may not, on its own, provide sufficient information for decision-making, as it provides no information on the distribution of possible outcomes within the interval, or on how far outside the interval the distribution extends.

In some areas of EFSA's work, assessments may be intended to overestimate the severity and/or frequency of an adverse outcome (e.g. overestimate exposure or hazard and consequently risk). Such assessments are sometimes described as 'conservative'. Generally it is intended that the degree of overestimation is sufficient to allow for uncertainty, such that the likelihood (probability) of outcomes that are more adverse than the estimated outcome is appropriately low. Thus an assertion of conservatism requires three elements: specification of the target quantity (what severity and frequency of outcome is of interest); specification of what probability of more adverse outcomes is acceptable (the required level of confidence); and estimation of the target quantity such that outcomes more adverse than the target level are expected with the specified probability. The first two elements should be determined by decision-makers, while the third element is the responsibility of assessors. Asserting that an estimate is conservative without specifying the target quantity and required level of confidence conflates the roles of decision-maker and assessor and is not transparent, because it implies acceptance of some likelihood of more adverse outcomes without making clear what that likelihood is. Therefore, if the decision-maker wishes to receive a single conservative estimate, they could specify the target quantity and required level of confidence when setting the terms of reference for the assessment, as has been proposed by IPCS (2014) for chemical hazard characterisation. Alternatively, the assessor could provide a range of estimates with different levels of confidence, so the final choice remains with the decision-maker.

6.5. Expert judgement

Assessing uncertainty relies on subjective judgement, because different people have different knowledge and experience and therefore different uncertainty. Indeed, this is true of science in general. Choosing a model or chain of reasoning for the assessment involves subjective judgements. The choice of assessment scenarios is subjective, as is the decision to use a default assessment factor or the choice of a non-standard factor specific to the case in hand. In probabilistic assessments, the choice of distributions and assumptions about their dependence or independence are subjective. Even when working with 'hard' data, assessing the suitability of those data is subjective. Even ideal data are rarely truly representative, so implicit or explicit judgements about extrapolation are needed (e.g. from one country to another or the EU as a whole, between age groups or sexes, and from the past to the present or future). When these various types of choices are made, the assessor implicitly considers the range of alternatives for each choice and how well they represent what is known about the problem in hand: in other words, their uncertainty. Thus the subjective judgement of uncertainty is fundamental, ubiquitous and unavoidable in scientific assessment.

The use of subjective judgement is not a weakness of science; on the contrary, well-reasoned judgements are a key ingredient of good science. However, subjective judgements are made by psychological processes that are vulnerable to various cognitive biases such as over-confidence (e.g. in small data sets), anchoring and adjustment and availability (e.g. the most familiar or recent publications)(Kahneman et al. 1982). Formal expert knowledge elicitation methods (see Section 9.2.1.3 and EFSA, 2014a) are designed to counter these biases and should be used when appropriate, especially for important uncertainties that have significant implications for decision-making. The principles on which those formal methods are based – e.g. the need to review and revise potentially over-confident judgements – should also be considered in more informal expert judgement, to reduce the risk of bias.

It has been demonstrated that people often perform poorly at judging combinations of probabilities (Gigerenzer, 2002). This implies they will perform poorly at judging how multiple uncertainties in an assessment combine. Therefore, this Guidance recommends that uncertainties should be combined by calculation when possible, even if the calculation is very simple (e.g. a series of what-if calculations with alternative assumptions), to help inform judgements about the overall uncertainty from the identified sources. When doing this, assessors should take account of the additional uncertainties associated with choosing the calculation model, and avoid using combinations of inputs that could not occur together in reality. If uncertainties are combined by expert judgement, then the assessor should try to take account of the added uncertainty that this introduces (e.g. widen their overall range or distribution until they judge that it represents the range of results they consider plausible).

6.6. Probability

When dealing with uncertainty, decision-makers need to know how different the outcomes might be and how likely they are. The natural quantitative measure for this is probability, which expresses the relative likelihood of different outcomes.

There are two major views about the scope of probability as a method for quantifying uncertainty. One, sometimes known as the frequentist view, considers that the use of probability should be restricted to uncertainties caused by variability and should not be applied to uncertainties caused by limitations in knowledge. As a result, it offers no solution for characterising many types of uncertainty. The other, subjectivist (Bayesian), view asserts that a probability is a direct personal statement of uncertainty and that all well-defined uncertainties can be quantified using probability. This Guidance takes the latter view.

A key advantage of subjective probability as a quantitative measure of uncertainty is that there are ways to enhance comparability when probabilities are expressed by different individuals. Informally, an individual can compare any particular uncertainty to situations where there is a shared understanding of what different levels of probability mean: tossing a fair coin, rolling fair dice, etc. Formally, an operational definition of probability was developed by de Finetti (1937) and Savage (1954), in part to ensure comparability. This formal definition leads to a second key advantage of probability. It shows that the extensive mathematical and computational tools of probability can legitimately be applied to subjective probabilities. In particular, those tools aid expression of judgements about combinations of uncertainties (e.g. in different parts of an assessment) which the human mind would otherwise find difficult. In other words, it can help the assessor make more rational judgements about questions such as: if I can express my uncertainty about hazard and exposure, then what should my uncertainty be about risk?

For these reasons, this Guidance encourages the use of probability to express uncertainty, except when qualitative expression of uncertainty or a quantitative range is sufficient for decision-making, or when it is felt that it is too difficult to quantify uncertainty (see Section 6.7).

Probabilities need not necessarily be expressed fully or precisely. More limited probability statements may be easier for assessors to provide, and may be sufficient for decision-making. A simple limited form is a *probability bound*, which states that the probability is greater than some specified value, and/or less than a specified value. It may be simpler for assessors to judge that an adverse outcome has less than a given probability, rather than giving a specific probability, and if that probability is low enough it may be sufficient for decision-making. As a result, probability bounds may be useful when using expert judgement to characterise overall uncertainty (see Section 10).

6.7. Unquantified uncertainties

In general, uncertainty should be quantified as far as is scientifically achievable (Codex, 2015). From the perspective of subjective probability it is always possible to quantify *well-defined* uncertainties (de Finetti 1937, Walley 1990). An uncertain quantity or proposition is well-defined if it is possible to specify it in such a way that it would be possible to determine it with certainty if an appropriate observation or measurement could be made, at least in principle (even if it making that observation would never be feasible in practice). In everyday language, it is possible to give a subjective probability for anything that one could bet on, that is, if it would be possible in principle to determine without ambiguity whether the bet was won or lost. For example, one can bet on the final score of a sports event, but not on whether it will be a 'good game' unless that could be defined without ambiguity. If this is not possible, then it is not appropriate to quantify the uncertainty using subjective probability. Such an uncertainty is literally *unquantifiable*.

Making probability judgements can be difficult, and training will be needed to facilitate the uptake of these approaches in EFSA. Sometimes assessors may find it difficult to give a distribution for a well-defined uncertainty, but nevertheless find it possible to give a range or bound, either with a specified probability (e.g. a 90% bound) or with a bounded probability (e.g. a limit with at least 90% probability). This may be sufficient, if the decision-maker considers that the bound excludes unacceptable outcomes with sufficient probability. This is conceptually similar to the default factors and conservative estimates used in many current assessments, which are interpreted as if they were bounds with sufficient (though unspecified) probability for decision-making.

An assessor may still be unable to quantify a well-defined uncertainty, if they cannot provide any quantitative expression of the magnitude of an uncertainty or its impact on the assessment. In such cases it is, for that assessor, not scientifically achievable to quantify the uncertainty, with the evidence available to them at the time of the assessment. Uncertainties that are not quantified for either reason (inability to define or inability to quantify) are sometimes referred to as 'deep' uncertainties and are most likely to arise in problems that are novel or very complex (Stirling, 2010).

It is important to note that it is not necessary to quantify every source of uncertainty individually in order to quantify overall uncertainty. Provided that all the uncertainties are at least *potentially* quantifiable individually, then it may be possible for the assessor to quantify their combined effect. However, if there is *even one* source of uncertainty that the assessor would be unable to quantify individually, then it is in principle not possible to include them when quantifying overall uncertainty. This is because the one uncertainty that cannot be quantified could potentially alter the assessment outcome to any extent and with unknown probability. Therefore it is very important for the assessor to identify any sources of uncertainty that they could not quantify, as they will not be able to include these when quantifying overall uncertainty. Their quantification of overall uncertainty will then be *conditional* on assumptions made in the assessment regarding the uncertainties that they could not quantify. All assessments are conditional to some degree, so this concept is discussed in more detail below.

6.8. Conditional assessments

Conditional assessment is an important option for dealing with identified uncertainties that are not quantified. Before considering this, it is important to recognise that all expressions of uncertainty are conditional to some extent. Because uncertainty is intrinsically personal and temporal, all expressions of uncertainty are conditional on the assessors who provide them and the knowledge available to them at the time of assessment. Decision-makers should be aware of this, and take account of it when comparing different assessments of the same issue. In additional, expression of overall uncertainty is always conditional on the assessor having identified all relevant uncertainties.

When one or more of the identified uncertainties are not quantified in the expression of overall uncertainty, this becomes conditional on the assumptions made for the uncertainties that remain unquantified. Often, these assumptions may take the form of a scenario. An approach of this type was used in EFSA's (2008) statement on melamine, which reported that exposure estimates for a high exposure scenario exceeded the Tolerable Daily Intake (TDI), but stated that it was unknown whether such a scenario may occur in Europe.

Conditional assessment is a potentially important strategy for helping EFSA Panels work towards more quantitative expression of uncertainty, as previously recommended by the Scientific Committee (EFSA, 2012a). Many of EFSA's assessments deal with uncertainty primarily through the use of default assessment factors and conservative assumptions or scenarios: the melamine statement (EFSA 2008) is an example of this. Full quantification of uncertainty for such assessments is challenging, because it requires considering not only uncertainties affecting the data being used in the assessment (which might be termed specific uncertainties), but also uncertainty about how the default factors, assumptions and scenarios and the calculation in which they are used relate to conditions and processes in the real world (which might be termed generic uncertainties). The generic uncertainties relate to standard procedures that are used in multiple assessments of the same type; therefore they may only need to be quantified once. It is clearly desirable to move towards quantifying the generic uncertainties, for the general reasons discussed in Section 4, however they are accepted by assessors and decision-makers as being covered by the assessment approaches currently used. Therefore, a practical strategy may be to start by quantifying specific uncertainties affecting data used in individual assessments, *conditional* on current assessment factors, assumptions and scenarios, and move towards quantifying the generic uncertainties in the medium term, e.g. when guidance documents are revised (for further discussion of these issues, see Section 8.3).

Conditional assessments provide an incomplete quantification of uncertainty but may still be useful for decision-making, especially if the conditional element is something the decision-maker can influence (e.g. the effectiveness of management measures). If an assessment is conditional, the assessor should state the conditions for which uncertainty has been quantified and describe the nature and causes of any uncertainties that remain unquantified, and explain why they were not quantified. This is essential information for the decision-maker, who will need to consider the implications for decision-making. However, assessors should avoid making assessments conditional on uncertainties that could in principle be quantified, since this is the assessors' responsibility and should not be transferred to the decision-maker (see Section 3).

The assessor should communicate clearly to the decision-maker that the likelihood of other conditions is unknown (as in the melamine statement), and that the impact of some identified uncertainties has not been quantified, and avoid any language that implies a probability judgement about those issues (e.g. 'outside chance', 'cannot exclude', etc.). If the assessor feels able to use such language, this implies that they are in fact able to make a probability judgement. If so, they should express it quantitatively – for transparency, to avoid ambiguity, and to avoid the risk management connotations that verbal expressions often imply (Section 4).

6.9. Question type and assessment structure

It is useful for later parts of this guidance to introduce some terms that will be used to distinguish different types of assessment question and different aspects of assessment structure.

Assessment questions may be of two main types:

- **Quantitative questions** concern estimation of a quantity. Examples of such questions include estimation of exposure or a reference dose, the level of protein expression for a GM trait, the infective dose for a pathogen, etc.
- **Categorical questions** concern choices between two or more categories. Examples of such questions include hazard identification (does chemical X have the capability to cause effect Y?), mode of action, human relevance, adversity, the equivalence of GM traits and their non-GM counterparts, whether an animal pathogen will infect humans, etc.

Quantitative questions are sometimes be answered by direct measurement or expert judgement of the quantity in question. In other cases, the assessment will be some form of calculation involving a mathematical or statistical model. When the assessment is a calculation or model, it will be useful to distinguish three **assessment components**:

- **Assessment inputs**: inputs to the calculation or model, including any data, assessment factors, assumptions, expert judgements, or other types of input.
- **Assessment structure**: the structure of the calculation or model, i.e. how the inputs are combined to generate the assessment output. This could generally be written down as a mathematical equation or sequence of equations.
- **Assessment output**: the output of the model or calculation, i.e. the estimate it provides in answer to the assessment question.

Note that the assessment inputs and outputs for a quantitative question may be either variables or parameters:

- A **variable** is a quantity that takes multiple values in the real world.
- A **parameter** is a quantity that has a single true value. Parameters include quantities that are considered constant in the real world, and also quantities that are used to describe variability in a population (e.g. mean, standard deviation and percentiles).

Uncertainty about a parameter can be quantified by a single distribution, representing uncertainty about its single true value, whereas uncertainty about a variable can be quantified by distributions for the parameters that describe it.

Categorical questions are often addressed by a **weight of evidence** approach, where the assessment inputs may alternatively be referred to as *lines of evidence*, which are weighed against each other, usually by expert judgement, to arrive at the assessment output. Weight of evidence approaches will be considered in more detail under a separate mandate[6]. However, since the mandate for the present Guidance extends to all areas of EFSA's work, a qualitative approach to uncertainty in categorical questions is included (see Section 9.1.5). Uncertainty in categorical questions can also be addressed by quantitative models, such as Bayesian Belief Nets (BBNs), which are briefly referred to in Section 9.2.4 and have the same components as the models for quantitative questions (inputs, outputs and assessment structure).

Many assessment questions are sufficiently complex that they are, explicitly or implicitly, broken down into **sub-questions** for assessment. This can apply to both quantitative and categorical questions. Separate assessments (or sub-assessments) are then needed for each of the sub-questions. The division of risk assessment into exposure assessment and hazard assessment is a common example of this. Each sub-assessment has its own inputs, structure and output, and the output of sub-assessments become inputs for subsequent stages of assessment that are needed to answer the overall question. Consequently, assessing uncertainty for the overall question requires first assessing uncertainty for the sub-questions, which is then treated as uncertainty in inputs to the overall question. Note that a single overall question may involve a mixture of quantitative and categorical sub-questions.

7. Identification of uncertainties

The first step of uncertainty analysis is to identify uncertainties affecting the assessment. Although it will generally be efficient to concentrate the subsequent analysis on the most important uncertainties, the initial identification needs to be as comprehensive as possible to minimise the risk that important uncertainties will be overlooked. It is therefore recommended that, in general, a systematic and structured approach is taken to identifying uncertainties. This can be facilitated by having a structured classification of uncertainties according to their characteristics, that is, a typology of uncertainties.

Various approaches to classify uncertainties into a typology exist, ranging from practically-oriented lists of types of uncertainties encountered in a particular domain (e.g. EFSA 2006a) to more theoretically-based typologies (e.g. Hayes 2011, Regan et al. 2002a, Walker et al. 2003 and Knol et al. 2009). Others include Morgan and Henrion 1990, IPCS 2008 and many more. The main purposes of using a typology of uncertainties in risk assessment are to help identify, classify and describe the different uncertainties that may be relevant. Another important role of a typology is that it provides a structured, common framework and language for describing uncertainties. This facilitates effective communication during the assessment process, when reporting the finished assessment and when communicating it to decision-makers and stakeholders, and therefore contributes to increasing both the transparency and reproducibility of the risk assessment.

It is recommended to take a practical approach to identifying uncertainties in EFSA's work, rather than seek a theoretical classification. It is therefore recommended that assessors should be systematic in

[6] Guidance on the use of the Weight of Evidence Approach in Scientific Assessments, EFSA-Q-2015-00007.

searching for uncertainties affecting their assessment, by considering each part or component of their assessment in turn and checking whether different types of uncertainty are present. This is intended to minimise the risk of overlooking important sources of uncertainty. It is consistent with the Codex Working Principles for Risk Analysis (2015), which state that 'Constraints, uncertainties and assumptions having an impact on the risk assessment should be explicitly considered at each step in the risk assessment'.

Component refers to the part of the assessment where the uncertainty arises, i.e. the assessment inputs, assessment structure and, where present, sub-assessments (see Section 6.8). The nature of the assessment components varies between different parts of EFSA, due to the differences in the nature, content and structure of the assessments they do. Therefore, this guidance does not offer a general classification of components, but rather recommends that each area of EFSA should consider establishing a list of components for the main types of assessment done in their area. Where no such list is applicable, the assessor is responsible for ensuring that they consider all parts of their assessment when searching for sources of uncertainty.

Type refers to the nature and/or source of the uncertainty. Two general lists of types are proposed (Tables 1 and 2) which are thought to be applicable to most areas of EFSA's work. Table 1 lists types of uncertainty that commonly affect assessment inputs, while Table 2 lists types of uncertainty that commonly arise in relation to the structure of the assessment (i.e., uncertainties about how the assessment inputs should be combined to generate the assessment output, and about any missing inputs). In developing these Tables, priority has been given to maximising their practical usefulness to assessors in helping them identify uncertainties in their work, rather than to the philosophical rigour of the differentiation between types. As a result, assessors may find that some uncertainties could be placed in more than one type: this was considered of less importance than ensuring that each uncertainty can be placed in at least one type. Tables 1 and 2 also contain lists of questions that may be helpful to assessors when considering whether each type of uncertainty is present in their assessment. Both Tables refer primarily to assessments for quantitative questions. Many of the same sources of uncertainty apply to categorical questions, especially to lines of evidence that are quantitative, but the tables could be extended to include other types of uncertainty that are particularly relevant to categorical questions, e.g. regarding the relevance and provenance or pedigree of evidence.

Tables 1 and 2 are not intended to be prescriptive. Another example of an approach using a series of questions to help identify uncertainties has been developed by the BfR and a translation of this to English is provided in Annex B. EFSA Panels and Units may use other typologies or question lists, for example those cited earlier in this section, if they consider them to be better suited for their work, or adapt Tables 1 and 2 to reflect the uncertainties commonly encountered in their assessments.

If Tables 1 and 2 are used to identify uncertainties, it may be helpful to proceed in the following manner:

1. List any sub-questions into which the overall assessment is divided (e.g. exposure and hazard assessment, and any further sub-questions within these).
2. List all the inputs for each sub-question.
3. For each input, list which types of uncertainties it may be affected by. To be systematic, consider all the types shown in Table 1.
4. Identify which types of uncertainty affect the structure of each sub-question and the overall assessment (where the sub-questions are combined). To be systematic, consider all the types shown in Table 2.

When using typologies such as Tables 1 and 2 it may sometimes be difficult to decide which type of uncertainty some sources belong to. However, this is less important than identifying as many as possible of the potential sources of uncertainty that are present.

In many assessments, the number of individual sources identified may be large. It will generally be necessary to prioritise them in some way, to make the subsequent steps of analysis practical. Such prioritisation implies an initial screening assessment of all the identified uncertainties (equivalent to steps 3-5 of uncertainty analysis, see Section 5), to decide which to prioritise. Assessors must document all the uncertainties that are identified at least briefly, together with their initial screening assessment. This is necessary to improve the reliability of this initial assessment (reduce the chance of missing or underestimating important uncertainties), inform the assessors judgement of the overall uncertainty (which should take all identified uncertainties into account) and ensure a transparent record of the assessment. However, if the full list of uncertainties is long it may be more practical to place it in an annex or separate document, and list only the major uncertainties in the main assessment report or Opinion.

Some areas of EFSA undertake multiple assessments of very similar nature, with the same structure and types of inputs but differing data. This is especially true for assessments of regulated products where the types of data and assessment structure are prescribed by regulations or formal guidance. In such cases, it may be possible to establish a generic list of uncertainties that can be used as a starting point for each assessment without needing to be re-created. However, the assessor should always check whether the case in hand is affected by any additional uncertainties, which would need to be added to the generic list.

Table 1: Example of a practical typology to assist in identifying uncertainties affecting *assessment inputs* for quantitative questions. Individual EFSA Panels and Units may adapt this or adopt alternative typologies as appropriate, to meet the needs of their assessments.

Type/source of uncertainty	Questions that may help to identify uncertainties*
1. Ambiguity	Are all necessary aspects of any data, evidence or assumptions used in the assessment (including the quantity measured, the subjects or objects on which the measurements are conducted, and the time and location where the measurements were conducted) *adequately described*, or is some interpretation required?
2. Measurement uncertainty	What is the precision and accuracy of any measurements that have been used? Are there any censored data (e.g. non-detects)?
3. Sampling uncertainty	Is the input based on measurements made on a sample from a larger population? If yes: How was the sample collected? Was randomisation conducted? Was stratification needed or applied? Was the sampling biased in any way, e.g. by intentional or unintentional targeting of sampling? How large was the sample? How does this affect the uncertainty of the estimates used in the assessment?
4. Assumptions incl. default values	Is the input partly or wholly based on assumption (including default values) or expert judgement? If yes: What is the nature, quantity, relevance, reliability and quality of evidence available to support the assumption or judgement? How many experts contributed to the assumption or judgement, how relevant and extensive was their expertise and experience for making it, and to what extent did they agree? How might the assumption or judgement be affected by psychological biases such as over-confidence, anchoring, availability, group-think, etc.? Was any formal elicitation methodology used to counter this?
5. Extrapolation uncertainty	Are any data, evidence or assumptions used in the assessment (including the quantity they address, and the subjects or objects, time and location to which that quantity refers) *directly relevant* to what is needed for the assessment, or is some extrapolation required? If the input is based on measurements on a sample from a population, how closely relevant is the sampled population to the population or subpopulation of interest for the assessment? Is some extrapolation implied?
6. Distribution uncertainty	Is the input a distribution representing a quantity that is variable in the real world? If so, how closely does the chosen form of distribution (normal, lognormal etc.) represent the real pattern of variation? What alternative distributions could be considered?
7. Other uncertainties	Where the input is the output from a sub-question, has uncertainty been adequately characterised in assessing the sub-question? Is the input affected by any other sources of uncertainty that you can identify, or other reasons why the input might differ from the real quantity it represents?

Table 2: Example of a practical typology to assist in identifying uncertainties affecting *how the assessment inputs are combined* for quantitative questions. Individual EFSA Panels and Units may adapt this or adopt alternative typologies as appropriate, to meet the needs of their assessments.

Type/source of uncertainty	Questions that may help to identify uncertainties*
1. Ambiguity	If the assessment includes mathematical or statistical model(s) that were developed by others, are all aspects of them *adequately described*, or is some interpretation required?
2. Excluded factors	Are any potentially relevant factors or processes excluded? (e.g. excluded modifying factors, omitted sources of additional exposure or risk, etc.)
3. Relationship between components	Regarding those inputs that are included in the assessment: How closely does the combination of assessment inputs represent the way in which the real process operates? Are there alternative models that could be considered? Are there dependencies between variables affecting the question of interest? How different might they really be from what is assumed in the assessment?
4. Distribution uncertainty	Does the model include some fixed values representing quantities that are variable in the real world, e.g. default values or conservative assumptions? If so, are the percentiles at which those fixed values are set appropriate for the needs of the assessment, i.e. so that when considered together they provide an appropriate and known degree of conservatism in the overall assessment?
5. Evidence for the structure of the assessment	What is the nature, quantity, relevance, reliability and quality of evidence available to support the assumption or judgement? How many experts contributed to developing the structure of the assessment or model, how relevant and extensive was their expertise and experience for making it, and to what extent did they agree? How might the choices made in developing the assessment structure or model be affected by psychological biases such as over-confidence, anchoring, availability, group-think, etc.? Was any formal elicitation methodology used to counter this? Where the assessment involves two or more sub-questions, is the division into sub-questions and the way they are linked appropriate?
6. Comparisons with independent data	Is there any independent information, not used in constructing the assessment, with which intermediate or final outputs of the assessment may be compared? If so, consider the following: What uncertainties affect the independent information? Assess this by considering all the questions listed above for assessing the uncertainty of inputs. How closely does the independent information agree with output of the assessment to which it pertains, taking account of the uncertainty of each? What are the implications of this for your uncertainty about the assessment outputs?
7. Dependency between uncertainties	Are there dependencies between any of the uncertainties affecting the assessment and/or its inputs, or regarding factors that are excluded? If you learned more about any of them, would it alter your uncertainty about one or more of the others?
8. Other uncertainties	Is the assessment structure affected by any other sources of uncertainty that you can identify?

8. Scaling uncertainty analysis to the needs of the assessment

8.1. General approach

All aspects of scientific assessment, including uncertainty analysis, should be conducted at a level of scale and complexity that is proportionate to the needs of the problem and within the time and resources agreed with the decision-maker. This is often achieved by starting with simple methods and progressively refining the assessment until it provides sufficient information to support decision-making. In many frameworks for risk assessment, refinement consists of progressing through a number of distinct 'tiers', in which different methods and data are used.

There are two main levels of uncertainty analysis, qualitative and quantitative, with quantitative assessments being subdivided further into those using sets, bounds, ranges and distributions (Section 4). However, there is a wide range of possible methods at each level, of varying complexity, and different sources of uncertainty in the same assessment may be treated at different levels.

> This Guidance therefore recommends a flexible, iterative approach, which refines the uncertainty analysis progressively as far as is needed, rather than a fixed set of tiers. The approach can be scaled to any type of assessment problem, including emergency situations where a response is required within hours or days.

The principles of the iterative refinement approach are as follows:

1. In general, uncertainty analysis should start with a simple approach, unless it is evident at the outset that more complex approaches are needed. However, contrary to what was implied by EFSA (2006), a simple starting point need not necessarily use qualitative methods, if quantitative methods have been implemented in a way that makes them simple to use.
2. Uncertainty analysis should be refined as far as is needed to inform decision-making. This point is reached either when there is sufficient certainty about the assessment outcome for the decision-maker to make a decision with the level of certainty they require, or if it becomes apparent that achieving the desired level of uncertainty is unfeasible or too costly and the decision-maker decides instead to manage the uncertainty without further refinement of the analysis.
3. Refinements of the uncertainty analysis should be targeted on those sources of uncertainty where refinement will contribute most efficiently to improving the characterisation of uncertainty, taking account of the cost and feasibility of the refinement. Sensitivity analysis can help to identify these (see Section 9.2.3). This targeting of refinement means that, *in most assessments, different uncertainties will be analysed at different levels of refinement*.
4. The overall assessment of uncertainty must integrate the contributions of identified sources of uncertainties that have been expressed in different ways (e.g. qualitatively, with ranges, or with distributions). After each stage of refinement, this assessment of overall uncertainty must be updated to take account of the results of the refined analysis.

The process of iterative refinement is illustrated in Figure 1. The whole process flows from the assessment question, at the top of the figure. The next 3 steps identify and describe uncertainties relevant to the assessment and assess them individually. Assessing overall uncertainty is essential, but assessing the contributions of individual uncertainties to overall uncertainty is shown as an optional step. This is because some methods (e.g. Monte Carlo simulation) allow overall uncertainty to be assessed directly from the individual uncertainties, and if the overall uncertainty is too small to influence decision-making then it may not be important to separate their individual contributions. Some other methods (e.g. uncertainty tables) assess overall uncertainty by first assessing the contributions of individual uncertainties and then considering how they combine. These alternative options are illustrated by the three dashed arrows in the centre of the Figure.

A key point in the process is where a decision is made on how to proceed. If the decision-maker was able to specify in advance what degree of certainty they require, the assessor will be able to determine whether this has been achieved and, if so, end the uncertainty analysis and report the results. If the decision-maker has not specified what degree of certainty is required, one option for the assessor is to continue refining the assessment as far as is possible within the agreed time and resources and then report the results. Options for refinement include refining the uncertainty analysis, or obtaining additional data or using more sophisticated models with the aim of reducing uncertainty. The choice of refinement option should weight the expected benefits of each option against its cost in terms of time and resources. If the preferred refinement option would involve exceeding the agreed time or resources the assessor will need to consult with the decision-maker before proceeding. In some cases, the results emerging from the assessment might lead the assessor or decision-maker to consider modifying the Terms of Reference or their interpretation. For example if it became apparent that the risk or uncertainty was likely to be unacceptable, the decision-maker might wish to change the ToR to include assessment of possible mitigation or precautionary actions. If a change in the ToR is required, or a substantial change in their interpretation, the assessor may need to consult with the decision-maker to agree the change.

It is emphasised that it is not necessary to treat all uncertainties at the same level of refinement. Rather, the process of iterative refinement should enable the assessor to target more refined methods on those uncertainties where refinement is most beneficial. The consequence of this is that, as already stated, in most assessments, different uncertainties will be treated at different levels of refinement. Methods for combining the contributions of uncertainties treated at different levels are described in Section 10.

It can be seen from this discussion and Figure 1 that uncertainty analysis plays an important role in decisions about whether and how far to refine the overall assessment, and in what way. Therefore, uncertainty analysis should be an integral part of the overall assessment from its beginning, not added at the end of the process.

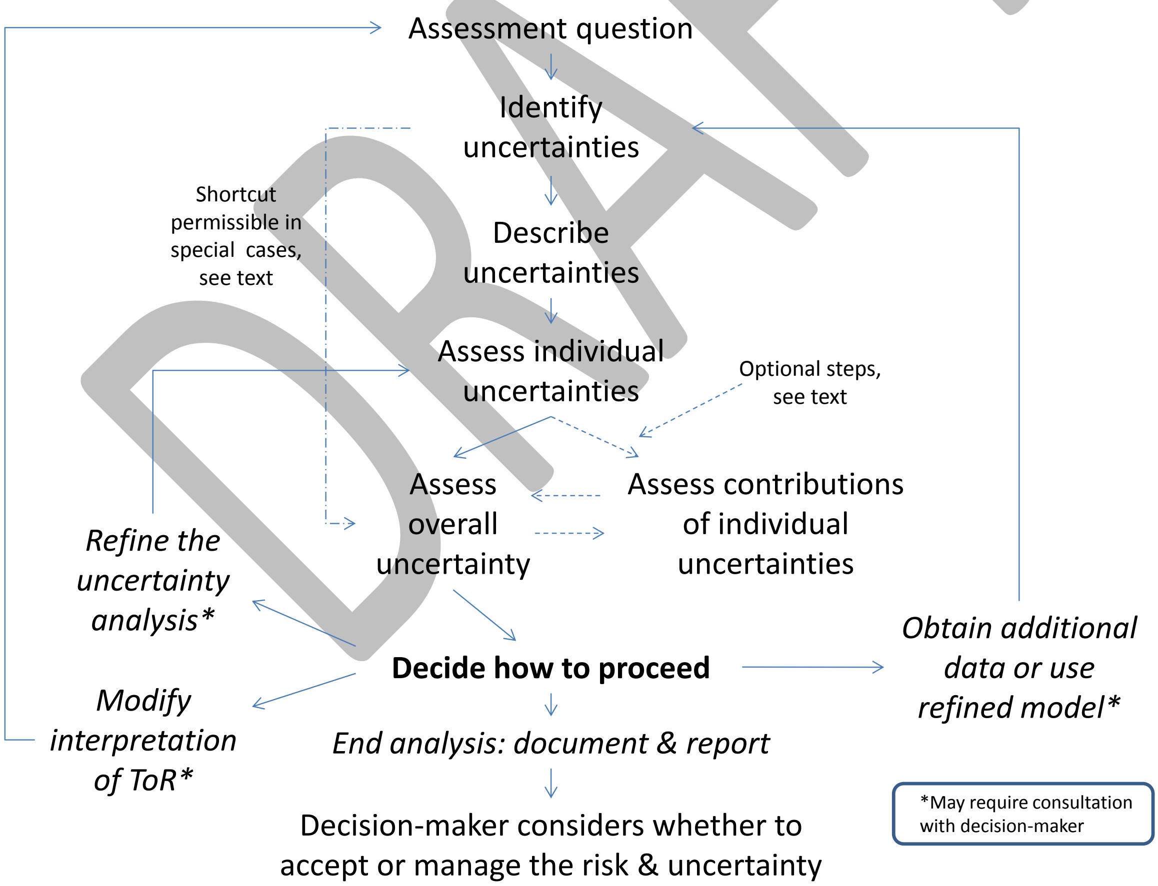


Figure 1: Iterative process for refining the uncertainty analysis, including shortcut for emergency situations and other special cases (see Section 8.1). ToR = Terms of Reference for the assessment.

8.2. Emergency situations

The iterative approach is highly flexible, enabling the scale and complexity of uncertainty analysis to be adapted to the needs of each assessment, including emergency situations where an initial assessment may be required within hours or days.

Every uncertainty analysis should include a systematic effort to identify all important uncertainties affecting the assessment, to reduce the risk of missing a major source of uncertainty that could substantially change the assessment conclusion. Even in emergency situations, some time should be spent on identifying uncertainties, and used in a manner that is most conducive to identifying the most important uncertainties (e.g. 'brainstorming' each of the main elements of the assessment in turn).

Every uncertainty analysis should quantify the combined impact of the identified uncertainties to the extent that is scientifically achievable. When time is severely limited, this may have to be done by expert judgement in which the contributions of individual uncertainties are assessed and combined without being individually expressed or documented. Note that such judgements are unavoidably implied when giving emergency advice, regardless of how the advice is expressed.

Provided the preceding requirements are met, uncertainty analysis in an emergency situation might *initially* be limited to a brief assessment by expert judgement of the overall impact of the identified uncertainties, without first assessing them individually. The overall impact should still be expressed quantitatively if scientifically achievable, in terms of the range of possible outcomes and their relative likelihoods expressed. This initial assessment should generally be followed by more detailed uncertainty analysis, including individual consideration of the most important uncertainties, after the initial assessment has been delivered to decision-makers.

8.3. Standard or default assessment procedures

Standard or default assessment procedures are common in many areas of EFSA's work, especially for regulated products, and are subject to periodic review. Some are agreed at international level. Most standard procedures these involve simple calculations using a combination of standard study data, default assessment factors and default values (see Annex B.7): for example, standard animal toxicity studies, default assessment factors for inter- and intra-species differences in toxicity, default values for body-weight, default values for consumption, and a legal limit or proposed level of use for concentration. These procedures are considered appropriate for routine use on multiple assessments because it is judged (implicitly or explicitly) that they are sufficiently conservative. This does not mean they will never underestimate risk, but that they will do so sufficiently rarely to be acceptable. This implies that, for each individual assessment, the probability of the standard procedure underestimating the risk is agreed by assessors and decision makers to be acceptable.

This approach is used, either implicitly or explicitly, in all areas of EFSA's work where standard procedures are used, including Thresholds of Toxicological Concern (TTC), first tier assessments of human and environmental risk for plant protection products, etc. Such procedures are compatible with the principles of uncertainty analysis described in the present Guidance, provided that the basis for them is justified and transparent. This requires that the level of conservatism provided by each standard procedure should be assessed by an appropriate uncertainty analysis following the procedure shown in Figure 1, quantified to the extent that is scientifically achievable, and documented. In addition, it is essential to specify what class of assessments each standard procedure is applicable to (similar to the domain of applicability for a QSAR). These steps can be regarded as 'calibrating' the level of conservatism for standard procedures, and a logical part of quality assurance in EFSA's work.

The documentation or guidance for a standard procedure should specify the assessment question, the standardised elements of the procedure (equation and default inputs), the type and quality of case-specific data to be provided, and the generic uncertainties considered when calibrating the level of conservatism. It is then the responsibility of assessors to check the applicability of all these elements to each new assessment. Any deviations, including provision of non-standard data, that would increase the uncertainties considered in the calibration or introduce additional uncertainties, will mean that it cannot be assumed that the calibrated level of conservatism and certainty will be achieved for that assessment. Assessing this requires identifying any increased or additional uncertainties, evaluating their impact on the overall uncertainty and conservatism of the assessment, and documenting that these things have been done. It therefore requires some of the steps in Figure 1, but not a full uncertainty analysis. However, in cases where this evaluation shows additional or increased uncertainties, the standard assessment procedure is not applicable, and the assessor will need to carry out a case-specific assessment and uncertainty analysis, following the procedure in Figure 1.

The principles outlined above were recognised by the Scientific Committee in their earlier Guidance on uncertainty in exposure assessment (EFSA, 2006) and also by WHO/IPCS (2008), both of which refer to calibrated standard procedures as 'Tier zero' screening assessments. EFSA (2006) included a recommendation that each Panel should review whether standard procedures in its area of work provided adequately for uncertainty. Where a standard procedure has not previously been calibrated by an appropriate uncertainty analysis, providing this may require substantial work. However, as noted in above, existing standard procedures are currently accepted by assessors and decision-makers. Therefore, it will be practical to start by quantifying specific uncertainties affecting data used in individual assessments, conditional on the existing standard procedure, and move towards quantifying the generic uncertainties and thus calibrating the procedure over a longer period as part of the normal process for progressive improvement of EFSA's approaches. Where the existing procedure is part of an internationally-agreed protocol, any changes will need to be made in consultation with relevant international partners and the broader scientific community.

9. Qualitative and quantitative methods for use in uncertainty analysis

Details of individual methods are to be found in Annex B, with special emphasis given to their strengths and weaknesses and situations where their application is more suitable. Tables summarising the detailed evaluations of the methods may be found at the end of the chapter.

9.1. Qualitative methods

Qualitative methods characterise uncertainty using descriptive expression or ordinal scales, without quantitative definitions (Section 4). They range from informal description of uncertainty to formal, structured approaches, aimed at facilitating consistency of approach between and within both assessors and assessments. In contrast to quantitative methods (see Section 9.2), they lack any well-developed or rigorous theoretical basis, relying instead on careful use of language and expert judgement.

The Scientific Committee identified the following broad types of qualitative methods that can be used in uncertainty analysis:

- **Descriptive methods**, using narrative phrases or text to describe uncertainties.
- **Ordinal scales**, characterising uncertainties using an ordered scale of categories with qualitative definitions (e.g. high, medium or low uncertainty).
- **Uncertainty matrices**, providing standardised rules for combining two or more ordinal scales describing different aspects or dimensions of uncertainty.
- **NUSAP method**, using a set of ordinal scales to characterise different dimensions of each source of uncertainty, and its influence on the assessment outcome, and plotting these together to indicate which uncertainties contribute most to the uncertainty of the assessment outcome.
- **Uncertainty tables for quantitative questions**, a template for listing sources of uncertainty affecting a quantitative question and assessing their individual and combined impacts on the uncertainty of the assessment outcome.
- **Uncertainty tables for categorical questions**, a template for listing lines of evidence contributing to answering a categorical question, identifying their strengths and weaknesses, and expressing the uncertainty of the answer to the question. (The difference between quantitative and categorical questions is explained in Section 6.8).

The first four methods could be applied to either quantitative or categorical assessment questions, whereas the fifth is specific to quantitative questions and the sixth to categorical questions. These 6 methods are described briefly in the following sub sections, and in more detail in Annexes B.1 to B.6. The section ends by identifying which steps of uncertainty analysis each method can contribute to, identifying which form of uncertainty expression they provide (using the categories listed in Section 4.1), evaluating them against the criteria established by the Scientific Committee, and making recommendations on when and how to use them.

9.1.1. Descriptive methods (Annex B.1)

Descriptive expression is currently the main approach to characterising uncertainty in EFSA assessments. Descriptive methods characterise uncertainty using verbal expressions only, without any defined ordinal scale, and without any quantitative definitions of the words. Whenever a descriptive expression of uncertainty is used, the inherent ambiguity of language means that care is needed to avoid misinterpretation. Dialogue between risk assessor and the risk manager could reduce ambiguity.

Even when uncertainty is quantified, the intuitive nature and general acceptance of descriptive expression make it a useful part of the overall communication. When quantification is not scientifically achievable, descriptive expression of the nature and causes of uncertainty is essential.

Verbal descriptions are important for expressing the nature or causes of uncertainty. They may also be used to describe the magnitude of an individual uncertainty, the impact of an individual uncertainty on the assessment outcome, or the collective impact of multiple uncertainties on the assessment outcome.

Descriptive expression of uncertainty may be explicit or implicit. Explicit descriptions refer directly to the presence, magnitude or impact of the uncertainty, for example 'the estimate of exposure is highly uncertain'. In implicit descriptions, the uncertainty is not directly expressed but instead implied by the use of words such as 'may', 'possible' or 'unlikely' that qualify, weaken or strengthen statements about data or conclusions in a scientific assessment, for example 'it is unlikely that the exposure exceeds the ADI'.

Special care is required to avoid using language that implies risk management judgements, such as 'negligible, unless accompanied by objective scientific definitions (EFSA, 2012b).

Potential role in main steps of uncertainty analysis: descriptive expression can contribute to qualitative characterisation of the nature and cause of uncertainties, their individual and combined magnitude, and their relative contribution to overall uncertainty.

Form of uncertainty expression: Descriptive.

Principal strengths: intuitive, requiring no special skills from assessor and accessible to audience.

Principal weaknesses: verbal expressions are ambiguous and mean different things to different people, leading to miscommunication, reduced transparency and decision-makers having to make quantitative inferences for themselves.

9.1.2. Ordinal scales (Annex B.2)

An ordinal scale is a scale that comprises two or more categories in a specified order without specifying anything about the degree of difference between the categories. For example, an ordinal scale of low – medium – high has a clear order but does not specify the magnitude of the differences between the categories (e.g. whether moving from low to medium is the same as moving from medium to high).

Categories in an ordinal scale should be defined, so that they can be used and interpreted in a consistent manner. Often the definitions refer to the causes of uncertainty (e.g. amount, quality and consistency of evidence, degree of agreement amongst experts), rather than degree of uncertainty, although the two are related: e.g., limited, poor quality evidence is likely to lead to larger uncertainty.

Ideally, ordinal scales for degree of uncertainty should represent the magnitude of uncertainty (an ordinal expression of the range and likelihood of alternative answers to the assessment question). Scales of this type are used in uncertainty tables (see Section 9.1.5 and 9.1.6 below).

Potential role in main steps of uncertainty analysis: can contribute to describing and assessing individual uncertainties and/or overall uncertainty, and inform judgements about the relative contributions of different uncertainties.

Form of uncertainty expression: Ordinal.

Principal strengths: provides a structured approach to rating uncertainties which forces assessors to discuss and agree the ratings (what is meant by e.g. low, medium and high).

Principal weaknesses: does not express how different the assessment outcome could be and how likely that is, or does so only in ambiguous qualitative terms.

9.1.3. Uncertainty matrices (Annex B.3)

'Risk matrices' are widely used as a tool for combining ordinal scales for different aspects of risk (e.g. likelihood and severity) into an ordinal scale for level of risk. Matrices have also been proposed by a number of authors as a means of combining two or more ordinal scales representing different sources or types of confidence or uncertainty into a third scale representing a combined measure of confidence or uncertainty. The matrix defines what level of the output scale should be assigned for each combination of the two input scales. Ordinal scales themselves are introduced in the preceding section; here the focus is on the use of matrices to combine them.

Matrices can be used to combine ordinal scales for different sources of uncertainty affecting the same assessment component. When used to combine ordinal scales for uncertainty in different parts of an assessment, the output expresses the uncertainty of the overall assessment.

The matrix shows how the uncertainties represented by the input scales contribute to the combined uncertainty represented by the output scale, but does not identify any individual contributions within each input.

Potential role in main steps of uncertainty analysis: matrices can be used to assess how (usually two) different uncertainties combine, but suffer from significant weaknesses that are likely to limit their usefulness as a tool for assessing uncertainty in EFSA's work (see Annex B.3).

Form of uncertainty expression: Ordinal.

Principal strength: Conceptually appealing and simple to use, aiding consistency in how pairs of uncertainties are combined.

Principal weakness: Shares the weaknesses of ordinal scales (see preceding section) and lacks theoretical justification for how it combines uncertainties.

9.1.4. NUSAP approach (Annex B.4)

NUSAP stands for: Numeral, Unit, Spread, Assessment and Pedigree. The first three dimensions are related to commonly applied quantitative approaches to uncertainty, expressed in numbers (N) with appropriate units (U) and a measure of spread (S) such as a range or standard deviation. Methods to address spread include statistical methods, sensitivity analysis and expert elicitation. The last two dimensions are specific to NUSAP and are related to aspects of uncertainty than can less readily be analysed by quantitative methods. Assessment (A) expresses qualitative expert judgments about the quality of the information used in the model by applying a Pedigree (P) matrix, which involves a multi-criteria evaluation of the process by which the information was produced.

A Pedigree matrix typically has four dimensions for assessing the strength of parameters or assumptions, and one dimension for the influence on results. The method is flexible, in that customized scales can be developed. In comparison to using single ordinal scales, the multi-criteria evaluation provides a more detailed and formalized description of uncertainty. These median scores over all experts for the strength and influence are combined for all uncertainty sources in a diagnostic diagram, which will help to identify the key uncertainties in the assessment, i.e. those sources with a low strength and a large influence on the model outcome. The NUSAP approach therefore can be used to evaluate uncertainties that are not quantified, but can also be useful in identifying the most important uncertainties for further quantitative evaluation and/or additional work to strengthen the evidence base of the assessment.

The NUSAP method is typically applied in a workshop involving multiple experts but in principle can also be carried out less formally with fewer experts.

Potential role in main steps of uncertainty analysis: contributes to describing uncertainties, assessing their individual magnitudes and relative influence on the assessment outcome, but does not assess their combined impact.

Form of uncertainty expression: Ordinal.

Principal strength: Systematic approach to describing the strength and influence of different elements in an assessment, even when these are not quantified, thus informing prioritisation of further analysis.

Principal weakness: Qualitative definition of pedigree criteria is abstract and ambiguous and may be interpreted in different ways by different people. It is questionable whether taking the median across multiple ordinal scales leads to an appropriate indication of uncertainty.

9.1.5. Uncertainty tables for quantitative questions (Annex B.5)

EFSA (2006) suggested using a tabular approach to list and describe uncertainties and evaluate their individual and combined impacts on the assessment outcome, using plus and minus symbols to indicate the direction and magnitude of the impacts. In early examples of the approach, the meaning of different numbers of plus and minus symbols was described qualitatively (e.g. small, medium, large impacts), but in some later examples they have quantitative definitions (e.g. +/-20%, <2x, 2x-5x, etc.). The quantitative version is discussed further in section 9.2.1.2.

The purpose of the table is three-fold: to provide an initial qualitative evaluation of the uncertainty that helps in deciding whether a quantitative assessment is needed; to assist in targeting quantitative assessment (when needed) on the most important sources of uncertainty; and to provide a qualitative assessment of those uncertainties that remain unquantified.

The approach is very general in nature and can be applied to uncertainties affecting any type of quantitative estimate. It is flexible and can be adapted to fit within the time available, including emergency situations. The most up-to-date detailed description of the approach is included in a paper by Edler et al. (2013, their section 4.2).

The table documents expert judgements about uncertainties and makes them transparent. It is generally used for informal expert judgements (see Annex B.11), but formal elicitation (see Annex B.12) could be incorporated where appropriate, e.g. when the uncertainties considered are critical to decision-making.

The method uses expert judgement to combine multiple uncertainties. The results of this will be less reliable than calculation, which can be done by applying interval analysis or probability bounds to the intervals represented by the +/- symbols. Calculations should be preferred when time permits and especially if the result is critical to decision-making. However, the method without calculation provides a useful option for two important needs: the need for an initial screening of uncertainties to decide which to include in calculations, and the need for a method to assess those uncertainties that are not included in calculations so that they can be included in the overall characterisation of uncertainty.

Potential role in main steps of uncertainty analysis: Structured format for describing uncertainties, evaluating their individual and combined magnitudes, and identifying the largest contributors to overall uncertainty.

Form of uncertainty expression: Ordinal (when used with a qualitative scale). For use with quantitative scales see Section 9.2.1.2.

Principal strength: Provides a concise, structured summary of uncertainties and their impact on the outcome of the assessment, which facilitates and documents expert judgements, increases transparency and aids decisions about whether to accept uncertainties or try to reduce them.

Principal weakness: Less informative than quantifying uncertainties on a continuous scale and less reliable than combining them by calculation.

9.1.6. Uncertainty tables for categorical questions (Annex B.6)

This method provides a structured approach for addressing uncertainty in weight of evidence assessment of categorical questions and expressing the uncertainty of the conclusion.

The method uses a tabular format to summarise the lines of evidence that are relevant for answering the question, their strengths, weaknesses, uncertainties and relative influence on the conclusion, and the likelihood or probability of the conclusion.

The tabular format provides a structured framework, which is intended to help the assessor develop the assessment and improve its transparency. The expression of conclusions as probabilities is intended to avoid the ambiguity of narrative forms. The approach relies heavily on expert judgement, which can be conducted informally or using formal elicitation techniques.

This approach is relatively new and would benefit from further case studies to evaluate its usefulness and identify improvements.

Potential role in main steps of uncertainty analysis: this approach addresses all steps of uncertainty analysis for categorical questions and could be the starting point for more quantitative assessment.

Form of uncertainty expression: Ordinal (for individual lines of evidence) and distribution (for probability of conclusion).

Principal strength: Promotes a structured approach to weighing multiple lines of evidence and taking account of their uncertainties, and avoids the ambiguity of narrative terms by expressing the conclusion as a probability.

Principal weakness: Relatively new method; very few examples and little experience of application so far.

9.2. Quantitative methods

This section describes: (i) the main available approaches to characterising uncertainty quantitatively; (ii) methods for implementing parts of those approaches; (iii) why some combinations of methods are more appropriate than others.

There are three basic approaches to addressing uncertainty quantitatively. One is to try to express quantitatively the uncertainty attached to the risk assessment output (section 9.2.1). A second is to construct a risk assessment procedure so that some uncertainties are already addressed by the risk assessment output, by including conservative assumptions of various types in a deterministic calculation (section 9.2.2). A third is to investigate the sensitivity of the risk assessment output to choices which have been made (section 9.2.3).

The three approaches are not mutually exclusive. Some form of scenario or sensitivity analysis is likely to be helpful at several stages: (i) when deciding how to approach quantification of uncertainty in a risk assessment; (ii) as a way of prioritising which of multiple sources of uncertainty to address carefully; and (iii) at the end of the process as a way of establishing confidence in the output. A quantitative assessment of uncertainty relating to a risk assessment protocol is a rational step in the process of deriving conservative assumptions and deterministic calculation procedures to be used for subsequent risk assessments (see Section 8.3).

9.2.1. Quantifying uncertainty

In most of what follows, it is envisaged that there is a clearly defined calculation for the assessment output based on the values of a number of numerical inputs. This will be called the risk calculation. If any of the inputs to the risk calculation is variable, then the output of the risk calculation is also variable and any method for quantifying uncertainty will need to take the variability into account (see section 6.2). In such situations it is important to define clearly the context/scope of the variability: population, time-period, etc. A value used as an estimate of a variable should be representative for that context.

It is also important to consider how best to treat variability. This is in part a risk management judgement to be exercised in the framing of the assessment: the risk manager(s) should state what aspect of the variability is of interest. The risk manager may be interested in the entire distribution of variability or want an estimate of some particular aspect, for example the true worst case or a specified percentile or other summary of variability. This decision will in part determine which methods are applicable. In discussing applicability, a distinction will be made between situations where the risk calculation involves variable inputs and situations where there are no variables or the true worst case is the focus.

If additional uncertainties are identified that are not quantified in the risk calculation, it is better to refine the risk calculation to include them, if possible, rather than address them qualitatively. Some uncertainties would not easily be addressed in this way, for example the family of distributions to use when modelling a variable statistically. Such uncertainties may be better addressed by sensitivity analysis.

9.2.1.1 Measures of uncertainty

For a single numerical input, the simplest quantitative description of uncertainty is a range of values or an upper or lower bound. A range specifies both a lower limit and an upper limit but does not express the relative likelihood of values within the range. A bound specifies just one of the limits. The benefits of quantifying uncertainty in this way are simplicity of the expression of uncertainty and apparent simplicity for the experts expressing uncertainty. In principle, it is possible to specify a disconnected set, for example made of two non-overlapping ranges.

If uncertainty is to be quantified in a way which makes it possible to express a judgement that some values of parameters or variables are more likely than others, the natural language to use is that of probability. As discussed in section 6.6, the subjectivist view of probability is particularly well suited to risk assessment.

When using probability to describe uncertainty about a numerical input or output, there is a choice between specifying a complete probability distribution and simplifying by making a more limited probability statement. A probability distribution quantifies the relative likelihood of all values whereas a limited statement reduces the amount of detail. As an example of the latter, a probability specification might be limited to a single number: the probability that the input or output falls in some specified range of values or exceeds some specified bound. A further simplification would be to avoid specifying the probability exactly and instead to specify an upper and/or lower limit for the probability. Clearly, making such limited specifications may be less onerous for experts but it also severely limits the scope of subsequent calculations. If limited probability statements are made for one or more inputs, there is no distribution representing uncertainty about the assessment output. Instead, a probability, or a bound on probability, can only be calculated for certain ranges of output values.

9.2.1.2 Uncertainty expressed as a bound or as a range of values

An upper or lower limit for a variable or a parameter may sometimes derive from theoretical considerations, for example that a concentration cannot exceed 100%. A bound or range may also derive from expert judgement by formal or informal elicitation (see section 9.2.1.3 and Annex B.8). Such expert judgements will often be informed by relevant data.

The methods in this section are suitable for quantitative assessment questions (see Section 6.9).

<u>Quantitative Uncertainty Tables (Annex B.6)</u>

Uncertainty tables for quantitative questions were described earlier in section 9.1.5. Here, more detail is provided about the case where quantitative definitions are made for the ranges, corresponding to the various +/- symbols, used in an uncertainty table. In practice, it will often be easiest to express each such range relative to some nominal value for the corresponding input or output.

In effect, judgements are being expressed as a range on an ordinal scale where each point on the ordinal scale corresponds to a specified range on a suitable numerical scale for the corresponding assessment input or output. The range on the ordinal scale translates directly into a range on the numerical scale. As well as recording judgements about assessment inputs, the table may also record ranges representing judgements about the combined effect of sub-groups of uncertainties and/or the combined effect of all the uncertainties considered in the table.

Judgements about the combined effect of multiple uncertainties can be made directly by experts. However, calculation should in principle be more reliable. Where the range for each input covers 100% of uncertainty, interval arithmetic (see below) can be used to find a range for the output which also covers 100% of uncertainty. Alternatively, experts might also assign a probability (or a lower bound for such a probability) for each input range. However, they would then be making a limited probability statement and it might be more appropriate to apply probability bounds analysis (section 9.2.1.3 and Annex B.12) to calculate a range of values for the output of the risk calculation and a lower bound for the probability attached to the range.

Potential role in main steps of uncertainty analysis: As for uncertainty tables for quantitative questions in general (section 9.1.5)

Form of uncertainty expression: Range or range with probability.

*Principal strength (*relative to non-quantitative uncertainty tables): provides numerical ranges for uncertainties.

Principal weaknesses: As for uncertainty tables for quantitative questions in general (section 9.1.5)

<u>Interval Analysis (Annex B.13)</u>

Interval analysis is a method to compute a range of values for the output of a risk calculation based on specified ranges for the individual inputs.

The output range includes all values which could be obtained from the risk calculation by selecting a single value for each input from its specified range. Implicitly, any combination of values from within individual ranges is allowed. If it was felt to be appropriate to make the range for one parameter depend on the value of another parameter, the effect would be to specify a two-dimensional set of values for the pair of parameters and a modified version of the interval arithmetic calculation would be needed.

If the range for each individual input covers all possibilities, i.e. values outside the range are considered impossible, then the resulting range for the output also covers all possibilities. The result may well be a range which is so wide that it does not provide sufficient information to support the risk management decision.

It is acceptable in such situations to narrow down the ranges if a probability is specified for each input range. However in such case interval analysis does not provide a meaningful output range. Instead, probability bounds analysis (section 9.2.1.3 and Annex B.15) could be applied to calculate a minimum value for the probability attached to the range. If ranges are narrowed without specifying any probabilities, for example using verbal descriptions such as "reasonable" or "realistic", it is then not possible to state precisely what the output range means.

One simplification which may sometimes have value is to avoid specifying both ends of the ranges, restricting instead to specifying a suitable bound for each input: the end, or intermediate point in more complex situations, which corresponds to the highest level of risk. Knowing whether to specify the lower or upper limit requires an understanding of how the individual inputs affect the output of the risk calculation.

Potential role in main steps of uncertainty analysis: assesses the combined impact of multiple uncertainties and contributes to assessing the magnitudes of individual uncertainties and their relative contributions.

Form of uncertainty expression: Range.

Principal strength: simplicity in the representation of uncertainty and in calculation of uncertainty for the output.

Principal weakness: provides no indication of relative likelihood of values within the output range which may well be very wide.

9.2.1.3 **Uncertainty expressed using probability**

When using probability to quantify uncertainty, there are many tools available. The most complex involve constructing a complete multivariate probability distribution for all the parameters from which the probability distribution for the risk calculation output can be deduced mathematically. The simplest require specifying only some limited aspects of the multivariate distribution, for example the probability of exceeding a specified threshold for each parameter combined with an assertion that uncertainties about parameters are independent. In the simpler cases, the probability information provided about the uncertain output is also limited.

Probability judgements can arise directly from expert elicitation or from statistical analysis of data. In the latter case, expert judgement is still required for selection of data and the statistical model. Once judgements are available for individual sources of uncertainty, they can be combined using the laws of probability. The remainder of this section is structured accordingly.

The methods in this section are all suitable for quantitative assessment questions. Expert knowledge elicitation is also applicable to categorical questions (see Section 6.9). Uncertainties for categorical questions could be combined by Monte Carlo simulations (see below), or using Bayesian Belief Nets (Section 9.2.4).

Obtaining probabilities by expert knowledge elicitation (Annex B.11 and B.12)

Expert knowledge elicitation (EKE) is a collection of methods for quantification of expert judgements of uncertainty, about an assessment input or output, using subjective probability. Usually, the initial elicitation provides a limited probability statement in the form of quantiles, instead of a full distribution. Subsequently, that specification may be extended to a full probability distribution which provides the relative likelihood of values between the quantiles.

The use of EKE is not restricted to eliciting uncertainty about inputs to the risk calculation or about parameters in statistical models of variability. It may sometimes also be used to directly elicit uncertainty about the risk assessment output or about intermediate quantities such as exposure or a tolerable intake.

Potential role in main steps of uncertainty analysis: provides probabilistic judgments about individual uncertainties and may also be applied to suitable combinations of uncertainties.

Formal and informal methods for EKE are distinguished in what follows. In practice, there is not a dichotomy between these, but rather a continuum. The informal method described in Annex B may be regarded as a minimal EKE methodology. Individual EKE exercises should be conducted at the level of formality appropriate to the needs of the assessment, considering the importance of the assessment, the potential impact of the uncertainty on decision-making, and the time and resources available.

Formal EKE (Annex B.12)

The EFSA (2014a) guidance on EKE specifies a protocol which provides procedures for: (i) choosing experts, (ii) eliciting selected probability judgements from the experts; (iii) aggregating and/or reconciling the different judgments provided by experts for the same question; (iv) feeding back the distributions selected for parameter(s) on the basis of the aggregated/reconciled judgments.

The formal EKE procedure is designed to reduce the occurrence of a number of cognitive biases affecting the elicitation of quantitative expert judgements.

Form of uncertainty expression: Primarily distributions, but can be applied using all forms.

Principal strength: provides a structured way to elicit expert uncertainty in the form of a probability distribution.

Principal weakness: doing it well is resource intensive.

Informal EKE (Annex B.11)

In practice, informal methods are also often used. Annex B.8 describes an approximation to the formal protocol for use when there is insufficient time/resource to carry out a formal EKE.

Form of uncertainty expression: All forms.

Principal strength (relative to formal EKE): informal methods offer greater flexibility of application since they are less resource intensive.

Principal weakness (relative to formal EKE): informal methods are more vulnerable than formal EKE to cognitive biases; and more subject to bias from expert selection since this is less formal and structured.

Obtaining probabilities by statistical analysis of data

<u>Statistical Inference from Data – Confidence Intervals (Annex B.8)</u>

Confidence intervals are the most familiar form of statistical inference for most scientists. They are a method for quantifying uncertainty about parameters in a statistical model of variability on the basis of data. The ingredients are a statistical model for the variability, data which may be considered to have arisen from the model, and a defined procedure for calculating confidence intervals for parameters of the statistical model from the data. The result is a range of values for each parameter having a specified level of confidence. By varying the confidence level, it is possible to build a bigger picture of the uncertainty.

For statistical models having more than one parameter, it is in principle possible to construct a confidence region which addresses dependence in the uncertainties about parameters. However, such methods are technically more challenging and are less familiar.

A confidence interval provides a limited quantification of uncertainty about a parameter. It does so with reference to the hypothetical outcomes of many repetitions of an experiment (or survey). The confidence level is a frequency-based probability. It is the chance, before the experiment is carried out, that the confidence interval from the experiment will contain the true value of the parameter. As such, it is not a direct probability statement, given the data from the experiment, about the uncertain value of the parameter. A confidence interval does not directly provide a probability for the chance that the parameter lies in the interval but in many cases it will be reasonable for expert judgement to be used to make such an interpretation of the confidence level.

With the exception of a small number of special cases, confidence interval procedures are only approximate, in the sense that the actual success rate of a confidence procedure differs from the nominal rate (often taken to be 95%) and the direction and/or magnitude of that difference are often unknown

Potential role in main steps of uncertainty analysis: provides limited probabilistic judgments about individual uncertainties relating to parameters in statistical models..

Form of uncertainty expression: Range with probability.

Principal strengths: very familiar method of statistical inference, often used to report uncertainty in literature and often easy to apply.

Principal weaknesses: does not quantify uncertainty about a parameter either as a probability distribution or as a probability that the parameter lies in the interval, and does not easily address dependence between parameters.

<u>Statistical Inference from Data – The Bootstrap (Annex B.9)</u>

The bootstrap is a method for quantifying uncertainty about parameters in a statistical model of variability on the basis of data. The ingredients are a statistical model for the variability, data which may be considered to have arisen from the model, and a choice of statistical estimator(s) to be applied to the data. The technical term “estimator” means a statistical calculation which might be applied to a dataset of any size: it may be something simple, such as the sample mean or median, or something complex such as the a percentile of an elaborate Monte Carlo calculation based on the data.

The basic output of the bootstrap is a sample of possible values for the estimator(s) obtained by applying the estimator(s) to hypothetical datasets, of the same size as the original dataset, obtained by re-sampling the original data with replacement. This provides a measure of the sensitivity of the estimator to the sampled data. It provides a measure of uncertainty for estimators for which standard confidence interval procedures are unavailable without requiring advanced mathematics. The bootstrap is often easily implemented using Monte Carlo.

Various methods can be applied to the basic output to obtain a confidence interval for the "true" value of each estimator: the value which would be obtained by applying the estimator to the whole distribution of the variable. Each of the methods is approximate and makes some assumptions which apply well in some situations and less well in others. As for all confidence intervals, they have the weakness that the confidence interval does not directly provide a probability distribution for the parameters of the statistical model.

Although the basic output from the bootstrap is a sample from a probability distribution, that distribution does not directly represent uncertainty. However, in many cases it will be reasonable for experts to make the judgement that the distribution does approximately represent uncertainty. In such situations, the bootstrap output can be used as an input to subsequent calculations to combine uncertainties, for example using either probability bounds analysis or Monte Carlo.

Potential role in main steps of uncertainty analysis: can be used to obtain limited probabilistic judgments, and in some cases full probability distributions representing uncertainty, about general summaries of variability.

Form of uncertainty expression: Distribution (represented by a sample).

Principal strengths: can be used to evaluate uncertainty for non-standard estimators, even in non-parametric models, and provides a probability distribution which may be an adequate representation of uncertainty for an estimator.

*Principal weakness*es: the distribution, from which the output is sampled, does not directly represent uncertainty and expertise is required to decide whether or not it does adequately represent uncertainty.

<u>Statistical Inference from Data – Bayesian Inference (Annex B.10)</u>

Bayesian inference is a method for quantifying uncertainty about parameters in a statistical model of variability on the basis of data and expert judgements about the values of the parameters. The ingredients are a statistical model for the variability, a prior distribution for the parameters of that model, and data which may be considered to have arisen from the model. The prior distribution represents uncertainty about the values of the parameters in the model prior to observing the data. The prior distribution may be obtained by expert elicitation or sometimes by formal mathematical arguments which suggest a particular form of prior distribution which experts may wish to adopt. The result of a Bayesian inference is a (joint) probability distribution for the parameters of the statistical model. That distribution combines the information provided by the prior distribution and the data and is called the posterior distribution. It represents uncertainty about the values of the parameters and incorporates both the information provided by the data and the prior knowledge of the experts expressed in the prior distribution.

The posterior distribution from a Bayesian inference is suitable for combination with probability distributions representing other uncertainties.

Potential role in main steps of uncertainty analysis: provides a quantitative assessment of uncertainty, in the form of a probability distribution, about parameters in a statistical model.

Form of uncertainty expression: Distribution.

Principal strengths: output is a probability distribution representing uncertainty and which may incorporate information from both data and expert judgement.

Principal weakness: lack of familiarity with Bayesian inference amongst risk assessors – likely to need specialist support.

Combining uncertainties by probability calculations

Bayesian inference provides a full probability distribution representing uncertainty for the parameters in each statistical model for which it is applied. In some situations, the bootstrap does the same. EKE provides either a limited probability statement or a full probability distribution representing uncertainty about each input to which it is applied.

The laws of probability dictate how probability distributions representing individual uncertainties should be combined to obtain a probability distribution representing the combined uncertainty. In some special situations, simple analytical calculations are available but Monte Carlo can be used instead. In most other situations, Monte Carlo is the only practical tool.

The laws of probability also govern the combination of limited probability statements and constrain the kinds of limited probability statement that can be made about combined uncertainty. Probability bounds analysis is a practical tool for doing such calculations. Since a full probability distribution can be used to deduce limited probability statements, probability bounds analysis also provides a way to combine uncertainties for which only limited probability statements have been made with uncertainties for which full probability distributions have been specified.

Probability Bounds Analysis (Annex B.15)

Probability bounds analysis is general method for combining limited probability specifications about inputs in order to make a limited probability specification about the output of a risk calculation.

In the simplest form, for calculations not involving any variables, the assessor specifies a threshold for each input and (a bound on) the probability that the input exceeds the threshold in the direction where the output of the risk calculation increases. A threshold for the output of the risk calculation is obtained by combining the threshold values for the inputs using the risk calculation. Probability bounds analysis then provides a bound on the probability that the output of the risk calculation exceeds that threshold. The method can also be applied using a range for each input rather than just a threshold value.

That simple form of probability bounds analysis includes interval arithmetic as a special case if the exceedance probabilities are all specified to be zero. It can be extended to handle a limited range of situations where variability is part of the risk calculation.

The calculation makes no assumptions about dependence or about distributions. Because no such assumptions are made, the bound on the final probability may be much higher than would be obtained by a more refined probabilistic analysis of uncertainty.

Potential role in main steps of uncertainty analysis: provides a way to combine limited probability statements about individual uncertainties in order to make a limited probability statement about the combined uncertainty.

Form of uncertainty expression: Bound with probability.

Principal strengths: relatively straightforward calculations which need only limited probability judgements for inputs and which makes assumptions about dependence or distributions.

Principal weaknesses: makes only a limited probability statement about the output of the risk calculation and that probability may be much higher than would be obtained by a refined analysis.

Monte Carlo (Annex B.14)

Monte Carlo simulation can be used for: (i) combining uncertainty about several inputs in the risk calculation by numerical simulation when analytical solutions are not available; (ii) carrying out certain kinds of sensitivity analysis. Random samples from probability distributions representing uncertainty for parameters and variability for variables, are used as approximations to those distributions. Monte Carlo calculations are governed by the laws of probability. In the risk assessment arena, distinction is often made between 2D Monte Carlo (2D MC) and 1D Monte Carlo (1D MC).

Potential role in main steps of uncertainty analysis: provides a way to combine uncertainties expressed as probability distributions in order to obtain a probability distribution representing overall uncertainty from those sources. Also useful as part of a method for quantifying contributions of individual uncertainties to overall uncertainty.

2D MC separates distributions representing uncertainty from distributions representing variability and allows the calculation of total uncertainty about any interesting summary of variability. The output from 2D MC is (i) a random sample of values from the joint distribution of all parameters, which represents total uncertainty; (ii) for each value of the parameters, a random sample of values for all variables, including the output of the risk calculation and any intermediate values, representing variability conditional on those parameter values. From the output, for each variability sample, one can calculate any summary statistic of interest such as the mean, standard deviation, specified percentile, fraction exceeding a specified threshold, etc. The result is a sample of values representing uncertainty about the summary. More than one summary can be considered simultaneously if dependence is of interest.

Form of uncertainty expression: Distribution (represented by a sample).

Principal strengths: rigorous probability calculations without advanced mathematics which provide a probability distribution representing uncertainty about the output of the risk calculation.

Principal weakness: requires understanding of when and how to separate variability and uncertainty in probabilistic modelling.

<u>1D MC</u> does not distinguish uncertainty from variability and is most useful if confined to either variability or uncertainty alone. In the context of uncertainty assessment, it is most likely to be helpful when variability is not part of the model. It then provides a random sample of values for all parameters, representing total uncertainty.

Form of uncertainty expression: Distribution (represented by a sample).

Principal strengths (relative to 2DMC): conceptually simpler and communication of results is more straightforward.

Principal weakness (relative to 2DMC): restricted in application to assessments where variability is not part of the model.

9.2.2. Deterministic calculations with conservative assumptions (Annex B.7)

A deterministic calculation uses fixed numbers as input and will always give the same answer, in contrast to a probabilistic calculation where one or more inputs are distributions and repeated calculations give different answers. Deterministic calculations for risk assessment are usually designed to be *conservative*, in the sense of tending to overestimate risk, and are among the most common approaches to uncertainty for quantitative assessment questions in EFSA's work.

Various types of conservative assumptions can be distinguished:

- **default assessment factors** such as those used for inter- and intra-species extrapolation in toxicology
- **chemical-specific adjustment factors** used for inter- or intra-species differences when suitable data are available
- **default values** for various parameters (e.g. body weight), including those reviewed by the Scientific Committee (EFSA, 2012c)
- **conservative assumptions specific to particular assessments**, e.g. for various parameters in the exposure assessment for BPA (EFSA, 2015)
- **quantitative decision criteria** with which the outcome of a deterministic calculation is compared to determine whether refined assessment is required, such as the trigger values for Toxicity Exposure Ratios in environmental risk assessment for pesticides (e.g. EFSA, 2009).

Some conservative assumptions represent only uncertainty, but many represent a combination of variability and uncertainty. Those described as *default* are intended for use as a standard tool in many assessments in the absence of specific relevant data. Those described as *specific* are applied within a particular assessment and are based on data or other information specific to that case. Default factors may be replaced by specific factors in cases where suitable case-specific data exist.

What the different types of conservative assumptions have in common is that they use a single number to represent something that in reality takes a range of values, and that the numbers are chosen in a one-sided way that is intended to make the assessment conservative.

Deterministic calculations generally involve a combination of several default and specific values, each of which may be more or less conservative in themselves. Assessors need to use a combination of values that results in an appropriate degree of conservatism for the assessment as a whole, since that is what matters for decision-making.

Potential role in main steps of uncertainty analysis: provide a way to represent individual sources of uncertainty and to account for their impact on the assessment outcome.

Form of uncertainty expression: Bound or bound with probability.

Principal strength: simple to use, especially default calculations and assumptions that can be applied to multiple assessments of the same type.

Principal weakness: difficulty of assessing the conservatism of individual assumptions, and the overall conservatism of a calculation involving multiple assumptions.

9.2.3. Investigating sensitivity

Sensitivity means the extent to which changes in the parameters and assumptions used in an assessment, produce a change in the results. Therefore it is concerned with the overall robustness of the risk calculation output with respect to input variability and uncertainty.

<u>Sensitivity Analysis (Annex B.16)</u>

Sensitivity Analysis (SA) comprises a suite of methods for assessing the sensitivity of the output of the risk calculation (or an intermediate value) to the inputs and to choices made expressing uncertainty about inputs. It has multiple objectives: (i) to help prioritise uncertainties for quantification: (ii) to help prioritise uncertainties for collecting additional data; (iii) to investigate sensitivity of final output to assumptions made; (iv) to investigate sensitivity of final uncertainty to assumptions made. Sensitivity analysis is most commonly performed for quantitative assessment questions, but can also be applied to categorical questions.

In the context of uncertainty assessment, sensitivity analysis allows the apportionment of the uncertainty in the output to the different sources of uncertainty in the inputs (Saltelli, 2008) and, consequently, the identification of inputs and assumptions mainly contributing to the uncertainty in the results. In its purpose it complements uncertainty analysis whose objective is instead attempting to provide a range of values for the output arising from uncertain inputs. Two possible approaches to sensitivity analysis have been developed. The first approach looks at the effects on the output of infinitesimal changes of default values of the inputs (local) while the second one investigates the influence on the output of changes of the inputs over their whole range of values (global). In the following the discussion will focus only on methods for global sensitivity analysis since the local one is considered of limited relevance in the risk assessment context.

Classification of methods for assessing sensitivity of the output can be performed according to various criteria. Frey and Patil (2004) suggest grouping the methodologies that can be used to perform a sensitivity analysis in three categories:

- Mathematical methods: these methods involve evaluating the variability of the output with respect to a range of variation of the input with no further consideration of the probability of occurrence of its values.
- Statistical methods: The input range of variation is addressed probabilistically so that not only different values of the inputs but also the probability that they occur are considered in the sensitivity analysis.
- Graphical methods: These methods are normally used to complement mathematical or statistical methodology especially to represent complex dependency and facilitate their interpretation.

Collectively, these methods have the capacity to reveal which datasets, assumptions or expert judgements deserve closer scrutiny and /or the development of new knowledge. Simple methods can be applied to simple risk calculations to assess the relative sensitivity of the output to individual variables and parameters. A key issue in sensitivity analysis is clear separation of the contribution of uncertainty and variability. 2D Monte Carlo sampling makes it possible in principle to disentangle the influence of the two components on output uncertainty. However, methodologies for sensitivity analysis in such situations are still under development.

Potential role in main steps of uncertainty analysis: sensitivity analysis provides a collection of methods for analysing the contributions of individual uncertainties to uncertainty of the assessment outcome.

Form of uncertainty expression: expresses sensitivity of assessment output, quantitatively and/or graphically, to changes in input.

Principal strengths: it provides a structured way to identify sources of uncertainty/variability which are more influential on the output.

Principal weakness: assessment of the sensitivity of the output to sources of uncertainty and variability separately is difficult and lacks well established methods.

9.2.4. Other methods not considered in detail

Uncertainty expressed using possibility

Possibility theory (Zadeh, 1978; Dubois and Prade, 1988) and the related theories of fuzzy logic and fuzzy sets have been proposed as an alternative way to quantify uncertainty.

Fuzzy set theory has been applied to quantify uncertainty in risk assessment (Arunraj and Maiti, 2013, Kentel and Aral, 2005). It has mostly been used in combination with stochastic methods such as Monte Carlo, often called hybrid approaches: Li et al. (2007) used an integrated fuzzy-stochastic approach in the assessment of the risk of groundwater contamination by hydrocarbons. Li et al. (2008) applied a similar approach to assessing the health-impact risk from air pollution. Matbouli (2014) reported the use of fuzzy logic in the context of prospective assessment of cancer risks.

However, it is not yet clear how much benefit there is from using Fuzzy methods as compared to methods that use the concept of probability. The WHO/IPCS (2008) Guidance Document on Characterizing and Communicating Uncertainty in Exposure Assessment discussed fuzzy methods briefly, concluding that they "can characterize non-random uncertainties arising from vagueness or incomplete information and give an approximate estimate of the uncertainties" but that they "cannot provide a precise estimate of uncertainty" and "might not work for situations involving uncertainty arising from random sampling error". Therefore, these methods are not covered in our overall assessment of methods.

Imprecisely specified probabilities

For all probabilistic methods, there is the possibility to specify probabilities imprecisely, i.e. rather than specifying a single number as the probability one would attach to a particular outcome, one specifies an upper and a lower bound. Walley (1991) gives a detailed account of the foundational principles, which extend those of de Finetti (1937) and Savage (1954). The basis of the de Finetti approach was to define a probability to be the value one would place on a contract which pays one unit (on some scale) if an uncertain outcome happens and which pays nothing if the event does not happen. The basic idea of Walley's extension is that one does not have a single value for the contract but that there is both some maximum amount one would be willing to pay to sign the contract and some minimum amount one would be willing to accept as an alternative to signing the contract. These maximum and minimum values, on the same scale as the contract's unit value, are one's lower and upper probabilities for the event. The implication of Walley's work is that the accepted mathematical theory of probability extends to a rational theory for imprecise probabilities. Computationally, imprecise probabilities are more complex to work with and so there is not yet a large body of applied work although there are clear attractions to allowing experts to express judgements imprecisely.

Bayesian modelling methodologies

Bayesian Belief Networks and Bayesian graphical models are modern tools which can both support the construction of probabilistic models of uncertainty and variability and provide a framework for computation for both quantitative and categorical assessment questions. There exist a number of software packages for both tools but they are not designed specifically for risk assessment applications. These methods have potential for application in food-related risk assessment in the future.

9.3. Selection of methods for use in uncertainty analysis

The types of assessment question (quantitative or categorical) that the different qualitative and quantitative methods can be applied to, and the types of uncertainty expression they produce, are summarised in Table 3. The applicability of each method to the different steps of uncertainty analysis is considered in Annex B and summarised in Table 4. Each method was also evaluated against performance criteria established by the Scientific Committee (see Section 2), and the results of this are summarised in Table 5. These tables are intended, together with other considerations, to assist readers in choosing which methods to consider for particular assessments. For a more detailed evaluation of each method, see the respective Annex.

It can be seen from Table 4 that, in general, each method addresses only some of the main steps required for a complete uncertainty analysis. The only exception to this is uncertainty tables for categorical questions. Most quantitative methods address 2-3 steps: evaluating individual and overall uncertainty from identified sources and assessing their relative contributions. In general, therefore, assessors will need to select two or more methods to construct a complete uncertainty analysis.

All of the approaches have stronger and weaker aspects, as can be seen from assessing them against the evaluation criteria (Table 5). Broadly speaking, qualitative methods tend to score better on criteria related to simplicity and ease of use but less well on criteria related to theoretical basis, degree of subjectivity, method of propagation, treatment of variability and uncertainty and meaning of the output, while the reverse tends to apply to quantitative methods.

Selecting from the wide array of available methods with differing applicability and quality is a challenging task. Most of the methods have not yet been tried on sufficient EFSA assessments to form a firm conclusion on their usefulness, so it would be premature to give prescriptive guidance on choice of methods, apart from the general principle that uncertainty should be quantified as far as is scientifically achievable. However, some suggestions can be offered to assist users in choosing combinations of methods to consider for particular assessments. These follow in the remainder of this section, after some initial observations on the context for choosing methods.

First, recall (from Section 4) that there are important differences between methods that quantify uncertainty using distributions (full probability specifications), methods that quantify uncertainty using bounds and ranges (partial probability specifications), methods that give alternative individual values (no specification of probability), and methods that express uncertainty in qualitative terms (no quantitative specification at all).

Second, it is likely that most assessments will use more than one form of uncertainty expression, with some uncertainties being characterised using distributions, some using bounds or ranges and some qualitatively.

Third, in most assessments some uncertainties will not be individually characterised in any way.

Fourth, as explained in Section 8, it is efficient to adopt an iterative approach to uncertainty analysis, starting with simple approaches and refining only as far as is needed to support decision-making. Methods using distributions tend to be more demanding than those using ranges, bounds or qualitative expression, unless standardised tools are available that are relevant to the case in hand. Consequently, the user is likely to start with many uncertainties not characterised individually, some uncertainties characterised qualitatively or with bounds or ranges, and few or none characterised probabilistically. This situation is illustrated graphically in the left half of Figure 2. If this initial assessment is not sufficient for decision-making, the user may progressively refine the assessment, by characterising more uncertainties individually, and by 'moving' the more important uncertainties from qualitative expression to bounds and ranges, and from bounds and ranges to distributions. This results in higher proportions being treated by the latter methods, and fewer by the former. This progression is illustrated by the right hand graphic in Figure 2. Note that other degrees of refinement are possible: e.g., in the initial assessment for an emergency situation, there may be insufficient time to assess any uncertainties individually (see Section 8.2).

Each form of uncertainty expression (listed above) can be generated by more than one method, some more complex or refined than others, from which the assessor must select the methods best suited for the assessment in hand. It seems likely that, in any particular assessment, one primary method will be used in each class. This seems likely for practical reasons of simplicity and reducing the need to combine uncertainties assessed by different methods in the same class, although there will be cases where using multiple methods is beneficial.

Finally, the choice of methods for some steps of uncertainty analysis combining uncertainties often constrains or dictates the choice of methods for other steps. For example, electing to use assessment factors as ranges implies that some form of interval analysis or probability bounds will be needed to combine those uncertainties, and narrows the choice of methods for analysing contributions.

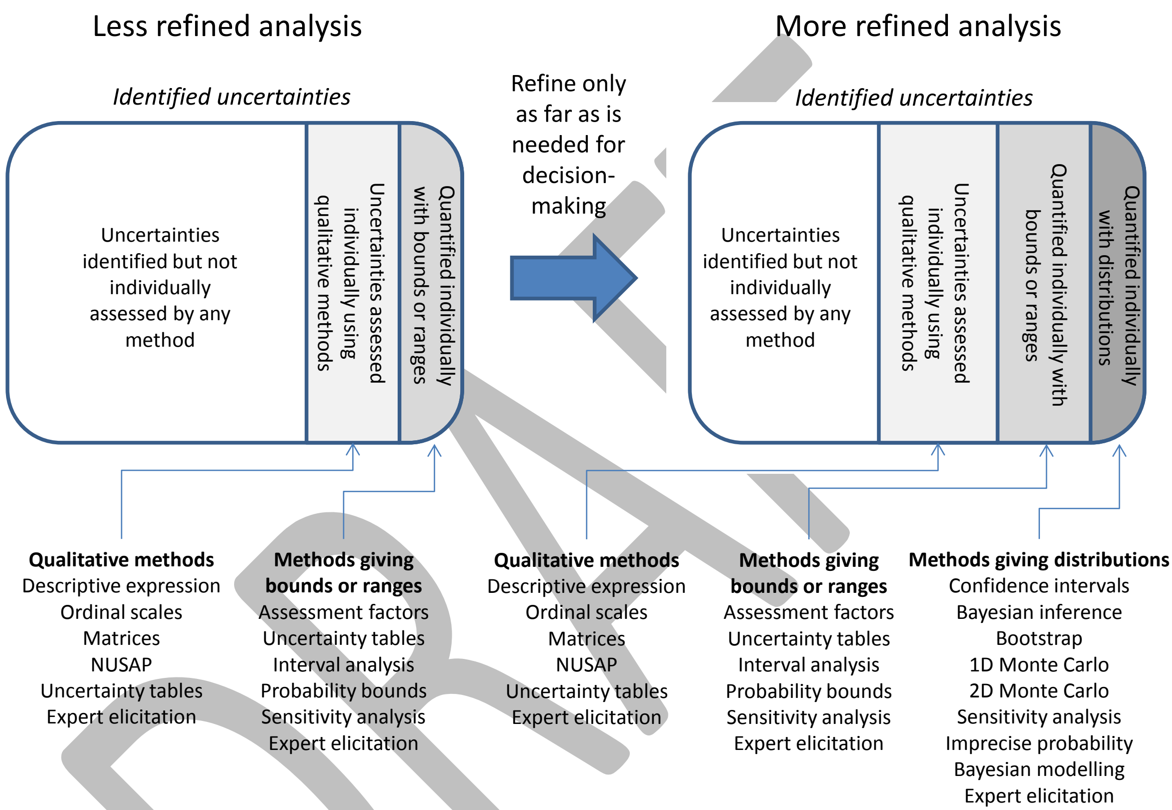


Figure 2: Illustration of change in the proportion of uncertainties assessed individually, the forms of uncertainty expression and the methods of assessment, as an uncertainty analysis is refined. Each rectangle represents the set of identified uncertainties, and sections of the rectangle represent the subset of uncertainties expressed in different forms. Each form of expression can be provided by multiple methods, from which the assessor must select those best suited for the assessment in hand.

Given the context outlined above and illustrated in Figure 2, the following sequence of steps is suggested for practical selection of methods:

1. Identify the uncertainties affecting the assessment. This should always include a systematic consideration of all parts of the assessment (see Section 7). Even in an emergency situation, some time should be reserved for this, possibly using a rapid brainstorming approach. In more complex or refined assessments, informal or formal NUSAP workshops could be considered.
2. Decide which classes of methods will be used in the initial assessment: usually this will include qualitative expression and ranges but sometimes also distributions.
3. Within each class of methods to be used, consider which of the available methods are best suited to the assessment in hand. In making this choice, take account of the relative strengths and

weaknesses of the alternative methods as indicated by the evaluation criteria in Table 222 and also the more detailed discussion in the respective Annexes. In addition, take account of the specific needs of the assessment, the nature of the evidence and uncertainties involved, and the time, resources and expertise available for the assessment.

4. Check which steps of uncertainty analysis (defined in Section 5) are addressed by the chosen methods in each class. Choose additional methods to address the remaining steps. For example, if it is decided to use Monte Carlo, it will be necessary to choose additional methods to derive input distributions and a method of sensitivity analysis for assessing their relative contributions.

5. Some methods can be implemented at different levels of refinement (e.g. formal or informal EKE). Decide what is proportionate for the needs of the assessment and the time and resources available.

6. Carry out the uncertainty analysis and review the results. If iterative refinement is needed, consider whether this can be achieved by characterising more uncertainties with ranges or distributions, and/or by selecting a more refined method within one or more of the classes (e.g. progressing from assessment factors to probability bounds or from 1D to 2D Monte Carlo). Continue iterative refinement until the uncertainty analysis is sufficient to support decision-making (see Section 8).

7. It is essential for transparency to document in a concise and clear way all of the uncertainties identified and how they have been addressed in the assessment. This may usefully be done in tabular form, with one column listing the uncertainties (organised in a suitable manner, e.g. by location in the assessment) and a second column stating how each uncertainty has been addressed, including at least the method used. This serves as a summary and should be accompanied by more detailed documentation of the rationale, methods and results in suitable formats. It is recommended to make a first version of this table in the first iteration of the uncertainty analysis, and update it each time the analysis is refined, as this will help the user to maintain an overview of the uncertainty analysis and identify options for further refinement.

At the present time, there is insufficient experience with applying the methods within EFSA's work to provide more prescriptive guidance. Therefore, it is recommended that EFSA Panels and Units apply the guidance provided above for an initial period, with suitable support from specialists in the different methods. Feedback from this experience may then be used to revise and refine this section and other parts of this guidance, and potentially form the basis for more specific and/or prescriptive guidance.

Table 3: Summary evaluation of which methods can be applied to which types of assessment question (defined in Section 6.9), and provide which forms of uncertainty expression (defined in Section 4).

Method	Types of assessment question	Forms of uncertainty expression provided
Descriptive expression	Quantitative and categorical	Descriptive
Ordinal scales	Quantitative and categorical	Ordinal
Matrices	Quantitative and categorical	Ordinal
NUSAP	Quantitative and categorical	Ordinal
Uncertainty table for quantitative questions	Quantitative	Ordinal, range or range with probability
Uncertainty table for categorical questions	Categorical	Ordinal and distribution
Interval Analysis	Quantitative	Range
Expert Knowledge Elicitation (EKE)	Quantitative and categorical	All
Confidence Intervals	Quantitative	Range with probability
The Bootstrap	Quantitative	Distribution
Bayesian Inference	Quantitative and categorical	Distribution
Probability Bounds Analysis	Quantitative and categorical	Bound with probability
Monte Carlo	Quantitative and categorical	Distribution
Conservative assumptions	Quantitative	Bound or bound with probability
Sensitivity Analysis	Quantitative and categorical	Sensitivity of output to input uncertainty

Table 4: Summary evaluation of which methods can contribute to which steps of uncertainty analysis. Yes/No = yes, with limitations, No/Yes = no, but some indirect or partial contribution. See Annex B for detailed evaluations.

Method	Identifying (finding) uncertainties	Describing uncertainties	Assessing the magnitude of individual uncertainties	Assessing the combined impact of multiple uncertainties on the assessment output, taking account of dependencies	Assessing the contribution of individual uncertainties to overall uncertainty
Descriptive expression	No	Yes	Yes	Yes	Yes
Ordinal scales	No	Yes	Yes	Yes	No/Yes
Matrices	No	No	No	Yes	Yes/No
NUSAP	Yes	Yes	Yes	No	No/Yes
Uncertainty table for quantitative questions	No	Yes	Yes	Yes	Yes
Uncertainty table for categorical questions	Yes	Yes	Yes	Yes	Yes
Interval Analysis	No	No	Yes	Yes	No
Informal Expert Knowledge Elicitation	No	No	Yes	Yes	Yes
Formal Expert Knowledge Elicitation	No	No	Yes	Yes	No
Confidence Intervals	No	No	Yes	No	No
The Bootstrap	No	No	Yes	No/Yes	No
Bayesian Inference	No	No	Yes	No	No
Probability Bounds Analysis	No	No	No	Yes	No
C Monte Carlo	No	No	No	Yes	Yes
Conservative assumptions	No	No	Yes	Yes	No
Sensitivity Analysis	No	No	No	No	Yes

Table 5: Summary evaluation of methods against the performance criteria established by the Scientific Committee. The entries A-E represent varying levels of performance, with A representing stronger characteristics and E representing weaker characteristics. See Table 6 for definition of criteria, Annexes B.1 to B.16 for detailed evaluations.

Method	Evidence of current acceptance	Expertise needed to conduct	Time needed	Theoretical basis	Degree/ extent of objectivity	Method of propagation	Treatment of uncertainty and variability	Meaning of output	Transparency and reproducibility	Ease of understanding for non-specialist
Descriptive	A	A	A	E	C, E	E	C, E	E	D, E	A, B
Ordinal	B	A, B	A	E	D	C,D	C	E	B	D
Matrix	A, D	B	A, B	E	C, D	B, C	C	E	B	B
NUSAP	C	C	A, B	C	D	B, C	C, E	E	B	B
Uncertainty tables for quantitative questions	B, D	B, C	A, B	D, E	C, D	B, C	B, C	C	B	B
Uncertainty tables for categorical questions	D	A, B	A, B	D, E	C, D	B, C	E	A	B	B
Expert Knowledge Elicitation (formal)	B	D	D	C	C	E	A	A	B	B
Expert Knowledge Elicitation (informal)	B	C	B	D	C	C		A	C	C, D
Bayesian Inference	C, D	D, E	A-E	A	A,B	A	A	A	A	C
Confidence Intervals	A	C	A	A	A	E	B	B	A	B
The Bootstrap	C	C-E	A-B	A	A	A, E	B	A	A	C
Conservative assumptions	A	A, B	A	C	B, C	A, D	C, E	A	B, C	B
Interval Analysis	C	B	A	C	B, C	A	E	C	B	A
Probability Bounds Analysis	C, D	C, D	A	A	A	A	A	A	A	B
1D Monte Carlo	A	D	A	A	A	A	B	A	A	C
2D Monte Carlo	B	E	A	A	A	A	A	A	A	D
Sensitivity Analysis (deterministic)	B	B	A	C	B	E	E	-	A	B
Sensitivity Analysis (probabilistic)	D	D, E	A, B	A	B	E	E	-	A	C

EFSA Journal

2159 **Table 6:** Criteria used in Table 5 for assessing performance of methods.

Criteria		Evidence of current acceptance	Expertise needed to conduct	Time needed	Theoretical basis	Degree/ extent of subjectivity	Method of propagation	Treatment of uncertainty and variability	Meaning of output	Transparency and reproducibility	Ease of understanding for non-specialist
Stronger character-istics	A	International guidelines or standard scientific method	No specialist knowledge required	Hours	Well established, coherent basis for all aspects	Judgement used only to choose method of analysis	Calculation based on appropriate theory	Different types of uncert. & var. quantified separately	Range and probability of alternative outcomes	All aspects of process and reasoning fully documented	All aspects fully understandable
	B	EU level guidelines or widespread in practice	Can be used with guidelines or literature	Days	Most but not all aspects supported by theory	Combination of data and expert judgment	Formal expert judgment	Uncertainty and variability quantified separately	Range and relative possibility of outcomes	Most aspects of process and reasoning well documented	Outputs and most of process understandable
	C	National guidelines, or well established in practice or literature	Training course needed	Weeks	Some aspects supported by theory	Expert judgment on defined quantitative scales	Informal expert judgment	Uncertainty and variability distinguished qualitatively	Range of outcomes but no weighting	Process well documented but limited explanation of reasoning	Outputs and principles of process understandable
	D	Some publications and/or regulatory practice	Substantial expertise or experience needed	A few months	Limited theoretical basis	Expert judgment on defined ordinal scales	Calculation or matrices without theoretical basis		Quantitative measure of degree of uncertainty	Limited explanation of process and/or basis for conclusions	Outputs understandable but not process
Weaker character-istics	E	Newly developed	Professional statistician needed	Many months	Pragmatic approach without theoretical basis	Verbal description, no defined scale	No propagation	No distinction between variability and uncertainty	Ordinal scale or narrative description for degree of uncertainty	No explanation of process or basis for conclusions	Process and outputs only understandable for specialists

10. Overall characterisation of uncertainty

10.1. The need to combine quantified and unquantified uncertainties

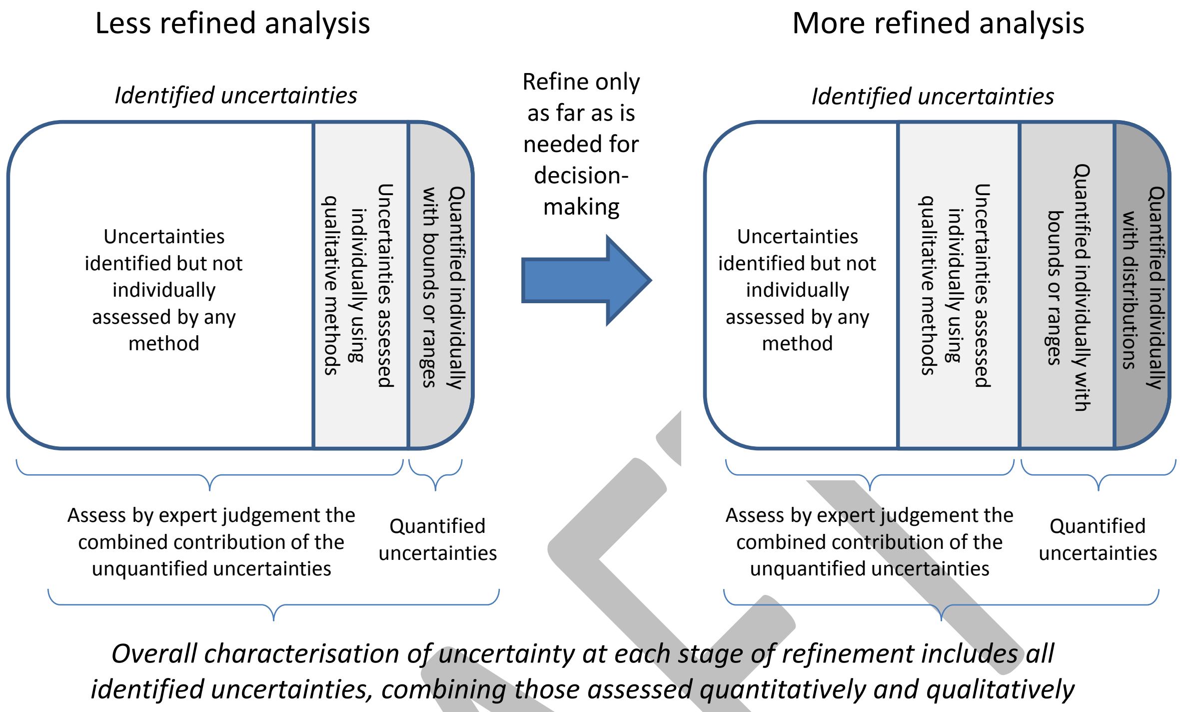
The final output of the uncertainty analysis should be an overall characterisation of the uncertainty of the assessment that takes all identified uncertainties into account. This is because decision-makers need as complete a picture as possible of the overall uncertainty to inform decision-making. As explained in Section 4, this should characterise overall uncertainty in terms of how different the outcome might be and how likely that is, and *quantify it to the extent that is scientifically achievable*.

As explained in Section 9, many assessments will use more than one type of method, for addressing different uncertainties. Therefore, in a single assessment, the impact of some uncertainties on the outcome may be expressed qualitatively, some deterministically and some probabilistically. These must be combined by the assessor, in order to produce an overall characterisation of uncertainty.

Deterministic and probabilistic treatments of uncertainty can be combined by calculation, repeating the probabilistic analysis using alternative assumptions or scenarios for the uncertainties that have been treated deterministically. An overall characterisation of the quantified uncertainty could then be constructed by reporting the two alternative median values, together with the higher of the two upper confidence bounds and the lower of the two lower confidence bounds. The resulting upper and lower values can then be regarded as outer bounds for the confidence interval for all the quantified uncertainties.

Although deterministic and probabilistic treatments of individual uncertainties can be combined by calculation, this will never provide a complete characterisation of identified uncertainties. This is because, even if all identified sources of uncertainty have been quantified individually and combined using deterministic or probabilistic methods, those methods themselves may introduce additional uncertainties (e.g. regarding the choice of distributions used and specification of dependence or independence). Therefore the overall characterisation of uncertainty must always include a final step in which the contribution of those uncertainties that have been quantified individually is combined with an assessment of the contribution of those that have not, including those that have been assessed qualitatively and those that have not been individually assessed by any method. This concept is illustrated graphically in Figure 3.

Figure 3: Illustration of the process for overall characterisation of uncertainty, in more and less refined uncertainty analyses.



10.2. Assessing overall uncertainty

For brevity, identified uncertainties that have not been quantified individually are referred to as *additional uncertainties* in this section. The contribution of these additional uncertainties can only be combined by expert judgement since, if they are quantified by other methods, those methods will themselves add further uncertainties. A final expert judgement is therefore required to avoid entering into an 'infinite regress' of uncertainty about the quantification of uncertainties. There are multiple ways in which that judgement could be made and incorporated into the assessment, which should be considered in the following sequence:

1. If the assessor considers that it would not be scientifically achievable to quantify some of the additional uncertainties, they should still quantify those that they do feel able to quantify and combine them with the uncertainties that have been quantified individually, using the methods described in the following steps (2-5). They should make clear to the decision-maker that the result from this is an *incomplete* picture of the identified uncertainties, and is *conditional* on whatever assumptions have been made about those uncertainties that remain unquantified. As explained in Section 6.8, conditional assessments may still be useful for decision-making. The assessor must describe the nature and causes of the uncertainties that remain unquantified. They should communicate clearly to the decision-maker that the impact of those uncertainties is not quantified, and avoid expressing their conclusions using words that imply a probability judgement about the effect or importance of the unquantified uncertainties (e.g. 'unlikely', etc.).

2. If the assessors judge that the additional uncertainties are so unimportant that, collectively, they would make no difference to the bound, range or distribution obtained for the uncertainties that have been quantified individually, then the latter can be taken as representing the overall uncertainty from those sources that have been identified. This should only be done if there is good reason to believe the additional identified uncertainties make no difference, and the basis for this should be documented and justified.

3. Quantify by expert elicitation the combined impact of the additional uncertainties as a distribution or range for the size of adjustment to the outcome of the assessment that would

be needed to allow for the effect of those additional uncertainties. A practical way to do this is to judge the impact of the additional uncertainties as an additive or multiplicative factor on the scale of the assessment output. Note that this is equivalent to the well-established and accepted practice of using additional assessment factors to allow for additional sources of uncertainty. For example, EFSA (2012c) endorses the use of case-by-case expert judgement to assign additional assessment factors to address uncertainties due to deficiencies in available data, extrapolation for duration of exposure, extrapolation from LOAEL to NOAEL and extrapolation from severe to less severe effects. If the contribution of the additional uncertainties would be large enough to have implications for decision-making, then it would be advisable to quantify it using formal rather than informal elicitation, as the former is more rigorous and reliable.

4. The distribution or range for the combined contribution of additional uncertainties from point 2 above needs to be combined with the contribution from those uncertainties that have been quantified individually. This should be done by calculation rather than expert judgement if possible, as people are known to perform poorly at judging how probabilities combine (Gigerenzer, 2002). Calculation requires a model for how the range or distribution for the additional uncertainties combines with those quantified individually. If the contribution of the additional uncertainties was elicited as an additive or multiplicative factor on the scale of the assessment output it can be combined additively or multiplicatively with the range or distribution for the individually-quantified uncertainties, in the same way as envisaged by EFSA (2012c). However, the assessor should consider whether there are dependencies between any of the uncertainties involved and account for them, either in the calculation or by expert judgement, if they are considered large enough to alter the overall uncertainty.

5. If the assessor is not able to combine the additional uncertainties with the rest of the uncertainty analysis by calculation, then this must be done by expert judgement. This would involve judging by how much the range or distribution for the individually-quantified uncertainties needs to be changed (usually increased) to represent the contribution of the additional uncertainties, taking account of any dependencies between them. This is much less rigorous and reliable than calculation, but still much better than ignoring the additional uncertainties, which would at best be untransparent and at worst negligent (if it caused a significant underestimation of risk). If assessors find it hard to express their judgement of the combined uncertainty as a distribution, it may be sufficient to give a limited probability statement, e.g. a bounded probability for the likelihood of an outcome of interest to the decision-maker (e.g. the likelihood of a specified adverse outcome is less than some stated probability). Possible approaches for doing this are discussed in the following section (10.3). If the outcome of this has implications for decision-making, then it would be advisable to make these judgements by a formal EKE process.

6. When assessors cannot provide even a conditional bound or range for overall uncertainty, they should consider carefully whether to offer any conclusion or estimate from the assessment at all, as they cannot say how different the outcome might be or how likely that is. One option might be to present quantitative estimates for one or more possible scenarios, but it should be made clear that these do not necessarily cover the plausible range and nothing can be said about their likelihoods, and care should be taken to avoid decision-makers anchoring excessively on those results. Another option is to characterise overall uncertainty qualitatively, using descriptive expression or ordinal scales. However, as in (1) above, the assessor should avoid any language that implies a probability judgement. If the assessor feels able to use such language, this implies that they are in fact able to make a probability judgement. If so, they should express it quantitatively – for transparency, to avoid ambiguity, and to avoid the risk management connotations that verbal expressions often imply (Section 4). Whether or not any estimates are offered, the nature and cause of any identified uncertainties that remain unquantified must be described clearly and unambiguously, so that decision-makers can consider what strategies to adopt.

In principle, the procedure above itself introduces additional uncertainties, in the assessment of the additional uncertainties, potentially leading to an 'infinite regress' in which each assessment creates the need for further assessment. The practical solution to this is to take the uncertainty of judging the additional uncertainties into account as part of that judgement. Although this sounds challenging, assessors can do this by first considering what range or distribution would represent their judgement of the additional uncertainties, and then considering whether that range or distribution needs to be inflated to represent their uncertainty in (a) making that judgement and (b) combining it with the individually-quantified uncertainties (whether by expert judgement or calculation).

10.3. Probability judgements for overall uncertainty

It is preferable to combine the contributions of individually-quantified and additional uncertainties by calculation when possible, as emphasised in the preceding section. When they are combined by expert judgement, as outlined in points 4 and 5 of the procedure in the preceding section, the judgement could be elicited in the form of a probability distribution expressing the overall impact of the identified uncertainties on the assessment outcome. However, a more limited alternative is to elicit a judgement of the probability of a specified outcome that is relevant for decision-making, for example, the probability that some measure of risk exceeds an acceptable limit. Assessors may find it difficult to express a precise probability, but a probability bound might be easier to express and may often be sufficient for decision-making.

In making this judgement, it may be helpful to use a standard scale of bounded probabilities, similar to that used by the IPCC (Mastrandrea et al. 2010). The Scientific Committee noted in a previous opinion that a scale of this type might be useful for expressing uncertainty in EFSA opinions (EFSA, 2012b). The IPCC scale as presented by Mastrandrea et al. (2010) was used in a recent opinion on bisphenol A, to express uncertainties affecting hazard characterisation (EFSA, 2015). A modified version of the scale is proposed for future use in EFSA, as shown in Table 7 below. In this version, the probability ranges have been changed to be non-overlapping. This was done because it is expected that experts will sometimes be able to bound their probability on both sides, rather than only on one side as in the IPCC scale. For example, when experts consider an outcome to be 'Likely' (more than 66% probability), they will sometimes be sure that the probability is not high enough to reach the 'Very likely' category (>90% probability). This was evident in the elicitation for the BPA opinion, where experts sometimes selected combinations of categories (e.g. 'As likely as not' to 'Likely') but chose not to extend this to the 'Very likely' category. The ranges in Table 7 overlap at the bounds, but if the expert was able to express their probability sufficiently precisely for this to matter, then they could express their probability directly without using an interval from the Table. Another change in Table 7, compared to the IPCC table, is that the title for the right hand column is given as 'Subjective probability range', as this describes the judgements more accurately than 'Likelihood of outcome', and avoids any confusion with other uses of the word 'likelihood' (e.g. in statistics). Finally, the terms for the first and last likelihood categories have been revised, because the Scientific Committee considered that the common language interpretation of the IPCC terms 'Virtually certain' and 'Exceptionally unlikely' is too strong for probabilities of 99% and 1% respectively.

Table 7: Scale proposed by this Guidance for harmonised use in EFSA to express the probability of uncertain outcomes. See text for details and guidance on use.

Probability term	Subjective probability range
Extremely likely	99-100%
Very likely	90-99%
Likely	66-90%
As likely as not	33-66%
Unlikely	10-33%
Very unlikely	1-10%
Extremely unlikely	0-1%

Table 7 is intended as an aid to expert knowledge elicitation (EKE), not an alternative to it: the principles of EKE should be followed when using it. Judgements should be made by the experts conducting the assessment, who should previously received general training in making probability judgements. Before making their judgements, the experts should review and discuss their assessment of the uncertainties that have been individually assessed either quantitatively or qualitatively, and those that have been identified but not individually assessed. The outcome to be elicited should be well-defined. If the experts are able to specify their judgements about the outcome directly as a precise probability or range of probabilities, without using Table 7, this is preferred. Otherwise, Table 7 may be used as an aid to support the development of judgements. The experts should be asked to select one or more categories from the table, to represent their judgement of the probability of the specified outcome. If they feel no one range covers their judgement of the probability, then they should choose two or more that do so. If an expert finds it difficult to express a judgement, it may be helpful to ask them whether they would like to select all 7 intervals (i.e., give a probability range from 0 to 100%, in effect complete uncertainty), or whether their judgement would be better represented by fewer of the individuals. The judgements of the experts might then be shared, discussed and aggregated to provide a group conclusion, depending on what type of EKE procedure is considered appropriate for needs and context of the assessment (see Annexes B.8 and B.9 and EFSA (2014a)).

It is not intended that experts should be restricted to using the probability ranges in Table 7. On the contrary, they should be encouraged to specify other ranges, or precise probabilities, whenever these express better their judgement of the question or outcome under assessment. However, they should then not use the terms in the left hand column of Table 7 when reporting their assessment, to avoid confusion with the harmonised use of those terms.

In principle, all well-defined uncertainties can be quantified with subjective probability, as explained in Section 6.6. Therefore, Table 7 can be used to express uncertainty for any well-defined outcome. This contrasts with the view of Mastrandrea et al. (2010), who advise that uncertainty may be quantified using the IPCC scale when there is either 'robust evidence' or 'high agreement' or both, which they assess on ordinal scales. The present Guidance shares instead the position of Morgan et al. (2009) who, when discussing the IPCC approach, state that all states of evidence and agreement can be appropriately handled through the use of subjective probability, so long as the question to be addressed is carefully specified. However, as discussed in Section 6.8, assessors may not be able to quantify some uncertainties. In such cases, they should make a conditional assessment, applying Table 7 to those uncertainties they can quantify and describing those they cannot.

Finally, it is emphasised that all probability judgements should be made in a structured and documented manner, complying with at least the minimal requirements for informal EKE (Annex B.8). When the outcome has implications for decision-making, a more formal EKE procedure should be considered (Annex B.9).

10.3.1. The role of qualitative methods in assessing overall uncertainty

The requirement to quantify overall uncertainty as far as scientifically achievable does not mean there is no role for qualitative methods. On the contrary, they will continue to play an important role.

First, there will be some assessments where overall uncertainty cannot be quantified, even in a conditional manner, as in point 6 of the procedure in Section 10.2. In such cases, qualitative approaches will play an important role in describing the source and nature of the uncertainty to decision-makers.

Second, in assessments where the overall uncertainty can be quantified, there will always be some individual uncertainties that remain unquantified. It will often be very helpful to characterise at least some of these qualitatively, as illustrated in Figure 3. This has two main benefits:

- informing judgements about which sources of uncertainty to prioritise for quantitative assessment, based on a qualitative evaluation of their relative impacts on the assessment output. This can be done using qualitative methods that assess relative influence directly, such as NUSAP, or indirectly such as uncertainty tables or ordinal scales.

- informing quantitative judgements about the impact of the combined effect of the unquantified uncertainties, as part of the assessment of overall uncertainty (section 10.2). Qualitative methods that express uncertainty in terms of impact on the assessment outcome (e.g. uncertainty tables and some types of ordinal scale) will be most useful for this because they relate more directly to the uncertainty of the outcome than measures of evidence, agreement, etc.

It is therefore expected that qualitative methods will continue to play an important role in EFSA assessments, in both simple and refined assessments (as indicated in Figure 2).

10.3.2. Overall uncertainty for categorical questions

The approach described above relates to assessments for quantitative questions, which produce quantitative outputs, for example measures of exposure, hazard or risk, where the overall uncertainty from the identified sources can be characterised as a bound, range or distribution around the estimate. For assessments of categorical questions where the output is qualitative, e.g. identification of hazard or mechanism of action, assessment of causality, etc., the overall characterisation of uncertainty should express the range of possible outcomes and their relative likelihoods. The likelihoods should be expressed as quantitative probabilities, to the extent that is scientifically achievable, for reasons discussed in Section 4. As for quantitative questions, bounded probabilities may be easier to judge, using the scale in Table 7. In qualitative risk assessments where the probabilities for alternative categories of outcome have been derived by calculation, the final step in characterising overall uncertainty will need to consider whether those probabilities need to be adjusted to take into account any other identified uncertainties that were not included in the calculations. Again, this final step could be undertaken by formal expert judgement, if informal expert judgement suggests the need for significant adjustment.

10.4. Documentation of overall characterisation

Whatever approach is used to address the additional uncertainties, it should be clearly documented and justified. If it is decided that no allowance is needed for the additional uncertainties, the basis for this should be documented (note that such a judgement implies the same solution to the problem of infinite regress as that described above). Uncertainty tables (see Annexes B.5 and B.6) provide one possible option for documenting the basis for these judgements, as they provide a format for listing the uncertainties that are being considered and showing (using plus and minus symbols or any other method the assessor finds effective) how their combined impact has been assessed. If informal expert judgement indicates that the collective impact may be significant, consideration should be given to making this final judgement using formal expert elicitation (option 3 in Section 10.2), or to identifying the most important additional uncertainties and quantifying them individually by suitable methods.

11. Reporting uncertainty analysis in scientific assessments

The methods and results of the uncertainty analysis should be reported fully and transparently, in keeping with EFSA's (2009) Guidance on Transparency. Wherever statistical methods have been used, reporting of these should follow EFSA's (2014) Guidance on Statistical Reporting.

It is recommended that the report of the uncertainty analysis should be presented as a separate section within the main document of the assessment it relates to. In some cases, several such sections may be needed in different parts of the report, relating to different parts of the overall assessment (e.g. as was done for bisphenol A, EFSA 2015).

Sections addressing uncertainty should be titled in a clear manner (e.g. 'Uncertainty analysis') so it is immediately recognised by the reader and placed at an appropriate location in the document: often, a logical position will be immediately preceding the overall conclusion of the document, since the uncertainty analysis takes account of other parts of the assessment and has direct consequences for the conclusions. If the uncertainty analysis is substantial, a summary could be placed in the main document with more detail presented in Annexes.

Reporting should always include the following elements, which may usefully be used as headings within a section on uncertainty to provide an organised structure for documenting the uncertainty analysis. It is intended to provide examples of this in Annex D of the final version of this Guidance.

1. **Assessment question**: Specify the assessment question for which uncertainty is to be considered.
2. **Description of potential sources of uncertainty**: the complete list of the potential sources of uncertainty that have been identified at the beginning of or during the assessment should be provided along with their qualitative description in terms that are, as far as possible, comprehensible to non-specialists. If it is decided to prioritize among sources of uncertainty for further assessment, methods and criteria used to screen the uncertainty sources should be specified.
3. **Methods used for expressing and assessing the magnitude of sources of uncertainty**
   a) **Individual sources of uncertainty and their impact on the assessment:** describe the methods used to express and assess the impact of the individual sources of uncertainty.
   b) **Multiple sources of uncertainties and their combined impact on the assessment output:** describe the method used to express and assess the impact (propagation) of multiple sources of uncertainty on the final assessment output, in terms of the alternative values the output might really take and how likely they are..
   c) **Overall summary of identified uncertainties and the methods used to address them**, presented in a concise and accessible form, e.g. list or table.
4. **Outcome of the uncertainty assessment**: The results of expressing and assessing the individual and combined sources of uncertainty on the output should be reported in terms of the alternative values the output might really take and how likely they are. The assessment question should be recalled at this stage. The final conclusion should be expressed quantitatively, if scientifically achievable, and also in narrative form using language comprehensible to non-specialists. If there are any sources of uncertainty that it is not scientifically possible to quantify, these should be highlighted and their nature and origin should be described.
5. **Relative contribution of individual uncertainties to their overall uncertainty**: the relative contribution of different sources of uncertainty to the overall uncertainty of the assessment outcome should be reported in order to provide decision-makers with information about factors that are more influential on the final conclusions and/or that require further data collection or investigation.

A layered approach to reporting is recommended, to address the needs of different audiences and enable each reader to access easily whatever level of information they require. A structured approach to this is presented in Table 8. It should, of course, be ensured that information provided in each layer is consistent with all the other layers.

Table 8: Layered approach to reporting of uncertainty analysis.

Location	Content	Audience
Abstract	One line summary of overall uncertainty from identified sources	All readers including the public
Summary Conclusion section	One paragraph including the conclusion on the overall uncertainty and short	All readers including the public

	explanation of the main sources of identified uncertainty. The same paragraph may appear in both locations or be expanded in the Conclusion section.	
Uncertainty section in main document	Summary of the uncertainty analysis including methods and results (typically 1-2 pages, but longer if proportional to the size and complexity of the overall assessment)	Scientists Members of the public, risk managers, stakeholders who want a summary of the basis for the conclusions
Annex	Full technical documentation and justification of uncertainty analysis	Scientists Others who want to see details on all or part of the uncertainty analysis

12. Communicating scientific uncertainties

12.1. EFSA's risk communication mandate

EFSA is mandated to "be an independent scientific source of advice, information and risk communication in order to improve consumer confidence". Creating and sustaining such confidence require coherence and co-ordination of all three outputs: advice, information and risk communication. The quality, independence and transparency of EFSA's scientific advice and information, supported by the robustness of the working processes needed to develop them, are critical for effective risk communication and for increasing public confidence. Equally, clear and unambiguous communication of assessment outcomes contextualises the scientific advice and information, aiding decision-makers to prioritise policy options and take informed decisions. Through multipliers (e.g. media, NGOs) this also forms a basis for consumers' greater confidence in their own choices and in risk management action.

Therefore, EFSA communicates the results of its scientific assessments to risk managers, stakeholders, and the public at large. Besides the huge cultural and social diversity in the European Union, there is also a vast spectrum of individual needs and technical knowledge among these target audiences. Decision-makers and stakeholders are also responsive to the perceptions of the general public. Effective risk communication, therefore, requires careful crafting of messages and selection of tools keeping in mind the target audience as well as the perceived sensitivities of the subject. These activities are generally conducted at the level of EFSA as an organisation rather than individual Panels or Units.

To be useful to decision-makers, ensure coherence and limit possible misinterpretation of its scientific assessments, EFSA communicates its scientific results in a manner that aims to be both meaningful to specialists and understandable to informed laypersons. To achieve this, EFSA uses a variety of communications channels and media, ranging from the simple to the complex, to communicate the same messages to different audiences (e.g. newsletters, frequently-asked questions (FAQs), infographics, videos, interactive tools, and images, as well as technical reporting through opinions, statements, etc.).

12.2. Risk perception and uncertainty

Perceptions of the risks or benefits for which EFSA is providing an assessment and the meaningful expression of the identified uncertainties, play paramount roles in how recipients of EFSA's communication act upon the results. This varies by target audience and their respective level of technical knowledge.

Understanding of the type and degree of uncertainties identified in the assessment helps to characterise the level of risk to the recipients and is therefore essential for informed decision-making. This is especially useful for risk managers and political decision-makers. As the level of technical knowledge among the target audiences decreases, however, increasing awareness of scientific uncertainties could in some cases undermine confidence in the recipient's individual decision-making. Yet, in some cultural contexts, communication of the uncertainties to non-technical audiences is received positively even if it makes decisions more difficult, because of the greater transparency of the process. As such, the potential decrease in confidence is offset by an increase in trust.

The roles of risk communication within this process are to contextualise the uncertainties in relation to the perceived risks, to underline the transparency of the process and to explain how scientists can address the information gaps in the future.

12.3. Challenges of communicating uncertainty in scientific assessments

Three combined factors affect the effectiveness of communicating food-related risks: complexity, uncertainty and ambiguity (Renn, 2005). Communicating scientific uncertainty requires both simplifying and complicating the normal scientific discourse (Fischhoff & Davis, 2013). In terms of the best methods, the literature is equivocal (Rowe, 2010) about the advantages and/or disadvantages of communicating uncertainty to stakeholders in qualitative or quantitative terms.

Various arguments have been made both for and against communicating uncertainty to the general public (Johnson & Slovic, 1995, 1998). Yet, there is little empirical evidence to support either view (Miles S & Frewer L, 2003).

From EFSA's organisational perspective, communicating scientific uncertainties is crucial to its core mandate, reaffirming its role in the scientific assessment process. The clear and unambiguous communication of scientific uncertainty is an enabling mechanism, providing decision-makers with the scientific grounds for risk-based decision-making. It increases transparency both of the assessments and of the resulting decision-making, ensuring that confidence in the scientific assessment process is not undermined.

As a consequence decision-makers are also better able to take account of the uncertainties in their risk management strategies and to explain, as appropriate, how scientific advice is weighed against other legitimate factors. Explaining how decisions or strategies take account of scientific uncertainties will contribute to increased public confidence in the EU food safety system as well.

Overall, while developing this Guidance document, EFSA has identified a need to differentiate more systematically the level of scientific technicality in the communications messages on uncertainties intended for different target audience. This more differentiated and structured approach marks a shift from the current one described in 12.1 above.

12.4. Towards best practice for communicating uncertainty

As indicated above the literature is equivocal about the most effective strategies to communicate scientific uncertainties. Although EFSA regularly communicates the scientific uncertainties related to its assessments in its scientific outputs and in its non-technical communication activities, it has not developed a model that is applied consistently across the organisation. According to IPCS, for example, "it would be valuable to have more systematic studies on how risk communication of uncertainties, using the tools presented […] functions in practice, regarding both risk managers and other stakeholders, such as the general public" (IPCS, 2014). Although some scientific assessment bodies have compiled case study information to develop a body of reference materials (BfR, 2013), on the whole there is a lack of empirical data in the literature on which to base a working model.

Therefore, while EFSA's scientific Panels are piloting this Guidance on uncertainty, EFSA will conduct target audience research among stakeholders on communicating scientific uncertainty and integrate the results in the final version of this document.

The development of effective communications messages requires an in-depth knowledge of target audiences including: their level of awareness and understanding of food safety issues; their attitudes to food in general and food safety in particular; the possible impact of communications on behaviour; and the appropriate channels for effective dissemination of messages.

EFSA proposes using the Clear Communication Index (CCI), a research-based tool to help develop and assess public communication materials, developed by the USA's Center for Disease Control and Prevention (CDC). Fundamental to the CCI, and thus the rationale for choosing this methodology, is that each communication output should only be tailored to one single target audience.

This will allow EFSA to identify how changes could be made to its current communications practices in relation to uncertainties and to tailor key messages to specific target audience needs.

13. Way forward and recommendations

This guidance document is intended to guide EFSA panels and staff on how to deal with uncertainties in scientific assessments by providing a toolbox of methods, from which assessors can select those methods which most appropriately fit the purpose of their individual assessment.

While leaving flexibility in the choice of methods, all EFSA scientific assessments must include consideration of uncertainties; for reasons of transparency, these assessments must clearly state all the uncertainties which have been identified and the overall impact of these on the assessment outcome. This must be reported clearly and unambiguously.

It is further recommended that:

The endorsed guidance document is introduced to EFSA panels and staff in an implementation period which gives sufficient time for testing the applicability of the guidance in mandates of different complexity and time constraints and covering all the different areas of EFSA's assessments.

When the testing period is completed and any resulting improvements to the Guidance Document have been agreed, uncertainty analysis will be unconditional for EFSA Panels and staff and must be embedded into scientific assessment in all areas of EFSA's work.

The final Guidance should be implemented in a staged process, starting by focussing on uncertainties specific to individual assessments. The implications for standardised assessment procedures should be considered over a longer period, as part of the normal process for evolving EFSA approaches. Where appropriate, this should be done in consultation with international partners and the wider scientific community.

A specific plan be drafted which will detail the responsibilities of panel members and EFSA staff in testing the guidance document and giving their feedback on the applicability. Such a plan should consider that:

- All Panels and relevant EFSA units appoint one or two members as ambassadors for ensuring the implementation of the guidance in their area of work.
- All panels and relevant EFSA units select at least one new opinion to try the guidance during the testing phase.
- Panels and relevant EFSA units consider whether it would be useful to develop lists of assessment components and uncertainties commonly encountered in their area of work, as an aid to identifying relevant uncertainties in their future individual assessments.
- EFSA´s secretariat facilitates dialogue between Panels and Risk managers.
- A targeted consultation with relevant stakeholders to be conducted by EFSA in parallel with the testing phase.

In addition, it is recommended that EFSA forms a competency network and a centralized support group which should also identify and support the initiation of the necessary training activities starting early in the testing phase. This should include:

- Making training on the guidance and its use available to both risk assessors and risk managers.
- Establishing a standing Working Group on Uncertainty analysis to provide expert technical support to the Panels at least in the initial phases of the implementation.

Furthermore EFSA should initiate (research) activities to explore best practices in the communication of uncertainties in scientific assessments targeted to the different audiences.

References

Codex, 2015. Codex Alimentarius Commission. Procedural Manual. 23rd edition.

Bundesinstitut für Risikobewertung (BfR), "Risk Perception and Crisis Communication – How to Deal with Uncertainty". Presentation by Dr. Suzan Fiack, Symposium "How to overcome a life-threatening crisis in the food chain", Berlin (2013) http://www.bfr.bund.de/cm/343/risk-perception-and-crisis-communication-how-to-deal-with-uncertainty.pdf

Dubois, Didier, and Henri Prade. Possibility theory. Springer, 1988.

EC (European Commission), 2000. Communication from the Commission on the precautionary principle. COM(2000) 1 final. European Commission, Brussels.

EC (European Commission), 2006. Report of the First Meeting of the Chairs of Scientific Committees of Community bodies involved in Risk Assessment (7 and 8 December 2005). C7/MM D(2006) 370004. Brussels, 28 February 2006.

EFSA (European Food Safety Authority), 2006a. Guidance of the Scientific Committee on a request from EFSA related to Uncertainties in Dietary Exposure Assessment. The EFSA Journal 438, 1-54.

EFSA (European Food Safety Authority), 2006b. Transparency in risk assessment carried out by EFSA: Guidance document on procedural aspects. The EFSA Journal (2006), 353, 1-16.

EFSA (European Food Safety Authority), 2009. Guidance of the Scientific Committee on Transparency in the Scientific Aspects of Risk Assessments carried out by EFSA. Part 2: General Principles. The EFSA Journal (2009), 1051, 1-22.

EFSA (European Food Safety Authority), 2010. Application of systematic review methodology to food and feed safety assessments to support decision making. EFSA Journal 2010; 8(6):1637.

EFSA PLH Panel (EFSA Panel on Plant Health), 2011. Guidance on the environmental risk assessment of plant pests. EFSA Journal 2011; 9(12): 2460.

EFSA BIOHAZ Panel (EFSA Panel on Biological Hazards), 2012a. Scientific Opinion on the development of a risk ranking framework on biological hazards. EFSA Journal 2012; 10(6): 2724.

EFSA Scientific Committee, 2012b. Scientific Opinion on risk assessment terminology. EFSA Journal 2012; 10(5): 2664

EFSA (European Food Safety Authority), 2012c. Guidance on selected default values to be used by the EFSA Scientific Committee, Scientific Panels and Units in the absence of actual measured data. The EFSA Journal, 2579, 1-32.

EFSA (European Food Safety Authority), 2014a. Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment. EFSA Journal 2014; 12(6):3734.

EFSA (European Food Safety Authority), 2014b. Guidance on Statistical Reporting. EFSA Journal 2014;12(12):3908, 18 pp.

EFSA (European Food Safety Authority), 2015. Scientific Opinion on the risks to public health related to the presence of bisphenol A (BPA) in foodstuffs – Part II: Toxicological assessment and risk characterisation. EFSA Journal 2015;13(1):3978

de Finetti, B. (1937), La prévision: ses lois logiques, ses sources subjectives, Ann. Inst. Henri Poincaré 7, 1–68.

Fischhoff B, Davis AL: Communicating scientific uncertainty PNAS 2014 111 (Supplement 4) 13664-13671; published ahead of print September 15, 2014,doi:10.1073/pnas.1317504111

Fischhoff B & Davis AL, Communicating scientific uncertainty, US Department of Health and Human Services Food and Drug Administration, 2011

Gigerenzer, G. 2002. Calculated risks: how to know when numbers deceive you. Simon & Schuster, New York.

Heads of Food Agencies WG, 2012. Heads of National Food Agencies Working Group on Transparent Use of Risk Assessment in Decision Making. Final report to the Heads of Agencies.

IPCS (International Programme on Chemical Safety), 2004. IPCS Risk Assessment Terminology. International Programme on Chemical Safety. WHO, Geneva.

Kahneman D, Slovic P and Tversky A (Eds), 1982. Judgment under uncertainty: heuristics and biases. Cambridge University Press, Cambridge, UK.

Madelin R, 2004. The importance of scientific advice in the Community decision making process. Opening address to the Inaugural Joint Meeting of the members of the Non-Food Scientific Committees. Directorate General for Health and Consumer Protection, European Commission, Brussels.

Mastrandrea, MD, CB Field, TF Stocker, O Edenhofer, KL Ebi, DJ Frame, H Held, E Kriegler, KJ Mach, PR Matschoss, G-K Plattner, GW Yohe, and FW Zwiers, 2010: Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties. Intergovernmental Panel on Climate Change (IPCC). Available at <http://www.ipcc.ch>.

Morgan MG and Henrion M, 1990. Uncertainty. A guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge University Press.

Morgan, M.G., H. Dowlatabadi, M. Henrion, D. Keith, R. Lempert, S. McBride, M. Small and T. Wilbanks. 2009. Best practice approaches for characterizing, communicating, and incorporating scientific uncertainty in climate decision making. U.S. Climate Change Science Program. Synthesis and Assessment Product 5.2.

Morgan MG, 2014. Use (and abuse) of expert elicitation in support of decision making for public policy. Proc. Nat. Acad. Sciences, 111: 7176–7184. www.pnas.org/cgi/doi/10.1073/pnas.1319946111

NRC (National Research Council), 1983. Risk assessment in the Federal Government: managing the process. National Academy Press, Washington DC.

NRC (National Research Council), 1996. Understanding risk: informing decisions in a democratic society. National Academy Press, Washington DC.

NRC (National Research Council), 2009. Science and decisions: advancing risk assessment. Committee on Improving Risk Analysis Approaches used by the US EPA. National Academies Press, Washington DC.

OMB/OST (Office of Management and Budget/Office of Science and Technology Policy of the Executive Office of the President), 2007. Updated principles for risk analysis. Memorandum for the Heads of Executive Departments and Agencies, September 19, 2007. US Office of Management and Budget and US Office of Science and Technology Policy.

Oxford Dictionaries, 2015. Accessed 17 February 2015. http://www.oxforddictionaries.com/definition/english/

Petersen AC, Cath A, Hage M, Kunseler E and van der Sluijs JP, 2011. Post-Normal Science in practice at the Netherlands Environmental Assessment Agency. Science, Technology and Human Values. 36: 362-388.

Phillips N, 2000. The BSE Inquiry Report. Lord Phillips of Worth Matravers. Published October 2000.

Renn O: Risk perception and communication: lessons for the food and food packaging industry. Food Addit Contam. 2005 Oct;22(10):1061-71.

Savage, L.J. (1954),The Foundations of Statistics. New York: John Wiley and Sons (second edition, 1972, New York: Dover)

SCHER, SCENIHR, SCCS, 2013. Making risk assessment more relevant for risk management. Scientific Committee on Health and Environmental Risks, Scientific Committee on Emerging and Newly Identified Health Risks, and Scientific Committee on Consumer Safety. March 2013.

Stirling A, 2010. Keep it complex. Nature, 468: 1029-1031.

Theil M, 2002. The role of translations of verbal into numerical probability expressions in risk management: a meta-analysis. Journal of Risk Research, 5 (2) 177 – 186.

US Environmental Protection Agency, 2000. Risk characterization handbook. Science Policy Council, U.S. EPA, Washington DC.

Vose D, 2008. Risk analysis: a quantitative guide. 3rd edition. Wiley.

Walker WE, Harremoes P, Rotmans J, van der Sluijs JP, van Asselt MB, Janssen PH, Krayer von Krauss MP: Defining uncertainty. A conceptual basis for uncertainty management in model-based decision support. Integr Assessment 2003, 4:5-17.

WHO/IPCS, 2005. Chemical-specific adjustment factors for interspecies differences and human variability: guidance document for use of data in dose/concentration–response assessment. IPCS Harmonization Project Document No. 2. World Health Organisation, Geneva.

Walley, Peter (1991). Statistical Reasoning with Imprecise Probabilities. London: Chapman and Hall. ISBN 0-412-28660-2.

Zadeh, Lotfi, "Fuzzy Sets as the Basis for a Theory of Possibility", Fuzzy Sets and Systems 1:3–28, 1978. (Reprinted in Fuzzy Sets and Systems 100 (Supplement): 9–34, 1999.

Glossary

(Note: the present glossary is a draft developed to support the public consultation process; it will be further revised following the outcome of the public consultation)

Term	Definition
Aleatory uncertainty	Uncertainty caused by variability, e.g. uncertainty about a single toss of a coin, or the exposure of a randomly-selected member of a population.
Assessment factor	A numerical factor used in quantitative assessment, to represent or allow for extrapolation or uncertainty.
Assessment input	Inputs to a calculation or model, including any data, assessment factors, assumptions, expert judgements, etc.
Assessment output	The output of a calculation or model, i.e. the estimate it provides in answer to the assessment question.
Assessment question	The question to be addressed by an assessment. Assessment questions may be *quantitative* (estimation of a quantity) or *categorical* (e.g. yes/no questions). Many questions may usefully be divided into sub-questions for assessment.
Assessment structure	The structure of a calculation or model, i.e. how the inputs are combined to generate the assessment output. Can generally be written down as a mathematical equation or sequence of equations.
Assessor	A person conducting an assessment.
Bayesian inference	A form of statistical inference in which probability distributions are used to represent uncertainty.
Bound	The upper or lower limit of a range of possible numbers, or of a probability interval.
Categorical question	An assessment question that concerns a choice between two or more categories, e.g. hazard identification, mode of action, human relevance, adversity, equivalence of a GM plant and its non-GM counterpart, etc.
Chemical-specific adjustment factor (CSAF)	A quantitative measurement or numerical parameter estimate that replaces a default uncertainty subfactor (WHO/IPCS, 2005).
Conditional assessment	An assessment which is made subject to specified assumptions or scenarios to to address uncertainties that have not been quantified. Because uncertainty is intrinsically personal and temporal, all expressions of uncertainty are conditional on the assessors who provide them and the knowledge available to them at the time of assessment.
Confidence	Levels of confidence (e.g. high, low, etc.) are often used to express the probability that a conclusion is correct. In frequentist statistics, a confidence interval is a range within which an estimated value would like in a specified proportion of occasions if the experiment and/or statistical analysis were repeated an infinite number of times. In Bayesian statistics it is replaced with a credibility interval, which is a range within which the real value would lie with specified probability. In a social science context, confidence is the expectation of an outcome based on prior knowledge or experience.
Conservative	Term used to describe assessments, or parts of assessments (e.g. assumptions, default factors, etc.), that tend to overestimate the severity and/or frequency of an adverse outcome (e.g. overestimate exposure or hazard and consequently risk). Conservatism is often introduced intentionally, as a method to allow for uncertainty (see Section 6.4 and Annex B15).
Decision criterion	Numerical criteria (sometimes called 'trigger values') used in some parts of EFSA for deciding what conclusion can be made on risk and/or whether further assessment is needed. In some cases (e.g. pesticides), provision for uncertainty is built into the trigger value instead of, or as well as, being built into the assessment or its inputs.
Decision-maker	A person with responsibility for making decisions; in the context of this document, a person making decisions informed by EFSA's scientific advice. Includes risk managers but also people making decisions on other issues, e.g. health benefits, efficacy, etc.
Deep uncertainty	Either not well-defined, or not able to quantify. Stirling.
Default value	Pragmatic, fixed or standard value used in the absence of relevant data (WHO/IPCS, 2005), implicitly or explicitly regarded as accounting appropriately for the associated uncertainty.
Deterministic	A deterministic calculation uses fixed numbers as input and will always give the same

	answer, in contrast to a probabilistic calculation where one or more inputs are distributions and repeated calculations give different answers.
Distribution parameters	Numbers which specify a particular distribution from a family of distributions.
Epistemic uncertainty	Uncertainty due to limitations in knowledge.
Expert knowledge elicitation (EKE)	A systematic, documented and reviewable process to retrieve expert judgements from a group of experts, often in the form of a probability distribution.
Frequency	The number of occurrences of something, expressed either as the absolute number or as a proportion or percentage of a larger population (which should be specified).
Generic uncertainty	Source of uncertainty arising in the same way in multiple assessments. If the magnitude of a generic uncertainty is consistent across many assessments, it may be efficient to assess it generically and develop a generic way of providing for it in assessments (e.g. a default distribution or uncertainty factor), rather than assessing it anew in each case.
Infinite regress	In relation to uncertainty, refers to the problem that assessment of uncertainty is itself uncertain, thus opening up the theoretical possibility of an infinite series of assessments, each assessing the uncertainty of the preceding one. See Section 10 for proposed solution.
Likelihood	In everyday language, refers to the chance or probability of something: used with this informal meaning in many places in this document. In statistics, maximum likelihood estimation is one option for obtaining confidence intervals (Annex B.10). In Bayesian statistics, the likelihood function encapsulates the information provided by the data (Annex B.12).
Limited probability statement	An incomplete specification of probability, i.e. not a precise value. A simple limited form is a *probability bound*, which states that the probability is greater than some specified value, or less than a specified value, or both (when a range is given). Limited probability statements may be easier for assessors to provide, and may be sufficient for decision-making in some cases.
Line of evidence	A collective term for multiple pieces of evidence of the same type, relating to the same question or parameter, and distinguished from other types of evidence relating to the same question or parameter. For example, human studies, animal studies, in vitro studies and in silico methods might be considered as different lines of evidence for assessing toxicity of a chemical.
Model	In scientific assessment, usually refers to a mathematical or statistical construct, which is a simplified representation of data or of real world processes, and is used for calculating estimates or predictions.
Monte Carlo	A method for making probability calculations by random sampling from distributions
Markov Chain Monte Carlo	A form of Monte Carlo where values are not sampled independently but instead are sampled from a Markov chain. In many situations where standard Monte Carlo is difficult or impossible to apply, MCMC provides a practical alternative.
Ordinal scale	A scale of measurement comprised of ordered categories, where the magnitude of the difference between categories is not quantified.
Parameter	A quantity that has a single true value. Parameters include quantities that are considered constant in the real world, and also quantities that are used to describe variability in a population (e.g. mean, standard deviation and percentiles).
Posterior distribution	In Bayesian inference, a probability distribution representing uncertainty about parameters in a statistical model after observing data from the model. The distribution combines information obtained from the data with any information used to derive the prior distribution
Prior distribution	In Bayesian inference, a probability distribution representing uncertainty about parameters in a statistical model prior to observing data from the model. The distribution may be derived from expert judgments based on other sources of information
Probabilistic	1) Representation of uncertainty and/or variability using probability distributions. 2) Calculations where one or more inputs are probability distributions and repeated calculations give different answers.
Probability	Defined depending on philosophical perspective: 1) the frequency with which samples arise within a specified range or for a specified category; 2) quantification of uncertainty as degree of belief regarding the likelihood of a particular range or category. See Section 6.3.

Propagation of uncertainty	Propagation refers to the process of carrying one or more uncertainties through an assessment in order to evaluate their impact on the assessment outcome. It may be done by calculation or expert judgement.
Probability bound	A limited probability statement which states that a probability is greater than some specified value, or less than a specified value, or lies between two specified values.
Quantity	A property or characteristic having a numerical scale.
Quantitative question	A question requiring estimation of a quantity. E.g., estimation of exposure or a reference dose, the level of protein expression for a GM trait, the infective dose for a pathogen, etc.
Range	A set of contiguous values or categories, specified by an upper and lower bound.
Risk analysis	A process consisting of three interconnected components: risk assessment, risk management and risk communication.
Risk assessment	A scientifically based process consisting of four steps: hazard identification, hazard characterisation, exposure assessment and risk characterisation.
Risk communication	The interactive exchange of information and opinions throughout the risk analysis process as regards hazards and risks, risk-related factors and risk perceptions, among risk assessors, risk managers, consumers, feed and food businesses, the academic community and other interested parties, including the explanation of risk assessment findings and the basis of risk management decisions.
Risk management	The process, distinct from risk assessment, of weighing policy alternatives in consultation with interested parties, considering risk assessment and other legitimate factors, and, if need be, selecting appropriate prevention and control options.
Risk manager	A type of decision-maker, responsible for risk management.
Severity	Description or measure of an effect in terms of its adversity or harmfulness.
Specific uncertainty	Source of uncertainty specific to a particular assessment, or which arises in a similar way in multiple assessments but is sufficiently different in nature or magnitude to warrant assessing it separately in each case.
Sub-question	A question whose answer is useful to address a subsequent question. Assessment of a complex question may be facilitated by dividing it into a series of sub-questions.
Target quantity	A quantity which it is desired to estimate, e.g., what severity and frequency of effects is of interest. See section 6.4.
Trust (in social science)	The expectation of an outcome taking place within a broad context and not based on prior knowledge or experience.
Typology of uncertainties	A structured classification of uncertainties according to their characteristics.
Uncertainty	In this document, uncertainty is used as a general term referring to all types of limitations in knowledge. (expand as per box in introduction) – explain is also used to refer to a source of uncertainty (or remove this usage from text)
Uncertainty analysis	A collective term for the processes used to identify, characterise, explain and account for uncertainties.
Variable	A quantity that takes multiple values in the real world (e.g. body weight).
Well-defined uncertainty	An uncertain quantity or proposition that is specified in such a way that it would be possible to determine it with certainty if an appropriate observation or measurement could be made, at least in principle (even if it making that observation would never be feasible in practice). See section 6.7.

Annex A – The melamine case study

A.1 Purpose of case study

Worked examples are presented in annexes to the Guidance Document, to illustrate the different approaches. To increase the coherence of the document and facilitate the comparison of different methods, a single case study was selected, which is introduced in the following section.

Presentation of the case study is arranged as follows:

- Introduction to the melamine example (this Annex, section A2)
- Definition of assessment questions for use in the case study (this Annex, section A3)
- Overview of outputs produced by the different methods (this Annex, section A4)
- Detailed description of how each method was applied to the example (subsections on 'Melamine example' within the sections on each method, in Annex B (1-16))
- Description of models used when demonstrating the quantitative methods (Annex C)
- Examples of complete assessments including characterisation of overall uncertainty, for three levels of refinement (Annex D) – this will be added after the public consultation.

A.2 Introduction to melamine example

The example used for the case study is based on an EFSA Statement on melamine that was published in 2008 (EFSA, 2008). This Statement was selected for the case study in this guidance because it is short, which facilitates extraction of the key information and identification of the uncertainties, and because it incorporates a range of uncertainties. However, it should be noted that the risk assessment in this statement has been superseded by a subsequent full risk assessment of melamine in food and feed (EFSA, 2010).

While this is an example from chemical risk assessment for human health, the principles and methodologies illustrated by the examples are general and could be applied to any other area of EFSA's work, although the details of implementation would vary.

> It is emphasised that the examples on melamine in this document are provided for the purpose of illustration only, and are based on information that existed when the EFSA statement was prepared in 2008. The examples were conducted only at the level needed to illustrate the principles of the approaches and the general nature of their outputs. They are not representative of the level of consideration that would be needed in a real assessment and must not be interpreted as a definitive assessment of melamine or as contradicting anything in any published assessment of melamine.

The case study examples were developed using information contained in the EFSA (2008) statement and other information cited therein, including a previous US FDA assessment (FDA, 2007). Where needed for the purpose of the examples, additional information was taken from EFSA (2012) opinion on default values for risk assessment or from EFSA's databases on body weight and consumption, as similar information would have been available in other forms in 2008.

The EFSA (2008) statement was produced in response to a request from the European Commission for urgent scientific advice on the risks to human health due to the possible presence of melamine in composite food products imported from China into the EU. The context for this request was that high levels of melamine in infant milk and other milk products had led to very severe health effects in Chinese children. The import of milk and milk products originating from China is prohibited into the EU, however the request noted that "Even if for the time being there is no evidence that food products containing melamine have been imported into the EU, it is appropriate to assess, based on the information provided as regards the presence of melamine in milk and milk products, the possible (worst case) exposure of the European consumer from the consumption of composite food products such as biscuits and confectionary (in particular chocolate) containing or made from milk and milk products containing melamine."

The statement identified a number of theoretical exposure scenarios for biscuits and chocolate containing milk powder both for adults and children.

In the absence of actual data for milk powder, the highest value of melamine (2,563 mg/kg) reported in Chinese infant formula was used by EFSA (2008) as the basis for worst case scenarios. The available data related to 491 batches of infant formula produced by 109 companies producing infant formula. Melamine at varying levels was detected in 69 batches produced by 22 companies. Positive samples from companies other than the one with the highest value of 2,563 mg/kg, had maximum values ranging from 0.09 mg/kg to 619 mg/kg. The median for the reported maximum values was 29 mg/kg. Tests conducted on liquid milk showed that 24 of the 1,202 batches tested were contaminated, with a highest melamine concentration of 8.6 mg/kg.

Milk chocolate frequently contains 15–25 percent whole milk solid. Higher amounts of milk powder would negatively influence the taste of the product and are unlikely in practice; therefore the upper end of this range (25%) was used in the worst case scenario of EFSA (2008).

Data on consumption of Chinese chocolate were not available. The high level consumption of chocolate used in the exposure estimates in the EFSA statement were based on the EU average annual per capita consumption of chocolate confectionary of 5.2 kg (equivalent to an average EU daily per capita consumption of 0.014 kg). The average daily consumption was extrapolated to an assumed 95th percentile of 0.042kg per day, based on information in the Concise European Food Consumption Database. In estimating melamine intake expressed on a body weight basis, a body weight of 20kg was used for children.

Because the request was for urgent advice (published 5 days after receipt of the request), the EFSA statement did not review the toxicity of melamine or establish a Tolerable Daily Intake (TDI). Instead it adopted the TDI of 0.5 mg/kg b.w. set by the former Scientific Committee for Food (SCF) for melamine in the context of food contact materials (EC, 1986). The primary target organ for melamine toxicity is the kidney. Because there is uncertainty with respect to the time scale for development of kidney damage, EFSA used the TDI in considering possible effects of exposure to melamine over a relatively short period, such as might occur with repeated consumption of melamine contaminated products.

The assessment in the EFSA (2008) statement used conservative deterministic calculations that addressed uncertainty and variability in a number of ways: through assessment factors used by the SCF in deriving the TDI (though documentation on this was lacking); assuming contaminated foods were imported into the EU and focussing on consumers of those foods; using alternative scenarios for consumers of individual foods or combinations of two contaminated foods; using mean/median and high estimates for 3 exposure parameters; and comparing short-term exposure estimates with a TDI that is protective for exposure over a lifetime.

The EFSA statement concluded that, for the scenarios considered, estimated exposure did not raise concerns for the health of adults in Europe, nor for children with mean consumption of biscuits. In worst case scenarios with the highest level of contamination, children with high daily consumption of milk toffee, chocolate or biscuits containing high levels of milk powder would exceed the TDI, and children who consumed both such biscuits and chocolate could potentially exceed the TDI by more than threefold. However, EFSA noted that it was unknown at that time whether such high level exposure scenarios were occurring in Europe.

A.3 Defining assessment questions for the case study

When preparing the case study for this document, it was noted that the Terms of Reference for the EFSA (2008) Statement included the phrase: "it is appropriate to assess…the possible (worst case) exposure of the European consumer from the consumption of composite food products such as biscuits and confectionary (in particular chocolate) containing or made from milk and milk products containing melamine". It appears from this that the decision-maker is interested in the actual worst case exposure, i.e. the most-exposed European consumer.

The 2008 Statement included separate assessments for adults and children, consuming biscuits and/or chocolate. For the purpose of illustration the following examples are restricted to children and chocolate because, of the single-food scenarios considered in the original Statement, this one had the highest estimated exposure.

On this basis, the first question for uncertainty analysis was defined as follows: *does the possible worst case exposure of high-consuming European children to melamine from consumption of chocolate containing contaminated Chinese milk powder exceed the relevant health-based guidance value, and if so by how much?*

In addition, a second question was specified, concerning a specified percentile of the exposed population. This was added in order to illustrate the application of methods that quantify both variability and uncertainty probabilistically. This second question was defined as follows: *does the 95th percentile of exposure for European children to melamine from consumption of chocolate containing contaminated Chinese milk powder exceed the relevant health-based guidance value, and if so by how much?* This question might be of interest to decision-makers if the answer to the first question raised concerns.

A.4 Identification of uncertainties

Each part of the EFSA (2008) risk assessment was examined for potential sources of uncertainty. Tables A.1 and A.2 below list the uncertainties that were identified in the case study for this guidance document, numbered to show how they relate to the types of uncertainty listed in Tables 1 and 2 in Section 7 of the guidance document.

A.5 Example output from each method described in Annex B

Table A.3 and the following subsections present a short summary of what each method contributes to uncertainty analysis, illustrated by examples for the melamine case study. Some methods provide inputs to the analysis (shown in italics in Table A.3), while others contribute to the output (shown in quotes).

Each subsection begins with a short statement of the principle of the method and a short summary statement of its contribution to the uncertainty analysis. Where the output of the method is a contribution to the output of the uncertainty analysis, this is expressed in a summary form that might be used as part of communication with decision-makers. Where the output of the method is an input to other parts of uncertainty analysis, e.g. a distribution for an assessment input, this is briefly described. These short summaries are presented together in Table A.3, to provide an overview of the types of contributions the different methods can make.

The subsections following Table A.3 also include a limited version of the assessment output behind the summary statement, such as might be provided as a first level of detail from the underpinning assessment, if this was wanted by the decision-maker. More details of how the outputs were derived are presented in the respective sections of Annex B, and the model of melamine exposure that was used with the quantitative methods is described in Annex C.

It is important to note that while it is unlikely that any single assessment would use all the methods listed in Table A.2, it will be common to use a combination of two or more methods to address different uncertainties affecting the same assessment. See sections 9.3 and 10 of the main document for further explanation of how the different methods can be combined to produce a characterisation of overall uncertainty.

Note: The results in Table A.3 are examples, the purpose of which is only to illustrate the forms of contribution that can be made by the different methods. They should not be interpreted as real evaluations of uncertainty for the EFSA (2008) assessment nor any other assessment. Apparent conflicts between results from different methods are due to differing assumptions that were made in applying them, including differences in which sources of uncertainty were considered.

It should also be noted that some of the methods were only applied to the exposure calculations in Annex B. For the purpose of comparison with other methods, the exposure estimates are expressed as ratios to the TDI of 0.5 mg.kg bw/day in this Annex, without any consideration of uncertainty about the TDI.

A number of observations may be made from Table A.3:

- Four of the methods (expert knowledge elicitation, confidence intervals, the bootstrap and Bayesian inference) provide *inputs to other parts of uncertainty analysis*. Expert knowledge elicitation can also be applied to the output of uncertainty analysis, as in the characterisation of overall uncertainty (see Section 10 of guidance document).
- The other methods in Table A.3 *contribute to the output of uncertainty analysis*. Many assessments will use a combination of methods addressing different sources of uncertainty, making complementary contributions to the uncertainty analysis. Also, in every assessment, some uncertainties will not be individually assessed by any method. Therefore, it will always be necessary to conclude with a characterisation of overall uncertainty, combining the results from different methods with expert judgements about the uncertainties were not individually quantified (see Section 10 of guidance document).
- It can be observed from Table A.3 that those methods contributing to the output of the uncertainty analysis differ markedly in the nature of the information they provide. The descriptive, ordinal and matrix methods provide only qualitative information, and do not express how different the exposure or risk might be or how likely that is. The quantitative methods do provide information of that sort, but in different forms. Deterministic calculations with conservative assumptions provide conservative (high end) estimates; the likelihood of those estimates was not quantified in the case study, although this could be added (e.g. by expert judgement). Interval analysis and the uncertainty table for quantitative questions both provide a range of estimates, but no indication of the probability of values outside that range. Probability bounds analysis provides an upper estimate and also information on the probability of higher values. None of the preceding methods provide information on where the most likely values might lie. The two Monte Carlo methods do provide that information, as well as both lower and upper estimates and the probability of lower or higher values. NUSAP provides ordinal information on the relative influence of different assessment inputs to the uncertainty of the assessment output, while sensitivity analysis provides quantitative information on this. Finally, the uncertainty table for categorical questions addresses a different aspect of the risk assessment, providing an expression of the probability that a hazard exists, based on weight-of-evidence considerations.
- The examples in Table A.3 illustrate the general types of contribution that the different methods can make to uncertainty analysis, and may be helpful in considering which methods to select for particular assessments. However, the case study was necessarily limited in scope, and does not illustrate the full potential of each method. Finally, it is emphasised again that most assessments will include more than one method, addressing different uncertainties, and all should end with a characterisation of overall uncertainty that provides an integrated evaluation of all the identified uncertainties.

Table A.1: List of uncertainties affecting *assessment inputs* for the EFSA (2008) statement on melamine, as identified in the case study for this document. Note that in some instances other assumptions were used in the different methods of uncertainty analysis (Annex B) in order to explore their applicability.

Assessment components		Types of uncertainty (from Table 1 in the Guidance Document)	Specific sources of uncertainty (and related types of uncertainty)
Assessment/ sub-assessment	**Assessment inputs**		
Hazard identification	Identification of toxic effects	1. Ambiguity (incomplete information) 2. Measurement 3. Sampling (e.g with respect to numbers of animals, power of the study) 4. Assumptions 5. Extrapolation 6. Distribution 7. Other	No details in the EFSA statement or SCF opinion on the critical studies and what effects were tested for (1). Possibility of more sensitive effects than the measure of kidney damage used in establishing the TDI (2) Lack of information on key study protocol (e.g numbers of animals, power of the study (3)
Hazard characterization	TDI	1. Ambiguity (incomplete information) 5. Extrapolation	No details available on type of study or derivation of TDI (1) Assumed that TDI of 0.5 mg/kg appropriately derived from adequate study (1,5) Assumed that uncertainty factor of 100 was used and is appropriate for inter- and intra-species differences (1, 5) Possibility that TDI would be lower if based on more sensitive endpoints or higher if uncertainty factor of less than 100 would be appropriate (1,5)
Exposure assessment	Maximum concentration of melamine in milk powder	1. Measurement 3. Sampling 4. Assumptions 5. Extrapolation	Unknown accuracy of the method used to measure melamine (1) 491 batches from 109 companies (3) Used maximum measured value 2563 mg/kg as proxy for the maximum actual value (4,5) Extrapolation from infant formula to milk powder (5)
	Maximum concentration of milk powder in chocolate	4. Assumptions 5. Extrapolation	Assumed 25%, based on information about industry practice for chocolate produced in EU (4) Extrapolation from EU chocolate to Chinese chocolate (5)
	Maximum daily consumption of Chinese chocolate	2. Measurement 3. Sampling 4. Assumptions 5. Extrapolation 6. Distribution	Estimates based on data for chocolate confectionery (2,3,5) Accuracy of per capita consumption data unknown (2,3,4) Representativeness of consumption data unknown (3,5,6) Used an estimate of 95th percentile daily consumption as proxy for maximum actual value (5,6) Extrapolation from daily average to 95th percentile based on a different database (5,6) Extrapolation from chocolate overall to Chinese chocolate (5)
	Body weight	4. Assumptions 6. Distribution	Default value of 20kg for children (4,6)

Table A.2: List of uncertainties affecting the *assessment structure* for the EFSA (2008) statement on melamine, as identified in the case study for this document. Note that in some instances other assumptions were used in the different methods of uncertainty analysis (Annex B) in order to explore their applicability.

Assessment output	Assessment structure	Types of uncertainty (from Table 2 in Guidance Document)	Specific sources of uncertainty (and related types of uncertainty)
Risk characterization	Model for estimating exposure as % of TDI	1. Ambiguity 2. Excluded factors 3. Relationship between components 4. Distribution 5. Evidence for the structure of the assessment 6. Comparisons of independent data 7. Dependency between uncertainties 8. Other	Lack of information on duration of exposure to melamine in chocolate, and how it compares to the timescale required for kidney damage to develop (1,3) Uncertainty about the relation between age, body weight and chocolate consumption (whether the daily chocolate consumption of 0.042 kg applies to children of 20 kg) (3,7)

Table A.3: Short summary of what each method contributes to uncertainty analysis, illustrated by examples for the melamine case study. Some methods provide inputs to the analysis (shown in italics), while others contribute to the output (shown in quotes). The right hand column provides a link to more detail.

Method	**Short summary of contribution** *Examples based on melamine case study. Apparent conflicts between results are due to differing assumptions made for different methods.*	Section No.
Descriptive expression	Contribution to output: "Exposure of children could potentially exceed the TDI by more than threefold, but it is currently unknown whether such high level scenarios occur in Europe."	B.1.
Ordinal scale	Contribution to output: "The outcome of the risk assessment is subject to 'Medium to high' uncertainty."	B.2.
Matrices for confidence/uncertainty	Contribution to output: "The outcome of the risk assessment is subject to 'Low to medium' to 'Medium to high' confidence."	B.3.
NUSAP	Contribution to output: "Of three parameters considered, consumption of Chinese chocolate contributes most to the uncertainty of the risk assessment."	B.4.
Uncertainty tables for quantitative questions	Contribution to output: "The worst case exposure is estimated at 269% of the TDI but could lie below 30% or up to 1300%".	B.5.
Uncertainty tables for categorical questions	Contribution to output: "It is Very likely (90-100% probability) that melamine has the capability to cause adverse effects on kidney in humans." (Hazard assessment)	B.6.
Interval analysis	Contribution to output: "The worst case exposure is estimated to lie between 11 and 66 times the TDI."	B.7.
Expert knowledge elicitation	Input to uncertainty analysis: *A distribution for use in probabilistic calculations, representing expert judgement about the uncertainty of the maximum fraction of milk powder used in making milk chocolate.*	B.8. & B.9.
Confidence intervals	Input to uncertainty analysis: *95% confidence intervals representing uncertainty due to sampling variability for the mean and standard deviation of the logarithm of body weight were (1.028, 1.046) and (0.054, 0.067) respectively.*	B.10.
The bootstrap	Input to uncertainty analysis: *A bootstrap sample of values for mean and standard deviation of log body-weight distribution, as an approximate representation of sampling uncertainty for use in probabilistic calculations.*	B.11.
Bayesian inference	Input to uncertainty analysis: *Distributions quantifying uncertainty due to sampling variability about the mean and standard deviation of log body weight, for use in probabilistic calculations.*	B.12.
Probability bounds	Contribution to output: "There is at most a 10% chance that the worst case exposure exceeds 37 times the TDI."	B.13.
1D Monte Carlo (uncertainty only)	Contribution to output: "There is a 95% chance that the worst case exposure lies between 14 and 30 times the TDI, with the most likely values lying towards the middle of this range."	B.14.
2D Monte Carlo (uncertainty and variability)	Contribution to output: "There is a 95% chance that the percentage of 1-2 year old children exceeding the TDI is between 0.4% and 5.5%, with the most likely values lying towards the middle of this range."	B.14.

Deterministic calculations with conservative assumptions	Contribution to output: "The highest estimate of adult exposure was 120% of the TDI, while for children consuming both biscuits and chocolate could potentially exceed the TDI by more than threefold."	B.15.
Sensitivity analysis (various methods)	Contribution to output: "Exposure is most sensitive to variations in melamine concentration and to a lesser extent chocolate consumption."	B.16.

A.5.1 Descriptive expression of uncertainty

Descriptive methods characterise uncertainty using only verbal expressions, without any defined ordinal scale, and without any quantitative definitions of the words that are used.

Short summary of contribution to uncertainty analysis: "Exposure of children could potentially exceed the TDI by more than threefold, but it is currently unknown whether such high level scenarios occur in Europe." (Contribution to output of uncertainty analysis)

This is an abbreviated version of part of the conclusion of the EFSA (2008) statement:

'Children who consume both such biscuits and chocolate could potentially exceed the TDI by more than threefold. However, EFSA noted that it is presently unknown whether such high level exposure scenarios may occur in Europe.'

The EFSA (2008) statement also includes descriptive expression of some individual sources of uncertainty that contribute to the uncertainty of the assessment outcome: '*There is uncertainty* with respect to the time scale for the development of kidney damage' and '*In the absence of actual data* for milk powder, EFSA used the highest value of melamin'. The words expressing uncertainty are italicised.

For more details on descriptive expression see Section 1 of Annex B.

A.5.2 Ordinal scale

An ordinal scale is a scale that comprises two or more categories in a specified order without specifying anything about the degree of difference between the categories.

Short summary of contribution to uncertainty analysis: "The outcome of the risk assessment is subject to 'Medium to high' uncertainty." (Contribution to output of uncertainty analysis)

This is based on evaluation of 3 sources of uncertainty as follows:

Source of uncertainty	**Level of uncertainty**
Hazard characterization (TDI)	'Low to medium' to 'Medium to high'
Concentration of melamine in milk powder	'Medium to high'
Consumption of Chinese chocolate	'Medium to high' to 'High'
Impact on risk assessment of these three sources of uncertainty combined.	'Medium to high'*

*The category 'Medium to high' uncertainty was defined as follows: "Some or only incomplete data available; evidence provided in small number of references; authors' or experts' conclusions vary, or limited evidence from field observations, or moderate data available from other species which can be extrapolated to the species being considered."

For more details on ordinal scales see Section 2 of Annex B.

A.5.3 Matrices for confidence and uncertainty

Matrices can be used to combine two ordinal scales representing different sources or types of confidence or uncertainty into a third scale representing a combined measure of confidence or uncertainty.

Short summary of contribution to uncertainty analysis: "The outcome of the risk assessment is subject to 'Low to medium' to 'Medium to high' confidence." (Contribution to output of uncertainty analysis)

This is based on evaluation of the *level of evidence* and *agreement between experts* supporting the assessment, as follows:

- Level of evidence (type, amount, quality, consistency): Low to medium
- Level of agreement between experts: High
- Level of confidence: 'Low to medium' to 'Medium to high'

Each aspect was rated on a four point scale: Low, Low to medium, Medium to high, High.

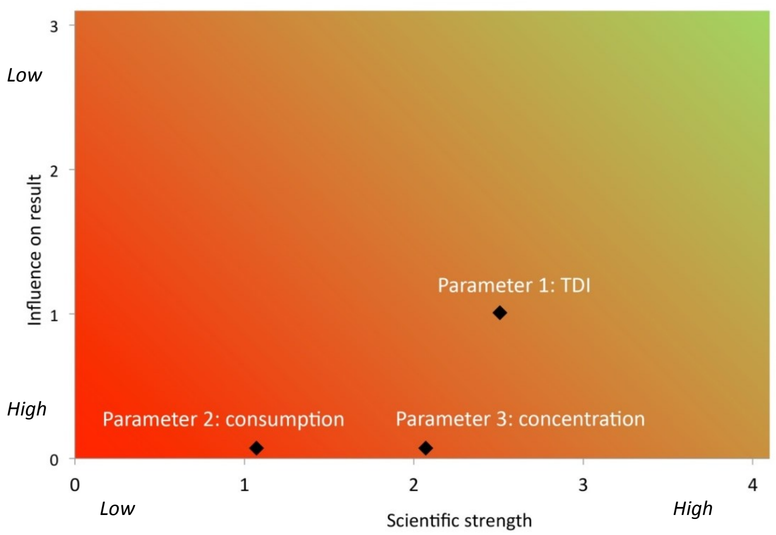
For more details on matrices see Section 3 of Annex B.

A.5.4 NUSAP

NUSAP stands for: Numeral, Unit, Spread, Assessment and Pedigree. A Pedigree matrix typically has four ordinal scales for assessing the strength of parameters or assumptions, and one ordinal scale for their influence on the assessment outcome.

Short summary of contribution to uncertainty analysis: "Of three parameters considered, consumption of Chinese chocolate contributes most to the uncertainty of the risk assessment." (Contribution to output of uncertainty analysis)

This is based on interpretation of the following 'diagnostic plot', showing that chocolate consumption has both poor scientific strength and high influence on the assessment outcome. Each point is the median of judgements by seven assessors on a 5-point ordinal scale.



For more details on NUSAP see Section 4 of Annex B.

A.5.5 Uncertainty tables for quantitative questions

Uncertainty tables for quantitative questions list uncertainties affecting the assessment together with expert judgements of their individual and combined impacts on the assessment outcome, using plus and minus symbols to indicate the direction and magnitude of the impacts.

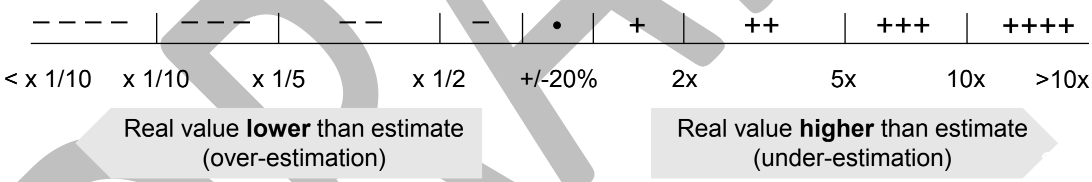
Short summary of contribution to uncertainty analysis: "The worst case exposure is estimated at 269% of the TDI but could lie below 30% or up to 1300%". This should be accompanied by the same caveat as in EFSA (2008): that it is unknown whether the exposure scenario occurs. (Contribution to output of uncertainty analysis)

This is based on expert judgement of uncertainties affecting 3 inputs to the assessment and their impact on the assessment outcome, using a defined scale of symbols, followed by conversion of the symbols for the output to quantitative estimates using the same scale.

Parameters		Value in EFSA (2008) assessment	Uncertainty range
Assessment inputs	TDI	0.5 mg/kg bw/day	---/++*
	Highest concentration of melamine in milk powder	2563 mg/kg	---/+
	Highest consumption of Chinese chocolate by children	0.044 kg	---/++
Assessment output	Ratio of the calculated exposure to the TDI	269%	----/++* (<30% - 1300%)

*One expert considered these uncertainties to be unquantifiable.

Scale for ranges shown in the table above:



For more details on uncertainty tables for quantitative questions see Section 5 of Annex B.

A.5.6 Uncertainty table for categorical questions

This method provides a structured approach for addressing uncertainty in weight of evidence assessment of categorical questions and expressing the uncertainty of the conclusion.

For the melamine case, it was applied to the question: does melamine have the capability to cause adverse effects on kidney in humans?

Short summary of contribution to uncertainty analysis: "It is Very likely (90-100% probability) that melamine has the capability to cause adverse effects on kidney in humans." (Contribution to output of uncertainty analysis)

This is based on four lines of evidence, as shown in the table below. Expert judgement was used to assess the influence of each line of evidence on the conclusion to the question, expressed using arrow symbols, and the likelihood of a positive conclusion.

Lines of evidence	Influence on conclusion*
Line of Evidence 1 – animal studies	↑↑↑
Line of Evidence 2 – information on effects in humans	↑/↑↑
Line of Evidence 3 – information on mode of action	↑/↑↑
Line of Evidence 4 – evidence of adverse effects in companion animals	↑/↑↑
CONCLUSION on whether melamine has the capability to cause adverse effects on kidney in humans	Very likely (90-100% probability)

*Key to symbols: ↑, ↑↑, ↑↑↑ represent minor, intermediate and strong upward influence on likelihood respectively. Pairs of symbols (↑/↑↑) represent variation of judgements between assessors.

For more details on uncertainty tables for categorical questions see Section 6 of Annex B.

A.5.7 Interval Analysis

Interval analysis is a method to compute a range of values for the output of a risk calculation based on specified ranges for the individual inputs. The output range includes all values which could be obtained from the risk calculation by selecting a single value from the specified range for each input.

Short summary of contribution to uncertainty analysis: "The worst case exposure is estimated to lie between 11 and 66 times the TDI." (Contribution to output of uncertainty analysis)

This was derived by interval analysis from minimum and maximum possible values for each input to the risk calculation, specified by expert judgement, as shown in the table below.

Parameters		Minimum possible value	Maximum possible value
Inputs	Maximum concentration (mg/kg) of melamine in milk powder	2563	6100
	Maximum fraction, by weight, of milk powder in milk chocolate	0.28	0.30
	Maximum consumption (kg/day) of milk chocolate in a single day by a child aged from 1 up to 2 years	0.05	0.1
	Minimum body-weight (kg) of child aged from 1 up to 2 years	5.5	6.5
Outputs	Maximum intake (mg/kg bw/day) of melamine in a single day, via consumption of milk chocolate, by a child aged from 1 up to 2 years	5.5	33.3
	Ratio of maximum intake to TDI for melamine	11	66.6

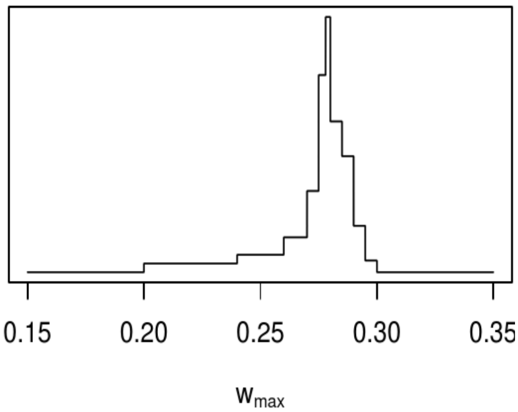
For more details on interval analysis see Section 7 of Annex B.

A.5.8 Expert Knowledge Elicitation (formal and informal)

Expert knowledge elicitation (EKE) is a collection of methods for quantification of expert judgements of uncertainty, about an assessment input or output, using subjective probability.

Short summary of contribution to uncertainty analysis: *A distribution for use in probabilistic calculations, representing expert judgement about the uncertainty of the maximum fraction of milk powder used in making milk chocolate.* (Input to uncertainty analysis)

For the purpose of the case study, an illustrative example was constructed, comprising judgements of 3 fictional experts for minimum, maximum and quartiles, from which the following aggregate distribution was derived.



For more details on formal and informal expert knowledge elicitation see Sections 8 and 9 of Annex B.

A.5.9 Statistical Inference from Data

Each of the methods in this section addresses uncertainty about the parameters of a statistical model for variability based on data. Examples are given in relation to (i) variability of (base 10) logarithm of body-weight and (ii) variability of consumption of chocolate for children aged from 1 up to 2 years.

Confidence Intervals

Confidence intervals representing uncertainty about the parameters for a statistical model describing variability are estimated from data. The result is a range of values for each parameter having a specified level of confidence.

Short summary of contribution to uncertainty analysis: *95% confidence intervals representing uncertainty due to sampling variability for the mean and standard deviation of the logarithm of body weight were (1.028, 1.046) and (0.054, 0.067) respectively.* (Input to uncertainty analysis)

This was calculated from the observed mean and standard deviation of a sample of body weights, assuming they were a random sample from a lognormal distribution.

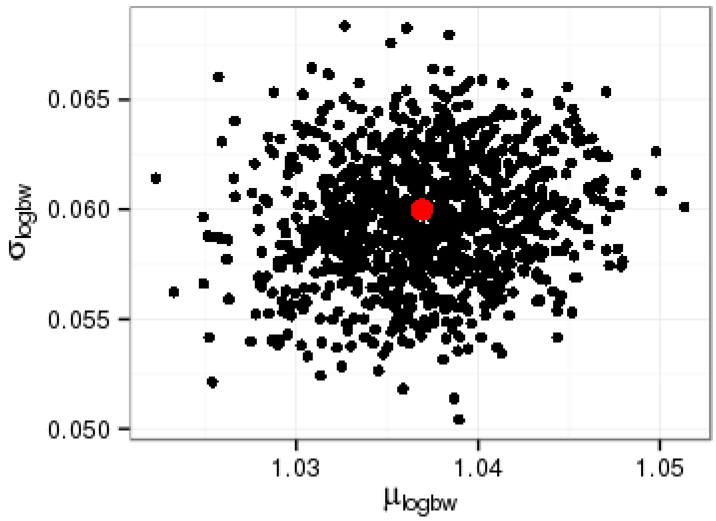
For more details on confidence intervals see Section 10 of Annex B.

The Bootstrap

The bootstrap is a method for obtaining an approximation of uncertainty for one or more estimates, in the form of a sample of possible values, by re-sampling data to create a number of hypothetical datasets of the same size as the original one.

Short summary of contribution to uncertainty analysis: *A bootstrap sample of values for mean and standard deviation of log body-weight distribution, as an approximate representation of uncertainty due to sampling for use in probabilistic calculations.* (Input to uncertainty analysis)

The means and standard deviations for log body weight in the original data and 999 bootstrap samples are plotted in the following Figure.



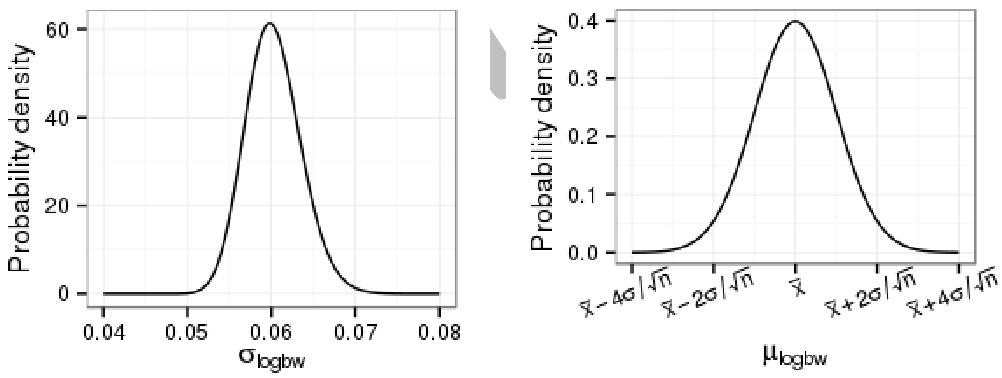
For more details on the bootstrap see Section 11 of Annex B.

Bayesian Inference

Bayesian inference is a method for quantifying uncertainty about parameters in a statistical model of variability on the basis of data and expert judgements about the values of the parameters.

Short summary of contribution to uncertainty analysis: *Distributions quantifying uncertainty due to sampling variability about the mean and standard deviation of log body weight, suitable for use in probabilistic calculations.* (Input to uncertainty analysis)

The distributions for the uncertainty of the standard deviation and mean of log body weight are plotted in the following Figures. The distribution for the mean is conditional on the standard deviation.



For more details on Bayesian inference see Section 12 of Annex B.

A.5.10 Probability Bounds Analysis

Probability bounds analysis is general method for combining limited probability statements (i.e. not complete probability distributions) about inputs in order to make a limited probability specification about the output of a risk calculation.

Short summary of contribution to uncertainty analysis: "There is at most a 10% chance that the worst case exposure exceeds 37 times the TDI." (Contribution to output of uncertainty analysis)

This is one of the outputs produced by probability bounds analysis, shown in the Table below. Also shown are the limited probability statements for each input to the calculation, which were specified by expert judgement.

Parameters		Threshold value	Probability parameter exceeds threshold value
Inputs	Maximum concentration (mg/kg) of melamine in milk powder	3750	≤ 3.5%
	Maximum fraction, by weight, of milk powder in milk chocolate	0.295	≤2%
	Maximum consumption (kg/day) of milk chocolate in a single day by a child aged from 1 up to 2 years	0.095	≤2.5%
	Minimum body-weight (kg) of child aged from 1 up to 2 years	1/(5.6)	≤2%
Outputs	Maximum intake (mg/kg bw/day) of melamine in a single day, via consumption of milk chocolate, by a child aged from 1 to 2 years	18.6	≤10%
	Ratio of maximum intake to TDI for melamine	37.2	≤10%

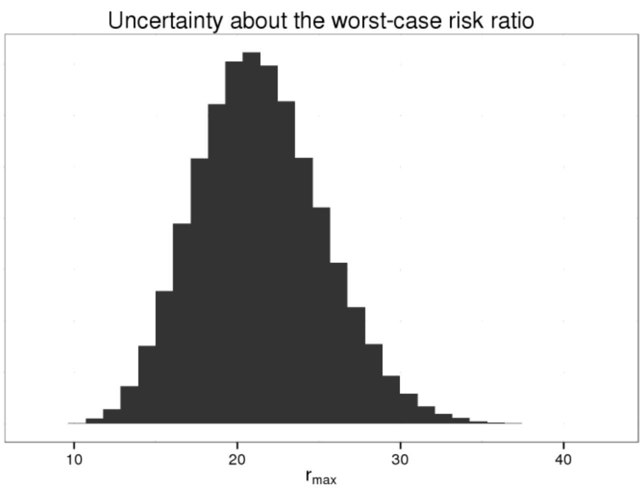
For more details on probability bounds analysis see Section 13 of Annex B.

A.5.11 1D Monte Carlo (Uncertainty only)

1-dimensional (1D) Monte Carlo simulation can be used for combining uncertainty about several inputs in the risk calculation by numerical simulation when analytical solutions are not available.

Short summary of contribution to uncertainty analysis: "There is a 95% chance that the worst case exposure lies between 14 and 30 times the TDI, with the most likely values lying towards the middle of this range." (Contribution to output of uncertainty analysis)

This is based on a distribution for the uncertainty of the worst case exposure produced by 1D Monte Carlo, shown in the following figure, calculated by sampling from distributions for the exposure parameters and the TDI of 0.5 mg/kg bw/day.



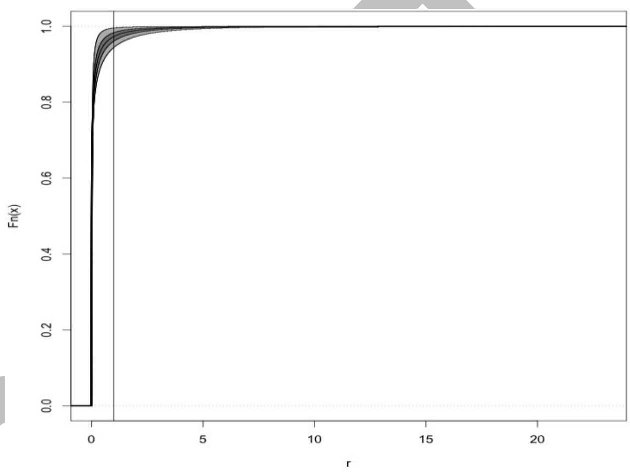
For more details on Monte Carlo for uncertainty only see Section 14 of Annex B.

A.5.12 2D Monte Carlo (Uncertainty and Variability)

2-dimensional (2D) Monte Carlo simulation separates distributions representing uncertainty from distributions representing variability and provides an uncertainty distribution for any interesting summary of variability, in this case the percentage of 1-2 year old children exceeding the TDI.

Short summary of contribution to uncertainty analysis: "There is a 95% chance that the percentage of 1-2 year old children exceeding the TDI is between 0.4% and 5.5%, with the most likely values lying towards the middle of this range." (Contribution to output of uncertainty analysis)

This is based on a 2D distribution quantifying variability and uncertainty of exposure for 1-2 year old children produced by 2D Monte Carlo, shown in the following figure, based on 2D distributions for the exposure parameters and the TDI of 0.5 mg/kg bw/day. The vertical line shows where exposure equals the TDI.



For more details on Monte Carlo for uncertainty and variability see Section 14 of Annex B.

A.5.13 Deterministic calculations with conservative assumptions

These methods deal with uncertainty by using deterministic calculations with assumptions that are conservative, in the sense of tending to overestimate risk.

Short summary of contribution to uncertainty analysis: "The highest estimate of adult exposure was 120% of the TDI, while for children consuming both biscuits and chocolate could potentially exceed the TDI by more than threefold." (Contribution to output of uncertainty analysis)

For more details see Section 15 of Annex B.

A.5.14 Sensitivity Analysis

Sensitivity Analysis is a suite of methods for assessing the sensitivity of the output of the risk calculation to the inputs and to choices made expressing uncertainty about inputs.

Short summary of contribution to uncertainty analysis: "Exposure is most sensitive to variations in melamine concentration and to a lesser extent chocolate consumption." (Contribution to output of uncertainty analysis)

This is based on outputs from several methods of sensitivity analysis for the melamine example, two of which are shown below. For both the nominal range sensitivity analysis index and Sobol first-order index, larger values indicated parameters with more influence on the exposure estimate: melamine concentration and chocolate consumption are more influential than milk powder fraction or body weight which hardly affects the model results.

Input parameters	Nominal range sensitivity analysis index	Sobol first-order index
Concentration (mg/kg) of melamine in milk powder	1.38	0.54
Fraction, by weight, of milk powder in milk chocolate	0.07	0.01
Consumption (kg/day) of milk chocolate in a single day by a child aged from 1 up to 2 years	1	0.19
Body-weight (kg) of child aged from 1 up to 2 years	0.17	0.00

For more details on sensitivity analysis see Section 16 of Annex B.

Annex B – Qualitative and quantitative methods to assess uncertainty

Annex B: Table of Contents

B.1 Descriptive expression of uncertainty

Purpose, origin and principal features

Descriptive expression of uncertainty in this document refers to a form of qualitative assessment of uncertainty using verbal expressions only, without any defined ordinal scale, and without any quantitative definitions of the words. It originates in everyday language rather than any formulated system or theory of uncertainty analysis.

Verbal descriptions are important for expressing the nature or causes of uncertainty. They may also be used to describe the magnitude of an individual uncertainty, the impact of an individual uncertainty on the assessment outcome, or the collective impact of multiple uncertainties on the assessment outcome.

Descriptive expression of uncertainty may be explicit or implicit. Explicit descriptions refer directly to the presence, magnitude or impact of the uncertainty, for example 'the estimate of exposure is highly uncertain'. In implicit descriptions, the uncertainty is not directly expressed but instead implied by the use of words such as 'may', 'possible' or 'unlikely' that qualify, weaken or strengthen statements about data or conclusions in a scientific assessment, for example 'it is unlikely that the exposure exceeds the ADI'.

Descriptive information on uncertainty may be presented at different points within a scientific assessment, Report or Opinion. Individual uncertainties may be described at the specific points of the assessment, where they arise. They may also be summarised and/or discussed together, as part of sections that discuss or interpret the assessment. In some cases, the assessment may include a separate section that is specifically identified as dealing with uncertainty.

Applicability in areas relevant for EFSA

Descriptive phrases are the most commonly-used method for expressing uncertainty in scientific assessment, by EFSA as well as other authorities. In documents produced by EFSA's Panels, such phrases are produced through an iterative drafting process in a Working Group and in its parent Panel or Scientific Committee. At each stage of this process, phrases that are regarded as important or controversial may attract detailed discussion. The Opinion is finalised and adopted by consensus of the Panel or Scientific Committee. If no consensus can be reached then the minority view(s) are recorded in the Opinion, although this is uncommon (about 14 instances up to October 2014).

In order to inform the development of an Opinion on risk assessment terminology (EFSA, 2012), EFSA commissioned a review by external contractors of the language used in the concluding and summary sections of 219 EFSA Opinions published between 2008 and the beginning of 2010. The review found 1199 descriptors which were interpreted by the review authors as expressing uncertainty, of which 1133 were qualitative and 66 quantitative (Table 4 in FERA, 2010). Separate sections dedicated to a type of uncertainty analysis were included in 30 of the 219 documents reviewed.

EFSA's guidance on transparency (EFSA, 2009) states that uncertainties and their relative importance and influence on the assessment outcome must be described. The Opinion of the EFSA Scientific Committee on risk assessment terminology (EFSA, 2012) recommends the use of defined terminology for risk and uncertainty. The Opinion also notes that some words (e.g. 'negligible', 'concern' and 'unlikely') have risk management connotations in everyday language and recommends that, when used in EFSA Opinions, they should be used carefully with objective scientific definitions so as to avoid the impression that assessors are making risk management judgments.

Selected examples from the review by FERA (2010) are presented in Table B.1.1 to provide an indication of the types of words that were used in different contexts in EFSA Opinions at that time. The 5 most frequent descriptors in each category are shown, taken from Tables 17.1-17.9 of FERA (2010). The words that were interpreted as the review authors as expressing possibility or probability are all referring to situations of uncertainty, since they all indicate the possibility of alternative outcomes. Words expressing difficulty of assessment also imply uncertainty (about what the conclusion of the assessment should be), as do words expressing lack of data or evidence. The data presented in the report do not distinguish the use of words to describe uncertainty from their use to describe benefit, efficacy or risk, therefore not all of the words in the Table B.1.1 refer exclusively to uncertainty. Even so, many of the words are ambiguous, in that they provide a relative description whose absolute magnitude is unspecified (e.g. High, Rare, Increase). Other words convey certainty, e.g. some of those relating to comparisons (e.g. Higher), change (e.g. Exceed), agreement (e.g. Agrees with), and absence (e.g. No/Not, which is the most frequent of all the descriptors reviewed).

Table B.1.1: Examples of descriptive terms used in EFSA Opinions.

Context as perceived by authors of FERA (2010).	Most frequent descriptors found by FERA (2010). Numbers are frequency of occurrence, out of 3882 descriptors identified in 219 Opinions.
Words perceived as expressing possibility or probability	May 104, Potential 92, Unlikely 79, Can 47, Likely 46
Words perceived as expressing difficulty or inability to assess or evaluate	Cannot 34, Not possible 30, Could not 18, Not appropriate 9, No conclusion(s) 7
Words perceived as expressing magnitude of benefit or efficacy or risk and/or uncertainty	High 105, Low 92, Safety concern(s) 78, Limit/Limited 52, Moderate 49
Words perceived as expressing comparison of benefit, efficacy or risk or uncertainty	Higher 48, Below 32, Increase/Increased/Increasing 26, Lower 25, Highest 23
Words perceived as expressing frequency relevant to the assessment of benefit or efficacy or risk or uncertainty	Rare/Rarely 15, Occasional/Occasionally/On occasion 5, Often 5, Usually 5, Most frequently 3
Words perceived as expressing change or no Change	Increase/Increased/Increasing 43, Reduce/Reduced 26, Exceed/Exceeded/Exceeding 10, Not exceed/Not be exceeded 8, No change/Not changed 5
Words perceived as expressing agreement or disagreement usually referring to a previous assessment	Agrees with 8, Concurs with 4, Does not agree 4, Confirm 3, Remain(s) valid 3
Words perceived as driving a definite yes/no Outcome	No/Not 225, Contributes 11, Cause/Caused/Causing 10, Demonstrated 8, Established 8
Words perceived as contributing in the characterisation of benefit or efficacy or risk and/or uncertainty, and did not belong to any of the above defined categories	No indication/Do not indicate 45, Controlled 39, No evidence 20, Associated with 12, No new data/information 9

The table shows the 5 most frequently-found descriptors found in 9 different contexts, as perceived by the authors of the FERA (2010) review. Note that several rows of the table refer to benefit, efficacy and risk as well as uncertainty, and the report does not indicate what proportion of occurrences of descriptors relate to each.

The FERA (2010) review considered Opinions published up to early 2010 and therefore does not indicate to what extent the recommendations of EFSA (2009) and EFSA (2012) have been implemented in EFSA's subsequent work.

Potential contribution to the main steps of uncertainty analysis

Potential contribution of descriptive expression to the main steps of uncertainty analysis, as assessed by the Working Group.

Steps in uncertainty analysis	Potential contribution of this approach
Identifying uncertainties	Not applicable
Describing uncertainties	Verbal description.

Assessing the magnitude of individual uncertainties	Verbal description
Assessing the combined impact of multiple uncertainties on the assessment output, taking account of dependencies	Verbal description
Assessing the contribution of individual uncertainties to overall uncertainty	Verbal description

Melamine example

Descriptive narrative is the main method that was used to express uncertainties in the EFSA (2008) statement on melamine. The summary of the statement includes the following phrases, in which the words indicating the presence of uncertainty have been italicised:

'*There is uncertainty* with respect to the time scale for the development of kidney damage.'

'*In the absence of actual data* for milk powder, EFSA used the highest value of melamine…'

'Children who consume both such biscuits and chocolate *could potentially* exceed the TDI by more than threefold. However, EFSA noted that it is *presently unknown* whether such high level exposure scenarios *may* occur in Europe.'

Many further examples can be identified within the detailed text of the EFSA (2008) statement.

Strengths

1. Intuitive, requires no special skills (for assessors proficient in the language used for the assessment).
2. Flexibility – language can in principle describe any uncertainty.
3. Single uncertainties and combined overall uncertainty and its rationale can be expressed in a narrative.
4. Requires less time than other approaches, except when the choice of words provokes extensive discussion (sometimes revisited in multiple meetings).
5. Accepted (or at least not challenged) in most contexts by assessors, decision-makers and stakeholders (but see below).

Weaknesses and possible approaches to reduce them

1. Verbal expressions without quantitative definitions are ambiguous: they are interpreted in different ways by different people. This causes a range of problems, discussed in Section 4 of the Guidance Document and by EFSA (2012).; These problems were recognised by some risk managers interviewed during the development of this guidance, who said they would welcome a move to less ambiguous forms of expression. Ambiguity could be reduced and consistency improved by providing precise (if possible, quantitative) definitions.
2. Where descriptive expression refers to the magnitude of uncertainty, ambiguous wording may leave the decision-maker to assess for themselves the range and likelihood of outcomes – which is a scientific question that should be addressed by assessors. Again, this can be avoided by providing precise definitions.
3. Some words that are used in situations of uncertainty imply risk management judgements, unless accompanied by objective scientific definitions.

4. Lack of transparency of the basis for conclusions that are presented as following from a combination of considerations involving descriptive expressions of uncertainty; this could be partially addressed by describing the relative weight given to each uncertainty.
5. Lack of repeatability due to incomplete recording of the individual experts' involvement and of the chain of arguments leading to the expression of risk and the associated uncertainties; this could in principle be addressed by appropriate recording.

Assessment against evaluation criteria

This method is assessed against the criteria in Table B.1.2.

Conclusions

1. Descriptive expression is currently the main approach to characterising uncertainty in EFSA and elsewhere. However, there are reasons to move towards more quantitative forms of expression, (see EFSA2012 and Chapter 4 of Guidance Document).
2. When a descriptive expression of uncertainty is used, the inherent ambiguity of language means that care is needed to avoid misinterpretation. Ambiguity can be reduced by providing precise definitions that are consistently used across Panels, and by increased dialogue between assessors and decision-makers.
3. When uncertainty is quantified, it may be useful to accompany it with descriptive expression, as the intuitive nature and general acceptance of descriptive expression make it a useful part of the overall communication.
4. Special care is required to avoid using language that implies value judgements, unless accompanied by objective scientific definitions.
5. Descriptive expression should be used to communicate the nature and causes of uncertainty. This is especially important when quantification of uncertainty is not scientifically achievable (see Section 6.7).

References

EFSA, 2009. Guidance of the Scientific Committee on transparency in the scientific aspects of risk assessment carried out by EFSA. Part 2: general principles. *The EFSA Journal (2009)* 1051, 1-22.

EFSA, 2012. Scientific Opinion on Risk Assessment Terminology. EFSA Journal 2012;10(5):2664.

FERA (Food and Environmental Research Agency, UK), 2010. Flari and Wilkinson: Terminology in risk assessment used by the scientific panels and scientific committee of EFSA. http://www.efsa.europa.eu/en/supporting/pub/101e.htm

Table B.1.2: Assessment of Descriptive expression of uncertainty (when applied well) against evaluation criteria.

Criteria	Evidence of current acceptance	Expertise needed to conduct	Time needed	Theoretical basis	Degree/ extent of subjectivity	Method of propagation	Treatment of uncertainty and variability	Meaning of output	Transparency and reproducibility	Ease of understanding for non-specialist
Stronger character-istics	International guidelines or standard scientific method	No specialist knowledge required	Hours	Well established, coherent basis for all aspects	Judgement used only to choose method of analysis	Calculation based on appropriate theory	Different types of uncert. & var. quantified separately	Range and probability of alternative outcomes	All aspects of process and reasoning fully documented	All aspects fully understandable
	EU level guidelines or widespread in practice	Can be used with guidelines or literature	Days	Most but not all aspects supported by theory	Combination of data and expert judgment	Formal expert judgment	Uncertainty and variability quantified separately	Range and relative possibility of outcomes	Most aspects of process and reasoning well documented	Outputs and most of process understandable
	National guidelines, or well established in practice or literature	Training course needed	Weeks	Some aspects supported by theory	Expert judgment on defined quantitative scales	Informal expert judgment	Uncertainty and variability distinguished qualitatively	Range of outcomes but no weighting	Process well documented but limited explanation of reasoning	Outputs and principles of process understandable
	Some publications and/or regulatory practice	Substantial expertise or experience needed	A few months	Limited theoretical basis	Expert judgment on defined ordinal scales	Calculation or matrices without theoretical basis		Quantitative measure of degree of uncertainty	Limited explanation of process and/or basis for conclusions	Outputs understandable but not process
Weaker character-istics	Newly developed	Professional statistician needed	Many months	Pragmatic approach without theoretical basis	Verbal description, no defined scale	No propagation	No distinction between variability and uncertainty	Ordinal scale or narrative description for degree of uncertainty	No explanation of process or basis for conclusions	Process and outputs only understandable for specialists

B.2 Ordinal scale

Purpose, origin and principal features

An ordinal scale is one that comprises two or more categories in a specified order without specifying anything about the degree of difference between the categories. For example, an ordinal scale of low – medium – high has a clear order, but does not specify the magnitude of the differences between the categories (e.g. whether moving from low to medium is the same as moving from medium to high). Ordinal scales provide more information than nominal scales (descriptive categories with no specified order), but less than interval and ratio scales, which quantify the distance between different values (Stevens, 1946). Ordinal scales may therefore be useful when the purpose is to describe the degree of uncertainty in relative terms, e.g. low, medium or high, but should be accompanied by quantitative expressions of uncertainty when possible.

Numerical values can be assigned to the categories as labels, but should then not be interpreted as representing the magnitude of differences between categories. Ordinal scales can be used to rank a set of elements, e.g. from lowest to highest; either with or without ties (i.e. some elements may have the same rank).

Ordinal scales can be used to describe the degree of uncertainty in a qualitative or quantitative risk assessment, e.g. low uncertainty, medium uncertainty, etc. Clearly it is desirable to provide a definition for each category, so that they can be used and interpreted in a consistent manner. In many cases, including the examples provided in the following section, the definitions refer to the causes of uncertainty (e.g. amount, quality and consistency of evidence, degree of agreement amongst experts, etc.). Strictly speaking, these are scales for the amount and quality of evidence rather than degree of uncertainty, although they are related to the degree of uncertainty: e.g., limited, poor quality evidence is likely to lead to larger uncertainty. This relationship is reflected in the approach used by IPCC (Mastrandrea et al., 2010), where 3-point scales for 'Evidence (type, amount, quality, consistency)' and 'Agreement' are combined to derive the 'Level of confidence', which is assessed on a 5-point scale from 'very low' to 'very high'. Level of confidence is inversely related to degree of uncertainty, as discussed in Section 6.

Ordinal scales for degree of uncertainty should ideally represent the magnitude of uncertainty, e.g., the degree to which the true value of a parameter could differ from its estimate. This could be expressed ordinally with categories such as low, medium, high, etc. However, it will usually be important also to provide information on the direction of the uncertainty, e.g., whether the true value is more likely to be higher or lower than the estimate. Perhaps the simplest way to represent this with ordinal scales would be to use a pair of ordinal scales, one indicating how much lower the true value could be, and the other indicating how much higher it could be. An example of this is the +/- scale suggested by EFSA (2006), described in the following section. For qualitative questions (e.g. whether an effect observed in animals can also occur in humans), uncertainty could be expressed on an ordinal scale for likelihood (ideally with quantitative definitions, e.g. Mastrandrea et al. 2010).

Applicability in areas relevant for EFSA

Some EFSA Panels have used ordinal scales that are described as scales for uncertainty, but defined in terms of evidence (e.g. type, amount, quality, consistency) and the level of agreement between experts. In a joint opinion in 2010, the Animal Health and Animal Welfare Panel (AHAW) and the BIOHAZ Panel defined three levels of uncertainty associated with the assessment of the effectiveness of different disease control options of *Coxiella burnetii*, the causative agent of Q-fever (EFSA, 2010).

> "*Low:* Solid and complete data available; strong evidence in multiple references with most authors coming to the same conclusions, or considerable and consistent experience from field observations
>
> "*Medium:* Some or only incomplete data available; evidence provided in small number of references; authors' or experts' conclusions vary, or limited evidence from field observations, or solid and complete data available from other species which can be extrapolated to the species being considered

> *"High:* Scarce or no data available; evidence provided in unpublished reports, or few observations and personal communications, and/or authors'/ or experts' conclusions vary considerably"

As can be seen in this example, different emphasis may be given to the different descriptors used in the definitions: some to the availability of data or the strength of evidence provided; and some to the level of agreement, either in the published literature or in expert's opinions.

The Plant Health (PLH) Panel uses ordinal scales for assessing both risk and uncertainty. Risk assessments are considered in sequential components: entry, establishment, spread and impact of the harmful organism. For each of these components there may be multiple pathways to consider. At each stage of the assessment risk ratings are made on a 5-category ordinal scale (e.g., very unlikely – unlikely – moderately likely – likely – very likely), where the descriptors for the categories must be specified and justified in advance. For each rating, a rating of the associated uncertainty (i.e. the level of confidence in the risk rating given) must also be made**.** Hence, for the risk assessment components – entry, establishment, spread and impact – the level of uncertainty has to be rated separately, usually on a 3-category scale with pre-specified definitions similar to those in the AHAW/BIOHAZ example above. An example of this approach is provided by the Opinion on the plant pest and virus vector *Bemisia* (EFSA,2013). For plants-for-planting the risk of entry of *Bemisia* was rated as *likely,* for cut flowers and branches *moderately likely*, and for fruits and vegetables *unlikely*. The uncertainty of each risk rating was assessed on a 3 point scale (low, medium and high, defined in terms of quality of evidence and degree of subjective judgement) and then consolidated across the three pathways by expert judgement to give an overall uncertainty of 'medium' for entry of *Bemisia* into the EU. This was accompanied by a narrative justification, summarising the rationale for the assessment of 'medium' uncertainty.

Ordinal scales defined in terms of the magnitude and direction of uncertainty, rather than amount or quality of evidence, have been used with 'uncertainty tables' in some EFSA opinions. The categories in these scales are often represented by different numbers of plus and minus symbols, e.g. +, ++, +++. Early examples provided qualitative definitions for the categories such as small, medium or large over-estimation of exposure (EFSA, 2006) and are therefore ordinal scales. Some later examples define the symbols by mapping them on to a quantitative scale, as in the exposure assessment for bisphenol A (EFSA, 2015). This makes the meaning of the categories less ambiguous, and opens the possibility of converting them to intervals for use in quantitative calculations (interval analysis or sensitivity analysis, see sections B.1 and B.2). However, since a scale of such categories is no longer strictly ordinal, they are not further discussed here (see instead section B.3).

Potential contribution to the main steps of uncertainty analysis

Potential contribution of ordinal scales to the main steps of uncertainty analysis, as assessed by the Working Group.

Steps in uncertainty analysis	Potential contribution of this approach
Identifying uncertainties	Not applicable
Describing uncertainties	Pre-definition of ordered categories for describing levels of uncertainty or confidence. Can also be used to describe factors that contribute to uncertainty, e.g. the type, amount, quality and consistency of evidence, or the degree of agreement.
Assessing the magnitude of individual sources of uncertainty	Provides an ordered set of descriptors for expressing magnitude of uncertainty. Categories defined in terms of evidence or agreement may provide indirect measures of magnitude of uncertainty. Assignation of individual uncertainties to the defined categories is assessed by expert judgement.
Assessing the combined impact of multiple uncertainties on the assessment output, taking account of dependencies	Ordinal scales can be used to express expert judgements about the combined impact of multiple uncertainties on the assessment output, but provide a more limited expression than quantitative judgements. No theoretically-justified methods available for propagating ordinal categories with

	qualitative definitions.
Assessing the contribution of individual uncertainties to overall uncertainty	Normally, not directly but through expert judgement can inform the assessment of relative contributions.

Melamine example

Members of the Working Group applied an ordinal scale to assess three uncertainties affecting the example assessment of melamine, based on the context described in Section 3 of the Guidance. They considered uncertainty of the answer to the following question: does the possible worst case exposure of high-consuming European children to melamine from consumption of chocolate containing contaminated Chinese milk powder exceed the relevant health-based guidance value, and if so by how much?

The group first defined an ordinal scale for use in the example, based on the 3-level scale with qualitative definitions in terms of level of evidence and agreement that is shown earlier in this section. The group expanded this to a 4-point scale, on the grounds that this avoids a potential tendency for assessors to pick the central category. For the purpose of illustration, the group retained wording similar to that of the original categories. The 4 categories used for the example were as follows:

- Low uncertainty (L): Solid and complete data available; strong evidence in multiple references with most authors coming to the same conclusions, or considerable and consistent experience from field observations.
- Low to medium uncertainty (LM): Moderate amount of data available; evidence provided in moderate number of references; moderate agreement between authors or experts, or moderate evidence from field observations, or solid and complete data available from other species which can be extrapolated to the species being considered.
- Medium to high uncertainty (MH): Some or only incomplete data available; evidence provided in small number of references; authors' or experts' conclusions vary, or limited evidence from field observations, or moderate data available from other species which can be extrapolated to the species being considered.
- High uncertainty (H): Scarce or no data available; evidence provided in unpublished (unverified) reports, or few observations and personal communications, and/or authors'/ or experts' conclusions vary considerably.

The group members were asked to use the above scale to assess three selected sources of uncertainty (content of melamine in milk powder, Chinese chocolate consumption of European children and appropriate health guidance value for melamine) individually, by expert judgement, and also to assess the combined impact of these three sources of uncertainty on the uncertainty of the assessment outcome. The evaluation was conducted in two rounds, with the scores from the first round being collated on-screen and discussed before the second round. This allowed assessors to adjust their scores in the light of the discussion, if they wished. The results are shown in Table B.2.1. If it was desired to arrive at a 'group' evaluation of uncertainty, this could be done either by seeking a consensus view by discussion, or by 'enveloping' the range of categories assigned for each source of uncertainty in the second round. In this example, the latter option would result in evaluations of LM/MH, MH and MH/H for the 3 individual sources of uncertainty and MH for the overall uncertainty in the second round.

Table B.2.1: Example of the use of an ordinal scale (defined in the text above) to evaluate 3 sources of uncertainty affecting the melamine example assessment.

Assessor	Hazard characterization (TDI)	Concentration of melamine in milk powder	Consumption of Chinese chocolate	Overall
1	LM/LM	MH/MH	H/MH	MH/MH
2	LM/LM	MH/MH	H/H	MH/MH

3	MH/LM	LM/MH	MH/MH	MH/MH
4	H/MH	LM/MH	MH/MH	MH/MH
5	H/MH	H/MH	MH/MH	MH/MH
6	LM/LM	MH/MH	MH/MH	LM/MH
7	MH/LM	MH/MH	MH/H	MH/MH

Pairs of scores (e.g. H/MH) show the first and second rounds of assessment respectively.

Strengths

1. Guidelines exist and the method is already used in certain EFSA Panels.
2. Structured approach to rating uncertainties which forces assessors to discuss and agree the ratings (what is meant by e.g. low, medium and high).
3. Ordinal expressions for sources of uncertainty that are not individually quantified may provide a useful summary to inform quantitative expert judgements about the overall uncertainty of the assessment outcome, and to help document the reasoning behind them.

Weaknesses and possible approaches to reduce them

1. Ordinal categories without definitions or with qualitative definitions are subject to linguistic ambiguity, and will be interpreted in different ways by different people. This can partly be avoided by the use of ordinal categories with quantitative definitions such as the IPCC scale for likelihood (Mastrandrea et al. 2010).
2. Ordinal categories with qualitative definitions are sometimes *labelled* with numbers rather than words. This increases the chance that they will be interpreted as expressing a quantitative definition of the degree of uncertainty, which is invalid.
3. Statistical approaches are sometimes used to combine and summate numerical ratings of uncertainty made on an ordinal scale (e.g. mean and variance), for different experts or different sources of uncertainty or both, but this is not valid. Use of the mode, median and percentiles may be appropriate, but are better applied to verbal category descriptors (e.g. the modal uncertainty category is 'high') to avoid invalid interpretation (see preceding point).
4. Although it is possible to devise rules or calculations for combining ordinal measures of uncertainty or propagating them through an assessment, there is no valid theoretical basis for this.
5. Ordinal scales are often defined in terms of evidence and level of agreement: these are measures of evidence and only an indirect indication of degree of uncertainty. Therefore interpreting such a scale as a measure of uncertainty is likely to be incomplete and misleading.
6. Ordinal scales defined in terms of confidence are more directly related to uncertainty, but generally lack a clear interpretation in terms of the range and likelihood of alternative outcomes.
7. Use of three categories in an ordinal scale might lead to a bias towards assigning the middle category. This can be avoided by using four categories.

Assessment against evaluation criteria

The use of ordinal scales for evaluating uncertainty is assessed against the Working Group's criteria in Table B.2.2. The evaluation is based on ordinal scales with qualitative definitions, since a scale with quantitative definitions is no longer ordinal and is closer to an interval approach (see section B.1). For some criteria a range of levels are ticked, as the assessment depends on how ordinal scales are used (with qualitative or quantitative definitions for categories) and where they are applied (to individual uncertainties or overall uncertainty).

Conclusions

1. Ordinal scales are often defined in terms of the nature, amount, quality and consistency of evidence or the degree of agreement between experts. When used in this way, they should be described as scales for evidence or agreement and not as scales for uncertainty, as they do not describe uncertainty directly. However, they may help to inform subsequent judgements about the degree of uncertainty.
2. Ordinal scales can also be used to describe the degree of uncertainty, if they are defined in terms of the range or likelihood of alternative outcomes.
3. Calculations which treat ordinal scales as if they were quantitative are invalid and should not be used.
4. Ordinal scales provide a useful way of summarising multiple sources of uncertainty to inform quantitative judgements about their combined impact, e.g. when assessing the combined effect of uncertainties which are for whatever reason not quantified individually in the assessment.

References

EFSA, 2006. Guidance of the Scientific Committee on a request from EFSA related to Uncertainties in Dietary Exposure Assessment. EFSA Journal 438, 1-54.

EFSA AHAW Panel (EFSA Panel on Animal Health and Welfare), 2010. Scientific Opinion on Q Fever. EFSA Journal 2010; 8(5):1595. [114 pp.]. doi:10.2903/j.efsa.2010.1595.

EFSA PLH Panel (EFSA Panel on Plant Health), 2013. Scientific Opinion on the risks to plant health posed by *Bemisia tabaci* species complex and viruses it transmits for the EU territory. EFSA Journal 2013;11(4):3162. [302 pp.] doi:10.2903/j.efsa.2013.3162.

EFSA CEF Panel (EFSA Panel on Food Contact Materials, Enzymes, Flavourings and Processing Aids), 2015. Scientific Opinion on the risks to public health related to the presence of bisphenol A (BPA) in foodstuffs: Part I – Exposure assessment. EFSA Journal 2015;13(1):3978, 396 pp. doi:10.2903/j.efsa.2015.3978.

Mastrandrea, MD, CB Field, TF Stocker, O Edenhofer, KL Ebi, DJ Frame, H Held, E Kriegler, KJ Mach, PR Matschoss, G-K Plattner, GW Yohe, and FW Zwiers, 2010: Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties. Intergovernmental Panel on Climate Change (IPCC). Available at <http://www.ipcc.ch>.

Stevens SS. 1946. On the Theory of Scales of Measurement, Science 7 June 1946: 103 (2684), 677-680. [DOI:10.1126/science.103.2684.677]

Table B.2.2: Assessment of Ordinal scales with qualitative definitions for expression of uncertainty (when applied well) against evaluation criteria.

Criteria	Evidence of current acceptance	Expertise needed to conduct	Time needed	Theoretical basis	Degree/ extent of subjectivity	Method of propagation	Treatment of uncertainty and variability	Meaning of output	Transparency and reproducibility	Ease of understanding for non-specialist
Stronger character-istics	International guidelines or standard scientific method	No specialist knowledge required	Hours	Well established, coherent basis for all aspects	Judgement used only to choose method of analysis	Calculation based on appropriate theory	Different types of uncert. & var. quantified separately	Range and probability of alternative outcomes	All aspects of process and reasoning fully documented	All aspects fully understandable
	EU level guidelines or widespread in practice	Can be used with guidelines or literature	Days	Most but not all aspects supported by theory	Combination of data and expert judgment	Formal expert judgment	Uncertainty and variability quantified separately	Range and relative possibility of outcomes	Most aspects of process and reasoning well documented	Outputs and most of process understandable
	National guidelines, or well established in practice or literature	Training course needed	Weeks	Some aspects supported by theory	Expert judgment on defined quantitative scales	Informal expert judgment	Uncertainty and variability distinguished qualitatively	Range of outcomes but no weighting	Process well documented but limited explanation of reasoning	Outputs and principles of process understandable
	Some publications and/or regulatory practice	Substantial expertise or experience needed	A few months	Limited theoretical basis	Expert judgment on defined ordinal scales	Calculation or matrices without theoretical basis		Quantitative measure of degree of uncertainty	Limited explanation of process and/or basis for conclusions	Outputs understandable but not process
Weaker character-istics	Newly developed	Professional statistician needed	Many months	Pragmatic approach without theoretical basis	Verbal description, no defined scale	No propagation	No distinction between variability and uncertainty	Ordinal scale or narrative description for degree of uncertainty	No explanation of process or basis for conclusions	Process and outputs only understandable for specialists

B.3 Matrices for confidence and uncertainty

Purpose, origin and principal features

'Risk matrices' are widely used as a tool for combining ordinal scales for different aspects of risk (e.g. likelihood and severity) into an ordinal scale for level of risk. Matrices have also been proposed by a number of authors as a means of combining two or more ordinal scales representing different sources or types of confidence or uncertainty into a third scale representing a combined measure of confidence or uncertainty. The matrix defines what level of the output scale should be assigned for each combination of the two input scales. Ordinal scales themselves are discussed in more detail in section B.2; here the focus is on the use of matrices to combine them.

An example of a matrix used to combine two ordinal scales is provided by Figure B.3.1, used by the Intergovernmental Panel on Climate Change (IPCC, Mastrandrea et al. 2010). The two input scales on the axes of the matrix relate to different sources of confidence in a conclusion: one scale for amount and quality of evidence and the other for degree of agreement (the latter refers to agreement across the scientific community, Kunreuther et al. 2014). These are combined to draw conclusions about the level of confidence in the conclusion. In this example, the relationship between the input and output scales is flexible. IPCC state that, for a given combination of evidence and agreement, different confidence levels could be assigned, but increasing levels of evidence and degrees of agreement are correlated with increasing confidence (Mastrandrea et al. 2010). They also state that level of confidence should be expressed using five qualifiers from 'very low' to 'very high', synthesising the assessors' judgments about the validity of findings as determined through evaluation of evidence and agreement. IPCC also state that confidence cannot necessarily be assigned for all combinations of evidence and agreement and, in such cases, the assessor should report only the individual assessments for evidence and agreement.

Searching for 'uncertainty matrix' on the internet reveals a substantial number of similar structures from other areas of application.

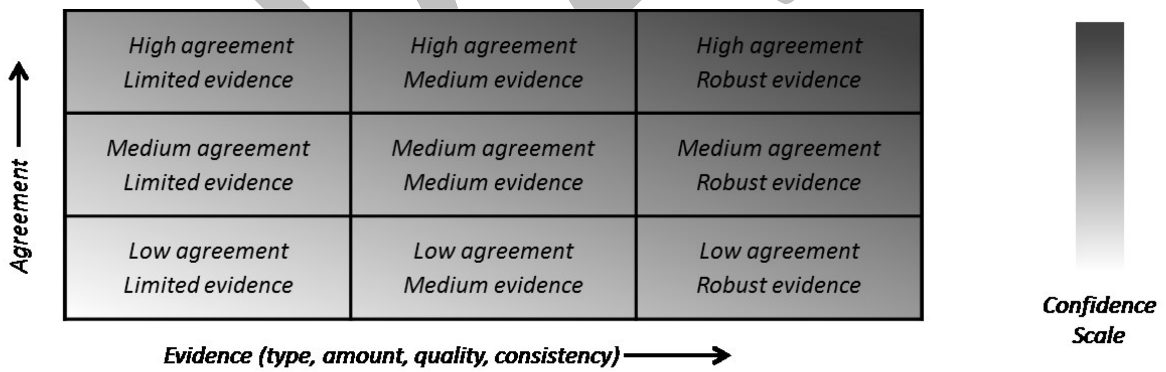


Figure B.3.1: Confidence matrix used by IPCC (Mastrandrea et al., 2010). Confidence increases towards the top-right corner as suggested by the increasing strength of shading. Generally, evidence is most robust when there are multiple, consistent independent lines of high-quality evidence.

Applicability in areas relevant for EFSA

The concept of using a matrix to combine ordinal scales representing different sources or types of uncertainty is a general one and could, in principle, be applied to any area of EFSA's work. For example, in an opinion on cattle welfare (EFSA, 2012), the EFSA Animal Health and Welfare Panel expressed the degree of uncertainty in their assessments of exposure and probability using two ordinal scales, and then used a matrix to derive a third ordinal scale for the uncertainty of the resulting risk (Figure B.3.2).

Exposure uncertainty

Probability uncertainty	**High**	**Medium**	**Low**
High	High	High	High
Medium	High	Medium	Medium
Low	High	Medium	Low

Figure B.3.2: Example of matrix used for combining two ordinal scales representing uncertainty. In this example the two input scales represent uncertainty in different parts of the assessment (uncertainty about exposure to welfare hazards, and uncertainty about the probability of adverse effects given that exposure occurs) and their combination expresses the uncertainty of the assessment as a whole.

Potential contribution to the main steps of uncertainty analysis

Steps in uncertainty analysis	Potential contribution of this approach
Identifying uncertainties	Not applicable
Describing uncertainties	Not applicable
Assessing the magnitude of individual uncertainties	Can be used to combine ordinal scales for different sources of uncertainty affecting the same assessment component, but cumbersome for more than 2 sources and lacks a theoretical basis (see below).
Assessing the combined impact of multiple uncertainties on the assessment output, taking account of dependencies	Can be used to combine ordinal scales for uncertainty in different parts of an assessment, the output expresses the uncertainty of the overall assessment, but cumbersome for more than 2 sources and lacks a theoretical basis (see below).
Assessing the contribution of individual uncertainties to overall uncertainty	The matrix shows how the uncertainties represented by the input scales contribute to the combined uncertainty represented by the output scale, but does not identify individual contributions within each input.

Melamine example

The use of an confidence matrix is illustrated here using a modified version of the IPCC matrix (Mastrandrea et al., 2010), in which each of the two input scales has been expanded from 3 to 4 ordinal categories (Table B.3.1). Note that, as discussed in chapter 6.4 of the main text and in section B.2 of this annex on ordinal scales, confidence is only a partial measure of uncertainty: it expresses the likelihood of a specified conclusion or outcome but provides no information on the range or relative likelihoods of alternative outcomes.

Table B.3.1: Confidence matrix combining ordinal scales for evidence and agreement, adapted from Mastrandrea et al. (2010).

Agreement	High agreement Limited evidence	High agreement Limited to Medium evidence	High agreement Medium to High evidence	High agreement High evidence
	Medium to High agreement Limited evidence	Medium to High agreement Limited to Medium evidence	Medium to High agreement Medium to High evidence	Medium to High agreement High evidence
	Low to Medium agreement Limited evidence	Low to Medium agreement Limited to Medium evidence	Low to Medium agreement Medium to High evidence	Low to Medium agreement High evidence
	Low agreement Limited evidence	Low agreement Limited to Medium evidence	Low agreement Medium to High evidence	Low agreement High evidence

Evidence (type, amount, quality, consistency)

Confidence is considered to increase diagonally across the table from bottom left to top right in a graded way (see Figure B.3.1).

The example considers the uncertainty of the ratio between the worst case exposure of the European children from contaminated chocolate and the TDI for melamine, as assessed in the EFSA (2008) melamine statement where the reported estimate was 269%. For the example, six assessors were asked to evaluate the levels Evidence and Agreement supporting the estimate of 269% and then combine these using Table B.3.1 to assess level of Confidence on the following scale: "very low," "low," "low to medium,", "medium to high", "high," "very high". In doing this, they were invited to make use of the assessment they had conducted immediately previously using a four-category ordinal scale reported in section B.2, where the categories were defined mainly in terms of evidence and the degree of agreement could be judged from the variation in scores between assessors. The assessors' judgements were collected and displayed on screen for discussion, after which the assessors were given the opportunity to amend their judgements if they wished. The results are shown in Table B.3.2. Note that although all the assessors gave identical scores for Evidence and Agreement, their assessments for Confidence varied. This is possible because, as in the IPCC matrix, the group did not assign fixed outputs for each cell in their matrix but, instead, assigned the output by expert judgement informed by the combination of inputs.

Table B.3.2: Evaluation of evidence, agreement and confidence for assessment of the ratio between the worst case exposure of the European children to melamine in contaminated chocolate and the TDI for melamine

Assessor	**Evidence**	**Agreement**	**Confidence**
1	LM	H	MH
2	LM	H	MH
3	LM	H	MH
4	LM	H	LM
5	LM	H	LM
6	LM	H	MH
Range for 6 assessors	LM	H	LM/MH

Key: LM = Low to medium, MH = Medium to high, H = High.

Strengths

1. Simplicity and ease of use: if the matrix gives defined outputs for each combination of inputs (as in Figure B.3.2), it can be used as a simple look-up table. If the matrix gives flexible outputs for each combination of inputs (as in Figure B.3.1), the user needs to make judgements about what outputs to assign, but these may be informed and facilitated by the matrix.

2. Using a matrix (of either type) provides structure for the assessment that should increase the consistency of the uncertainty analysis and also its transparency (it is easy for others to see what has been done, although not necessarily the reasons for it).

Weaknesses and possible approaches to reduce them

1. Using matrices becomes increasingly cumbersome when more than two inputs are involved.
2. The output of the matrix will only be useful if it has meaning. Bull et al. (2013) have demonstrated vastly different evaluations of risk matrices by different individuals and concluded that "*It appears that risk matrices may be creating no more than an artificial and even untrustworthy picture of the relative importance of hazards, which may be of little or no benefit to those trying to manage risk effectively and rationally*". This requires that unambiguous (preferably quantitative) definitions are provided for the meaning of the output. Ideally, the meaning of each level of the output scale should be defined in terms of of its implications for the outcome of the assessment that is being considered. For example, in the melamine example above, how much higher might the true worst case exposure be relative to the relevant health based guidance value, given that confidence in the estimate has been assessed as being in the range 'Low to medium' to 'Medium to high'?
3. Even when the meaning of the output is defined, its reliability will depend on whether the matrix combines the inputs in an appropriate way. Therefore it is essential that the reasoning for the structure of the matrix should be carefully considered and documented, and take account of the nature and relative importance of the inputs and how they should properly be combined to generate the output. Ideally, it should have an appropriate theoretical basis, e.g. in terms of probability theory. Alternatively, it could be based on subjective judgements about how the inputs combine to produce a meaningful measure of the degree of uncertainty. The latter is likely to be less reliable than the former, because of limitations in human ability to make subjective judgements about probability combinations. The IPCC state that the relation between the inputs and outputs of their matrix is flexible, so the user has to judge it case by case.
4. Superficially, a matrix such as that in Figure B.3.2 could be applied to any problem, which would be a major strength. However, defining the matrix structure and output scale sufficiently well to have meaning is likely to limit its applicability to the particular problems and uncertainties for which it was designed. The example in Figure B.3.1 is more generally applicable, but the outputs are not precisely defined and have to be considered by the user, case by case.
5. Even if the matrix structure has a sound basis in probability theory, it will be subject to similar problems to those demonstrated by Cox (2008) for risk matrices. Cox showed that the ordinal input scales discretise the underlying continuous quantities in ways that will cause the matrix outputs to differ, sometimes substantially, from the result that would be obtained by calculation.
6. A matrix does not provide information on the relevant importance of the different sources of uncertainty affecting each of its inputs. If this is needed it should be used in conjunction with other methods.

Assessment against evaluation criteria

The use of uncertainty matrices is assessed against the criteria in Table B.3.3.

Conclusions

1. Matrices with ordinal input and output scales that lack quantitative definitions are ambiguous and will be interpreted in different ways by different users.
2. Matrices that specify a fixed relation between input and output should not be used unless a clear justification, based on theory or expert judgement, can be provided for the relationships involved.

3. Matrices that do not specify a fixed relation between input and output might be regarded as a guide for expert judgement, reminding the user of the factors that should be considered when making judgements. However, users may be tempted to apply them as if they represented fixed rules, leading to inappropriate conclusions.

4. Even when the above issues are avoided, matrices become cumbersome when more than two sources or aspects of uncertainty are involved, which is usual in EFSA assessment.

The issues in (1-4) above are likely to limit the usefulness of matrices as a tool for assessing uncertainty in EFSA's work.

References

Ball DJ, Watt J, 2013. Further thoughts on the utility of risk matrices. Risk Anal. 33(11):2068-78.

Cox LA Jr 2008. What's wrong with risk matrices? Risk analysis 28, 497–511.

EFSA, 2012. Scientific Opinion on the welfare of cattle kept for beef production and the welfare in intensive calf farming systems. The EFSA Journal, 10(5):2669.

Kunreuther H., S. Gupta, V. Bosetti, R. Cooke, V. Dutt, M. Ha-Duong, H. Held, J. Llanes-Regueiro, A. Patt, E. Shittu, and E. Weber, 2014: Integrated Risk and Uncertainty Assessment of Climate Change Response Policies. In: Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Edenhofer, O., R. Pichs-Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Seyboth, A. Adler, I. Baum, S. Brunner, P. Eickemeier, B. Kriemann, J. Savolainen, S. Schlömer, C. von Stechow, T. Zwickel and J.C. Minx (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Mastrandrea, MD, CB Field, TF Stocker, O Edenhofer, KL Ebi, DJ Frame, H Held, E Kriegler, KJ Mach, PR Matschoss, G-K Plattner, GW Yohe, and FW Zwiers, 2010. Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties. Intergovernmental Panel on Climate Change (IPCC). Available at http://www.ipcc.ch.

Table B.3.3: Assessment of Uncertainty matrices (when applied well) against evaluation criteria.

Criteria	Evidence of current acceptance	Expertise needed to conduct	Time needed	Theoretical basis	Degree/ extent of subjectivity	Method of propagation	Treatment of uncertainty and variability	Meaning of output	Transparency and reproducibility	Ease of understanding for non-specialist
Stronger character-istics	International guidelines or standard scientific method	No specialist knowledge required	Hours	Well established, coherent basis for all aspects	Judgement used only to choose method of analysis	Calculation based on appropriate theory	Different types of uncert. & var. quantified separately	Range and probability of alternative outcomes	All aspects of process and reasoning fully documented	All aspects fully understandable
	EU level guidelines or widespread in practice	Can be used with guidelines or literature	Days	Most but not all aspects supported by theory	Combination of data and expert judgment	Formal expert judgment	Uncertainty and variability quantified separately	Range and relative possibility of outcomes	Most aspects of process and reasoning well documented	Outputs and most of process understandable
	National guidelines, or well established in practice or literature	Training course needed	Weeks	Some aspects supported by theory	Expert judgment on defined quantitative scales	Informal expert judgment	Uncertainty and variability distinguished qualitatively	Range of outcomes but no weighting	Process well documented but limited explanation of reasoning	Outputs and principles of process understandable
	Some publications and/or regulatory practice	Substantial expertise or experience needed	A few months	Limited theoretical basis	Expert judgment on defined ordinal scales	Calculation or matrices without theoretical basis		Quantitative measure of degree of uncertainty	Limited explanation of process and/or basis for conclusions	Outputs understandable but not process
Weaker character-istics	Newly developed	Professional statistician needed	Many months	Pragmatic approach without theoretical basis	Verbal description, no defined scale	No propagation	No distinction between variability and uncertainty	Ordinal scale or narrative description for degree of uncertainty	No explanation of process or basis for conclusions	Process and outputs only understandable for specialists

B.4 NUSAP

Purpose, origin and principal features

The purpose of this method is to provide a structured approach to deal with uncertainties in model-based health risk assessments. The NUSAP acronym stands for: Numeral, Unit, Spread, Assessment and Pedigree. The first three dimensions are related to commonly applied quantitative approaches to uncertainty, expressed in numbers (N) with appropriate units (U) and a measure of spread (S) such as a range or standard deviation. Methods to address spread include statistical methods, sensitivity analysis and expert elicitation. The last two dimensions are specific to NUSAP and are related to aspects of uncertainty than can less readily be analysed by quantitative methods. Assessment (A) expresses qualitative expert judgments about the quality of the information used in the model by applying a Pedigree (P) matrix, implying a multi-criteria evaluation of the process by which the information was produced.

The method was first proposed by Funtowicz and Ravetz (1993) and further developed by Van der Sluijs et al. (2005) to evaluate the knowledge base in model-based assessment and foresight studies of complex environmental problems. Such assessments are often characterized by uncertainties in the knowledge base, differences in framing the problem, and high stakes involved in decisions based on these assessments, often with conflicting views between different stakeholders.

The principal features of this method are to consider the background history by which the information was produced, in combination with the underpinning and scientific status of the information. Qualitative judgments about uncertainties are supported by so-called pedigree matrices, which are then translated in a numerical, ordinal scale. Typically, a pedigree matrix has four dimensions for assessing the strength of parameters or assumptions, and one dimension for their influence on results (e.g.Table B.4.1).

Table B.4.1: Example of NUSAP pedigree matrix for scoring parameter strength and influence.

	Strength				Effect
Score	**Proxy**	**Empirical basis**	**Methodological rigor**	**Validation**	**Influence on results**
4	Exact measure of the desired quantity (e.g. from the same geographical area)	Large sample, direct measurements (recent data, controlled experiments)	Best available practice (accredited method for sampling / diagnostic test)	Compared with independent measurements of the same variable (long domain, rigorous correction of errors)	
3	Good fit or measure (e.g. from another but representative area)	Small sample, direct measurements (less recent data, uncontrolled experiments, low non-response)	Reliable method (common within established discipline)	Compared with independent measurements of closely related variable (shorter time periods)	No or negligible impact on the results
2	Well correlated (e.g. large geographical differences , less representative)	Very small sample, modelled/derived data (indirect measurements, structured expert opinion)	Acceptable method (limited consensus on reliability)	Compared with measurements of non-independent variable (proxy variable, limited domain)	Little impact on the results
1	Weak correlation (e.g. very large geographical differences, low representativity)	One expert opinion, rule of thumb	Preliminary method (unknown reliability)	Weak, indirect validation	Moderate impact on the end result
0	Not clearly correlated	Crude speculation	No discernible rigor	No validation	Important impact on the end result

The NUSAP output is a score per uncertainty source for the scientific strength of the information and its influence on the model outcome. In NUSAP, scientific strength expresses the methodological and epistemological limitations of the underlying knowledge base (Van der Sluijs et al., 2005). In comparison to using single ordinal scales, the multi-criteria evaluation provides a more detailed and formalized description of uncertainty. These median scores over all experts for the strength and influence are combined for all uncertainty sources in a diagnostic diagram, which will help to identify the key uncertainties in the assessment, i.e. those sources with a low strength and a large influence on the model outcome. The NUSAP approach therefore can be used to evaluate uncertainties that cannot be quantified, but can also be useful in identifying the most important uncertainties for further quantitative evaluation and/or additional work to strengthen the evidence base of the assessment. Pedigree matrices have been developed to evaluate model parameters and input data as well as assumptions. The method is flexible, in that customized scales can be developed.

The NUSAP method is typically applied in a workshop involving multiple experts with various backgrounds in the subject matter of the assessment. The workshop would build on previous efforts to identify and characterize uncertainties using an appropriate typology. An introductory session would include presentations on the NUSAP methodology, the risk assessment to be evaluated and an open discussion about the identified uncertainties, followed by an introduction to the evaluation methodology and a discussion about the scoring methods. For each assumption, all experts would then be asked to write down their scores on a score-card and to also describe their rationale. Scores and rationales are then reported by all experts to the group and are the basis for a discussion. Experts are then given the opportunity to adjust their scores and invited to submit their results. Computer-assisted tools may help to show the key findings of the workshop directly after completing scoring of all uncertainties. The group discussions and iterative process are an important characteristic of the NUSAP process that helps to create a better and collective understanding of uncertainties. However, the method can also be applied by a small number of experts, see e.g. Bouwknegt et al. (2014) for an example in which only 2 experts provided scores. Data analysis after the workshop involves developing diagnostic diagrams and possibly other data analysis. Also in this respect, the method is flexible and can be adapted to the needs of the risk assessment body.

Applicability in areas relevant for EFSA

The NUSAP methodology has been developed mainly in the environmental sciences, including environmental health risk assessments but is in principle applicable in of EFSA's work. Published examples include an assessment of uncertainties in a Quantitative Microbial Risk Assessment (QMRA) models for Salmonella in the pork chain (Boone et al., 2009) and comparing QMRA-based and epidemiologic estimates of campylobacteriosis in the Netherlands (Bouwknegt et al., 2014). The method has also been applied in two outsourced projects to support BIOHAZ opinions (Vose Consulting, 2010; Vose Consulting, 2011).

The EFSA BIOHAZ Panel has performed a pilot study with the NUSAP methodology in the context of a Scientific Opinion on risk ranking. The Panel concluded that "the combination of uncertainty typology and NUSAP helped to systematically identify and evaluate the uncertainty sources related to model outcomes and to assess their impact on the end results" and that "applying the NUSAP method requires training of the experts involved to overcome ambiguity of language in the pedigree scales" . The Panel recommended that "a framework encompassing uncertainty typology and evaluation (for example by NUSAP) should be part of each risk ranking process to formalize discussions on uncertainties, considering practicality and feasibility aspects".

Potential contribution to the main steps of uncertainty analysis

Steps in uncertainty analysis	Potential contribution of this approach
Identifying uncertainties	Indirectly, by offering a standardized template
Describing uncertainties	Yes, by standardized pedigree matrices

Assessing the magnitude of individual uncertainties	Yes, by expert judgment using a standardized score
Expression of the impact of individual uncertainties on the assessment output	Yes, by standardized pedigree matrices and diagnostic diagrams, qualitatively or using ordinal numbers
Expression of the combined impact of multiple uncertainties on the assessment output	No
Assessing the relative contribution of different sources of uncertainties to the overall uncertainty	Not directly: diagnostic diagrams show the strength and influence of different assumptions, which can be used to judge the relative impact of different sources of uncertainty.

Melamine example

The NUSAP method was applied to evaluate three uncertain parameters in the melamine example. These were: the relevant health-based guidance value for melamine (referred to below as parameter 1), Chinese chocolate consumption (parameter 2) and melamine concentration in milk powder (parameter 3). The model outcome to be evaluated was defined as: does the possible worst case exposure of high-consuming European children to melamine from consumption of chocolate containing contaminated Chinese milk powder exceed the relevant health-based guidance value, and if so by how much?

When considering the results, it must be borne in mind that the main goal of this exercise was to illustrate the methodology, and not to provide a full evaluation of all uncertainties in the melamine risk assessment. Time to prepare and execute the NUSAP workshop was limited, and the results must be considered indicative only. The strength of the three parameters is shown in Figure B.4.1. According to the experts' judgments, the median strength of the parameter health-based guidance value was higher than that of melamine concentration in milk powder, which was higher than that for Chinese chocolate consumption. 50% of all scores for the latter two parameters were between 1 and 2. In particular, the strength of the parameter Chinese chocolate consumption was judged low on proxy and validation (both median scores of 1). The strength and influence diagram (Fig. B.4.2) shows that according to the experts, among the two most uncertain parameters, the consumption of chocolate was most influential on the assessment result.

Considering the group's experience, there needs to be a common understanding of interpretation of the risk management question before the NUSAP session starts. The four dimensions to evaluate parameter strength reflected different aspects of the knowledge base, but were also related and personal interpretations of the exact nature of these dimensions and their scales differed between group members. Therefore, precise definitions and training of experts to understand these definitions are prerequisites to a standardized application of the NUSAP methodology. The influence of a parameter on the risk assessment outcome can be evaluated by only considering the impact of changes in the parameter value on the risk assessment outcome (comparable to local sensitivity analysis, see section B.16). Alternatively, the plausible range over which a parameter may vary and parameter interactions can also be taken into account (comparable to global sensitivity analysis). These two interpretations may lead to different conclusions about parameter influence, and experts need to agree on the interpretation before scoring.

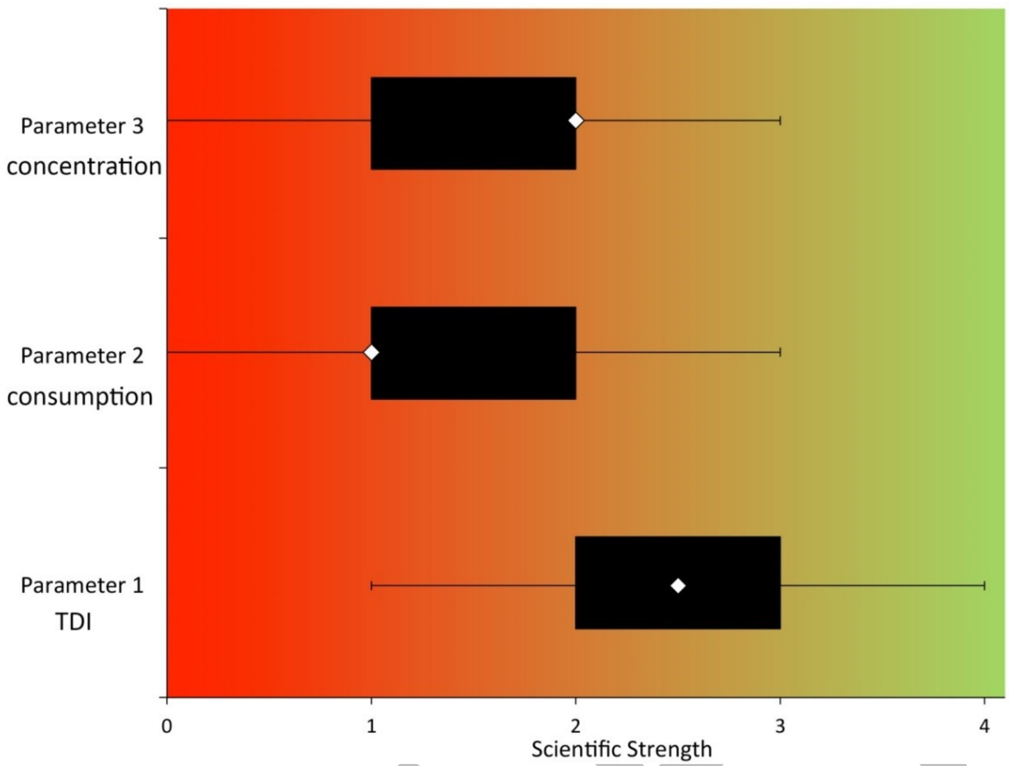


Figure B.4.1: Strength of the information for parameter estimation in the melamine risk assessment. The diamond shows the median of scores of all seven experts on all four dimensions, the black box the interquartile range and the error bars the range of all scores. Colour shading ranges from green to reflect high parameter strength to red to reflect low parameter strength.

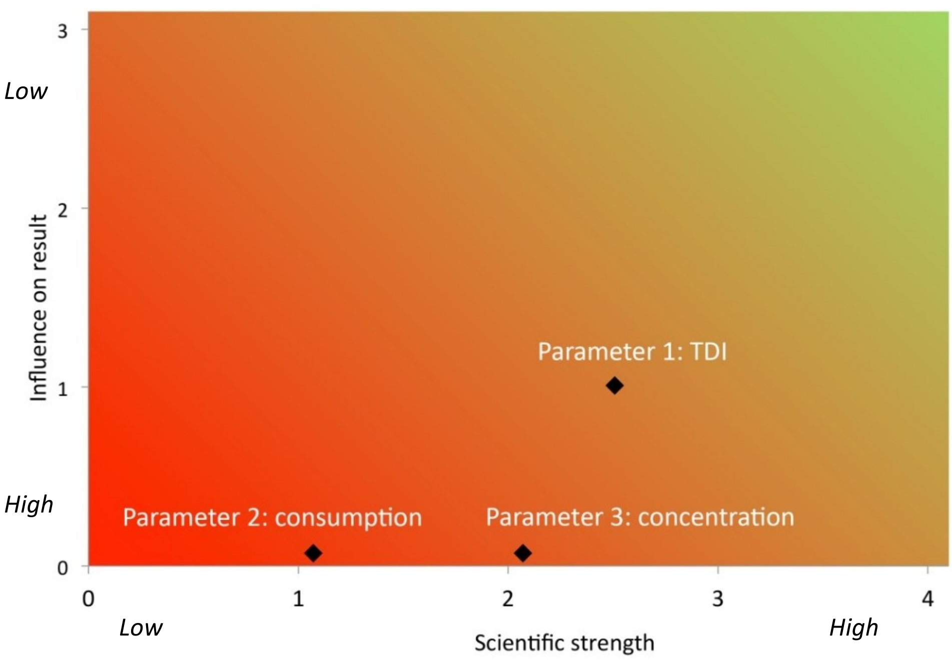


Figure B.4.2: Strength and influence diagram for parameter uncertainty in the melamine risk assessment. The diamond shows the median of scores of all seven experts on all four dimensions for strength and the median score of all seven experts for influence. Colour shading ranges from green to reflect high parameter strength and low influence to red to reflect low parameter strength and high influence.

Strengths

1. Pedigree criteria encourage systematic and consistent consideration of different aspects of uncertainty for each element of an assessment, providing a relative measure of its scientific strength.
2. Can inform the prioritization of uncertain elements in the risk assessment by combining the assessment of scientific strengths with an evaluation of the influence of each element on the assessment outcome using expert judgment.
3. As for other structured judgement approaches, when used in a workshop format NUSAP provides a framework for involving additional experts in an iterative process which should improve the quality of the uncertainty analysis.
4. The NUSAP method could in principle be applied in any area of EFSA's work provided that training is given.

Weaknesses and how to address them

1. The pedigree criteria may be interpreted in different ways by different participants due to ambiguity of the verbal definitions.
2. The current pedigree matrices may not be fully applicable to EFSA's work. However users are free to adapt it to their own purposes.
3. Applying the NUSAP method is more complex than working with ordinal scales.
4. The NUSAP method does not provide an evaluation of the combined effect of multiple uncertainties and therefore needs to be used in conjunction with other methods.
5. Combining scores for different criteria and different experts by taking median lacks theoretical basis and produces an ordinal scale for strengths without defined meaning. They can nevertheless be used as relative measure of strength of evidence.
6. Holding workshops to apply the NUSAP method has costs and time implications. In principle this could be reduced (but not eliminated) by using pedigree matrices and diagnostic diagrams within a normal working group procedure.

Assessment against evaluation criteria

This method is assessed against the criteria in Table B.4.2.

Conclusions

1. The NUSAP method can be used as a qualitative approach to help prioritize uncertain elements in risk assessment for quantitative analysis by other methods.
2. NUSAP may be especially useful as a structured approach for qualitative characterisation of uncertainties for which quantification is not scientifically achievable.
3. NUSAP practitioners encourage its use in a structured workshop format with groups of experts. As for other formal approaches, this requires additional time and resources but increases the chance of detecting relevant uncertainties and provides a more considered characterisation of their impact on the assessment.
4. The NUSAP method should be further evaluated in a series of case studies for EFSA.

5. A common terminology should be developed for use in NUSAP assessments, which is understood by all involved.

References

Boone I, Van der Stede Y, Bollaerts K, Vose D, Maes D, Dewulf J, Messens W, Daube G, Aerts M, Mintiens K. NUSAP method for evaluating the data quality in a quantitative microbial risk assessment model for Salmonella in the pork production chain. Risk Anal 2009;29:502-517.

Bouwknegt M, Knol AB, van der Sluijs JP, Evers EG. Uncertainty of population risk estimates for pathogens based on QMRA or epidemiology: a case study of Campylobacter in the Netherlands. Risk Anal 2014;34:847-864.

Funtowicz SO, Ravetz JR. Science for the post-normal age. Futures 1993;25:735-755.

Van der Sluijs JP, Craye M, Funtowicz S, Kloprogge P, Ravetz J, Risbey J. Combining quantitative and qualitative measures of uncertainty in model-based environmental assessment: the NUSAP system. Risk Anal. 2005;25:481-492.

Vose Consulting. 2010. A quantitative microbiological risk assessment of Campylobacter in the broiler meat chain. Boulder, CO (updated report 2011). EFSA Technical Report 132. http://www.efsa.europa.eu/it/search/doc/132e.pdf

Vose Consulting . 2011. A Quantitative Microbiological Risk Assessment of Salmonella spp. in broiler (Gallus gallus) meat production. Boulder, CO. EFSA Technical report 183. http://www.efsa.europa.eu/it/supporting/doc/183e.pdf

Table B.4.2: Assessment of NUSAP approach (when applied well) against evaluation criteria.

Criteria	Evidence of current acceptance	Expertise needed to conduct	Time needed	Theoretical basis	Degree/ extent of subjectivity	Method of propagation	Treatment of uncertainty & variability	Meaning of output	Transparency and reproducibility	Ease of understanding for non-specialist
Stronger character-istics	International guidelines or standard scientific method	No specialist knowledge required	Hours	Well established, coherent basis for all aspects	Judgement used only to choose method of analysis	Calculation based on appropriate theory	Different types of uncert. & var. quantified separately	Range and probability of alternative outcomes	All aspects of process and reasoning fully documented	All aspects fully understandable
	EU level guidelines or widespread in practice	Can be used with guidelines or literature	Days	Most but not all aspects supported by theory	Combination of data and expert judgment	Formal expert judgment	Uncert. & var. quantified separately	Range and relative possibility of outcomes	Most aspects of process and reasoning well documented	Outputs and most of process understandable
	National guidelines, or well established in practice or literature	Training course needed	Weeks	Some aspects supported by theory	Expert judgment on defined quantitative scales	Informal expert judgment	Uncert. & var. distinguished qualitatively	Range of outcomes but no weighting	Process well documented but limited explanation of reasoning	Outputs and principles of process understandable
	Some publications and/or regulatory practice	Substantial expertise or experience needed	A few months	Limited theoretical basis	Expert judgment on defined ordinal scales	Calculation or matrices without theoretical basis		Quantitative measure of degree of uncertainty	Limited explanation of process and/or basis for conclusions	Outputs understandable but not process
Weaker character-istics	Newly developed	Professional statistician needed	Many months	Pragmatic approach without theoretical basis	Verbal description, no defined scale	No propagation	No distinction between uncert. & var.	Ordinal scale or narrative description for degree of uncertainty	No explanation of process or basis for conclusions	Process and outputs only understandable for specialists

B.5 Uncertainty tables for quantitative questions

Purpose, origin and principal features

An EFSA guidance document on dealing with uncertainty in exposure assessment (EFSA, 2006) suggested using a tabular approach to identify and qualitatively evaluate uncertainties. Three types of tables were proposed, serving complementary functions in the assessment. The first two tables were designed to help assessors identify uncertainties in different parts of exposure assessment. The third table provided a template for assessors to evaluate the individual and combined impacts of the identified uncertainties on their assessment, using plus and minus symbols to indicate the direction and magnitude of the impacts. This section is focussed on this last type of table.

The original purpose of the table was three-fold: to provide an initial qualitative evaluation of the uncertainty to assist in deciding whether a quantitative assessment is needed; to assist in targeting quantitative assessment (when needed) on the most important sources of uncertainty; and to provide a qualitative assessment of those uncertainties that remain unquantified. In practice it has mostly been applied for the latter purpose, at the end of the assessment.

The approach is very general in nature and can be applied to uncertainties affecting any type of quantitative estimate. Therefore, although it was originally designed for evaluating uncertainties in human dietary exposure assessment, it is equally applicable to quantitative estimates in any other area of scientific assessment. It is less suitable for uncertainties affecting categorical questions, for which different tabular approaches have been devised (see section B.6).

The principal features of the method are the listing of uncertainties and evaluation of their individual and combined impacts on the quantitative estimate in question, presented in a table with two or more columns. The impacts are usually expressed using plus and minus symbols, indicating the direction and, in some cases, the magnitude of the impact. In early examples of the approach, the meaning of the plus and minus symbols was described qualitatively (e.g. small, medium, large impacts), but in some later examples a quantitative scale is provided (see below). The most up-to-date detailed description of the approach is included in a paper by Edler et al. (2013, section 4.2).

Applicability in areas relevant for EFSA

EFSA (2006) introduced the tabular approach and provided a simple example, but no detailed guidance. The most frequent user has been the CONTAM Panel, which has used a version of the third type of table in almost all of their Opinions since 2008, and extended it to include uncertainties affecting hazard and risk as well as exposure. CONTAM's version of the table lists the uncertainties affecting their assessment, and indicates the direction of the impact of each individual uncertainty on the assessment outcome: + for uncertainties that cause over-estimation of exposure or risk, and – for those that cause under-estimation. CONTAM initially attempted to indicate the magnitude of the uncertainty by using one, two or three + or – signs, but ultimately decided to use only one + or -, or a combination of both (+/-), due to the difficulty in assigning magnitude. CONTAM provide a qualitative (verbal) evaluation of the combined impact of the uncertainties in text accompanying the table.

The ANS Panel have for some years used uncertainty tables similar to those of EFSA (2006) and the CONTAM Panel and the Scientific Committee have included an uncertainty table in one of their Opinions (EFSA, 2014). Variants of the tabular approach have been used in Opinions and Guidance Documents by PPR Panel (e.g. EFSA 2007, 2008, 2012), a CEF Panel Opinion on bisphenol A (EFSA 2015) and an Opinion of the PLH Panel (EFSA 2013b). Some of these included scales defining quantitative ranges for the + and – symbols (see example below). In some cases the meaning of the + and – symbols was reversed (+ meaning the real exposure or risk may be higher than the estimate, rather than that the estimate is an overestimate).

The EFSA (2006) approach has been taken up in modified form by other EU risk assessment authorities. The ECHA (2008) guidance on uncertainty analysis includes two types of uncertainty table, adapted from those in EFSA (2006). One type of table is used for identifying uncertainties in exposure and effect assessment, while the other is used for evaluating the individual and combined impact of the identified uncertainties on exposure, hazard and risk. The latter table uses + symbols to indicate over-estimation and – for underestimation. One, two or three symbols indicate low, moderate and high magnitude respectively. Similarly, a SCENIHR (2012) memorandum on weight of evidence includes a table for evaluating uncertainty that is closely related to the EFSA (2006) tables. Aspects of uncertainty are listed together with evaluations of their nature, their magnitude and direction, and their importance for the risk assessment.

Edler et al. (2013) describe the application of uncertainty tables for evaluating unquantified uncertainties (those not quantified by the BMDL) in benchmark dose modelling for genotoxic carcinogens. They use uncertainty tables similar to those of EFSA (2006), with + and – symbols defined on a quantitative scale and expressing how much higher or lower the BMDL would be, if adjusted to take account of the unquantified uncertainties. Edler et al. (2013) provide step-by-step guidance on both forms of uncertainty table. Their instructions emphasise the importance of guarding against cognitive biases that tend to affect expert judgement, drawing on ideas from expert elicitation methodology. Annexes to the paper include case studies for the dye Sudan 1 and for PhIP, which is produced during the grilling and frying of meat and fish.

Potential contribution to the main steps of uncertainty analysis

Potential contribution of the uncertainty tables approach described in this section to the main steps of uncertainty analysis.

Steps in uncertainty analysis	Potential contribution of this approach
Identifying uncertainties	Not applicable (provides a framework within which identified uncertainties may be summarised)
Describing uncertainties	Verbal/narrative description.
Assessing the magnitude of individual uncertainties	In most cases this is not shown explicitly in the uncertainty table, but considered by the assessor when judging the impact of each uncertainty on the assessment output.
Assessing the combined impact of multiple uncertainties on the assessment output, taking account of dependencies	Combinations of plus and minus symbols on a defined (preferably quantitative) scale. Alternatively, ranges could be expressed numerically, without the use of symbols.
Assessing the contribution of individual uncertainties to overall uncertainty	The relative contribution of individual uncertainties can be assessed by comparing their evaluations in the uncertainty table.

Melamine example

Members of the Working Group used a modified form of uncertainty table to assess uncertainties affecting three parameters in the example assessment of melamine, based on the context described in section B.2. The group evaluated the individual and combined impacts of these parameters on the uncertainty of the following question: does the possible worst case exposure of high-consuming European children to melamine from consumption of chocolate containing contaminated Chinese milk powder exceed the relevant health-based guidance value, and if so by how much?

The group evaluated the uncertainties on a scale that was previously used in an opinion on BPA (EFSA, 2015). This scale uses plus and minus symbols with quantitative definitions in terms of how much lower or higher a real value might plausibly be compared to its estimate, as shown in Figure B.5.1. Note that the size of the intervals can be adjusted for different assessments, depending on the scale of uncertainties that are present (Edler et al. 2013).

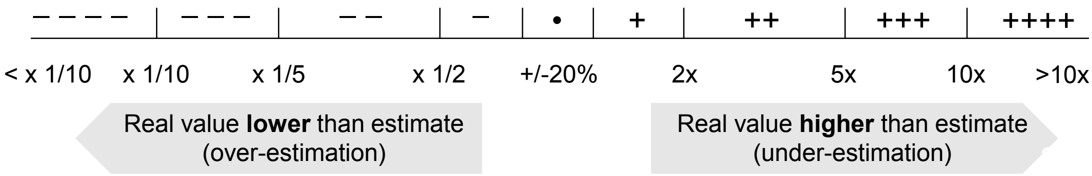


Figure B.5.1: Scale used for assessing uncertainty in example evaluation (Table B.5.1).

The group members were asked to assess the uncertainty of each individual parameter, and also to assess the combined impact of all three parameters on the uncertainty of the assessment output (ratio of exposure to TDI). The evaluation was conducted in two rounds, with the results from the first round being collated on-screen and discussed before the second round. This allowed assessors to adjust their evaluations in the light of the discussion, if they wished. The results of the second round are shown in Table B.5.1. The third column in Table B.5.1 shows the range of evaluations given by the assessors for the extent to which the real value of each individual parameter could be lower than its estimate, while the fourth column shows the range of evaluations for how much the real value of the assessment output (ratio of exposure to TDI) could exceed its estimate based on the uncertainty of that parameter alone. In the bottom row, the fourth column shows the range of evaluations for how much the real value of the assessment output (ratio of exposure to TDI) could exceed its estimate based on the uncertainty of all three parameters considered together. Various methods could be considered for aggregating the judgements of the individual experts. In this example, the overall range spans the set of ranges provided by the individual assessors, and thus expresses the range of values that were considered plausible by one or more of the assessors.

One assessor was unable to quantify the uncertainty of the TDI in either direction, and one was able to quantify the upwards uncertainty but not the downwards uncertainty. These assessments are shown in the table B.5.1 as NQ (not quantified). The results affected by this show first the range including all assessors, and then the range excluding the 'NQ' assessments.

Table B.5.1. Example of uncertainty table for the melamine case study.

Parameter	Value in EFSA (2008) assessment	Range for uncertainty of individual parameters	Range for uncertainty of assessment output
TDI	0.5 mg/kg bw/day	NQ/NQ or ---/++	NQ/NQ or --/+++
Highest concentration of melamine in milk powder	2563 mg/kg	---/+	---/+
Highest consumption of Chinese chocolate by children	0.044 kg	---/++	---/++
Assessment output: ratio of the calculated exposure to the TDI	269%		----/NQ or ----/++

NQ = not quantified. See Figure B.5.1 for definition of scale for plus and minus symbols. See text for further explanation.

The overall range for the output of the assessment (bottom right corner of Table B.5.1) can be converted to numeric form, using the scale in Figure B.5.1 (note this conversion uses the full width of each interval on the scale and may overstate the assessors' actual uncertainty). One expert considered that it was not possible to quantify how much higher the real ratio of exposure to TDI could be compared to the EFSA (2008) estimate of 269%, because they were not able to quantify how different the appropriate TDI could be than that used by EFSA (2008) based on the information available in the EFSA statement. The range of uncertainty for the remaining experts was from more than 10 x below the estimated ratio to 5x above it, i.e. the real worst case exposure for EU children eating contaminated chocolate could be below 30% of the TDI at the lower bound (or even 0 if there was no contamination), and about 13x the TDI at the upper bound (rounding to avoid over-precision).

In this example, the approach was modified to be feasible within the time reserved for it (1-2 hours). This illustrates how it can be adapted for situations when time is short. If more time were available, it would be good practice to document briefly (in the table or in accompanying text) the uncertainties that were considered for each parameter and the reasoning for the evaluation of their impact. If a parameter was affected by several different uncertainties, it might be useful to evaluate them separately and show them in separate rows of the table. In addition, it might be desirable for the assessors to discuss the reasons for differences between their individual ranges, and if appropriate seek a consensus on a joint range (which might be narrower than the range enveloping the individual judgements).

One assessor preferred to express their judgement of the uncertainty for each parameter as a quantitative range and then derive a range for the overall uncertainty by calculation: a form of interval analysis (see section B.7). Interval analysis can also be applied when using the +/- scale, by converting the scores to numeric form for calculation, as was done by EFSA (2015, page 107) when combining evaluations of uncertainty for different sources of internal BPA exposure. These examples suggest that a tabular format similar to uncertainty tables could be used to facilitate and document judgements on ranges for interval analysis.

Strengths

1. The uncertainty table makes transparent many subjective judgements that are unavoidably present in risk assessment, thus improving the quality of group discussion and the reliability of the resulting estimates, and making the judgements open to challenge by others.
2. Concise and structured summary of uncertainties facilitates evaluation of their combined impact by the assessor, even though not based on theory.
3. The approach can be applied to any area of scientific assessment.
4. The approach can be applied to all types of uncertainty, including ambiguity and qualitative issues such as study quality. Anything that the assessor identifies as a factor or consideration that might alter their answer to the assessment question can be entered in the table.
5. The approach facilitates the identification of unquantifiable uncertainties, which can be recorded in the table (a question mark or NQ for not quantifiable in the right hand column).
6. The tabular format is highly flexible. It can be expanded when useful to document the evaluation more fully, or abbreviated when time is short.
7. Using a quantitative scale reduces the ambiguity of purely score-based or narrative approaches. The symbols for the overall assessment can be converted into an approximate, quantitative uncertainty interval for use in interval analysis and to facilitate interpretation by risk managers.
8. The overall assessment helps to inform decision-making, specifically whether the combined effect of uncertainties is clearly too small to change the decision, or whether

more refined risk or uncertainty assessment is needed. But it may also suggest a false precision.

9. The main contributors to overall uncertainty are identified in a structured way, enabling their prioritisation for more quantitative assessment when required (e.g. sensitivity analysis or probabilistic modelling).
10. Tabular format provides a concise summary of the evidence and reasoning behind the assessment of overall uncertainty, increasing transparency for the reader when compared to scoring systems and narrative discussion of uncertainties.

Weaknesses and possible solutions to them

1. For some people, the approach seems not to be immediately intuitive. Therefore, training should be provided.
2. Some users find it difficult to assess the magnitude of uncertainties. EFSA is developing e-training in making probability judgements, which may help with this. Where assessors consider an uncertainty to be unquantifiable, this can be documented in the table.
3. People are bad at making judgements about how uncertainties combine. For this reason, it is better for users to assess plausible intervals for the individual uncertainties and derive their impacts on the assessment output by interval analysis (section B.7).
4. The scales used to define the + and - symbols can be prone to misunderstanding. Therefore they should be designed and communicated carefully. An alternative is for the assessors This is also beneficial when assessors are able to judge the uncertainty more finely than provided for in the scale.
5. Transparency will be impaired if insufficient information is given about the reasoning for the judgements in the table, or if readers cannot easily locate supporting information provided outside the table. This can be addressed by providing more information within the table, if necessary by adding extra columns, and by including cross-references in the table to additional detail in accompanying text and ensuring that this is clearly signposted.
6. The approach relies on expert judgement, which is subject to various psychological biases (see Section 9.2.1.3). Techniques from formal expert elicitation methodology can be used to improve the robustness of the judgements that are made; optionally, fully formal expert elicitation can be used to evaluate the overall uncertainty and/or the contribution of the most important individual uncertainties (see sections B.8 and B.9).

Assessment against evaluation criteria

This method is assessed against the evaluation criteria in Table B.5.2.

Conclusions

1. This method is applicable to all types of uncertainty affecting quantitative questions or estimates, in all areas of scientific assessment. It is flexible and can be adapted to fit within the time available, including emergency situations.
2. The method is a framework for documenting expert judgements and making them transparent. It is generally used for informal expert judgements, but formal techniques (see section B.9) could be incorporated where appropriate, e.g. when the uncertainties considered are critical to decision-making.
3. The method uses expert judgement to combine multiple uncertainties. The results of this will be less reliable than calculation, it would be better to use uncertainty tables as a

technique for facilitating and documenting expert judgement of quantitative ranges for combination by interval analysis. However, uncertainty tables using +/- symbols are a useful option for two important purposes: the need for an initial screening of uncertainties to decide which to quantify individually, and the need for a method to assess uncertainties that are not quantified individually in the overall characterisation of uncertainty (see chapter 10 of main document).

References

ECHA, 2008. Guidance on information requirements and chemical safety assessment. Chapter R.19: Uncertainty analysis. European Chemicals Agency (ECHA).

Edler L, Hart A, Greaves P, Carthew P, Coulet M, Boobis A, Williams GM, Smith B. 2013. Selection of appropriate tumour data sets for Benchmark Dose Modelling (BMD) and derivation of a Margin of Exposure (MoE) for substances that are genotoxic and carcinogenic: Considerations of biological relevance of tumour type, data quality and uncertainty assessment. Food Chem. Toxicol., http://dx.doi.org/10.1016/j.fct.2013.10.030

EFSA 2006. Guidance of the Scientific Committee on a request from EFSA related to Uncertainties in Dietary Exposure Assessment. The EFSA Journal, 438, 1-54.

EFSA 2007. Opinion of the Scientific Panel on Plant protection products and their Residues on a request from the Commission on acute dietary intake assessment of pesticide residues in fruit and vegetables. The EFSA Journal, 538, 1-88.

EFSA 2008. Scientific Opinion of the Panel on Plant protection products and their Residues (PPR) on the Science behind the Guidance Document on Risk Assessment for birds and mammals. The EFSA Journal (2008) 734: 1-181.

EFSA 2012. Guidance on the use of probabilistic methodology for modelling dietary exposure to pesticide residues. The EFSA Journal, 10: 1-95.

EFSA 2013a. Draft Scientific Opinion on the risks to public health related to the presence of bisphenol A (BPA) in foodstuffs – Part: exposure assessment. Published for public consultation.

EFSA 2013b. Scientific Opinion on the risk of *Phyllosticta citricarpa* (*Guignardia citricarpa*) for the EU territory with identification and evaluation of risk reduction options. EFSA Journal 2014; 12(2):3557.

EFSA 2014. Scientific Opinion on the safety assessment of carvone, considering all sources of exposure. EFSA Journal 2014; 12(7):3806.

EFSA 2015. Scientific Opinion on the risks to public health related to the presence of bisphenol A (BPA) in foodstuffs. Part I - Exposure. EFSA Journal 2015;13(1):3978.

Hart A, Gosling JP, Boobis A, Coggon D, Craig P, Jones D. 2010. Development of a framework for evaluation and expression of uncertainties in hazard and risk assessment. Research Report to Food Standards Agency, Project No. T01056.

SCENIHR Committee, 2012. Memorandum on the use of the scientific literature for human health risk assessment purposes – weighing of evidence and expression of uncertainty. DG SANCO, European Commission, Brussels.

Table B.5.2: Assessment of Uncertainty tables for quantitative questions (when applied well) against evaluation criteria.

Criteria	Evidence of current acceptance	Expertise needed to conduct	Time needed	Theoretical basis	Degree/ extent of subjectivity	Method of propagation	Treatment of uncertainty & variability	Meaning of output	Transparency and reproducibility	Ease of understanding for non-specialist
Stronger character-istics	International guidelines or standard scientific method	No specialist knowledge required	Hours	Well established, coherent basis for all aspects	Judgement used only to choose method of analysis	Calculation based on appropriate theory	Different types of uncert. & variab. quantified separately	Range and probability of alternative outcomes	All aspects of process and reasoning fully documented	All aspects fully understandable
	EU level guidelines or widespread in practice	Can be used with guidelines or literature	Days	Most but not all aspects supported by theory	Combination of data and expert judgment	Formal expert judgment	Uncert. & var. quantified separately	Range and relative possibility of outcomes	Most aspects of process and reasoning well documented	Outputs and most of process understandable
	National guidelines, or well established in practice or literature	Training course needed	Weeks	Some aspects supported by theory	Expert judgment on defined quantitative scales	Informal expert judgment	Uncert. & var. distinguished qualitatively	Range of outcomes but no weighting	Process well documented but limited explanation of reasoning	Outputs and principles of process understandable
	Some publications and/or regulatory practice	Substantial expertise or experience needed	A few months	Limited theoretical basis	Expert judgment on defined ordinal scales	Calculation or matrices without theoretical basis		Quantitative measure of degree of uncertainty	Limited explanation of process and/or basis for conclusions	Outputs understandable but not process
Weaker character-istics	Newly developed	Professional statistician needed	Many months	Pragmatic approach without theoretical basis	Verbal description, no defined scale	No propagation	No distinction between uncert. & var.	Ordinal scale or narrative description for degree of uncertainty	No explanation of process or basis for conclusions	Process and outputs only understandable for specialists

B.6 Uncertainty tables for categorical questions

Purpose, origin and principal features

The purpose of this method is to provide a structured approach for addressing uncertainty in weight of evidence assessment of categorical questions and expressing the uncertainty of the conclusion. Weight of evidence as an overall process will be considered in more detail in a separate mandate[7]. This section focusses specifically on the treatment of uncertainty for weight of evidence questions.

The method described here was developed by Hart et al. (2010), who noted that uncertainty tables of the type described by EFSA (2006) address uncertainty in quantitative estimates (e.g. exposure, reference dose) and are not well suited to addressing uncertainty in categorical questions. Categorical questions concern choices between two or more categories and are often addressed by a weight of evidence approach. Examples of such questions in chemical risk assessment include hazard identification (does chemical X have the capability to cause effect Y?), mode of action (through which mode of action does chemical X cause effect Y?), human relevance (is effect Y of chemical X in animals relevant to humans?) and adversity (if effect Y occurred in humans would it be adverse?). Examples in other areas of EFSA's work might include equivalence of GM traits and their non-GM counterparts, whether an animal pathogen will infect humans, etc.

The principal features of this method are the use of a tabular approach to summarise weight of evidence assessment, and the expression of conclusions in terms of their likelihood or probability rather than, or in addition to, the more common approach of using narrative phrases. The tabular approach provides a structured framework, which is intended to help the assessor develop the assessment and improve its transparency. The expression of conclusions as probabilities is intended to avoid the ambiguity of narrative forms, and also opens up the possibility of using probability theory to help form overall conclusions when an assessment comprises a series of linked categorical and/or quantitative questions.

The main steps of the approach can be summarised as follows:

1. Define clearly the question(s) to be answered.
2. Identify and describe relevant lines of evidence (LoE).
3. Organise the LoE into a logical sequence to address the question of interest.
4. Identify their strengths, weaknesses & uncertainties.
5. Evaluate the weight of each LoE and its contribution to answering the question.
6. Take account of any prior knowledge about the question.
7. Make an overall judgement about the balance of evidence, guarding against cognitive biases associated with expert judgement, and use formal elicitation methods if appropriate.
8. Express the conclusion as a probability or range of probabilities, if possible, and explain the reasoning that led to it.

Applicability in areas relevant for EFSA

The approach is, in principle, applicable to any two-category question in any area of EFSA's work. It would be possible to adapt it for questions with multiple categories (e.g. choices

[7] "Guidance on the use of the Weight of Evidence Approach in Scientific Assessments", EFSA-Q-2015-00007

between 3 or more modes of action), although this would be more complex. It provides a more structured approach to weight of evidence than the traditional approach of a reasoned argument in narrative text, and a less ambiguous way of expressing the conclusion. However, it is intended to complement those approaches rather than completely replace them, because it will always be desirable to accompany the tabular summary of the assessment with a detailed narrative description of the evidence and reasoning, and it may aid communication to accompany numerical likelihoods with narrative statements of the conclusion.

The approach has so far been used in only a few assessments. The original research report contains a simplified example of hazard identification for caffeine (Hart et al, 2010). Edler et al. (2014) provide step-by-step instructions for applying the method to assess the likelihood that chemicals are genotoxic carcinogens, and detailed case studies for Sudan 1 and PhIP. It was used for hazard identification in the EFSA (2015) Opinion on bisphenol A (BPA), assessing the likelihood that BPA has the capability to cause specific types of effects in animals based on evidence from a wide variety of studies. In the same Opinion, likelihood was also used to express judgements about the relevance to humans of effects seen animals and whether, if they occurred in humans, they would be adverse. Evidence for the judgements about relevance and adversity were discussed in the text of the opinion, rather than by tabulated lines of evidence.

Potential contribution to the main steps of uncertainty analysis

Steps in uncertainty analysis	Potential contribution of this approach
Identifying uncertainties	Structured approach promotes identification of uncertainties affecting individual lines of evidence and overall conclusion.
Describing uncertainties	Concise narrative description of each line of evidence including strengths, weaknesses and uncertainties.
Assessing the magnitude of individual uncertainties	Strengths, weaknesses and uncertainties of individual lines of evidence are assessed by expert judgement.
Expression of the combined impact of multiple uncertainties on the assessment output, taking account of dependencies	The combined impact of all the lines of evidence and their uncertainties is assessed by expert judgement and expressed as a probability or range of probabilities for a positive conclusion.
Assessing the contribution of individual uncertainties to overall uncertainty	The relative importance of uncertainties affecting individual lines of evidence can be assessed by considering the weaknesses identified in the table. The ordinal scale for influence indicates what each line of evidence contributes to the balance of likelihood (uncertainty) for the conclusion.

Melamine example

The EFSA (2008) Statement states that 'the primary target organ for melamine toxicity is the kidney'. Here, the use of uncertainty tables for categorical questions is illustrated by applying the approach to summarise the evidence that melamine causes kidney effects. Although the evidence in this case is rather one-sided, it serves to illustrate the principles of the approach.

The first step is to specify in precise terms the question to be considered. In this case the question was defined as follows: does melamine have the capability to cause adverse effects on kidney in humans?

The assessment was carried out by 3 toxicologists in the Working Group. First, they were asked to identify the main lines of evidence for assessing the potential for melamine to cause kidney effects, which were available at the time of the EFSA (2008) statement. Four lines of evidence were identified, as listed and briefly described in Table B.6.1. The assessors were then asked to consider the influence of each line of evidence on their judgement about the answer to the question, and to express this using a scale of arrow symbols which are defined in Table B.6.2. Upward arrows indicate an upward influence on the likelihood that melamine causes kidney effects, and the number of arrows indicates the strength of the influence. Next, the assessors were asked to make a judgement about the overall likelihood that melamine causes kidney effects, considering all lines of evidence together. They were asked to express this likelihood using another scale, defined in Table B.6.3. The assessors made their judgements for both influence and likelihood individually. The judgements were then collected and displayed on screen for discussion, and the assessors were given the opportunity to adjust their judgements if they wished. Table B.6.1 shows the range of judgements between assessors. In this case there was little variation between assessors in their assessment of influence, and all three gave the same overall conclusion: that it is very likely (probability 90-100%) that melamine has the potential to cause adverse effects kidney in humans.

Due to the limited time that was set for developing this example, Table B.6.1 provides only very limited explanation for the judgements made in assessing individual lines of evidence and the overall conclusion. More explanation should be provided in a real assessment, including an indication of the relevance and reliability of each line of evidence, and the reasoning for the overall conclusion. This may be done either within the table (adding extra content and/or columns, e.g. Annex C of EFSA, 2015), or in accompanying text. However, more abbreviated formats may sometimes be justified (e.g. in emergency situations).

The procedure adopted for making judgements in this example may be regarded as semi-formal, in that a structured approach was used in which experts considered their judgements individually and then reviewed them after group discussion. Ideally, it would be preferable to use a fully formal expert elicitation procedure (see section B.9), especially for weight of evidence questions that have a large impact on the assessment outcome.

Table B.6.1. Assessment of evidence and uncertainty for the question: does melamine have the capability to cause adverse effects on kidney in humans?

Lines of evidence	Influence on conclusion
Line of Evidence 1 – animal studies Same effect on more than one species	↑↑↑
Line of Evidence 2 – information on effects in humans Severe health effect in humans but unspecified in the EFSA statement	↑/↑↑
Line of Evidence 3 – information on mode of action Information on crystal formation in kidneys. Effect not dependent on metabolism indicating similar effects are likely in different species.	↑/↑↑
Line of Evidence 4 – Evidence of adverse effects in companion animals Kidney toxicity in cats with crystal formation resulting from melamine adulterated pet food.	↑/↑↑
CONCLUSION (by semi-formal expert judgement, see text) Based on the consistency from the different lines of evidence.	Very likely (90-100% probability)

See Table B.6.2 for key to symbols and Table B.6.3 for likelihood scale. Pairs of symbols separated by a slash (↑ / ↑↑) represent variation of judgements between assessors.

Table B.6.2. Key to scale of symbols used to express the influence of lines of evidence on the answer to the question in Table B.6.1.

Symbol	Influence on likelihood of positive answer to question
↑↑↑	strong upward influence on likelihood
↑↑	intermediate upward influence on likelihood
↑	minor upward influence on likelihood
●	no influence on likelihood
↓	minor downward influence on likelihood
↓↓	intermediate downward influence on likelihood
↓↓↓	strong downward influence on likelihood
?	unable to evaluate influence on likelihood

Table B.6.3. Scale used for expressing the likelihood of a positive answer to the question addressed in Table B.6.1, After Mastrandrea et al. (2010).

Term	Likelihood of outcome
Virtually certain	99-100% probability
Very likely	90-100% probability
Likely	66-100% probability
As likely as not	33-66% probability
Unlikely	0-33% probability
Very unlikely	0-10% probability
Exceptionally unlikely	0-1% probability

Strengths

1. Promotes a structured approach to weighing multiple lines of evidence and taking account of their uncertainties, which should help assessors in making their judgements and potentially lead to better conclusions.
2. Expressing the (uncertainty of the) conclusion in terms of likelihood or probability avoids the ambiguity of narrative conclusions, though care is needed to avoid suggesting false precision.
3. Compatible with formal approaches to eliciting expert judgements on the probability of the conclusion.
4. The judgements involved can be made by formal EKE, which would ideally be preferable. When judgements are made less formally, the process can still be designed to encourage assessors to guard against common cognitive biases.
5. Tabular structure is intended to make the evidence and reasoning more accessible, understandable and transparent for scientific peers, risk managers and stakeholders.

Weaknesses and possible approaches to address them

1. Tabular structure can become cumbersome if there are many lines of evidence and/or extensive detail is included. This can be addressed by careful management of the quantity, organisation (e.g. grouping similar studies) and format of table content, and by providing necessary additional detail in accompanying text.
2. For some types of question, probabilities may be misinterpreted as frequencies or risks (e.g. probability of chemical X having a carcinogenic mode of action may be misinterpreted as the probability of an individual getting cancer). This should be avoided by good communication practice.

3. Some assessors may be unwilling to give numerical probabilities. Can be addressed by using a scale of likelihood terms (e.g. EFSA, 2014), preferably with quantitative definitions.

Assessment against evaluation criteria

This method is assessed against the criteria in Table B.6.4.

Conclusions

1. This approach is potentially applicable to any type of binary question in all areas of EFSA's work, and to all types of uncertainty affecting those questions.
2. The approach is new and would benefit from further case studies to evaluate its usefulness and identify improvements.

References

Edler L, Hart A, Greaves P, Carthew P, Coulet M, Boobis A, Williams GM, Smith B. 2014. Selection of appropriate tumour data sets for Benchmark Dose Modelling (BMD) and derivation of a Margin of Exposure (MoE) for substances that are genotoxic and carcinogenic: Considerations of biological relevance of tumour type, data quality and uncertainty assessment. Food Chem. Toxicol., 70: 264–289. http://dx.doi.org/10.1016/j.fct.2013.10.030

EFSA 2006. Guidance of the Scientific Committee on a request from EFSA related to Uncertainties in Dietary Exposure Assessment. The EFSA Journal, 438, 1-54.

EFSA, 2008. Statement of EFSA on risks for public health due to the presences of melamine in infant milk and other milk products in China. EFSA Journal (2008) 807, 1-10.

EFSA, 2015. Scientific Opinion on the risks to public health related to the presence of bisphenol A (BPA) in foodstuffs. Part II - Toxicological assessment and risk characterisation. EFSA Journal 2015;13(1):3978.

Hart A, Gosling JP, Boobis A, Coggon D, Craig P, Jones D. 2010. Development of a framework for evaluation and expression of uncertainties in hazard and risk assessment. Research Report to Food Standards Agency, Project No. T01056.

Mastrandrea, MD, CB Field, TF Stocker, O Edenhofer, KL Ebi, DJ Frame, H Held, E Kriegler, KJ Mach, PR Matschoss, G-K Plattner, GW Yohe, and FW Zwiers, 2010: Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties. Intergovernmental Panel on Climate Change (IPCC). Available at http://www.ipcc.ch.

Table B.6.4. Assessment of Uncertainty tables for categorical questions (when applied well) against evaluation criteria.

Criteria	Evidence of current acceptance	Expertise needed to conduct	Time needed	Theoretical basis	Degree/ extent of subjectivity	Method of propagation	Treatment of uncertainty & variability	Meaning of output	Transparency and reproducibility	Ease of understanding for non-specialist
Stronger character-istics	International guidelines or standard scientific method	No specialist knowledge required	Hours	Well established, coherent basis for all aspects	Judgement used only to choose method of analysis	Calculation based on appropriate theory	Different types of uncert. & var. quantified separately	Range and probability of alternative outcomes	All aspects of process and reasoning fully documented	All aspects fully understandable
	EU level guidelines or widespread in practice	Can be used with guidelines or literature	Days	Most but not all aspects supported by theory	Combination of data and expert judgment	Formal expert judgment	Uncert. & var. quantified separately	Range and relative possibility of outcomes	Most aspects of process and reasoning well documented	Outputs and most of process understandable
	National guidelines, or well established in practice or literature	Training course needed	Weeks	Some aspects supported by theory	Expert judgment on defined quantitative scales	Informal expert judgment	Uncert. & var. distinguished qualitatively	Range of outcomes but no weighting	Process well documented but limited explanation of reasoning	Outputs and principles of process understandable
	Some publications and/or regulatory practice	Substantial expertise or experience needed	A few months	Limited theoretical basis	Expert judgment on defined ordinal scales	Calculation or matrices without theoretical basis		Quantitative measure of degree of uncertainty	Limited explanation of process and/or basis for conclusions	Outputs understandable but not process
Weaker character-istics	Newly developed	Professional statistician needed	Many months	Pragmatic approach without theoretical basis	Verbal description, no defined scale	No propagation	No distinction between uncert. & var.	Ordinal scale or narrative description for degree of uncertainty	No explanation of process or basis for conclusions	Process and outputs only understandable for specialists

B.7 Interval analysis

Origin, purpose and principal features

Interval analysis is a method to obtain a range of values for the output of a calculation based on specified ranges for the inputs to a calculation. If each input ranges expresses uncertainty about the corresponding input value, the output range is an expression of uncertainty about the output.

Interval analysis (also "interval arithmetic, "interval mathematics", "interval computation") was developed by mathematicians since the early 50s (Dwyer, 1951, as one of the first authors) to propagate errors or account for parameter variability. Modern interval analysis was introduced by Ramon E. Moore in 1966. Ferson & Ginzburg, 1996 proposed interval analysis for the propagation of ignorance (epistemic uncertainty) in conjunction with probabilistic evaluation of variability. The interval method is also discussed in the WHO-harmonisation document, 2008, along the concept of Ferson (1996).

Interval analysis is characterized by the application of upper and lower bounds to each parameter, instead of using a fixed mean or worst-case parameter (e.g. instead of the fixed value 1.8 for mean body height of Northern males one can use the interval 1.6 to 2.0 to account for the variability in the population). To yield a lower bound of an estimate all parameter bounds are combined in the model that result in the lowest estimate possible. To yield the upper bound of an estimate analogously the parameter bounds are combined that yield the highest estimate possible. The interval between the lower and the upper bound estimate is then considered to characterize the uncertainty and variability around the estimate.

For uncertainty assessment, where the range for each input covers all values considered possible, the range for the output then also covers all possible values. If it is desired to specify an input range covering a subset of possible values and accompanied by a probability, the method of probability bounds analysis (section B.13) is more likely to be useful.

Applicability in areas relevant for EFSA

Within EFSA the method is often used for the treatment of left-censored data (e.g. in the exposure analysis for chemical risk assessment, EFSA, 2010). If samples are included in a statistical analysis that have concentrations below the limit of detection (LOD), a lower bound estimate can be constructed by assuming that all sample concentrations <LOD are 0, and a higher bound by assuming that all sample concentrations are equal to the LOD. The true value will lie in between those values (e.g. EFSA, 2015).

Potential contribution to the main steps of uncertainty analysis

Steps in uncertainty analysis	Potential contribution of this approach
Identifying uncertainties	Not applicable.
Describing uncertainties	Not applicable.
Assessing the magnitude of individual uncertainties	Yes, the uncertainty is expressed for each individual uncertainty as a lower and as an upper bound.
Assessing the combined impact of multiple uncertainties on the assessment output, taking account of dependencies	Yes, range of output values, taking into account the range of all input parameters at the same time and making no assumptions about dependencies
Assessing the contribution of individual uncertainties to overall uncertainty	Not applicable.

Melamine example

As described in more detail in Annex C, exposure e is calculated according to

$$e = \frac{c \times w \times q}{bw}$$

where

c: concentration of melamine in adulterated milk powder (mg/kg)

w: weight fraction of milk powder in chocolate

q: consumption of chocolate in a day (kg/day)

bw: bodyweight of consumer (kg)

The variables q and bw are both expected to be positively correlated with the age of the child and as a result to be correlated with each other. As a simple example of an approach to address dependencies in an interval analysis, the method was applied to two subpopulations of children that might be expected to have higher exposure: children aged 1 and children aged 6. These groups two were selected for illustration because of the low body-weight of the younger group and a judgement that the older age group might consume as much as older children but have lower body-weight. A full assessment would in principle apply the method separately to each age from 1 to 10.

For the concentration c, the highest observed level in the data used in the melamine statement was 2563 mg/kg. This value however will not be the highest of the whole ensemble of possible values, because only a subsample has been analysed and not all samples in the ensemble. Knowing that melamine is used to mimic the N-content of milk that should be contained in the samples, but is not, it can be assumed that the higher bound for the melamine content is the amount needed to mimic 100% milk that should be contained in the sample. Multiplying the ratio between the N-content of milk protein and melamine (0.13/0.67=0.22) and the protein content in dry milk (3.4 g protein in cow milk/130 g dry matter=26 g/kg) the maximal content of melamine in dry milk yields a higher bound of 6100 mg/kg melamine in adulterated milk powder. The lower bound for melamine will be 0 mg/kg, because it is not naturally occurring, but the result of adulteration.

For the weight fraction of milk powder in milk chocolate w, the legally-required minimum of 0.14 is chosen as the lower bound, and the highest value found in an internet search (0.28) as the higher bound.

For q no data were available for high chocolate consumption. The assessors made informal judgements of 50 g and 300 g, for a 1 year old and a 10 year old child, respectively. In a real situation, expert knowledge elicitation (section B.8 and B.9) would be used to obtain these numbers.

For the lower and higher bound for bodyweight (bw) in both age groups, the assessors used low and high percentiles from WHO growth charts as a starting point for choosing more the more extreme values in the tables below to be absolute lower and upper bounds. Again, in a real situation, expert knowledge elicitation would be used to obtain these numbers.

Child 1 year old

Parameter/Estimate	Value	Lower bound	Higher bound
c (mg/kg)	29	0	5289 (highest observed level: 2563)
w (-)	0.25	0.14	0.28
q (kg/d)	0.042	0	0.05
bw (kg)	20	6	13
e (mg/d kg-bw)	**0.015225**	**0**	**14.2**

Child 6 years

Parameter/Estimate	Value	Lower bound	Higher bound
c (mg/kg)	29	0	6100 (highest observed level: 2563)
w (-)	0.25	0.14	0.28
q (kg/d)	0.042	0	0.3
bw (kg)	20	12	34
e (mg/d kg-bw)	**0.015225**	**0**	**42.7**

In the tables above the intervals cover both uncertainty and variability in the parameters. Below we aim to demonstrate how also within the interval method uncertainty and variability might be treated separately (example for the 1 year old child).

Child 1 year old, mainly variability

Parameter/Estimate	Value*	Lower bound	Higher bound
c (mg/kg)	29	0	**2563**
w (-)	0.25	0.14	**0.28**
q (kg/d)	0.042	0	**0.05**
bw (kg)	20	**6**	13
e (mg/d kg-bw)	**0.015**	0	**6.0**

* These values are not part of the interval analysis, only demonstrate the values around which the variability/uncertainty assessment is constructed
**the higher bound exposure is calculated by using the higher bound for the first three parameters and the lower bound for the bodyweight, denoted in bold

Child 1 year old, uncertainty about the worst case (wc) values for parameters

Parameter/Estimate	Favored value* for wc	Lower bound for wc value	Higher bound for wc value
c (mg/kg)	2563	2563	6100
w (-)	0.28	0.28	0.30
q (kg/d)	0.05	0.05	0.1
bw (kg)	6	5.5	6.5
e (mg/d kg-bw)	**6.0**	**5.5**	**33.3**

* These values are not part of the interval analysis, only demonstrate the values around which the variability/uncertainty assessment is constructed

Strengths

1. The method is relatively easy to perform and straightforward. It is particularly useful as a screening method to quickly assess whether more sophisticated quantitative uncertainty assessments are needed or whether, even for an upper bound, for example of an exposure, no concern exists. Ferson & Ginzburg, 1996 recommend it as an alternative method to probabilistic uncertainty assessments when the shape of the distribution is not known (e.g. for assessing uncertainty due to ignorance, see above).

2. When used with real upper and lower limits the method covers all possible scenarios.

Weaknesses and possible approaches to reduce them

1. Only quantifies range not probabilities within range. Therefore useful as initial screen to determine whether probabilistic assessment is needed.
2. Most of the time it is not made clear what the ranges really are meant to represent (minimum/maximum, certain percentiles, …). This can be cured by transparent communication in the text and by attempting to be as consistent as possible.
3. The method does not incorporate dependencies between variables, so that the interval of the final estimate will be larger than the range of the true variability and uncertainty, if dependencies between variables occur. This limitation can be partly addressed by using scenarios representing different combinations of input variables to explore the potential impact of dependencies, as illustrated in the example above.
4. The more parameters are involved the larger will become the uncertainty range, and the more likely it is that a probabilistic assessment taking account of dependencies will be required for decision-making. Nevertheless, since interval analysis is much simpler to perform, it is still useful as a screening method to determine whether more sophisticated analysis is needed. .
5. Variability and uncertainty are not separated by the concept behind this method and it is easy to forget that both uncertainty and variability are included in the range when it is applied to uncertain variability. However, because the interval method is a special case of probability bounds analysis, the method described in section B.13 for addressing problems with uncertain variability could be used in conjunction with interval analysis.

Assessment against evaluation criteria

This method is assessed against the criteria in Table B.7.1.

Conclusions

1. Interval analysis provides a simple and rigorous calculation of bounds for the output. However, it provides only extreme upper and lower values for the output resulting from combinations of inputs and gives no information on relative likelihood of values within the output range.
2. It has the potential to be very useful because it can be used to check quickly whether the output range includes both acceptable and unacceptable outcomes. If it does, a more sophisticated analysis of uncertainty is needed.

References

EFSA, 2015. Scientific Opinion on the risks to public health related to the presence of bisphenol A (BPA) in foodstuffs: Part I – Exposure assessment. EFSA Journal 13(1):3978.

EFSA, 2010, Scientific Report on Management of left-censored data in dietary exposure assessment of chemical substances., EFSA Journal 2010; 8(3):1557.

P S Dwyer, 1951, "Linear computations", John Wiley, New York.

S Ferson & L R Ginzburg "Different methods are needed to propagate ignorance and variability, Reliability Engineering and system safety 54 (1996) 133-144.

Ferson S (1996) What Monte Carlo methods cannot do. Human and Ecological Risk.

Assessment, 2(4): 990–1007.

R Moore, 1966, Interval analysis, Prentice-Hall.

Table B.7.1: Assessment of Interval analysis (when applied well) against evaluation criteria.

Criteria	Evidence of current acceptance	Expertise needed to conduct	Time needed	Theoretical basis	Degree/ extent of subjectivity	Method of propagation	Treatment of uncertainty and variability	Meaning of output	Transparency and reproducibility	Ease of understanding for non-specialist
Stronger character-istics	International guidelines or standard scientific method	No specialist knowledge required	Hours	Well established, coherent basis for all aspects	Judgement used only to choose method of analysis	Calculation based on appropriate theory	Different types of uncert. & var. quantified separately	Range and probability of alternative outcomes	All aspects of process and reasoning fully documented	All aspects fully understandable
	EU level guidelines or widespread in practice	Can be used with guidelines or literature	Days	Most but not all aspects supported by theory	Combination of data and expert judgment	Formal expert judgment	Uncert. & var. quantified separately	Range and relative possibility of outcomes	Most aspects of process and reasoning well documented	Outputs and most of process understandable
	National guidelines, or well established in practice or literature	Training course needed	Weeks	Some aspects supported by theory	Expert judgment on defined quantitative scales	Informal expert judgment	Uncert. & var. distinguished qualitatively	Range of outcomes but no weighting	Process well documented but limited explanation of reasoning	Outputs and principles of process understandable
	Some publications and/or regulatory practice	Substantial expertise or experience needed	A few months	Limited theoretical basis	Expert judgment on defined ordinal scales	Calculation or matrices without theoretical basis		Quantitative measure of degree of uncertainty	Limited explanation of process and/or basis for conclusions	Outputs understandable but not process
Weaker character-istics	Newly developed	Professional statistician needed	Many months	Pragmatic approach without theoretical basis	Verbal description, no defined scale	No propagation	No distinction between uncert. & var.	Ordinal scale or narrative description for degree of uncertainty	No explanation of process or basis for conclusions	Process and outputs only understandable for specialists

B.8 Informal Expert Knowledge Elicitation applied to uncertainty in risk assessments

This section describes the essential elements of an Expert Knowledge Elicitation (EKE) which are necessary in applications judging any uncertainties in risk assessments. The full process, so called formal Expert Knowledge Elicitation, is described in section B.9. Between the informal and formal Expert Knowledge Elicitation is a continuum of alternatives, which could be used to fit the process to the specific needs of the problem, e.g. reframe the problem into the language of practitioners – as described in the formal EKE – but using an existing network of experts – as described in the informal EKE.

Purpose, origin and principal features

Scientific evidence generated from appropriate empirical data or extracted from systematically reviewed literature should be the source of information to use in risk assessments. However, in practice empirical evidence is often limited and main uncertainties may not be quantified in the data analysis or literature. "In such cases it is necessary to turn to expert judgements. Psychological research has shown that unaided expert judgement of the quantities required for risk modelling - and particularly the uncertainty associated with such judgements - is often biased, thus limiting its value." (EFSA, 2014) Therefore EFSA developed Guidance on Expert Knowledge Elicitation which recommends a formal process to elicit expert judgements for use in quantitative risk assessments in the remit of EFSA. The Guidance document focusses on judgements about parameters in quantitative risk models.

Therefore judgements on qualitative aspects in the uncertainty assessment, e.g. the selection of the risk model / assessment method, or the complete identification of inherent sources of uncertainties, are not covered by the Guidance. These qualitative questions often arise at the beginning of a risk assessment when decisions have to be taken on the assessment method, e.g. the interpretation of the mandate, the definition of the scenario, the risk model, the granularity of the risk assessment, or the identification of influencing factors for use in the model. They further appear during the uncertainty assessment when the sources of uncertainties have to be identified. Expert judgement is used to develop a complete set of appropriate, alternative approaches, or a description of possible sources of uncertainties. The result is often a pure list which could be enriched by a ranking and/or judgements on the relevance for answering the mandate.

Another typical judgement is about the unknown existence of specific circumstances, e.g. causal relationships between an agent and a disease. Here the expert elicitation will result in a single subjective probability that the circumstance exist.

There is no sharp difference between qualitative and quantitative questions, as subjective probabilities could be used to express the appropriateness of different alternatives in a quantitative way. In addition what-if scenarios could be used to give quantitative judgements on the influence of factors or sources on the final outcome and express their relevance.

Furthermore the Guidance on Expert Knowledge Elicitation acknowledges that due to restrictions in resources, e.g. time and personnel, it may not be feasible to formally judge on uncertainties of all quantitative parameters in a risk assessment with a full EKE process. Procedures are given to identify most influencing parameters for which a formal elicitation process is recommended. A simplified elicitation process for quantitative parameters is also mentioned in the Guidance. For less influencing parameters qualitative as well as quantitative the expert knowledge elicitation can be done in a minimal assessment, the Informal Expert Knowledge Elicitation.

Table B.8.1: Types Expert Knowledge Elicitations

Method	**Topic to elicit**	
	qualitative, e.g. the selection of a risk model / assessment method, identification of sources of uncertainty	**quantitative**, e.g. parameters in the risk assessment, the resulting risk, and the magnitude of uncertainties
Informal (cp.this section)	Expert elicitation following the minimal requirements (predefined question and expert board, fully documented) resulting in a verbal reasoning, scoring or ranking on a list of identified alternatives, influencing factors or sources.	Expert elicitation following the minimal requirements (predefined question and expert board, fully documented) resulting in a description of uncertainties in form of subjective probabilities, probability bounds, or subjective probability distributions.
Formal (cp. section B.9)	Elicitation following a predefined protocol with essential steps: initiation, pre-elicitation, elicitation and documentation, resulting in a verbal reasoning, scoring or ranking on a list of identified alternatives, influencing factors or sources.	Elicitation following a predefined protocol with essential steps: initiation, pre-elicitation, elicitation and documentation, resulting in a description of uncertainties in form of a subjective probabilities, or subjective probability distributions.

The following section will describe the minimal requirements needed for this informal procedure:

1. Predefined question guaranteeing an unambiguous framing of the problem with regard to the intended expert board.
2. Questions for expert elicitation have *"to be framed in such a manner that the expert is able to think about it. Regional or temporal conditions have to be specified. The wording has to be adapted to the expert's language. The quantity should be asked for in a way that it is in principle observable and, preferably, familiar to the expert. (…) The metrics, scales and units in which the parameter is usually measured have to be defined."* (EFSA 2014)*.*
3. Clearly defined expert board guaranteeing the equal involvement of all experts of the board.
4. The elicitation of the question may need involvement of experts with different expertise profiles. To enable a review on the quality of the elicitation the appropriate constitution and equal involvement of all experts of the board should be documented.
5. Clearly documented elicitation method guaranteeing as much as possible unbiased and balanced elicitation of the expert board including the aggregation of the individual judgements.
6. Expert elicitation methods are developed to ensure an unbiased and balanced elicitation of the expert board. Different types of analysis can be used to aggregate the answers of the experts within the board expressing the individual uncertainty as well as within the board. To enable a review on the quality of the elicitation the elicitation and aggregation method should be documented.
7. Clearly expressed result of the elicitation to the question guaranteeing a description of uncertainties and summarizing the reasoning.
8. Each expert elicitation should result in an explicit statement on the outcome. This includes an expression of the inherent uncertainties, in a quantitative or qualitative way, and a summary of the reasoning. Further conversions of the results should be visible for later review.

Applicability in areas relevant for EFSA

Performing Informal Expert Knowledge Elicitation within an EFSA working group will already result in some short-cuts compared to the formal process.

The working group is already aware about the context and background of the problem. Therefore the question for the elicitation has not to be re-framed in such a manner that the experts are able to think about it. However questions should be asked in way, that avoids ambiguity about the objective, that the answer would be in principle observable / measurable, and that the expert is familiar with metrics and scales of the answer.

The working group is selected in order to answer the EFSA mandate. Therefore a general expertise is available to judge on the risk assessment question. Nevertheless it should be guaranteed that all experts are equally involved in the informal elicitation and all relevant aspects of the mandate are covered by the working group.

Members of the working group are already trained in steering an expert elicitation according to EFSAs Guidance, and are educated in judging uncertainties. Following the elicitation protocols and aggregation methods discussed in the guidance will ensure unbiased and accurate judgements as far as possible. During a regular working group meeting the application of e.g. the Sheffield protocol (EFSA, 2014) could result in a consensual judgement, so called behavioural aggregation method. Nevertheless most EKE processes will gain by the involvement of a specialized facilitator (elicitor for the selected protocol), who is able to moderate between deviating judgements within the working group.

Nevertheless also the Informal Expert Knowledge Elicitation should be completely documented in accordance with the Guidance to allow a review of the method by the corresponding EFSA panel, selected external reviewers or through the public after publication. The internal review of the elicitation via steering and working group will be omitted.

In summary Informal Expert Elicitation has a high applicability in EFSAs risk assessments, especially when empirical evidence is limited or not retrievable due to constraints in time and resources.

Potential contribution to the main steps of uncertainty analysis

Steps in uncertainty analysis	**Potential contribution of this approach**
Identifying uncertainties	Maybe, when discussing the question
Describing uncertainties	Maybe, when discussing the question
Assessing the magnitude of individual uncertainties	Yes
Assessing the combined impact of multiple uncertainties on the assessment output, taking account of dependencies	Yes
Assessing the contribution of individual uncertainties to overall uncertainty	Yes

Melamine example

To answer the question:

“What is the maximum fraction of milk power [dry milk solids in %], which have to be used to produce saleable milk chocolate?”

the working group calculated the sensitivity of this parameter in the risk assessment model. It was concluded that the influence on the uncertainty of the final outcome is minor and does not justify a Formal Expert Knowledge Elicitation. Instead the full working group was discussing the available evidence and performed a Sheffield-type approach. Each member was asked to individually judge on the uncertainty distribution of the parameter using the quartile method (cp. with section B.9). The individual results were reviewed and discussed. Finally the working group agreed on a common uncertainty distribution:

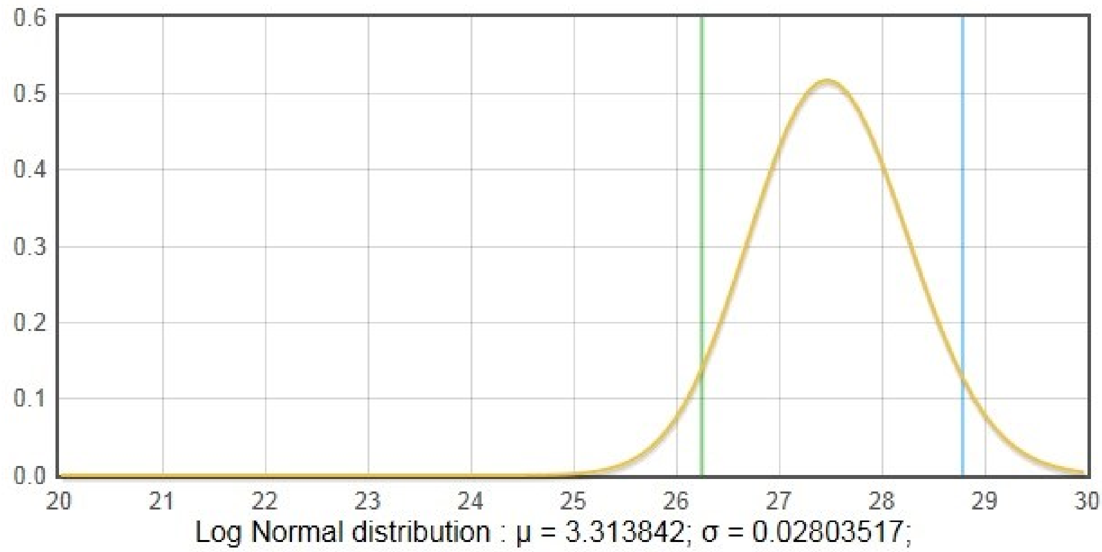
Input judgements:

Lower limit: 20%, upper limit 30%

Median: 27.5%

1st quartile: 27%, 3rd quartile: 28%

Best fitting distribution: Log-normal (μ=3.314, σ=0.02804) with 90% uncertainty bounds (5th and 95th percentile): 26.3–28.8



(Calculated with the MATCH elicitation tool, ref: David E. Morris, Jeremy E. Oakley, John A. Crowe, A web-based tool for eliciting probability distributions from experts, Environmental Modelling & Software, Volume 52, February 2014, Pages 1-4)

Strengths

1. This approach of uncertainty analysis could be used in situations where other methods are not applicable due to restricted empirical data, literature, other evidence, or due to limited resources.
2. The essential elements of the Expert Knowledge Elicitation reduce the impact of known psychological problems in eliciting expert judgements and ensure a transparent documentation and complete reasoning.
3. Using informal Expert Knowledge Elicitation will it be possible to express uncertainties in a quantitative manner, e.g. by probability distributions, In almost all situations.

Weaknesses and possible approaches to reduce them

1. Even when this approach is able to identify and quantify uncertainties, it is not able to increase the evidence from data, e.g. experiments/surveys and literature.
2. EKE is not a substitute for data. Rather, it provides a rigorous and transparent way to express what is known about a parameter from existing evidence, and can provide a good basis for deciding whether to request additional data.
3. In comparison to the Formal Expert Knowledge Elicitation the definition of the question, the selection of the expert board and the performance of the elicitation protocol are restricted to the competencies in the working group.

4. No internal, independent review is foreseen to validate the quality of the elicitation, and finally the result.

Assessment against evaluation criteria

This method is assessed against the criteria in Table B.8.2.

Conclusions

1. The method has a high applicability in working groups and boards of EFSA and should be applied to quantify uncertainties in all situations
    a. where empirical data from experiments / surveys, literature are limited
    b. where the purpose of the risk assessment does not require the performance of a full Formal Expert Knowledge Elicitation
    c. or where restrictions in the resources (e.g. in emergency situations) forces EFSA to apply a simplified procedure.
2. The method is applicable in all steps of the risk assessment, esp. to summarise the overall uncertainty of the outcome. Decisions on the risk assessment methods (e.g. risk models, factors, sources of uncertainties) could be judged qualitatively with quantitative elements (e.g. subjective probabilities on appropriateness, what-if scenarios).
3. The method should not substitute the use of empirical data, experiments, surveys or literature, when these are already available or could be retrieved with corresponding resources.
4. In order to enable a EFSA working group to perform expert elicitations all experts should have basic knowledge in probabilistic judgements and some experts of the working group should be trained in steering expert elicitations according to the EFSA Guidance.

References

EFSA (European Food Safety Authority), 2014. Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment. EFSA Journal 2014;12(6):3734. [219 pp.] doi:10.2903/j.efsa.2014.3734

Morgan MG and Henrion M, 1990. Uncertainty – A Guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge University Press (UK), 1990.

Table B.8.2. Assessment of Informal expert knowledge elicitation (when applied well) against evaluation criteria.

Criteria	Evidence of current acceptance	Expertise needed to conduct	Time needed	Theoretical basis	Degree/ extent of subjectivity	Method of propagation	Treatment of uncertainty and variability	Meaning of output	Transparency and reproducibility	Ease of understanding for non-specialist
Stronger character-istics	International guidelines or standard scientific method	No specialist knowledge required	Hours	Well established, coherent basis for all aspects	Judgement used only to choose method of analysis	Calculation based on appropriate theory	Different types of uncert. & var. quantified separately	Range and probability of alternative outcomes	All aspects of process and reasoning fully documented	All aspects fully understandable
	EU level guidelines or widespread in practice	Can be used with guidelines or literature	Days	Most but not all aspects supported by theory	Combination of data and expert judgment	Formal expert judgment	Uncertainty and variability quantified separately	Range and relative possibility of outcomes	Most aspects of process and reasoning well documented	Outputs and most of process understandable
	National guidelines, or well established in practice or literature	Training course needed	Weeks	Some aspects supported by theory	Expert judgment on defined quantitative scales	Informal expert judgment	Uncertainty and variability distinguished qualitatively	Range of outcomes but no weighting	Process well documented but limited explanation of reasoning	Outputs and principles of process understandable
	Some publications and/or regulatory practice	Substantial expertise or experience needed	A few months	Limited theoretical basis	Expert judgment on defined ordinal scales	Calculation or matrices without theoretical basis		Quantitative measure of degree of uncertainty	Limited explanation of process and/or basis for conclusions	Outputs understandable but not process
Weaker character-istics	Newly developed	Professional statistician needed	Many months	Pragmatic approach without theoretical basis	Verbal description, no defined scale	No propagation	No distinction between variability and uncertainty	Ordinal scale or narrative description for degree of uncertainty	No explanation of process or basis for conclusions	Process and outputs only understandable for specialists

B.9 Formal process on Expert Knowledge Elicitation (EKE) as described in the corresponding EFSA Guidance

This section summarises the process on Expert Knowledge Elicitation which is fully described and discussed in the corresponding EFSA Guidance. Because the remit of the Guidance is limited to the elicitation of main quantitative parameters in EFSAs risk assessments, a more general approach is described in section B.8. Between the informal and formal Expert Knowledge Elicitation is a continuum of alternatives, which could be used to fit the process to the specific needs of the problem, e.g. reframe the problem into the language of practitioners – as described in the formal EKE – but using an existing network of experts – as described in the informal EKE.

Purpose, origin and principal features

Formal techniques for eliciting knowledge from specialised persons were introduced in the first half of the 20th century (e.g. Delphi method in 1946 or Focus groups in 1930—Ayyub Bilal, 2001) and after the sixties they became popular in risk assessments in engineering (EFSA, 2014).

Since then, several approaches were further developed and optimised. Regarding the individual expert judgement on uncertainties of a quantitative parameter the use of subjective probabilities is common.

Nevertheless alternatives exist like fuzzy logic (Zimmermann, 2001), belief functions (Shafer, 1976), imprecise probabilities (Walley, 1991), and prospect theory (Kahneman and Tversky, 1979). The authors claim that these concepts better represent the way experts think about uncertainties than the formal concept of probabilities. On the other hand probabilities have a clear and consistent interpretation. They are therefore proposed in the EFSA Guidance on EKE (EFSA, 2014).

Formal techniques describe the full process of EKE beginning with its initiation (problem definition) done by the working group, the pre-elicitation phase (protocol definition: framing the problem, selecting the experts and method) done by a steering group, the main elicitation phase (training and elicitation) done by the elicitation group, and the post-elicitation phase (documentation) as common task.

Each phase has a clearly defined output which will be internally reviewed and passed to the next phase. The working group is responsible to define the problem to be elicited, summarize the risk assessment context and the existing evidence from empirical data and literature. The steering group will develop the elicitation protocol from the question by framing the problem according to the intended expert board, selecting the experts for the elicitation and the elicitation method to be applied. Finally the elicitation group will perform the elicitation and analyse the results. The separation of the elicitation from the working group allows EFSA to outsource the elicitation to an external contractor with professional experience in the selected elicitation method, to guarantee full confidentiality to the board of external experts, and third to enable the working group to perform an independent review of the results.

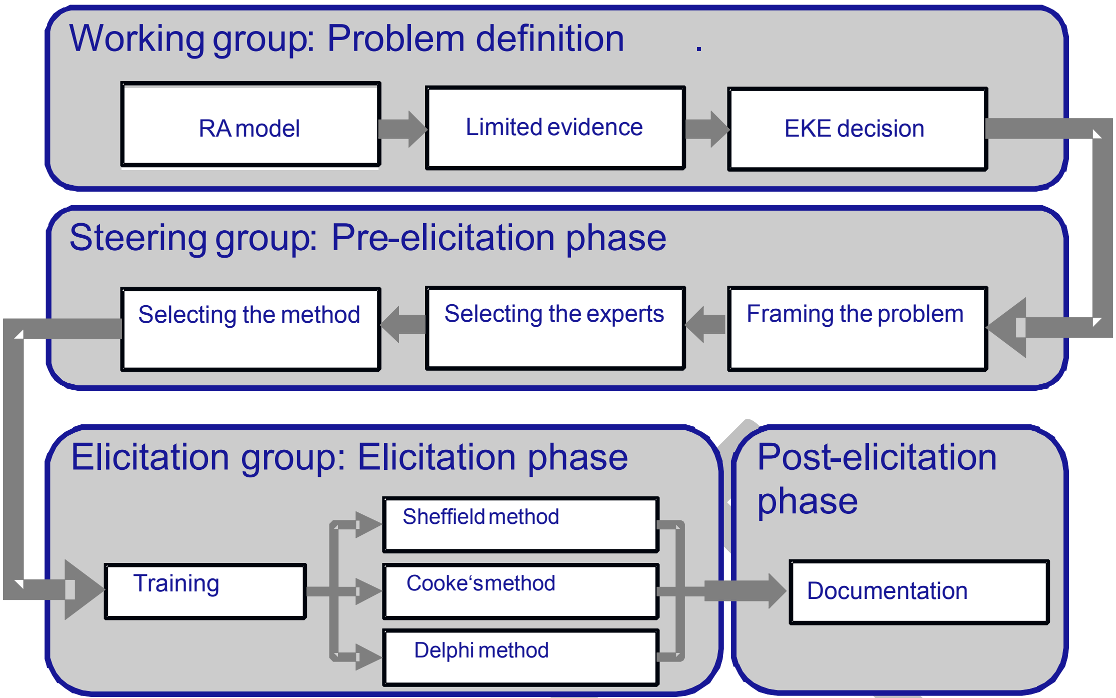


Figure B.9.1. The process of expert knowledge elicitation (EFSA, 2014a)

The elicitation methods differ in the way the judgements of several experts are aggregated. In general three types of methods can be distinguished:

1. Behavioural aggregation: Individual judgements will be aggregated by group interaction of the experts, e.g. using the Sheffield method (O'Hagan et al., 2006)
2. Mathematical aggregation: Individual judgements will be aggregated by a weighted average using e.g. seed questions to calibrate the experts, e.g. the Cooke method (Cooke, 1991)
3. Mixed methods: Individual judgements will be aggregated by moderated feedback loops avoiding direct interactions in the group, e.g. the Delphi protocol as described in EFSA, 2014

The result is in all methods a probability distribution describing the uncertainty of a quantitative parameter in risk assessment, like an influencing factor or the final risk estimate.

Applicability in areas relevant for EFSA

Formal Expert Knowledge Elicitation is applicable in all areas where empirical data from experiments / surveys or literature are limited or missing, and theoretical reasoning is not available, e.g. on future, emerging risks. It is an additional alternative to involve a broad range of stakeholders. In complex, ambiguous risk assessments it is also a possibility to pass the elicitation of detailed questions to independent institutions to gather evidence in broader communities of expertise.

Potential contribution to the main steps of uncertainty analysis

Steps in uncertainty analysis	Potential contribution of this approach
Identifying uncertainties	No, question must be defined beforehand
Describing uncertainties	No, question must be defined beforehand

Assessing the magnitude of individual uncertainties	Yes, by a clearly defined process
Assessing the combined impact of multiple uncertainties on the assessment output, taking account of dependencies	Yes, by a clearly defined process
Assessing the contribution of individual uncertainties to overall uncertainty	No

Melamine example

The problem was divided into two parts: The determination of technical limits in the fraction of milk power [dry milk solids in %], which can be used to produce saleable milk chocolate (without unacceptable changes in taste, consistence or other features of the chocolate). These are handled in questions 1 and 2. And finally the variation in the fraction of milk power [dry milk solids in %] in chocolate imported from China. For the final third question another board of experts was defined.

Question 1: *What is the* ***maximum fraction of milk power [dry milk solids in %],*** *which can be used to produce saleable milk chocolate (without unacceptable changes in taste, consistence or other features of the chocolate)?*

Question 2: *What is the* ***minimum fraction of milk power [dry milk solids in %],*** *which have to be used to produce saleable milk chocolate (without unacceptable changes in taste, consistence or other features of the chocolate)?*

Experts to ask:

Profile: Product developers in big chocolate production companies (including milk chocolate products)

Number of experts: 2-3, because of standardised production processes.

Elicitation methods: Written procedure using adapted Delphi approach. This approach is asking the experts to describe their uncertainty by five numbers:

Steps	Parameter	Explanation
Procedure		To avoid psychological biases in estimating quantitative parameters please give the requested numbers in the right queueing:
1st step:	Upper (U)	**Upper limit of uncertainty** of the maximum fraction of milk powder in saleable chocolate: “You should be really surprised, when you would identify a chocolate with a fraction of milk powder above the upper limit on the market.”
2nd step:	Lower (L)	**Lower limit of uncertainty** of the maximum fraction of milk powder in saleable chocolate: “You should be really surprised, when a person is claiming that a chocolate with a fraction of milk powder below the lower limit is not saleable because of too high milk powder content.”
3rd step:	Median (M)	**Median (or second quartile of uncertainty)** of the maximum fraction of milk powder in saleable chocolate: “Regarding your uncertainty about the true answer this is your best estimate of the maximum fraction of milk powder in saleable chocolate: in the sense that if you would get the true answer (by a full study/experiment) it is equal likely that the true value is above the median (M≤ true value ≤U) as it is below the median (L≤ true value ≤M).”
4th step:	3rd quartile (Q3)	**Third quartile of uncertainty** of the maximum fraction of milk powder in saleable chocolate: “Assuming that the true answer is **above** the median this is the division of the upper interval (between median and the upper limit: [M, U]) into two parts which are again equal likely: 1) between the median and the third quartile: [M, Q3]

		2) between the third quartile and the upper limit: [Q3, U]
5th step:	1st quartile (Q1)	**First quartile of uncertainty** of the maximum fraction of milk powder in saleable chocolate: "Assuming that the true answer is **below** the median this is the division of the upper interval (between lower limit and the median: [L, M]) into two parts which are again equal likely: 1) between the lower limit and the first quartile: [L, Q1] 2) between the first quartile and the median: [Q1, M]
Restrictions:		The five **numbers are ordered** from low to high as: L ≤ Q1 ≤ M ≤ Q3 ≤ U
Consistency check:		Finally please **check** if the following four intervals will have equal likelihood (of 25% or one quarter) to include the true maximum fraction of milk powder in saleable chocolate: 1) between the lower limit and the first quartile: [L, Q1] 2) between the first quartile and the median: [Q1, M] 3) between the median and the third quartile: [M, Q3] 4) between the third quartile and the upper limit: [Q3, U] This can be visualized by a bar chart on the four intervals, where each bar contains the same area of 25%, which is an expression of the subjective distribution of uncertainty.

First round with initial answers and reasoning (asked with a specific EXCEL file giving more explanations and setting restrictions to the answers) was performed during the first week involving 3 experts (hypothetical example for illustration):

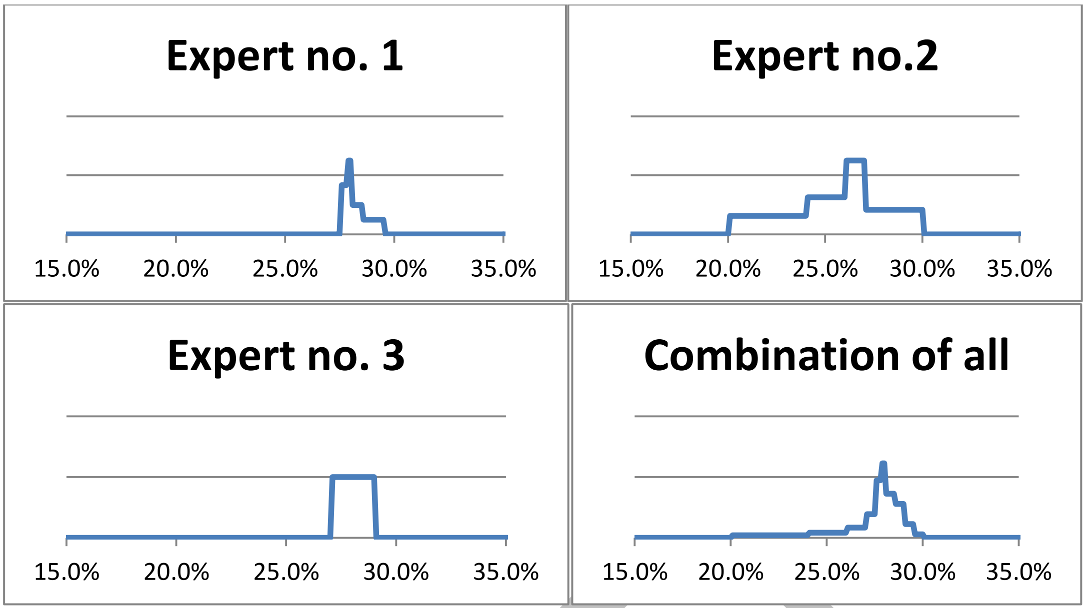
- Mrs. White, Chocolate Research Inc. (UK);
- Mrs. Argent, Chocolatiers Unis (France);
- and Mr. Rosso, Dolce International (Italy)

	Lower	**1st Quart**	**Median**	**3rd Quart**	**Upper**	**Reasoning**
Expert no1	24.5%	24.8%	25%	25.5%	26.5%	Variation in our production line of the product with highest content of milk power
Expert no 2	20%	24%	26%	27%	30%	Depending on the sugar content there will be an aftertaste of the milk powder
Expert no 3	27%	27.5%	28%	28.5%	29%	We recognized problems in the production line when higher the milk powder content.

After feedback of the answers to the experts they revised in the second week their answers:

	Lower	**1st Quart**	**Median**	**3rd Quart**	**Upper**	**Reasoning**
Expert no1	27.5%	27.8%	28%	28.5%	29.5%	Higher contents are possible, but not used by my company
Expert no 2	20%	24%	26%	27%	30%	
Expert no 3	27%	27.5%	28%	28.5%	29%	

As result of the procedure the judgements of all three experts were combined by using equal weights to each expert.



At the same time the expert board was asked about the minimum content of milk powder in milk chocolate. The experts concluded that milk chocolate needs by legal requirements a minimum of 14% milk powder (dry milk solids obtained by partly or wholly dehydrating whole milk, semi- or full-skimmed milk, cream, or from partly or wholly dehydrated cream, butter or milk fat; EC Directive 2000/36/EC, Annex 1, A4 of 23rd June 2000). The risk assessment is therefore restricted to the consumption of chocolate following the legal requirements. Illegal trade (in this sense) is not included. The minimum was set to 14%.

To assess the variability of Melamine content in chocolate imported from China an additional Question 3was asked to another board of experts:

Question 3: Assuming that milk chocolate was produced in and imported from China.

Part 3A: Consider a producer using a high content of milk powder in the chocolate that only in 5% (one of twenty) of the products from China will be with a higher content. What is the **fraction of milk power [in %]** contained in this chocolate? (Please specify your uncertainty)

Part 3B: Consider a producer using a low content of milk powder in the chocolate that only in 5% (one of twenty) of the products from China will be with a lower content. What is the **fraction of milk power [in %]** contained in this chocolate? (Please specify your uncertainty)

Part 3C: Consider a producer using an average content of milk powder in the chocolate that half of the products from China will be with higher and half with lower content. What is the **fraction of milk power [in %]** contained in this chocolate? (Please specify your uncertainty)

Experts to ask:

Profile: Quality controller (laboratory) of food importing companies / food control in importing regions with relevant import of chocolate or similar products (containing milk powder) from China.

Number of experts: 4, because of the limited number of experts with this profile.

Elicitation methods (hypothetical example)**:** The expert board was invited to a one-day physical meeting, summarizing the identified evidence on the topic. After a training session on the elicitation method, the Sheffield protocol was performed on Question 3, part A to C.

Strengths

1. Applicable in absence of empirical data or theoretical reasoning
2. Reproducible with regard to the pre-defined protocol
3. Transparent in the documentation
4. Applicable for emerging (future) risks / participation of stakeholders in complex, ambiguous RA

Weaknesses and possible approaches to reduce them

1. Time and resource intensive, should be primarily used for the most sensitive parameters in a risk assessment
2. Little previous experience of this approach in EFSA's areas of risk assessment. However, there is a substantial literature by expert practitioners, and it is better established in other areas (e.g. nuclear engineering, climate change).

Assessment against evaluation criteria

This method is assessed against the criteria in Table B.9.1.

Conclusions

1. The method has a high applicability in working groups and boards of EFSA and should be applied to quantify uncertainties in situations where empirical data from experiments / surveys, literature are limited and the purpose of the risk assessment is sensitive and need the performance of a full Formal Expert Knowledge Elicitation.
2. The method is applicable in steps of the risk assessment, where quantitative parameters have to be obtained.
3. The method should not substitute the use of empirical data, experiments, surveys or literature, when these are already available or could be retrieved with corresponding resources.
4. In order to initiate a Formal Expert Knowledge Elicitation some experts of the working group should be trained in steering expert elicitations according to the EFSA Guidance. In case of complex or sensitive questions the elicitation should be perform by professional elicitation groups.

References

EFSA (European Food Safety Authority), 2014. Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment. EFSA Journal 2014;12(6):3734. [278 pp.] doi:10.2903/j.efsa.2014.3734

Cooke RM, 1991. Experts in uncertainty. Oxford University Press, Oxford, UK.

Meyer MA and Booker JM, 2001. Eliciting and analyzing expert judgment: a practical guide. American Statistical Association and the Society for Industrial and Applied Mathematics, Alexandria, Virginia, USA.

O'Hagan A, Buck CE, Daneshkhah A, Eiser JR, Garthwaite PH, Jenkinson D, Oakley JE and Rakow T, 2006. Uncertain judgements: eliciting experts' probabilities. John Wiley and Sons, Chichester, UK.

Rowe G, 1992. Perspectives on expertise in the aggregation of judgments. In: Expertise and decision support. Eds Wright G and Bolger F. Plenum, Press, London, UK, 155–180.

Rowe G and Wright G, 1999. The Delphi technique as a forecasting tool: issues and analysis. International Journal of Forecasting, 15, 353–375.

Table B.9.1: Assessment of Formal expert knowledge elicitation (EKE) (when applied well) against evaluation criteria.

Criteria	Evidence of current acceptance	Expertise needed to conduct	Time needed	Theoretical basis	Degree/ extent of subjectivity	Method of propagation	Treatment of uncertainty and variability	Meaning of output	Transparency and reproducibility	Ease of understanding for non-specialist
Stronger character-istics	International guidelines or standard scientific method	No specialist knowledge required	Hours	Well established, coherent basis for all aspects	Judgement used only to choose method of analysis	Calculation based on appropriate theory	Different types of uncert. & var. quantified separately	Range and probability of alternative outcomes	All aspects of process and reasoning fully documented	All aspects fully understandable
	EU level guidelines or widespread in practice	Can be used with guidelines or literature	Days	Most but not all aspects supported by theory	Combination of data and expert judgment	Formal expert judgment	Uncertainty and variability quantified separately	Range and relative possibility of outcomes	Most aspects of process and reasoning well documented	Outputs and most of process understandable
	National guidelines, or well established in practice or literature	Training course needed	Weeks	Some aspects supported by theory	Expert judgment on defined quantitative scales	Informal expert judgment	Uncertainty and variability distinguished qualitatively	Range of outcomes but no weighting	Process well documented but limited explanation of reasoning	Outputs and principles of process understandable
	Some publications and/or regulatory practice	Substantial expertise or experience needed	A few months	Limited theoretical basis	Expert judgment on defined ordinal scales	Calculation or matrices without theoretical basis		Quantitative measure of degree of uncertainty	Limited explanation of process and/or basis for conclusions	Outputs understandable but not process
Weaker character-istics	Newly developed	Professional statistician needed	Many months	Pragmatic approach without theoretical basis	Verbal description, no defined scale	No propagation	No distinction between variability and uncertainty	Ordinal scale or narrative description for degree of uncertainty	No explanation of process or basis for conclusions	Process and outputs only understandable for specialists

B.10 Statistical inference from data — Confidence intervals

This section is only concerned with standard calculations for confidence intervals. The bootstrap is discussed in a separate section of this annex (section B.11).

Purpose, origin and principal features

A confidence interval is the conventional expression of uncertainty, based on data, about a parameter in a statistical model. The basic theory (Cox, 2006) and methodology was developed by statisticians during the first half of the 20^{th} century. Confidence intervals are used by the majority of scientists as a way of summarizing inferences from experimental data and the training of most scientists includes some knowledge of the underlying principles and methods of application. See, for example, Moore (2009).

A confidence interval provides a range of values for the parameter together with a level of confidence in that range (commonly 95% or 99%). Formally, the confidence level indicates the success rate of the procedure under repeated sampling and assuming that the statistical model is correct. However, the confidence level is often interpreted for a specific dataset, as the probability that the calculated range actually includes the true value of the parameter, i.e. a 95% confidence interval becomes a 95% probability interval for the parameter. That interpretation is reasonable in many cases but requires for each specific instance that the user of the confidence interval make a judgement that it is a reasonable interpretation. This is in contrast to Bayesian inference (section B.9) which sets out to produce probability intervals from the outset. The judgement the user needs to make is that the confidence interval does not convey additional information which would make the user want to alter the probability to be ascribed to the interval.

To use this method, one requires a suitable statistical model linking available data to parameters of interest and an appropriate procedure for calculating the confidence interval. For many standard statistical models, such procedures exist and are often widely known and used by scientists. Developing new confidence interval calculations is generally a task for theoretical statisticians.

Many standard confidence interval procedures deliver only an approximation to the stated level of confidence and the accuracy of the approximation is often not known explicitly although it usually improves as the sample size increases. When the statistical model does not correctly describe the data, the confidence level is affected, usually by an unknown amount.

Most statistical models have more than one parameter and in most cases the resulting uncertainty about the parameters will involve dependence. Unless there is very little dependence, it is inappropriate to express the uncertainty as a separate confidence interval for each parameter. Instead the uncertainty should be expressed as a simultaneous confidence region for all the parameters. This is often technically challenging for non-statisticians and it may be preferable in practice to use another statistical approach to representing uncertainty, especially one which can represent uncertainty as a Monte Carlo sample, each realisation of which provides a value for each of the parameters.

Applicability in areas relevant for EFSA

The methodology is applicable in principle to all areas where data from experiments or surveys are used in risk assessment.

However, unless data are being used to make inference about a single parameter of interest in statistical model, addressing dependence between parameters is likely to be challenging and this may reduce the usefulness of confidence intervals as an expression of uncertainty.

Standard confidence interval procedures, such as those for means of populations, regression coefficients and dose-response estimates, are used throughout EFSA'a work.

Potential contribution to the main steps of uncertainty analysis

Steps in uncertainty analysis	Potential contribution of this approach
Identifying uncertainties	Not applicable.
Describing uncertainties	Not applicable.
Assessing the magnitude of individual uncertainties	Yes/No. Limited to uncertainties relating to parameters in statistical models. For many statistical models, there is a clear procedure based on empirical data
Assessing the combined impact of multiple uncertainties on the assessment output, taking account of dependencies	Not applicable.
Assessing the contribution of individual uncertainties to overall uncertainty	Not applicable.

Melamine example

Confidence intervals and regions will be illustrated by application to uncertainty about two of the sources of variability considered in the version of the melamine example which considers uncertainty about variability of exposure. Further supporting details about both versions of the melamine example may be found in Annex C. The variables considered here are body-weight and consumption in a day.

Data for both variables for children aged from 1 up to 2 years old were obtained from EFSA. Annex C gives details of the data and some data analysis supporting the choice of distribution family for each variable. The variables are treated as independent in what follows and the reasoning for doing so is included in Annex C.

Both variables are considered in detail below because there are important differences between the statistical models used. The normal distribution used for log body-weight is the most commonly used model for continuous variability and the confidence interval procedures are well known. The gamma distribution used for consumption requires more advanced statistical calculations and also shows the importance of addressing dependence between distribution parameters.

<u>Body-weight (bw)</u>

For bw, the statistical model is that: (i) bw follows a log-normal distribution, so that log bw follows a normal distribution; (ii) the uncertain distribution parameters are the mean μ_{logbw} and standard deviation σ_{logbw} of the distribution of log bw (base 10); (iii) the data are a random sample from the distribution of bw for the population represented by the data.

For the mean and standard deviation of a normal distribution, there are standard confidence interval procedures which assume that the data are a random sample.

For the mean the confidence interval is $\bar{x} \pm t^* s/\sqrt{n}$ where $\bar{x}$ denotes the sample mean, s is the sample standard deviation and n is the sample size. t^* is a percentile of the t-distribution having $n-1$ degrees of freedom. The percentile to be chosen depends on the confidence level: for example, for 95% confidence, it is the 97.5th percentile; for 99% confidence, the 99.5th percentile. For the standard deviation, the confidence interval is $\left(s/\sqrt{\chi_u^2/(n-1)}, s/\sqrt{\chi_l^2/(n-1)}\right)$ where again s is the sample standard deviation and n is the sample size. χ_l^2 and χ_u^2 are lower and upper percentiles of the chi-squared distribution having $n-1$ degrees of freedom. The percentiles to be used depend on the required confidence level: for example, for 95% confidence, they are the 2.5th and 97.5th percentiles. Values for t^*, χ_l^2 and χ_u^2 are easily obtained from tables or using standard statistical software.

For the body-weight data used in the example, $n_{\text{logbw}} = 171$, $\bar{x}_{\text{logbw}} = 1.037$ and $s_{\text{logbw}} = 0.060$. Taking 95% as the confidence level, $t^* = 1.974$, $\chi_l^2 = 135.79$ and $\chi_u^2 = 208.00$. Consequently, the confidence interval for μ_{logbw} is $1.037 \pm 1.974 \times 0.060/\sqrt{171} = 1.037 \pm 0.009 = (1.028, 1.046)$ and the confidence interval for σ_{logbw} is $\left(0.060/\sqrt{208.00/170}, 0.060/\sqrt{135.79/170}\right) = (0.054, 0.067)$.

Because the mean of the underlying normal distribution is the logarithm of the geometric mean (and median) of a log-normal, we can convert the confidence interval for μ_{logbw} into a 95% confidence interval for the geometric mean of body-weight: $(10^{1.028}, 10^{1.046})$ = (10.67, 11.12) kg. Similarly, the standard deviation of the underlying normal is the logarithm of the geometric standard deviation of the log-normal and so a 95% confidence interval for the geometric standard deviation of body-weight is $(10^{0.054}, 10^{0.067})$ = (1.13, 1.17).

Each of these confidence intervals is an expression of uncertainty about the corresponding uncertain parameter for variability of body-weight. However, they do not express that uncertainty in a form which is directly suitable for use in a probability bounds analysis or Monte Carlo uncertainty analysis. In the absence of further information about body-weight, experts may be willing to make a probabilistic interpretation of the confidence level.

In principle, given data, there is dependence in the uncertainty about the two parameters of a normal distribution. That dependence may be substantial when the sample size is small but decreases for larger samples.

<u>Consumption (q)</u>

For q, the statistical model is that: (i) q follows a gamma distribution with uncertain distribution parameters being the shape α_{q} and rate β_{q}; (ii) the data are a random sample from the distribution of q.

Like the normal and log-normal distributions, the gamma family of distributions has two distribution parameters. The most common choice of how to parameterise the distribution is the mathematically convenient one of a shape parameter α and a rate parameter β so that the probability density for q is $p(q) \propto \frac{\beta^\alpha}{\Gamma(\alpha)} q^{\alpha-1} e^{-\beta q}$.

There are a number of ways to get approximate confidence intervals for both distribution parameters. Of those the one which has the best performance is maximum likelihood estimation (Whitlock and Schluter, 2014) combined with large sample approximation confidence interval calculations. However, the main practical difficulty is that the sampling distributions of estimates of the parameters are strongly correlated and so it is not very useful to consider uncertainty about each parameter on its own. The large sample theory for maximum likelihood estimation shows how to compute a simultaneous confidence region for both parameters. Figure B.10.1 shows the maximum likelihood estimate and 95% and 99% confidence regions for α and β based the consumption data used in the example; the dotted vertical and horizontal lines show respectively the ends of the 95% confidence intervals for α and β.

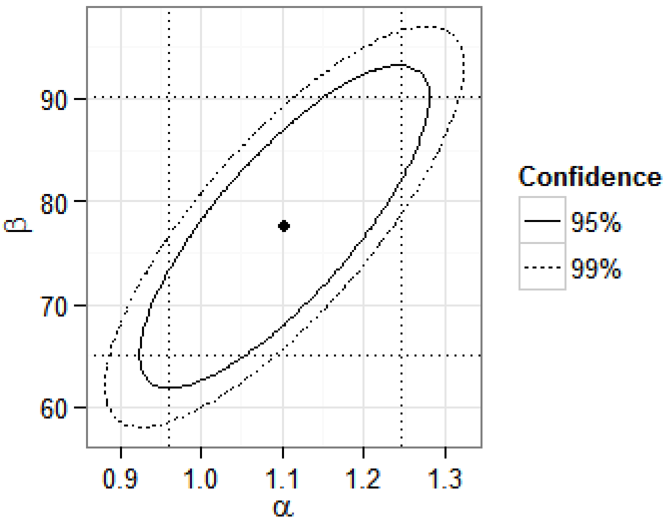


Figure B.10.1: Confidence regions for distribution parameters for gamma distribution used to model variability of consumption by one-year-old children.

Strengths

1. For many survey designs or study designs and corresponding statistical models, there is familiar methodology to obtain confidence intervals for individual statistical model parameters.
2. Widely available software for computing confidence intervals (Minitab, R, Systat, Stata, SAS, ...)
3. Computations are based on the generally accepted mathematical theory of probability although probability is only used directly to quantify variability.

Weaknesses and possible approaches to reduce them

1. Confidence intervals only address uncertainties relating to parameters in statistical models.
2. Requires specification of a statistical model for data, the model depending on parameters which be estimated. Specifying and fitting non-standard models can be time-consuming and difficult for experts and may often require the involvement of a professional statistician.
3. Results are expressed in the language of confidence rather than of probability. Uncertainties expressed in this form can only be combined in limited ways. They can only be combined with probabilistic information if experts are willing to make probability statements on the basis of their knowledge of one or more confidence intervals.
4. Dependence in the uncertainties about statistical model parameters is usual when a statistical model having more than one parameter is fitted to data. This can be addressed in principle by making a simultaneous confidence statement about multiple parameters. However, such methods are much less familiar to most scientists and generally require substantial statistical expertise.

Assessment against evaluation criteria

This method is assessed against the criteria in Table B.10.1.

Conclusions

1. Confidence intervals are suitable for application across EFSA in situations where standard statistical models are used in order to quantify uncertainty separately about individual statistical model parameters using intervals.
2. The quantification provided is not directly suitable for combining with other uncertainties in probabilistic calculations although expert judgement may be applied in order to support such uses.

References

Cox DR (2006). Principles of Statistical Inference. Cambridge University Press.

Moore DS (2009) The Basic Practice of Statistics, 5th ed. WH Freeman.

Whitlock M, Schulter D (2014). The Analysis of Biological Data, 2nd ed. Roberts and Company Publishers.

Table B.10.1: Assessment of Confidence intervals (when well applied) against evaluation criteria.

Criteria	Evidence of current acceptance	Expertise needed to conduct	Time needed	Theoretical basis	Degree/ extent of subjectivity	Method of propagation	Treatment of uncertainty and variability	Meaning of output	Transparency and reproducibility	Ease of understanding for non-specialist
Stronger character-istics	International guidelines or standard scientific method	No specialist knowledge required	Hours	Well established, coherent basis for all aspects	Judgement used only to choose method of analysis	Calculation based on appropriate theory	Different types of uncert. & var. quantified separately	Range and probability of alternative outcomes	All aspects of process and reasoning fully documented	All aspects fully understandable
	EU level guidelines or widespread in practice	Can be used with guidelines or literature	Days	Most but not all aspects supported by theory	Combination of data and expert judgment	Formal expert judgment	Uncertainty and variability quantified separately	Range and relative possibility of outcomes	Most aspects of process and reasoning well documented	Outputs and most of process understandable
	National guidelines, or well established in practice or literature	Training course needed	Weeks	Some aspects supported by theory	Expert judgment on defined quantitative scales	Informal expert judgment	Uncertainty and variability distinguished qualitatively	Range of outcomes but no weighting	Process well documented but limited explanation of reasoning	Outputs and principles of process understandable
	Some publications and/or regulatory practice	Substantial expertise or experience needed	A few months	Limited theoretical basis	Expert judgment on defined ordinal scales	Calculation or matrices without theoretical basis		Quantitative measure of degree of uncertainty	Limited explanation of process and/or basis for conclusions	Outputs understandable but not process
Weaker character-istics	Newly developed	Professional statistician needed	Many months	Pragmatic approach without theoretical basis	Verbal description, no defined scale	No propagation	No distinction between variability and uncertainty	Ordinal scale or narrative description for degree of uncertainty	No explanation of process or basis for conclusions	Process and outputs only understandable for specialists

B.11 Statistical inference from data — The bootstrap

Purpose, origin and principal features

The bootstrap is a tool for quantifying uncertainty due to sampling variability. It is both a basic sensitivity analysis tool and a method for producing approximate confidence intervals. It has the advantage that it is often easy to implement using Monte Carlo (see section B.14).

The bootstrap was originally proposed by Efron (1981). Davison and Hinkley (1997) give an account of theory and practice aimed at statisticians while Manly (2006) is aimed more at biologists and other scientists.

The problem it addresses is that it is usually uncertain how much the result of a calculation based on a sample of data might differ from the result which would be obtained by applying the calculation to the statistical population from which the data were drawn. For some statistical models, there is a well-known mathematical solution to that problem. For others, there is not. The bootstrap provides an approximate answer which is often relatively easily calculated. The underlying principle is that, for many situations, sampling variability when sampling from the statistical population is similar to sampling variability when re-sampling from the data. It is often easy to re-sample from the data and repeat the calculation. By repeating the re-sampling process many times it is possible to quantify the uncertainty attached to the original calculation.

The bootstrap can be applied in many ways and to a wide variety of parametric and non-parametric statistical models. However, it is most easily applied to situations where data are a random sample or considered to be equivalent to a random sample. In such situations, the uncertainty attached to any statistical estimator(s) calculated from the data can be examined by repeatedly re-sampling from the data and repeating the calculation of the estimator(s) for each new sample. The estimator may be something simple like the sample mean or median or might be something much more complicated such as a percentile of exposure from estimated from data on consumption and concentrations. The re-sampling procedure is to take a random sample from the data, with replacement and of the same size as the data. Although from a theoretical viewpoint it is not always necessary, in practice the bootstrap is nearly always implemented using Monte Carlo sampling.

When applying an estimator to a particular dataset, one is usually trying to estimate the population value: the value which would have been obtained by applying the estimator to the statistical population from which the data were drawn. There are many approaches to obtaining an approximate confidence interval, quantifying uncertainty about the population value, based on bootstrap output. The differences originate in differing assumptions about the relationship between re-sampling variability and sampling variability, some attempting to correct for potential systematic differences between sampling and re-sampling. All the approaches assume that the sample size is large. Further details are provided by Davison and Hinkley (1997).

The bootstrap can be used in relation to either a parametric or non-parametric statistical model of variability. The advantage of the latter is that no parametric distribution family need be assumed but it has the potential disadvantage that, if the whole distribution is being used in any subsequent calculation, the only values which will be generated for the variable are those in the original data sample. The advantage of working with a parametric statistical model is that, if one bootstraps estimates of all the parameters, one obtains an indication of uncertainty about all aspects of the distribution.

The bootstrap will not perform well when the sample size is low or is effectively low. One example of an effectively low sample size would be when estimating non-parametrically a percentile near the limit of what could be estimated from a given sample size. Another would be when a large percentage of the data take the same value, perhaps as values below a limit of detection or limit of quantification.

One very attractive feature of the bootstrap is that it can readily be applied to situations where there is no standard confidence interval procedure for the statistical estimator being used. Another is that it is possible to bootstrap more than one variable at the same time: if the data for two variables were obtained independently, then one takes a re-sample from each dataset in each re-sampling iteration. The frequency property of any resulting confidence interval is then with respect to repetition not of a single survey/experiment but is with respect to repeating all of them.

Because the output of the bootstrap is a sample of values for parameters, it is computationally straightforward to use the output as part of a 2D Monte Carlo analysis (section B.14) of uncertainty. Such an analysis could use bootstrap output for some uncertainties and distributions obtained by EKE and/or Bayesian inference for other uncertainties. However, the meaning of the output of the Monte Carlo calculation is unclear unless an expert judgement has been made that the bootstrap output is a satisfactory probabilistic representation of uncertainty for the parameters on the basis of the data to which the bootstrap has been applied.

Applicability in areas relevant for EFSA

The bootstrap is a convenient way to make an assessment of uncertainty due to sampling variability in situations which involve a random sample of data and where it is difficult to calculate a standard confidence interval or make a Bayesian inference. As such, it has particular applicability to data obtained from random surveys which are used in complex statistical calculations, for example estimation of percentiles of exposure using probabilistic modelling.

The bootstrap has been recommended as part of the EFSA (2012) guidance on the use of probabilistic methodology for modelling dietary exposure to pesticide residues. However, that guidance recognises its limitations and recommends that it be used alongside other methods. Bootstrapping was used frequently in microbial dose-response assessment but it has now largely been replaced by Bayesian inference (e.g. Medema et al. 1996, Teunis PFM et al. 1996).

Potential contribution to the main steps of uncertainty analysis

Steps in uncertainty analysis	Potential contribution of this approach
Identifying uncertainties	Not applicable.
Describing uncertainties	Not applicable.
Assessing the magnitude of individual uncertainties	Yes/No. Quantifies sampling variability but not other types of uncertainty.
Assessing the combined impact of multiple uncertainties on the assessment output, taking account of dependencies	No/Yes. Can be used to address multiple sources of uncertainty due to sampling variability in a single Monte Carlo calculation, thereby providing the combined impact of those, but not other, sources of uncertainty.
Assessing the contribution of individual uncertainties to overall uncertainty	Not applicable.

Melamine example

The bootstrap will be illustrated by application to uncertainty about one of the sources of variability considered in the version of the melamine example which considers uncertainty about variability of exposure. Further supporting details about both versions of the melamine example may be found in Annex C. The variable considered here is body-weight. The body-weight example is followed by a short discussion of the potential to apply the bootstrap to consumption: the other variable for which sample data were available

<u>Body-weight (bw)</u>

Data for body-weight for children aged from 1 up to 2 years old were obtained from EFSA. Annex C gives details of the data and some data analysis supporting the choice of distribution family.

For bw, the statistical model is that: (i) bw follows a log-normal distribution, so that log bw follows a normal distribution; (ii) the uncertain distribution parameters are the mean μ_{logbw} and standard deviation σ_{logbw} of the distribution of log bw (base 10); (iii) the data are a random sample from the distribution of bw for the population represented by the data.

Firstly, consider uncertainty attached to the estimates of parameters for the log-normal statistical model of variation in body-weight. These parameters are the mean μ_{logbw} and standard deviation σ_{logbw} of $\log_{10}\mathrm{bw}$. They are estimated simply by calculating the sample mean and sample standard deviation of the observed data for $\log_{10}\mathrm{bw}$. Figure B.8.1 plots the values of these estimates for the original data and for 999 datasets re-sampled from the original data:

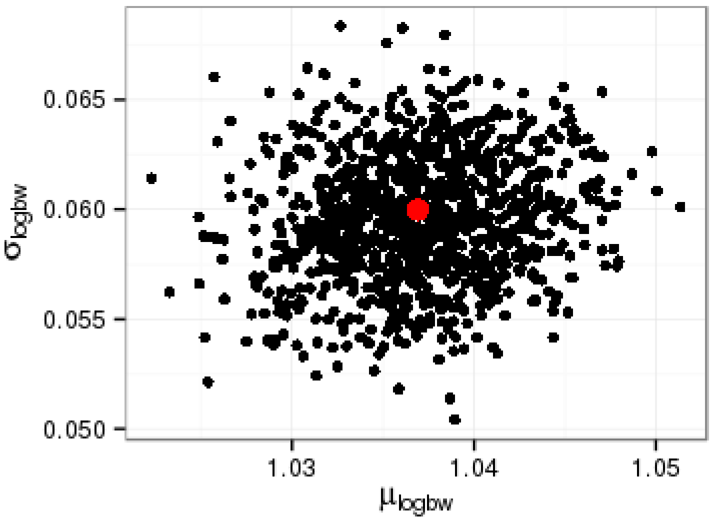


Figure B.11.1: Estimates of parameters of log-normal distribution fitted to datasets obtained by re-sampling the body-weight data. The red point shows the estimates for the original data.

The most commonly used methods for deriving a confidence interval from bootstrap output all give very similar answers for this example: an approximate 95% confidence interval for $\mu_{\mathrm{logbw}}$is (1.028, 1.046) and for σ_{logbw} the approximate 95% confidence interval using the "percentile" method is (0.0540, 0.0652) while other methods give (0.0548, 0.0659). There are two reasons why different methods give very similar answers here: the original sample size is large and the mean and standard deviation are both estimators for which the bootstrap performs reasonable well.

If a specific percentile, say the 99$^{\mathrm{th}}$, of variability of body-weight was of interest, there are two quite different approaches:

- For each bootstrap re-sample, the estimates of $\mu_{\mathrm{logbw}}$and σ_{logbw} can be calculated and then the estimated 99th percentile then μ_{logbw}+2.33*σ_{logbw} using the log-normal model, Doing so provides 999 bootstrap values for the 99th percentile to which a bootstrap confidence interval calculation can be applied: the percentile method gives (1.158, 1.192) for 99th percentile of $\log_{10}\mathrm{bw}$ which becomes (14.38, 15.56) as a CI for the 99th percentile of bw.

- Alternatively, the assumption of the log-normal parametric statistical model can be dropped and a non-parametric model for variability of body-weight used instead. For each re-sampled dataset, a non-parametric estimate of the 99th percentile is computed and a bootstrap confidence interval calculation is then applied to the 999 values of the 99th percentile: the percentile method gives (14.00, 15.42) and other methods give somewhat slightly lower values for both ends of the confidence interval.

<u>Other variables</u>

The bootstrap cannot be applied to variability of concentration (c) or weight fraction (w) because no sample of data is available for either source of variability.

For consumption (q), the bootstrap could be applied. If uncertainty about the parameters alpha and beta of the gamma distribution model was required, it would be necessary to estimate the distribution parameters α_q and β_q for each re-sampled dataset. This could be done by maximum likelihood estimation or, less optimally, by estimation using the method of moments.

Note that it would not be appropriate to carry out independent re-sampling of q and bw in this example. In the surveys from which the data were obtained, values for both variables come from the same individuals. The appropriate way to implement the bootstrap, to simultaneously address uncertainty about both q and bw, would be to re-sample entire records from the surveys. Doing so would also address dependence between q and bw.

Strengths

1. Computations are based on the generally accepted mathematical theory of probability although probability is only used directly to quantify variability.
2. Often does not require a lot of mathematical sophistication to implement.
3. Allows the user to decide what statistical estimator(s) to use.
4. Easily applied using Monte Carlo
5. Specialist software exists for a number of contexts (CrystalBall, MCRA, Creme, …) as well as the possibility to use some general purpose statistical software, e.g. R.

Weaknesses and possible approaches to reduce them

1. The bootstrap only addresses random sampling uncertainty whereas other statistical inference methods can address a wider range of uncertainties affecting statistical models.
2. The performance of the bootstrap is affected both by the original sample size and by the estimator used. Larger samples generally improve the performance. Estimators which are not carefully designed may be badly biased or inefficient. This can be avoided by consulting a professional statistician.
3. The non-parametric bootstrap never produces values in a re-sample which were not present in the data and consequently the tails of the distribution will be under-represented.
4. Bootstrap confidence interval procedures are only approximate and in some situations the actual confidence may differ greatly from the claimed level. This can sometimes be ameliorated by carrying out a suitable simulation study.
5. Deciding when the method works well or badly often requires sophisticated mathematical analysis.

Assessment against evaluation criteria

This method is assessed against the criteria in Table B.8.1. The two extremes of the "Method of propagation" column have both been selected because the method can combine uncertainties due to sampling variability for multiple variables but cannot combine those uncertainties with other kinds of uncertainty.

Conclusions

1. The bootstrap is suitable for application across EFSA in situations where data are randomly sampled and it is difficult to apply other methods of statistical inference.
2. It provides an approximate quantification of uncertainty in such situations and is often easy to apply using Monte Carlo.
3. The results of the bootstrap need to be evaluated carefully, especially when the data sample size is not large or when using an estimator for which the performance of the bootstrap has not been previously considered in detail.

References

Efron B. (1981). "Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods". Biometrika 68 (3): 589–599.

Davison AC and Hinkley DV (1997). Bootstrap Methods and their Application. Cambridge University Press.

EFSA (2012). Guidance on the Use of Probabilistic Methodology for Modelling Dietary Exposure to Pesticide Residues. EFSA Journal 2012;10(10):2839

Manly BFJ (2006). Randomization, Bootstrap and Monte Carlo Methods in Biology, 3rd ed. Chapman and Hall/CRC.

Medema GJ, Teunis PF, Havelaar AH, Haas CN. Assessment of the dose-response relationship of Campylobacter jejuni. Int J Food Microbiol. 1996;30(1-2):101-11.

Teunis PFM, Van der Heijden OG, Van der Giessen JWB, Havelaar AH. The dose-response relation in human volunteers for gastro-intestinal pathogens. Bilthoven: National Institute of Public Health and the Environment, 1996 Report nr. 284550002.

Table B.11.1: Assessment of The bootstrap (when applied well) against evaluation criteria.

Criteria	Evidence of current acceptance	Expertise needed to conduct	Time needed	Theoretical basis	Degree/ extent of subjectivity	Method of propagation	Treatment of uncertainty and variability	Meaning of output	Transparency and reproducibility	Ease of understanding for non-specialist
Stronger character-istics	International guidelines or standard scientific method	No specialist knowledge required	Hours	Well established, coherent basis for all aspects	Judgement used only to choose method of analysis	Calculation based on appropriate theory	Different types of uncert. & var. quantified separately	Range and probability of alternative outcomes	All aspects of process and reasoning fully documented	All aspects fully understandable
	EU level guidelines or widespread in practice	Can be used with guidelines or literature	Days	Most but not all aspects supported by theory	Combination of data and expert judgment	Formal expert judgment	Uncertainty and variability quantified separately	Range and relative possibility of outcomes	Most aspects of process and reasoning well documented	Outputs and most of process understandable
	National guidelines, or well established in practice or literature	Training course needed	Weeks	Some aspects supported by theory	Expert judgment on defined quantitative scales	Informal expert judgment	Uncertainty and variability distinguished qualitatively	Range of outcomes but no weighting	Process well documented but limited explanation of reasoning	Outputs and principles of process understandable
	Some publications and/or regulatory practice	Substantial expertise or experience needed	A few months	Limited theoretical basis	Expert judgment on defined ordinal scales	Calculation or matrices without theoretical basis		Quantitative measure of degree of uncertainty	Limited explanation of process and/or basis for conclusions	Outputs understandable but not process
Weaker character-istics	Newly developed	Professional statistician needed	Many months	Pragmatic approach without theoretical basis	Verbal description, no defined scale	No propagation	No distinction between variability and uncertainty	Ordinal scale or narrative description for degree of uncertainty	No explanation of process or basis for conclusions	Process and outputs only understandable for specialists

B.12 Statistical inference from data — Bayesian inference

Purpose, origin and principal features

Bayesian inference is a methodology for expressing and calculating uncertainty about parameters in statistical models, based on a combination of expert judgments and data. The resulting uncertainty is expressed as a probability distribution for the statistical model parameters and is therefore well-suited for combining with other uncertainties using the laws of probability.

The principle underlying Bayesian inference has a long history in the theoretical development of statistical inference. However, it was not until the advent of modern computing that it started to be widely applied and new methodology developed. Since around 1990, there has been an explosion in Bayesian research and in application to all areas of natural and social sciences and to quantification of uncertainty in various financial sectors of business. Between them, Berry (1995), Kruschke (2010) and Gelman et al (2013) cover a wide range from elementary Bayesian principles to advanced techniques.

It differs in two key features from other methods of statistical inference considered in this guidance. Firstly, with Bayesian approaches, uncertainty about the parameter(s) in a statistical model is expressed in the form of a probability distribution so that not only a range of values is specified but also the relative likelihoods of values. Secondly, the judgments of experts based on other information can be combined with the information provided by the data. In the language of Bayesian inference, those expert judgments must be represented as a *prior distribution* for the parameter(s). The statistical model applied to the observed data provides the *likelihood function* for the parameter(s). The likelihood function encapsulates the information provided by the data. The prior distribution and likelihood function are then combined mathematically to calculate the *posterior distribution* for the parameter(s). The posterior distribution is the probabilistic representation of the uncertainty about the parameter(s), obtained by combining the two sources of information. When expert judgements are not available with which to form the prior distribution, for many statistical models standard prior distributions are available which are often described being non-informative or as representing prior lack of knowledge.

As with other methods of statistical inference, calculations are straightforward for some statistical models and more challenging for others. A common way of obtaining a practically useful representation of uncertainty is by a large random sample from the distribution, i.e. Monte Carlo (see section B.14). For some models, there is a simple way to perform Monte Carlo to sample from the posterior distribution; for others, it may be necessary to use some form of Markov Chain Monte Carlo. Markov Chain Monte Carlo is more complex to implement but has the same fundamental benefit that uncertainty can be represented by a large sample of possible values for the statistical model parameter(s).

Applicability in areas relevant for EFSA

It is applicable to any area where a statistical model with uncertain parameters is used as a model of variability. However, training in Bayesian statistics is not yet part of the standard training of scientists and so it will often be the case that some specialist assistance will be needed, for example from a statistician.

EFSA Scientific Opinion and guidance documents have proposed the use of Bayesian methods for specific problems (EFSA 2006, EFSA 2012, and EFSA 2015). They have also been applied in EFSA internal and external scientific reports (EFSA 2009, Hald et al 2012). However, at present they are not widely used by EFSA.

The use of Bayesian methods has been proposed in many scientific articles concerning risk assessment in general and also those addressing particular applications. They have been adopted by some organisations for particular applications. For example, Bayesian methods have been used in microbial risk assessment by RIVM (Netherlands), USDA (USA) and IFR (UK) (Teunis and Havelaar 2000). Bayesian methods are also widely used in epidemiology and clinical studies which are fields with close links to risk assessment (e.g. Teunis et al. 2008).

Potential contribution to the main steps of uncertainty analysis

Steps in uncertainty analysis	Potential contribution of this approach
Identifying uncertainties	Not applicable.
Describing uncertainties	Not applicable.
Assessing the magnitude of individual uncertainties	Yes. For each source, uncertainty is expressed as a probability distribution. Where there is dependence between uncertainties about two or more parameters, the joint uncertainty is expressed using a multivariate probability distribution.
Assessing the combined impact of multiple uncertainties on the assessment output, taking account of dependencies	Not applicable. However, the results of EKE and/or Bayesian inferences for multiple uncertainties may be combined using the mathematics of probability. This is seen by some as being part of an overarching Bayesian approach to uncertainty.
Assessing the contribution of individual uncertainties to overall uncertainty	Not applicable. However, there exist methods of sensitivity analysis which are proposed from a Bayesian perspective and which are seen by some as being particularly appropriate for use in conjunction with Bayesian inference.

Melamine example

Bayesian inference will be illustrated by application to uncertainty about two of the sources of variability considered in the version of the melamine example which considers uncertainty about variability of exposure. Further supporting details about both versions of the melamine example may be found in Annex C. The variables considered here are body-weight and consumption in a day.

Data for both variables for children aged from 1 up to 2 years old were obtained from EFSA. Annex C gives details of the data and some data analysis supporting the choice of distribution family for each variable. The variables are treated as independent in what follows and the reasoning for doing so is included in Annex C.

Both variables are considered in detail below because there are important differences between the models used. For body-weight, the model is mathematically tractable and it is straightforward to use ordinary Monte Carlo to obtain a sample from the posterior distribution of the distribution parameters whereas for consumption it is necessary to use Markov Chain Monte Carlo for the same purpose. Moreover, for body-weight the posterior uncertainty involves very little dependence between the distribution parameters whereas for consumption there is strong dependence.

<u>Body-weight (bw)</u>

For bw, the statistical model is that: (i) bw follows a log-normal distribution, so that log bw follows a normal distribution; (ii) the uncertain distribution parameters are the mean μ_{logbw} and standard deviation σ_{logbw} of the distribution of log bw (base 10); (iii) the data are a random sample from the distribution of bw for the population represented by the data.

In the absence of expert input, the widely accepted prior distribution, proposed by Jeffreys, representing prior lack of knowledge is used. That prior distribution has probability density function $p(\sigma_{\mathrm{logbw}}, \mu_{\mathrm{logbw}}) \propto 1/\sigma_{\mathrm{logbw}}$ (O'Hagan and Forster, 2004).

For this choice of statistical model and prior distribution, the posterior distribution is known exactly and depends only on the sample size n_{logbw}, sample mean $\bar{x}_{\mathrm{logbw}}$ and sample standard deviation s_{logbw} of the log bw data. Let $\tau_{\text{logbw}} = 1/\sigma^2_{\text{logbw}}$. Then the posterior distribution of τ_{logbw} is a Gamma distribution. The Gamma distribution has two parameters: a shape parameter which here takes the value $\frac{1}{2}(n_{\text{logbw}} - 1)$ and a rate parameter which here takes the value $\frac{1}{2}(n_{\text{logbw}} - 1)s^2_{\text{logbw}}$. Conditional on a given value for σ_{logbw}, the posterior distribution of μ_{logbw} is normal with mean $\bar{x}_{\text{logbw}}$ and standard deviation $\sigma_{\text{logbw}}/\sqrt{n_{\text{logbw}}}$. Note that the distribution of μ_{logbw} depends on the value of σ_{logbw}, i.e. uncertainty about the two distribution parameters includes some dependence so that the values which are most likely for one of the parameters depend on what value is being considered for the other parameter.

For the data being used, n_{logbw}=171, $\bar{x}_{\text{logbw}}$ =1.037 and s_{logbw}=0.060. The posterior probability density of σ_{logbw}, is shown in Figure B.12.1a and the conditional probability density of μ_{logbw} given σ_{logbw}, is shown in Figure B.12.1b. The dependence between the parameters cannot be observed here.

However, when using these distributions in the exposure assessment, it is convenient to take a Monte Carlo sample from the posterior distribution to represent the uncertainty about μ_{logbw} and σ_{logbw}. This can be done as follows:

- Sample the required number of values of τ_{logbw} from the gamma distribution with shape=(171-1)/2=85 and rate = 85*0.060^2=0.306.
- For each value of τ_{logbw} in the previous step, calculate the corresponding value for $\sigma_{\text{logbw}} = 1/\sqrt{\tau_{\text{logbw}}}$
- For each value of σ_{logbw}, sample a single value of $\mu_{\text{logbw}}$from the normal distribution with mean 1.037 and standard deviation $\sigma_{\text{logbw}}/\sqrt{171}$.

The result of taking such a Monte Carlo sample is shown in Figure B.12.2 with the original sample mean and standard deviation for log bw shown respectively as dashed grey vertical and horizontal lines. The dependence between the two parameters is just visible in Figure B.9.2 (the mean is more uncertain when the standard deviation is high) but is not strong because the number of data n_{logbw} is large. Note that this Monte Carlo sampling process can easily be carried out in any standard spreadsheet software, for example Microsoft Excel or LibreOffice Calc.

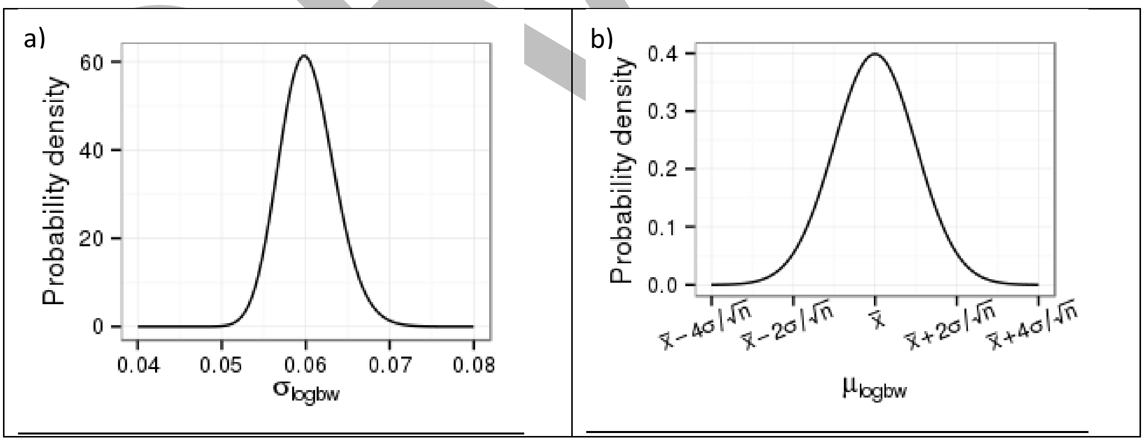


Figure B.12.1: Posterior distributions of parameters of log-normal distribution for body-weight of one-year-old children. The left panel shows the probability density for σ_{logbw}, the standard deviation of log bw. The panel on the right shows the conditional probability density for μ_{logbw}, the mean of log bw, given a value for the standard deviation σ_{logbw} .

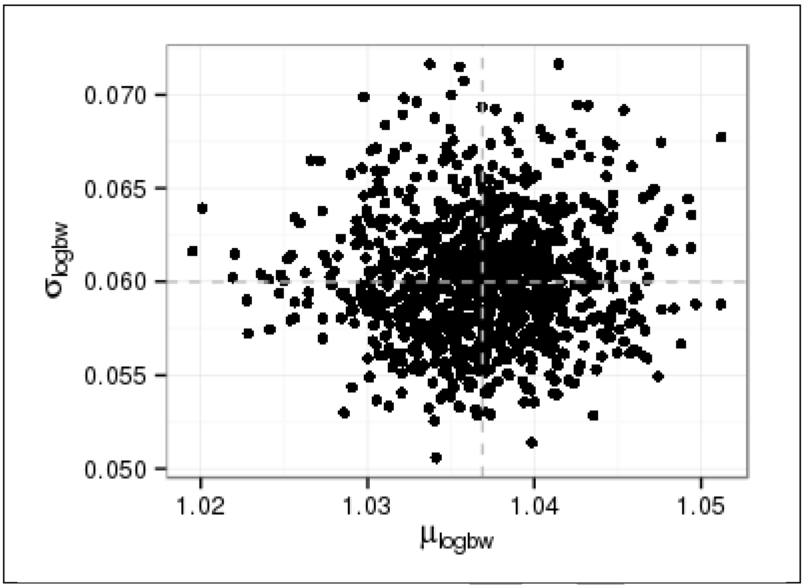


Figure B.12.2: Monte Carlo sample of 1000 values representing posterior uncertainty about $\boldsymbol{\sigma}_{\mathrm{logbw}}$ and $\boldsymbol{\mu}_{\mathrm{logbw}}$, given the data.

Consumption (q)

For q, the statistical model is that: (i) q follows a gamma distribution with uncertain distribution parameters being the shape α_{q} and rate β_{q}; (ii) the data are a random sample from the distribution of q.

Again, no expert judgements were provided with which to inform the choice of prior distribution for the parameters. Instead Jeffreys' general prior is used (O'Hagan and Forster 2004) which for this model has probability density function $p(\alpha_{\mathrm{q}}, \beta_{\mathrm{q}}) \propto \left(\sqrt{\alpha_{\mathrm{q}}\Psi(\alpha_{\mathrm{q}}) - 1}\right)/\beta_{\mathrm{q}}$.

For this model and choice of prior distribution, there is no simple mathematical representation of the posterior distribution. However, it is still quite possible to obtain a Monte Carlo sample from the posterior distribution by various methods. The results below were obtained using the Metropolis random walk version of Markov Chain Monte Carlo (Gelman et al, 2015) to sample from the posterior distribution of α_{q}. Values for the rate parameter β_{q} were directly sampled from the conditional distribution of β_{q} given α_{q}, for which there is a simple mathematical representation. Markov Chain Monte Carlo sampling of this kind is not easy to implement in a spreadsheet but takes only a few lines of code in software such as Matlab or R. This model is also easy to implement in software specializing in Bayesian inference, for example WinBUGS, OpenBUGS or JAGS

The results of taking a Monte Carlo sample representing uncertainty about the parameters are shown in Figure B.9.3a. This figure clearly shows the dependence between α_{q} and β_{q}. Figure B.9.3b shows the same uncertainty for the mean and coefficient of variation of the consumption distribution. The mean is $\alpha_{\mathrm{q}}/\beta_{\mathrm{q}}$ and the coefficient of variation is $1/\sqrt{\alpha_{\mathrm{q}}}$. Values for these alternative parameters can be computed directly from the values of α_{q} and β_{q} in the Monte Carlo sample. In figure B.9.3b, the mean and coefficient of variation of the data are shown respectively as dashed grey vertical and horizontal lines.

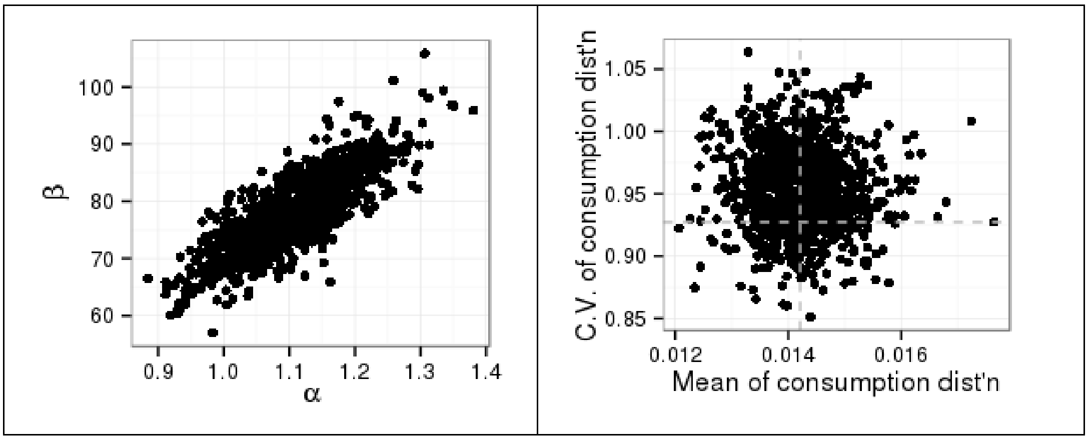


Figure B.12.3. Monte Carlo sample representing posterior uncertainty about parameters for the gamma distribution describing variability of consumption. The left panel shows uncertainty about the shape and rate parameters. The panel on the right shows uncertainty about the mean (kg/day) and coefficient of variation of the consumption distribution.

Strengths

1. Uncertainty about each parameter in a statistical model is quantified as a probability distribution for the possible values of the parameter. Therefore, the relative likelihood of different values of the parameter is quantified and this information can be taken into consideration by decision-makers. Probability distributions for multiple uncertainties may be combined using the laws of probability.
2. Dependence of uncertainty for one or more parameters is expressed using a multivariate probability distribution. This is the most complete and theoretically based treatment of dependence that is possible with methods available today.
3. The statistical uncertainty due to having a limited amount of data is fully quantified.
4. Knowledge/information about parameter values from sources other than the data being modelled can be incorporated in the prior distribution by using expert knowledge elicitation (EKE).
5. The output of a Bayesian inference is usually most easily obtained as a Monte Carlo sample of possible parameter values and is ideally suited as an input to a 2D Monte Carlo analysis of uncertainty.
6. Bayesian inference can be used with all parametric statistical models.

Weaknesses and possible approaches to reduce them

1. Bayesian inference is an unfamiliar form of statistical inference in the EFSA community and may require the assistance of a statistician. By introducing this method in training courses for statistical staff at EFSA this weakness can effectively be remediated.
2. When it is required to do so, obtaining a prior distribution by EKE (see sections B.8 and B.9) can require significant time and resources.
3. When the prior distribution is not obtained by EKE, one must find another way to choose it and for most models there is not a consensus about the best choice. However, there is a substantial literature and one can also investigate the sensitivity of the posterior distribution to the choice of prior distribution. Moreover, the influence of the choice of prior on the posterior distribution diminishes at larger sample sizes.
4. There is less software available than for other methods of statistical inference and there is less familiarity with the available software. Training in the use of software could be included in training on Bayesian inference.

5. As with other methodologies for statistical inference, an inappropriate choice of statistical model can undermine the resulting inferences. It is important to consider carefully the (sampling) process by which the data were obtained and to carry traditional statistical model validation activities such as investigation of goodness of fit and looking for influential data values.

Assessment against evaluation criteria

This method is assessed against the criteria in Table B.12.1. All entries in the "Time Needed" column have been highlighted because the time required for Bayesian inference is highly dependent on the complexity of the model.

Conclusions

1. The method is suitable for application across EFSA, subject only to availability of the necessary statistical expertise.
2. It can be used for quantification of parameter uncertainty in all parametric statistical models.
3. For all except the simplest models, incorporating expert judgments in prior distributions is likely to require the development of further guidance on expert knowledge elicitation (EKE).

References

Barker GC, Goméz-Tomé N (2013). A Risk Assessment Model for Enterotoxigenic Staphylococcus aureus in Pasteurized Milk: A Potential Route to Source-Level Inference. Risk Analysis 33(2): 249-269.

Berry DA (1995). Statistics: a Bayesian Perspective. Brooks/Cole.

EFSA (2006). Opinion of the Scientific Panel on Plant health, Plant protection products and their Residues on a request from EFSA related to the assessment of the acute and chronic risk to aquatic organisms with regard to the possibility of lowering the uncertainty factor if additional species were tested.

EFSA (2009). Meta-analysis of Dose-Effect Relationship of Cadmium for Benchmark Dose Evaluation. EFSA Scientific Report 254, 1-62

EFSA (2012). Guidance on the Use of Probabilistic Methodology for Modelling Dietary Exposure to Pesticide Residues. EFSA Journal 2012;10(10):2839

EFSA (2015). Scientific Opinion addressing the state of the science on risk assessment of plant protection products for non-target arthropods. EFSA Journal 2015;13(2):3996.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013). Bayesian Data Analysis, 3rd ed. Chapman & Hall/CRC.

Hald T, Pires SM, deKnegt L (2012). Development of a Salmonella source attribution model for evaluating targets in the turkey meat production. EFSA Supporting Publications 2012:EN-259.

Kruschke JK (2010). Doing Bayesian Data Analysis: A Tutorial with R and BUGS. Academic Press.

O'Hagan A, Forster JJ (2004). Kendall's advanced theory of statistics, volume 2B: Bayesian inference (Vol. 2). Arnold.

Teunis PFN, Havelaar AH (2000). The beta poisson dose-response model is not a single-hit model. Risk Analysis 20(3): 513-520.

Teunis PF, Ogden ID, Strachan NJ. Hierarchical dose response of E. coli O157:H7 from human outbreaks incorporating heterogeneity in exposure. Epidemiol Infect. 2008;136:761-70.

Williams MS, Ebel ED, Vose D (2011). Framework for microbial food-safety risk assessments amenable to Bayesian modeling. Risk Anal. 2011 31(4):548-65.

Table B.12.1. Assessment of Bayesian inference (when applied well) against evaluation criteria.

Criteria	Evidence of current acceptance	Expertise needed to conduct	Time needed	Theoretical basis	Degree/ extent of subjectivity	Method of propagation	Treatment of uncertainty and variability	Meaning of output	Transparency and reproducibility	Ease of understanding for non-specialist
Stronger character-istics	International guidelines or standard scientific method	No specialist knowledge required	Hours	Well established, coherent basis for all aspects	Judgement used only to choose method of analysis	Calculation based on appropriate theory	Different types of uncert. & var. quantified separately	Range and probability of alternative outcomes	All aspects of process and reasoning fully documented	All aspects fully understandable
	EU level guidelines or widespread in practice	Can be used with guidelines or literature	Days	Most but not all aspects supported by theory	Combination of data and expert judgment	Formal expert judgment	Uncertainty and variability quantified separately	Range and relative possibility of outcomes	Most aspects of process and reasoning well documented	Outputs and most of process understandable
	National guidelines, or well established in practice or literature	Training course needed	Weeks	Some aspects supported by theory	Expert judgment on defined quantitative scales	Informal expert judgment	Uncertainty and variability distinguished qualitatively	Range of outcomes but no weighting	Process well documented but limited explanation of reasoning	Outputs and principles of process understandable
	Some publications and/or regulatory practice	Substantial expertise or experience needed	A few months	Limited theoretical basis	Expert judgment on defined ordinal scales	Calculation or matrices without theoretical basis		Quantitative measure of degree of uncertainty	Limited explanation of process and/or basis for conclusions	Outputs understandable but not process
Weaker character-istics	Newly developed	Professional statistician needed	Many months	Pragmatic approach without theoretical basis	Verbal description, no defined scale	No propagation	No distinction between variability and uncertainty	Ordinal scale or narrative description for degree of uncertainty	No explanation of process or basis for conclusions	Process and outputs only understandable for specialists

B.13 Probability bound analysis

Purpose, origin and principal features

Probability bounds analysis provides a way of computing a bound (an upper or lower limit) on a probability relating to a combination of uncertainties. This allows the use of probability to quantify uncertainty while at the same time allowing assessors to make limited probability statements rather than having to specify full probability distributions. The simplest useful form of probability statement is to specify an upper or lower bound on the probability that a parameter exceeds some specified level. From limited probability statements for individual uncertainties, probability bounds analysis applies the laws of probability to make probability statements about the combined uncertainty. It is also in principle possible to incorporate bounds on dependence between uncertainties.

There is a long history in the theory of probability concerning methods for this kind of problem. It first appears in Boole (1854). A modern account of more complex approaches in the context of risk assessment is given by Tucker and Ferson (2003).

It is a generalisation of the interval analysis method (section B.7) but has the specific advantage that it incorporates some probability judgements and produces a limited form of probabilistic output. The key advantage compared to Monte Carlo (section B.14) is that experts do not have to specify complete probability judgements; the least they must provide is an upper bound on the probability of exceeding (or falling below) some threshold for each source of uncertainty. A second advantage is that no assumptions are made about dependencies unless statements about dependence are specifically included in the calculation.

There are many possible ways in which it might be applied. The examples below show minimalist versions, based on the Frechet (1935, 1951) inequalities, for problems involving only uncertainty and problems involving both uncertainty and variability.

The simplest version allows one to place an upper bound on the probability that a calculated quantity, which depends on individual components, exceeds a specified value. In order to apply the simplest version: (i) the calculated quantity must increase as each component increases and; (ii) a value must be specified for each component, together with an upper limit on the probability that the component exceeds that value.

Applicability in areas relevant for EFSA

Potentially applicable to all areas of EFSA's work but most obviously advantageous for assessments (or parts of assessments) for which probabilistic methods are considered to be too challenging.

It is not known to have been used by EFSA. Examples of use outside EFSA in risk assessment include Dixon (2007) and Regan et al (2002).

Potential contribution to the main steps of uncertainty analysis

Steps in uncertainty analysis	Potential contribution of this approach
Identifying uncertainties	Not applicable.
Describing uncertainties	Not applicable.
Assessing the magnitude of individual uncertainties	Not applicable.
Assessing the combined impact of multiple uncertainties on the assessment output, taking account of dependencies	Yes. However, simple versions do not involve quantification of dependencies but do allow for their possible existence in computing the bound on the combined impact.

Assessing the contribution of individual uncertainties to overall uncertainty	Not applicable.

Melamine example

In normal practice, the limited probability statements required for probability bounds analysis would be obtained in most cases by expert knowledge elicitation (Sections B.8 and B.9). However, for the purpose of illustrating calculations based on probability bounds in the examples which follow, values specified for parameters, and bounds on probabilities of exceeding those values, were obtained from probability distributions used for Monte Carlo analyses (section B.14).

The melamine example (details in Annex C) has two versions: a worst-case assessment and an assessment of uncertainty about variability. Both are considered below but require different approaches as only the second version directly involves variability.

Worst-case exposure

The focus of this example is to make a limited probability statement about worst-case exposure for children aged 1 up to 2 years, based on limited probability statements about individual parameters.

When increased, each of the following parameters increases the worst-case exposure: $c_{\max}$, $w_{\max}$, $q_{\max}$. When decreased, $bw_{\min}$ increases the worst-case exposure and so increasing $1/bw_{\min}$ increases the worst-case exposure

The following table shows a limited probability statement for each of the input parameters. The statements were derived from distributions used in sections B.8 and B.9 but it is likely that expert knowledge elicitation would be used in many cases in real assessments.

Parameter	Specified value	Probability parameter exceeds specified value
c_{max}	3750 mg/kg	$\leq$ 3.5%
w_{max}	0.295	$\leq$2%
q_{max}	0.095 kg	$\leq$2.5%
$1/bw_{min}$	1/(5.6 kg)	$\leq$2%

Note that the judgement for $\frac{1}{bw_{\min}}$ was actually arrived by considering the probability that $bw_{\min} \leq 5.6\text{kg}$.

The value being considered for $e_{\max}$ can then simply be calculated from the specified values for individual parameters which increase exposure: 3750*0.295*0.095/5.6 = 18.8

Based on the judgments in the preceding table, the laws of probability then imply that the probability that $e_{\max}$ exceeds 18.8 is less than (3.5+2+2.5+2)% = 10%. This is the simplest form of probability bounds analysis. No simulations are required.

As indicated earlier, the values specified for parameters and bounds on probabilities of exceeding those were obtained for illustrative purposes from the distributions used to represent in sections B.8 and B.9. If the method were being applied using expert judgements about the parameters we would be likely to end up with simpler probability values such as <=10%, <=5% or <=1% and the values specified for parameters would also be different having been specified directly by the experts. The method of computation would remain the same.

Uncertainty about variability of exposure

When variability is involved, the simplest approach to applying probability bounds analysis is to decide which percentile of the output variable will be of interest. The probability bounds method can then be applied twice in order to make an assessment of uncertainty about variability: once to variability and then a second time to uncertainty about particular percentiles.

For illustrative purposes, assessment will be made of uncertainty about the 95th percentile of exposure: e_{95}. In order to apply probability bounds analysis, for each input parameter a percentile needs to be chosen on which to focus. For illustrative purposes, it was decided to focus on the 98th percentile of variability of concentration, denoted c_{98}, and the 99th percentile of variability of each of the other input parameters which increase the exposure when increased: w_{99}, q_{99} and $(1/bw)_{99}$. Note that $(1/bw)_{99} = bw_{01}$.

Applying probability analysis first to variability, the laws of probability imply that

$$e_{95} \geq c_{98} \times w_{99} \times q_{99} \times (1/bw)_{99} = c_{98} \times w_{99} \times q_{99} / \, bw_{01}$$

where 95 is obtained as

$$95 = 100 - [(100 - 98) + (100 - 99) + (100 - 99) + (100 - 99)]$$

The following table shows a limited probability statement of uncertainty about the chosen percentile for each of the input variables. As before, the statements were derived from distributions used in sections B.8 and B.9 but it is likely that expert knowledge elicitation would be used in many cases in real assessments.

Parameter	Specified value	Probability parameter exceeds value specified
c_{98}	4400mg/kg	≤2.5%
w_{99}	0.295	≤2.5%
q_{99}	0.075kg	≤2.5%
$(1/bw)_{99}$	1/(7kg)	≤2.5%

Computing exposure using the values specified for the input parameters s leads to the following value to be considered for exposure: 4400*0.295*0.075/7=13.9. From this, by the same calculation as for worst-case example, the laws of probability imply that the probability that $c_{98} \times w_{99} \times q_{99} / \, bw_{01}$ exceeds 13.9 is less than 2.5%+2.5%+2.5%+2.5%=10%.

Since $e_{95} \geq c_{98} \times w_{99} \times q_{99} / \, bw_{01}$, the probability that e_{95}exceeds 13.9 is also less than 10%.

Various choices were made here:

- The choice of percentiles could have been made differently. It was assumed for illustrative purposes that the 95th percentile of exposure is of interest, although other percentiles could equally be considered. Given the focus on the 95th percentile, percentiles for the individual components were chosen so that the total variability not covered by them was less than or equal to 5%. Because there is reason to believe that the greatest source of variability is concentration, a lower percentile was chosen for concentration than for the other three parameters.
- Values specified for the percentiles of input parameters and probabilities of exceeding those values were obtained from the distributions used for the 2D Monte Carlo example in sections B.8 and B.9. The total limit of the exceedance probability was chosen to be 10% and this was divided equally between the 4 parameters to illustrate the calculation. Any other division would have been valid and would have led to different values for the parameters.
- If expert knowledge elicitation were used instead to make a limited probability statement about each of the 4 percentiles, it is likely that simpler probability values such as ≤10%, ≤5% or ≤1% would have resulted, and the values specified for the percentiles would therefore also be different having been specified directly by the experts. The method of computation would remain the same.

Strengths

1. Simple version provides an easily calculated bound on the probability that a calculated parameter exceeds a specified value. The method applies when a limited probability statement has been made about each input parameter.
2. Requires only limited probability judgements from experts. This greatly reduces the burden of elicitation compared to fully probabilistic methods.
3. Simple version makes no assumption about dependence between components of either uncertainty or variability.
4. More complex versions can exploit more detailed probability judgements and/or statements about dependence of judgements.

Weaknesses and possible approaches to reduce them

1. For the simple version, the calculated bound will be larger and may be much larger than would be obtained by a more refined probabilistic assessment. Nevertheless, it may sometimes be sufficient for decision-making, and can indicate whether a more refined probabilistic assessment is needed.
2. Provides only a limited quantification of uncertainty about the calculated value. Nevertheless, that may sometimes be sufficient for decision-making,
3. More complex versions involve more complex calculations and it is likely that professional mathematical/statistical advice would be needed.

Assessment against evaluation criteria

This method is assessed against the criteria in Table B.13.1. In evaluating time needed, only the simple form of probability bounds analysis was considered, as used in the two examples for melamine. Time needed to conduct EKE is not included.

Conclusions

1. This is potentially an important tool for EFSA as it provides a way to incorporate probabilistic judgements without requiring the specification of full probability distributions and without making assumptions about dependence.. In so doing, it provides a bridge between interval analysis and Monte Carlo. It allows the consideration of less extreme cases than interval analysis and involves less work than full EKE for distributions followed by Monte Carlo.
2. Judgements and concept are rather similar to what EFSA experts do already when using assessment factors and conservative assumptions. Probability bounds analysis provides a transparent and mathematically rigorous calculation which results in an unambiguous quantitative probability statement for the output.

References

Boole G (1854). An Investigation of the Laws of Thought on which are Founded the Mathematical Theories of Logic and Probabilities. Walton and Maberley.

Dixon WJ. (2007). The use of Probability Bounds Analysis for Characterising and Propagating Uncertainty in Species Sensitivity Distributions. Technical Report Series No. 163, Arthur

Rylah Institute for Environmental Research, Department of Sustainability and Environment. Heidelberg, Victoria, Australia

Fréchet M (1935). Généralisations du théorème des probabilités totales. Fundamenta Mathematica 25: 379–387.

Fréchet M (1951). Sur les tableaux de corrélation dont les marges sont données. Annales de l'Université de Lyon. Section A: Sciences mathématiques et astronomie 9: 53–77.

Regan, H.M., B.K. Hope, and S. Ferson (2002). Analysis and portrayal of uncertainty in a food web exposure model. Human and Ecological Risk Assessment 8: 1757–1777

Tucker WT, Ferson S (2003). Probability bounds analysis in environmental risk assessment. Applied Biomathematics, Setauket, New York, http://www. ramas. com/pbawhite. pdf.

5602 **Table B.13.1.** Assessment of Probability bound analysis (when applied well) against evaluation criteria.

Criteria	Evidence of current acceptance	Expertise needed to conduct	Time needed	Theoretical basis	Degree/ extent of subjectivity	Method of propagation	Treatment of uncertainty and variability	Meaning of output	Transparency and reproducibility	Ease of understanding for non-specialist
Stronger character-istics	International guidelines or standard scientific method	No specialist knowledge required	Hours	Well established, coherent basis for all aspects	Judgement used only to choose method of analysis	Calculation based on appropriate theory	Different types of uncert. & var. quantified separately	Range and probability of alternative outcomes	All aspects of process and reasoning fully documented	All aspects fully understandable
	EU level guidelines or widespread in practice	Can be used with guidelines or literature	Days	Most but not all aspects supported by theory	Combination of data and expert judgment	Formal expert judgment	Uncertainty and variability quantified separately	Range and relative possibility of outcomes	Most aspects of process and reasoning well documented	Outputs and most of process understandable
	National guidelines, or well established in practice or literature	Training course needed	Weeks	Some aspects supported by theory	Expert judgment on defined quantitative scales	Informal expert judgment	Uncertainty and variability distinguished qualitatively	Range of outcomes but no weighting	Process well documented but limited explanation of reasoning	Outputs and principles of process understandable
	Some publications and/or regulatory practice	Substantial expertise or experience needed	A few months	Limited theoretical basis	Expert judgment on defined ordinal scales	Calculation or matrices without theoretical basis		Quantitative measure of degree of uncertainty	Limited explanation of process and/or basis for conclusions	Outputs understandable but not process
Weaker character-istics	Newly developed	Professional statistician needed	Many months	Pragmatic approach without theoretical basis	Verbal description, no defined scale	No propagation	No distinction between variability and uncertainty	Ordinal scale or narrative description for degree of uncertainty	No explanation of process or basis for conclusions	Process and outputs only understandable for specialists

B.14 Monte Carlo simulation (1D-MC and 2D-MC)

Purpose, origin and principal features

In the context of assessing uncertainty, Monte Carlo (MC) is primarily a computational tool for (i) calculations with probability distributions representing uncertainty and/or variability and (ii) those methods of sensitivity analysis (section B.16) which require sampling random values for parameters. In the case of (i), it provides a means to compute the combined effect of several sources of uncertainty, each expressed as a probability distribution, providing a probability distribution representing uncertainty about an assessment output. MC software often also provides modelling tools.

Monte Carlo simulation was developed in the 1940s, primarily by Stanislav Ulam in collaboration with Nicholas Metropolis and John von Neumann in the context of the Manhattan project to develop atomic bombs, and first published in 1949 (Ferson, 1996). Currently, the method is widely applied in science, finance, engineering, economics, decision analysis and other fields where random processes need to be evaluated. Many papers have been written about the history of MC simulation, the reader is referred to Bier and Lin (2013) and Burmaster and Anderson (1994).

In a MC simulation model, variable and/or uncertain parameters are represented by probability distributions. Those probability distributions are the ”input parameters” to a MC calculation. The model is recalculated many times, each time taking a random value for each parameter from its distribution, to produce numerous scenarios or iterations. Each set of model results or ”outputs” from single iteration represents a scenario that could occur. The joint distribution of output parameters, across all the iterations, is a representation of the variability and/or uncertainty in the outputs.

Risk assessment models may include parameters that are correlated in some way. For example, the food consumption of a child will typically be less than that of an adult. Therefore, food consumption estimates are correlated with age and body weight. A cardinal rule to constructing a valid model is that “Each iteration of a risk analysis model must be a scenario that can physically occur” (Vose, 2008, p. 63). If samples are drawn independently for two or more parameters in an MC model, when in fact there should be dependence this may result in selecting combinations that are not plausible. Ferson (1996) argues that the risk to exceed a particular threshold concentration depends strongly on the presence or absence of dependencies between model parameters. If there are positive correlations, the exceedance risk may be underestimated whereas negative correlations may lead to overestimation. Burmaster and Anderson (1994) suggest to consider correlations with a Pearson product-moment correlation coefficient with magnitude >= 0.6. A simple approach to addressing dependence is to stratify the population into subgroups within which the inputs can be assumed not to be strongly correlated, but this may result in *ad-hoc* solutions and tedious calculations. Different software packages offer different approaches to including correlations such as by specifying a correlation coefficient. However, even then only a small space of possible dependencies between the two variables may be sampled (US EPA, 1997). More advanced approaches include the use of copulas to specify the joint probability distribution of model inputs.

For assessments in which variability is not considered directly, for example worst-case assessments, MC can be used with all input distributions being representations of uncertainty. The MC output distribution will then also be a representation of uncertainty. However, for assessments involving variability and uncertainty about variability (see Chapter 6.2), it is important to differentiate between variable and uncertain factors when building MC models, in order to allow a more informative interpretation of the output distributions. Two-dimensional

Monte Carlo (2D MC) simulation was proposed by Frey (1992) as a way to construct MC models taking this separation into account. First, input parameters are assigned to be either variable or uncertain. Uncertainty about variability can then be represented using a nested approach in which the distribution parameters, of probability distributions representing variability of input parameters, are themselves assigned probability distributions representing uncertainty. For example, a dose-response model may be fitted to a dataset involving a limited number of individuals, and the uncertainty of the fitted dose-response model might be represented by a sample from the joint distribution representing uncertainty about the dose-response parameters. The simulation model is then constructed in two loops. In each iteration of the outer loop, a value is sampled for each uncertain parameter, including distribution parameters. The inner loop samples a value for each variable parameter and is evaluated as a standard MC model, using the values sampled for distribution parameters in the outer loop to determine the probability distribution to use for each variable. This process will generate one possible realisation of all output values. The simulation is then repeated numerous times, usually repeating the inner loop many times per outer loop iteration. The outer loop iterations provide a sample of values for all uncertain parameters. For each outer loop iteration, the inner loop iterations provide a sample of values for variable parameters. In combination, they generate numerous possible realisations of all output distributions.

The results of a 2D MC model can be shown graphically as “spaghetti plots ”, in which probability density functions (PDFs) or cumulative density functions (CDFs) of all simulated variability distributions of an input or output parameter are plotted together. The spread in these distributions demonstrates the impact of uncertainty on the model results. Other commonly used outputs are probability bands (e.g. the median CDF and surrounding uncertainty intervals, see melamine example) or a combination of line- and box-plots.

Software for MC simulation is commercially available as add-ins to Excel such as @RISK, Crystal Ball, and ModelRisk; and dedicated software such as Analytica. MC modeling can also be done in statistical software such as R, especially the distrfit and mc2d packages which support 2D MC (Pouillot and Delignette-Muller, 2010), or SAS or mathematical software such as Mathematica or Matlab.

Applicability in areas relevant for EFSA

MC simulation models are used in many domains of risk assessment including food safety. In EFSA, they are widely used in the area of microbial risk assessment and there is an EFSA guidance document on their application to pesticide exposure assessment, which includes use of 2D MC (EFSA, 2012).

Specific software applications are available to support MC modeling in different domains relevant for EFSA. These include FDA-iRISK, sQMRA and MicroHibro for microbial risk assessment (reviewed in EFSA, 2015), MCRA and Creme for human dietary exposure to chemicals, and Webfram for some aspects of environmental risk of pesticides.

The BIOHAZ Panel has commissioned several outsourced projects to develop complex models including Salmonella in pork (Hill et al, 2011) and BSE prions in bovine intestines and mesentery (EFSA, 2014). The importance of 2D simulation was underlined, for example by Nauta (2011) who demonstrated that a simple model for the growth of *Bacillus cereus* in pasteurised milk without separation of uncertainty and variability may predict the (average) risk to a random individual in an exposed population. By separating variability and uncertainty, the risk of an outbreak can also be identified, as cases do not occur randomly in the population but are clustered because growth will be particularly high in certain containers of milk.

Pesticide intake rate for certain bee species was modelled by EFSA's PRAS Unit using MC simulation techniques. The 90th percentile of the residue intake rate and its 95% confidence interval were derived from the empirical joint distribution of the feed consumption and residue level in pollen and nectar.

Trudel et al. (2011) developed a 2D MC model to investigate whether enhancing the data sets for chemical concentrations would reduce uncertainty in the exposure assessment for the Irish population to polybrominated diphenyl ethers and concluded that "by considering uncertainty and variability in concentration data, margins of safety (MOS) were derived that were lower by a factor of 2 compared to MOS based on dose estimates that only consider variability". Based on the simulation results, they also suggested that "the datasets contained little uncertainty, and additional measurements would not significantly improve the quality of the dose estimates".

MC models are used by FAO/WHO committees supporting the work of the Codex Alimentarius Commision (JECFA, JMPR, JEMRA), as well as by national risk assessment agencies (RIVM, BfR, ANSES, and others). They are commonly used for exposure assessment in chemical risk assessment (US FDA), but not yet common in toxicology. In the USA, an interagency guideline document (USDA/FDIS and US EPA 2012) for microbial risk assessment features MC models prominently for exposure assessment and risk characterization.

There are many guidelines and books that provide detailed instructions on how to set up MC simulation models. Burmaster and Anderson (1994), Cullen and Frey (1999) and Vose (2008) all have an emphasis on the risk assessment domain. USEPA (1997) have published Guiding Principles on the use of MC analysis, which are very relevant to applications in EFSA.

Potential contribution to the main steps of uncertainty analysis

Steps in uncertainty analysis	**Potential contribution of this approach**
Identifying uncertainties	Not applicable
Describing uncertainties	Not applicable
Assessing the magnitude of uncertainties	Not applicable (required as input).
Assessing the combined impact of multiple uncertainties on the assessment output, taking account of dependencies	Yes, rigorous quantification of the impact of quantified input uncertainties on the output uncertainty, subject to model assumptions
Assessing the contribution of individual uncertainties to overall uncertainty	Yes, rigorous quantification of the contribution of individual uncertainties to overall combined uncertainty

Melamine example

Two examples are presented of the use of MC for assessment of uncertainty. The first illustrates how ordinary (1D) MC may be used, for assessments where variability is not modeled, to calculate uncertainty about assessment outputs based on probability distributions representing uncertainty about input parameters. It assesses uncertainty about the worst-case exposure for children aged from 1 up to 2 years. The second example illustrates how 2D MC may be used as a tool in assessing uncertainty about variability in assessments where that is an issue. It considers uncertainty about variability of exposure for those children in the same age group who consume contaminated chocolate from China.

Details of the models used may be found in annex C together with details and some analysis of data which were the basis for some distributions used in the 2D example.

<u>Worst-case assessment</u>

For simplicity, this example focuses only on selected uncertainties affecting the estimate of worst-case exposure for children aged from 1 up to 2 years. In particular, any uncertainties affecting the TDI are not considered. An overall characterization of uncertainty would need to include these and additional uncertainties affecting exposure. Distributions used to represent uncertainty about parameters were not obtained by careful elicitation of judgements from relevant experts. Rather, they are provided so that the MC calculations and output can be illustrated. Consequently, only a limited amount of reasoning is provided as it is likely that a real assessment would make different choices.

The worst-case exposure is obtained by

$$e_{max} = \frac{c_{max} \times w_{max} \times q_{max}}{bw_{min}}$$

and the worst-case risk ratio is then $r_{max} = e_{max}/\mathrm{TDI}$.

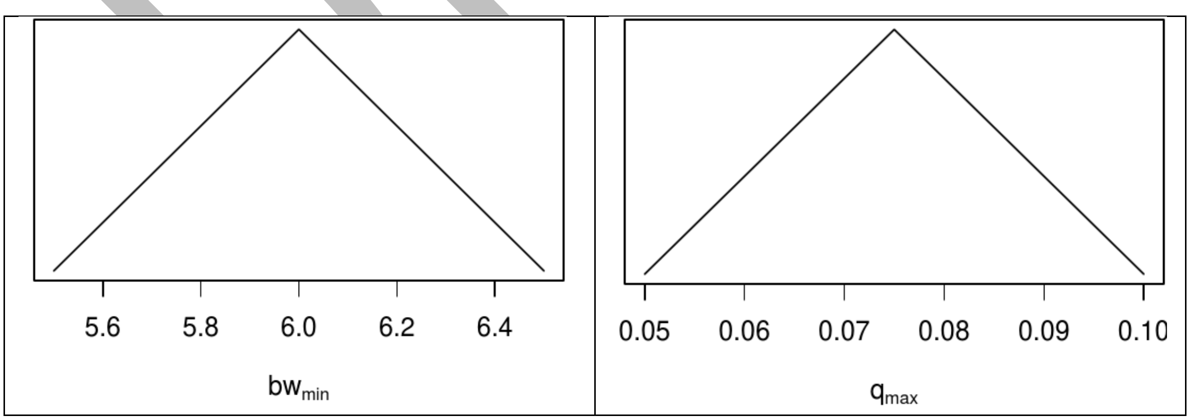
To build a MC model, a distribution must be provided for each uncertain input parameter. The distributions used for this example are shown in Figure B.14.1. For each parameter, the distribution is over the range of values used for the parameter in the final table of the Interval Analysis (section B.7) example.

The triangular distribution with 5.5 and 6.5 as endpoints and peak at 6 was selected to represent uncertainty about bw_{min}.

The triangular distribution with 0.05 and 0.10 as the endpoints and with peak at 0.075 was selected to represent uncertainty about q_{max}.

For uncertainty about w_{max}, the distribution obtained in the hypothetical example of expert knowledge elicitation example (sections B.8 and B.9) was used.

For uncertainty about c_{max}, a beta distribution was selected. Like the triangular distribution family, the beta distribution family only assigns non-zero probability to a finite range of values. However, it has the additional possibility for the probability density function to descend more quickly to zero near the end-points. This was felt to be particularly desirable for the upper endpoint since there would actually be no milk in the dried matter at that endpoint and so such values would be very unlikely.



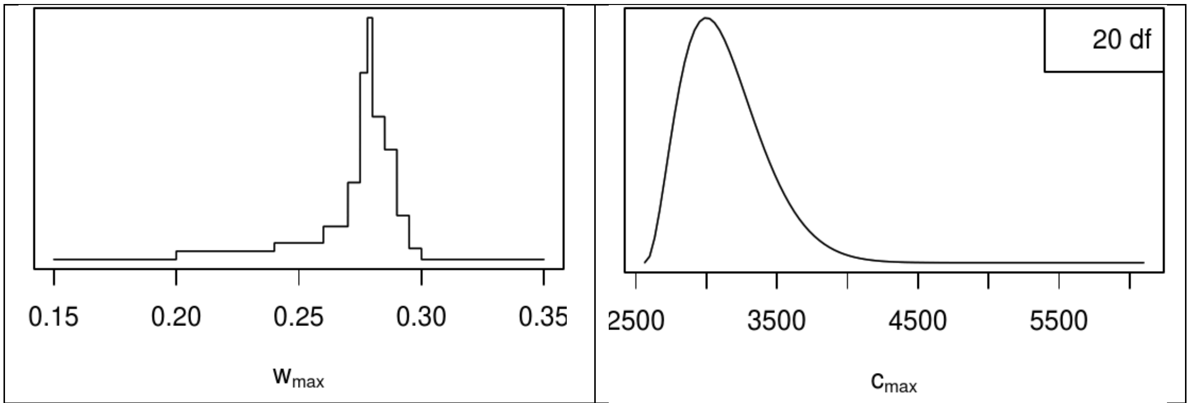


Figure B.14.1: Distributions used to represent uncertainty about input parameters in worst-case exposure assessment for children aged from 1 up to 2 years.

The MC model was built in R version 3.1.2 (R Core team, 2014), using the package mc2d (Pouillot and Delignette-Muller, 2010).

The output of the MC model is a distribution, shown in Figure B.14.2, representing uncertainty about e_{max}. The output is calculated from the distributions selected to represent uncertainty about input parameters. Table B.14.1 summarises the output and compares it to the TDI. The benefit of carrying out a MC analysis is that there is a full distribution representing uncertainty. This provides greater detail than other methods.

Table B.14.1: Uncertainty, calculated by MC, about the worst case exposure and ratio to TDI for children aged from 1 up to 2 years.

		Worst case exposure (e_{max})	**Risk ratio (r) (e_{max}/TDI)**
Summary of uncertainty distribution	Median	10.6	21.2
	Mean	10.7	21.4
	2.5%-ile	7.7	14.3
	97.5%-ile	14.8	29.5

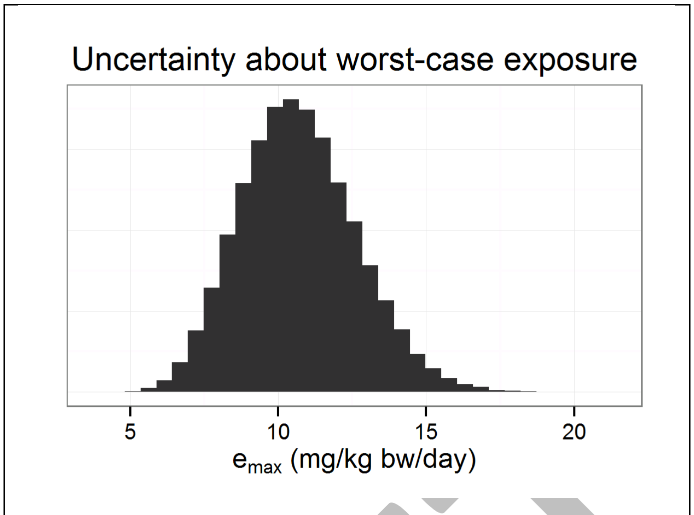


Figure B.14.2. Uncertainty, calculated by MC, about worst-case exposure for children aged from 1 up to 2 years.

Uncertainty about variability of exposure

For simplicity, this example focuses only on selected uncertainties affecting the estimate of worst-case exposure for children aged from 1 up to 2 years who consume contaminated chocolate from China. In particular, no consideration is given to (i) uncertainties affecting the TDI; (ii) uncertainties about the relevance of data used; (iii) uncertainties about distribution family choices. An overall characterization of uncertainty would need to include these and any other additional uncertainties. Distributions used to represent uncertainty about parameters are not considered to be the best possible choices. Rather, they are provided so that the MC calculations and output can be illustrated. Consequently, only a limited amount of reasoning is provided as it is likely that a real assessment would make different choices.

The assessment model (further details in annex C), in which all inputs are variable, is

$$e = \frac{c \times w \times q}{bw}$$

To carry out a 2D MC simulation for this model, it is necessary first, for each input, to choose a suitable distribution to model variability. The approach taken here is to choose a parametric distribution family for each input. It would also be possible to proceed non-parametrically if suitable data were to be available for a variable; in that situation, uncertainty about variability might be addressed by using the bootstrap (section B.8).

Table B.14.2. Summary of distribution families used to model variability of input parameters and of distributions used to represent uncertainty about variability distribution parameters.

Parameter	Model for variability (distribution family)	Uncertainty about distribution parameters
Body-weight (bw, kg)	Log-normal (restricted to a minimum of 5.5kg)	Posterior distribution from Bayesian inference (section B.12) applied to data described in annex C. See example in section B.12
Consumption (q, kg/day)	Gamma (restricted to a maximum of at 0.1kg)	Posterior distribution from Bayesian inference (section B.12) applied to data described in annex C. See example in section B.12
Concentration (c, mg/kg)	Log-normal (restricted to a maximum of	Median fixed at 29mg/kg. Beta(22,1) distribution used to represent uncertainty about percentile to which maximum

	6100mg/kg)	data value 2563mg/kg corresponds.
Weight-fraction (w, -)	Uniform	Lower end of uniform distribution fixed at 0.14. Uncertainty about upper end represented by distribution for w_{max} used in the worst-case example above.

The distribution family choices are shown in the second column of Table B.14.2. For body-weight (bw) and consumption (q), they were based on analysis of data described in annex C. For concentration (c) and weight-fraction (w), they are purely illustrative. The restrictions applied to the range of variability of bw, q and c derive from the worst-case limits used in the Interval Analysis example (section B.7) .

Having chosen distribution families to represent variability, the next step is to specify distributions representing uncertainty about distribution parameters and to decide how to sample from them. The choices made are summarized in the third column of Table B.14.2 and some further details follow.

1. The EFSA statement refers to data on concentrations in infant formula. Those data were not obtained by random sampling and only summaries are available. The median of those data was 29mg/kg and the maximum value observed was 2563mg/kg. In the 2D MC model, the median of the log-normal distribution for concentrations was taken to be 29 mg/kg. In reality, the median concentration is uncertain and so this choice introduces an additional uncertainty which is not addressed by the MC analysis. The percentile of the concentration distribution corresponding to the maximum data value of 2563 mg/kg is considered to be uncertain. Treating the maximum data value as having arisen from a random sample of size 22, both Bayesian and non-Bayesian arguments lead to a beta(22, 1) distribution for the percentile to which 2563 corresponds. When implementing 2D MC, a value is sampled from the beta distribution in each iteration of the outer loop; from that value, it is possible to calculate the standard deviation for the underlying normal distribution which would place 2563 at the specified percentile.
2. Sampling from the posterior distribution for the parameters of the log-normal distribution for body-weight was carried out by the MC method described in the example in section B.14.
3. Sampling from the posterior distribution for the parameters of the gamma distribution for consumption was carried out by Markov Chain MC as described in the example in section B.14.
4. Sampling from the distribution for w_{max} could be carried out several ways. The method used in producing the results shown below was to treat the distribution as a 12 component mixture of uniform distributions and to sample accordingly.

A by-product of the 2D MC calculation is that the samples can be used to summarise the input variables in various ways. For each variable, Table B.14.3 summarises uncertainty about 5 variability statistics: mean, standard deviation and 3 percentiles of variability. Uncertainty is summarized by showing the median estimate, the mean estimate and upper and lower 2.5th and 97.5th percentiles of uncertainty for each variability statistic. The two percentiles of uncertainty together make up a 95% uncertainty interval. For example, if one is interested in the mean body-weight of children aged 1 up to 2 years, the median estimate is 11.0kg and the 95% uncertainty interval is (10.8, 11.2)kg.

Table A14.3. Summaries, based on 2DMC output, of uncertainty about variability for each of the assessment inputs.

Parameter	Uncertainty	Variability				
		mean	st. dev.	2.5%	50%	97.5%
c (mg/kg)	50%	225.2	617	0.262	27.8	2059
	2.5%	83.7	198	0.002	14.9	509
	97.5%	377.3	947	1.629	29.9	3791
w (-)	50%	0.209	0.039	0.143	0.209	0.275
	2.5%	0.176	0.021	0.142	0.176	0.211
	97.5%	0.217	0.044	0.144	0.217	0.290
q (kg/day)	50%	0.014	0.013	0.00042	0.010	0.050
	2.5%	0.013	0.012	0.00031	0.0091	0.045
	97.5%	0.016	0.015	0.00069	0.0114	0.056
bw (kg)	50%	11.0	1.53	8.30	10.9	14.3
	2.5%	10.8	1.37	7.98	10.7	13.8
	97.5%	11.2	1.72	8.59	11.1	14.8

Turning to uncertainty about assessment outputs, the results of the 2D MC model are shown in Tables B.14.4 and B.14.5. Table B.14.4 shows summaries of uncertainty about 4 exposure variability statistics: the mean and three percentiles. For each variability statistic, the median estimate is shown along with two percentiles which together make up a 95% uncertainty interval. For example, for mean exposure, the median estimate is 0.0605 mg/kg bw/day and the 95% uncertainty interval ranges between 0.022 and 0.105 mg/kg bw/day. Table B.14.5 summarises uncertainty about the percentage of person-days for which exposure exceeds the TDI of 0.5mg/kg bw.

Table B.14.4: Summaries of uncertainty, based on 2DMC output, of uncertainty about variability of exposure for children aged from 1 up to 2 years.

Uncertainty	Variability			
	Mean	2.5%-ile	Median	97.5%-ile
Median	0.0605	2.0e-5	0.0045	0.527
2.5%-ile	0.0224	3.7e-7	0.0023	0.154
97.5%-ile	0.1052	9.0e-5	0.0054	1.037

Table B.14.5: Uncertainty, based on 2D MC output, about the percentage of child-days (1 year olds consuming contaminated chocolate from China) exceeding the TDI of 0.5mg/kg/day.

	Percentage of child-days exceeding TDI
Median estimate	2.7%
95% uncertainty interval	(0.4, 5.5)%

The results can also be presented graphically as a series of cumulative density functions. Figures B.14.3 and B.14.4 show uncertainty about variability of the risk ratio r. In these figures, the spread of the curve along the x-axis represents the variability dimension, whereas the spread along the y-axis (the grey-shaded areas) represents the uncertainty dimension. From these graphs, it is clear that, subject to the assumptions made in building the 2D MC model, there is major variability in the exposure to melamine, and hence in the risk ratio. The majority of 1 year old children consuming chocolate from China contaminated with melamine will be exposed to low levels but it is estimated that 2.7% (95% CI 0.4-5.5%) of those child-days have melamine exposure above TDI.

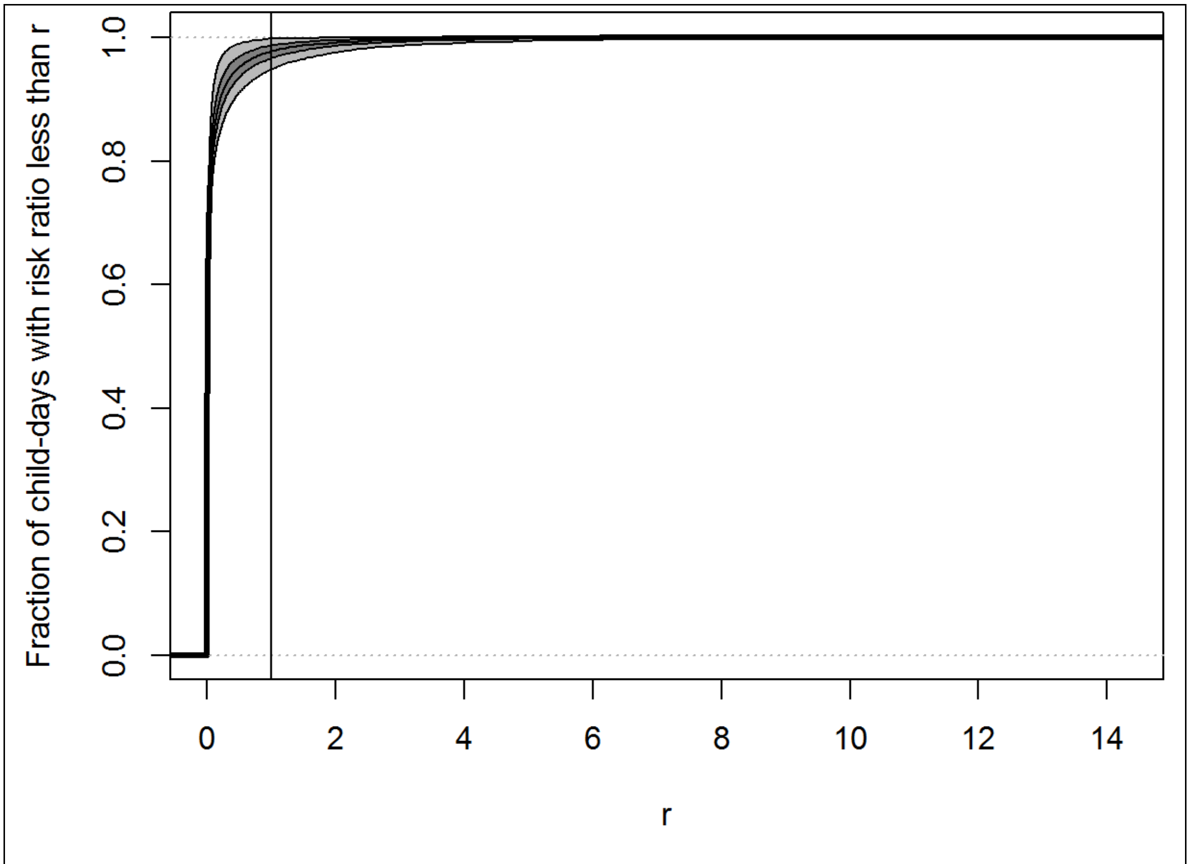


Figure B.14.3: Plot of estimated cumulative distribution of ratio of exposure to the TDI for melamine, for 1-year-olds consuming contaminated chocolate from China. Uncertainty about the cumulative distribution is indicated: the light grey band corresponds to 95% uncertainty range, and dark grey band corresponds to 50% uncertainty range.

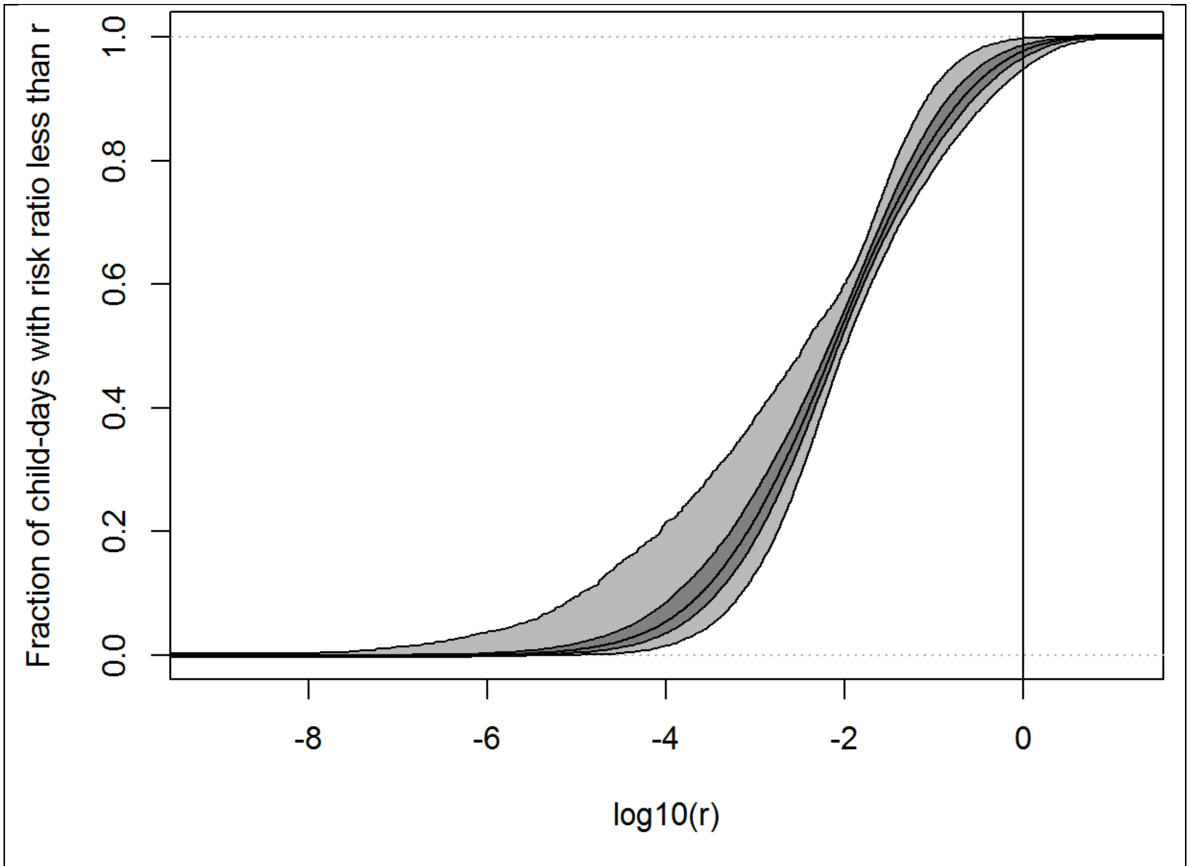


Figure 14.4: Plot, as in figure B.14.3 but with logarithmic scale for r, of cumulative distribution of ratio of exposure to the TDI for melamine, for 1-year-olds consuming contaminated chocolate from China. Uncertainty about the cumulative distribution is indicated: the light grey band corresponds to 95% uncertainty range, and dark grey band corresponds to 50% uncertainty range.

Strengths

1. Provides a fully quantitative method for propagating uncertainties, which is more reliable than semi-quantitative or qualitative approaches or expert judgement.
2. Is a valid mathematical technique, subject to the validity of the model and inputs.
3. Can model complex systems and changes to the model can be made quickly and results compared with previous models.
4. Level of mathematics required is quite basic, but complex mathematics can be included.
5. 2D-MC is capable of quantifying uncertainty about variability
6. Model behaviour can be investigated relatively easily.
7. Time to results is reasonably short with modern computers.
8. Correlations and other dependencies can be modelled (but it can be difficult in some software, and is often not done).

Weaknesses and possible approaches to reduce them

1. If the input distributions are uncertain MC needs to be combined with sensitivity analysis (section B.16).
2. Obtaining appropriate data to define input distributions may be data-intensive (but structured expert elicitation is an alternative).
3. MC requires estimates or assumptions for the statistical dependencies among the variables. Uncertainty affecting these may be substantial and, if not quantified within the model, must be taken into account when characterising overall uncertainty. Sensitivity analysis may help.
4. 1D-MC does not distinguish between variability and uncertainty. 2D MC addresses this.

The relationship between inputs and outputs is unidirectional. New data can only be used to update the probability distribution of one input factor but not the joint distribution of all input factors. However, this is possible using more advanced forms of Bayesian modelling and inference (section B.9).

Assessment against evaluation criteria

This method is assessed against the criteria in Table B.14.6.

Conclusions

1. MC is the most practical way to carry fully probabilistic assessments of uncertainty and uncertainty about variability and is therefore a very important tool.
2. Application of MC is demanding because it requires full probability distributions. 2D MC is particularly demanding because it requires modelling choices (distribution families) and quantification of uncertainty about distribution parameters using statistical inference from data and/or expert knowledge elicitation.
3. It is likely that MC will be used to quantify key uncertainties in some assessments, especially in assessments where variability is modeled, with other methods being used to address other uncertainties.
4. MC output can be used to make limited probability statements concerning selected parameters which can then be combined with other limited probability statements using probability bounds analysis.

References

Bier VM, Lin SW. On the treatment of uncertainty and variability in making decisions about risk. Risk Anal. 2013;33(10):1899-907.

Burmaster DE, Anderson PD. Principles of good practice for the use of Monte Carlo techniques in human health and ecological risk assessments. Risk Analysis. 1994;14(4):477-81.

Cullen AC, Frey HC. Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs. New York: Plenum Press; 1999.

EFSA Scientific Committee; Guidance on selected default values to be used by the EFSA Scientific Committee, Scientific Panels and Units in the absence of actual measured data. EFSA Journal 2012;10(3):2579.

EFSA. Statement of EFSA on risks for public health due to the presences of melamine in infant milk and other milk products in China. EFSA Journal 2008;807: 1-10.

EFSA 2012. Guidance on the use of probabilistic methodology for modelling dietary exposure to pesticide residues. The EFSA Journal, 10: 1-95.

EFSA BIOHAZ Panel (EFSA Panel on Biological Hazards), 2014. Scientific Opinion on BSE risk in bovine intestines and mesentery. EFSA Journal 2014;12(2):3554, 98 pp. doi:10.2903/j.efsa.2014.3554

Ferson S. What Monte Carlo methods cannot do. Human and Ecological Risk Assessment. 1996;2:990-1007.

Hill A, Simons R, Ramnial V, Tennant J, Denman S, Cheney T, et al. Quantitative Microbiological Risk Assessment on Salmonella in slaughter and breeder pigs: final report. Parma, Italy: European Food Safety Authority, 2011 EN-46.

Nauta, M. J. (2000). Separation of uncertainty and variability in quantitative microbial risk assessment models. International Journal of Food Microbiology, 57(1), 9-18.

Pouillot R, Delignette-Muller ML. Evaluating variability and uncertainty separately in microbial quantitative risk assessment using two R packages. Int J Food Microbiol. 2010;142(3):330-340.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Trudel D, Tlustos C, von Goetz N, Scheringer M, Reichert P, Hungerbuhler K. Exposure of the Irish population to PBDEs in food: consideration of parameter uncertainty and variability for risk assessment. Food Addit Contam A. 2011;28(7):943-55.

US EPA, 1997. Guiding Principles for Monte Carlo Analysis. EPA/630/R-97/001. US EPA Risk Assessment Forum.

USDA/FDIS and US EPA, 2012. U.S. Department of Agriculture/Food Safety and Inspection Service and U.S. Environmental Protection Agency. Microbial Risk Assessment Guideline: Pathogenic Organisms with Focus on Food and Water. FSIS Publication No. USDA/FSIS/2012-001; EPA Publication No. EPA/100/J12/001.

Vose D. Risk analysis - a quantitative guide. Third ed. Chicester, England: John Wiley & Sons; 2008.

Table B.14.6: Assessment of 1D-MC (grey) and 2D-MC (dark grey, where different from 1D-MC), when applied well against evaluation criteria.

Criteria	Evidence of current acceptance	Expertise needed to conduct	Time needed	Theoretical basis	Degree/ extent of subjectivity	Method of propagation	Treatment of uncertainty & variability	Meaning of output	Transparency and reproducibility	Ease of understanding for non-specialist
Stronger character-istics	International guidelines available	No specialist knowledge required	Hours	Well established, coherent basis for all aspects	Fully data based	Calculation based on appropriate theory	Different types of uncert. & var. quantified separately	Range and probability of alternative outcomes	All aspects of process and reasoning fully documented	All aspects fully understandable
	EU level guidelines available	Can be used with guidelines or literature	Days	Most but not all aspects supported by theory	Combination of data and expert judgment	Formal expert judgment	Uncertainty and variability quantified separately	Range and relative possibility of outcomes	Most aspects of process and reasoning well documented	Outputs and most of process understandable
	National guidelines, or well established in practice or literature	Training course needed	Weeks	Some aspects supported by theory	Expert judgment on defined quantitative scales	Informal expert judgment	Uncertainty and variability distinguished qualitatively	Range of outcomes but no weighting	Process well documented but limited explanation of reasoning	Outputs and principles of process understandable
	Some publications and/or regulatory practice	Substantial expertise or experience needed	A few months	Limited theoretical basis	Expert judgment on defined ordinal scales	Calculation or matrices without theoretical basis		Quantitative measure of degree of uncertainty	Limited explanation of process and/or basis for conclusions	Outputs understandable but not process
Weaker character-istics	Newly developed	Professional statistician needed	Many months	Pragmatic approach without theoretical basis	Verbal description, no defined scale	No propagation	No distinction between variability and uncertainty	Ordinal scale or narrative description for degree of uncertainty	No explanation of process or basis for conclusions	Process and outputs only understandable for specialists

B.15 Deterministic calculations with conservative assumptions

Purpose, origin and principal features

This section addresses a set of related approaches to dealing with uncertainty that involve deterministic calculations using assumptions that aim to be *conservative*, in the sense of tending to overestimate risk.

A deterministic calculation uses fixed numbers as input and will always give the same answer, in contrast to a probabilistic calculation where one or more inputs are distributions and repeated calculations give different answers.

In deterministic calculation, uncertain elements are represented by single numbers. Various types of these can be distinguished:

- default assessment factors such as those used for inter- and intra-species extrapolation in toxicology
- chemical-specific adjustment factors used for inter- or intra-species differences when suitable data are available
- default values for various parameters (e.g. body weight), including those reviewed by the Scientific Committee (EFSA, 2012)
- conservative assumptions specific to particular assessments, e.g. for various parameters in the exposure assessment for BPA (EFSA, 2015)
- decision criteria with which the outcome of a deterministic calculation is compared to determine whether refined assessment is required, such as the Toxicity Exposure Ratio in environmental risk assessment for pesticides (e.g. EFSA, 2009).

Those described as *default* are intended for use as a standard tool in many assessments in the absence of specific relevant data. Those described as *specific* are applied within a particular assessment and are based on data or other information specific to that case. Default factors may be replaced by specific factors in cases where suitable case-specific data exist.

These are among the most common approaches to uncertainty in EFSA's work. They have diverse origins, some dating back several decades (see EFSA, 2012). What they have in common is that they use a single number to represent something that could in reality take a range of values, and that the numbers are chosen in a one-sided way that is intended to make the assessment conservative.

Deterministic calculations generally involve a combination of several default and specific values, each of which may be more or less conservative in themselves. Assessors need to use a combination of values that results in an appropriate degree of conservatism for the assessment as a whole, since that is what matters for decision-making.

The remainder of this section introduces the principles of this class of approaches, in four steps. The first two parts introduce the logic of default and specific values, using inter- and intra-species extrapolation as an example. The third part shows how similar principles apply to other types of default factors, assumptions and decision criteria, and the fourth part discusses the conservativism of the output from deterministic calculations. The subsequent section then provides an overview of how these approaches are applied within EFSA's human and environmental risk assessments.

Default factors for inter- and intra-species differences in toxicity

Default factors for inter- and intra-species differences are used to allow for the possible difference between a specified point of departure from an animal toxicity study and the dose for a corresponding effect in a sensitive human. The size of this difference (expressed as a ratio) varies between chemicals, as illustrated by the distribution in Figure B.15.1. If there are no specific data on the size of the ratio for a particular chemical, then the size of the ratio for that chemical is uncertain and a default factor is required. The default factor is intended to be high enough that the proportion of chemicals with higher values is small, as illustrated by the grey shaded area in Figure B.15.1. This default factor is conservative in the sense that, for most chemicals, the true ratio will be lower than the default (white area of distribution in Figure B.15.1). Thus if the default factor is applied to a particular chemical, there is a high probability that the true ratio for that chemical is lower than the default. Thus the distribution in Figure B.15.1 represents variability of the ratio in the population of chemicals, but uncertainty for a single chemical.

The same default value is used for different chemicals in the population because, in the absence of specific data, the same distribution applies to them all. If their true ratios became known, it would be found that the default factor was conservative for some and unconservative for others. However, in the absence of chemical-specific data, the ratios could lie anywhere in the distribution. Therefore, the same default factor is therefore equally conservative for all chemicals that lack specific data at the time they are assessed.

In order to specify the distribution in Figure B.15.1, it is necessary to define the starting and ending points for extrapolation. The animal endpoint is generally a NOAEL or BMDL. 'Sensitive human' could be defined as a specified percentile of the human population, as in the 'HDMI', the human dose at which a fraction I of the population shows an effect of magnitude M or greater, an effects metric proposed by WHO/IPCS (2014).

In practice, the distribution for variability between chemicals is not known perfectly: there is at least some uncertainty about its shape and parameters (e.g. mean and variance) which could quantified in various ways (e.g. Bayesian inference, sensitivity analysis or expert judgement, see sections B.9, B.9 and B.16). This uncertainty about the distribution for the population of chemicals adds to the uncertainty for an individual chemical. This can be taken into account by basing the default factor on a single distribution that includes both sources of uncertainty (uncertainty about the shape of the distribution, and about where a given chemical lies within it). In general, this will be wider than the best estimate of the distribution for variability between chemicals, and consequently a larger default factor will be needed to cover the same proportion of cases, i.e. to achieve the same degree of conservatism. This is illustrated graphically in Figure B.15.2. If the uncertainty about the distribution is not taken into account within the default factor, then it should either be quantified separately or taken into account in the overall characterisation of uncertainty for the assessment as a whole (see section 10 of main document).

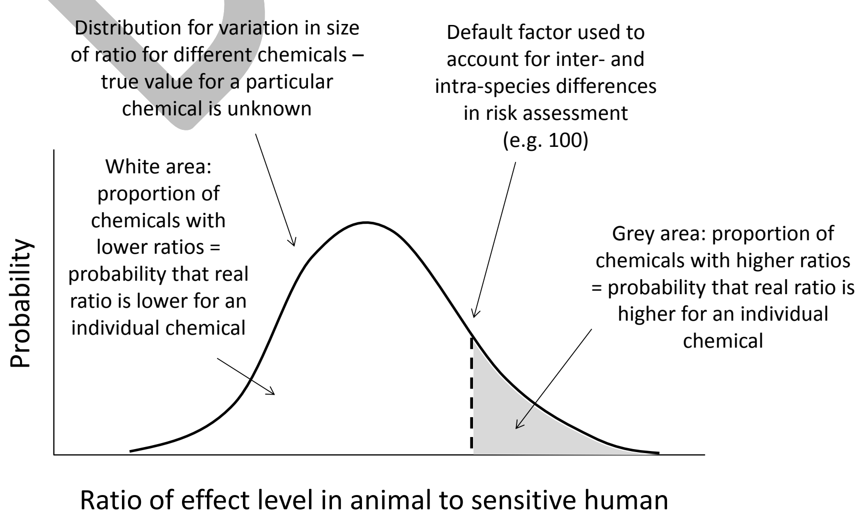


Figure B.15.1: Graphical representation of the general concept for default assessment factors for inter- and intra-species differences in toxicity.

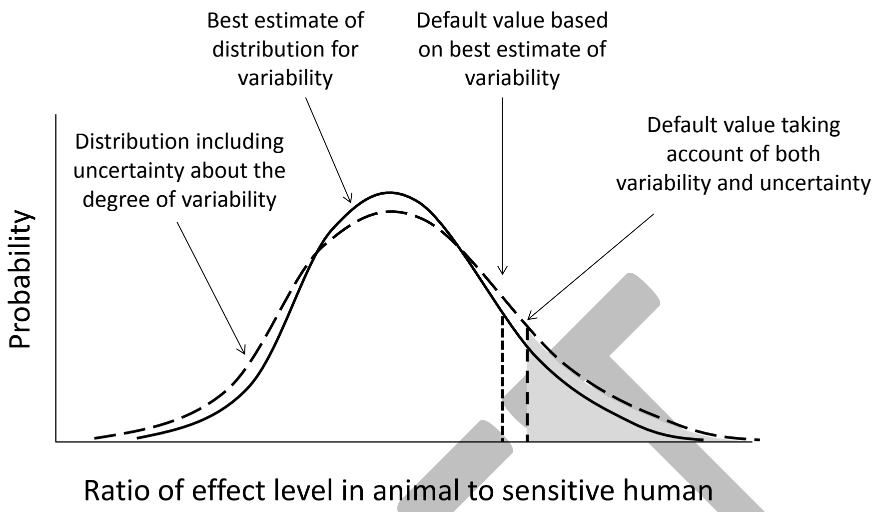


Figure B.15.2: Graphical representation of how uncertainty about the distribution for variability between chemicals can be taken into account when setting a default assessment factor.

<u>Specific factors for inter- and intra-species differences in toxicity</u>

When chemical-specific data are available to reduce uncertainty about part of the extrapolation for inter- and intra-species differences, this can be used to replace the corresponding part of the default assessment factor, as summarised by EFSA (2012). The default factor of 100 was introduced in the 1950s and later interpreted as reflecting extrapolation from experimental animals to humans (factor 10 for inter-species variability) and a factor of 10 to cover inter-individual human variability. A further division of these inter- and intra-species factors into 4 subfactors based on specific quantitative information on toxicokinetics and toxicodynamics was proposed by WHO/IPCS (2005). If specific data on toxicokinetics or toxicodynamics are available for a particular chemical, this can be used to derive chemical-specific adjustment factors (CSAF), which can then be used to replace the relevant subfactor within the overall default factor of 100.

WHO/IPCS (2005) provides detailed guidance on the type and quality of data required to derive CSAFs. For the inter-species differences, this includes guidance that the standard error of the mean (which represents sampling and measurement uncertainty in the data) should be less than approximately 20% of the mean. The guidance is designed to limit the various uncertainties affecting the data to a level that is small enough that the mean can be used as the basis for the CSAF.

The treatment of uncertainty for the CSAF is illustrated graphically in Figure B.15.3. The distribution represents all the uncertainty in deriving the CSAF. The value taken as the CSAF is the mean of the data. If this is near the median of the distribution, as illustrated in Figure B.15.3, then there is about a 50% chance that the true CSAF is higher. However, the criteria recommended in the guidance to reduce uncertainty mean that the true value is unlikely to be much higher than the mean of the data.

This illustrates an important general point, which is that *the choice of an appropriately conservative value to represent an uncertain or variable quantity depends not only on the chance that the true value is higher, but also on how much higher it could be*.

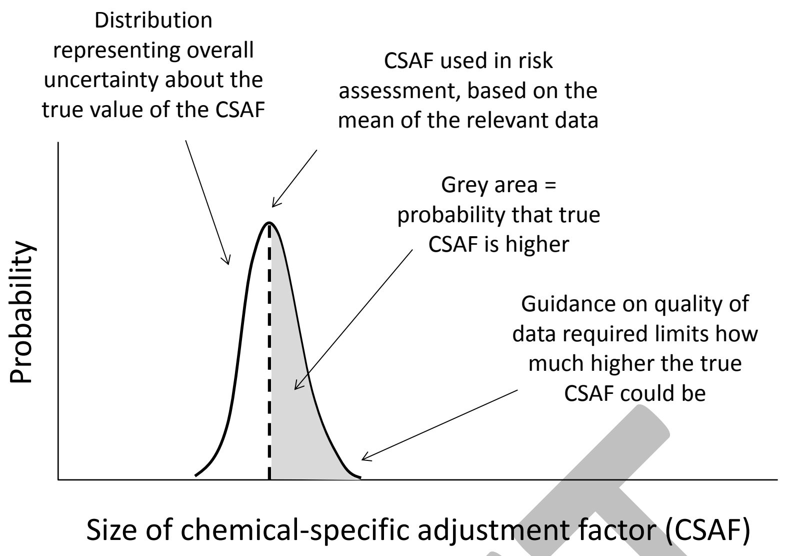


Figure B.15.3. Graphical illustration of treatment of uncertainty for a chemical-specific adjustment factor for inter- or intra-species differences in toxicokinetics or toxicodynamics.

<u>Default and specific values for other issues</u>

The principles and logic that are involved when using default or specific factors for inter- and intra-species differences, as illustrated in Figures B.15.1, B.15.2 and B.15.3, apply equally to other types of default and specific values used in risk assessment. This includes default values recommended by the Scientific Committee (EFSA, 2012), some of which refer to toxicity (including inter- and intra-species differences and extrapolation from subchronic to chronic endpoints) while others refer to exposure assessment (e.g. default values for consumption and body weight). For several other issues, EFSA (2012) does not propose a default factor but instead states that specific uncertainty factors should be derived case-by-case.

The same principles and logic also apply to all other values used in deterministic assessment, including conservative assumptions (which may be defaults applied to many assessments, or specific to a particular assessment) and decision criteria (which are usually defaults applied to many assessments). For example, in the melamine statement (EFSA, 2008), variability and uncertainty are addressed by repeating the assessment calculation with both central and high estimates for several parameters (described in more detail in the example at the end of this section).

What all of these situations have in common is that, in each assessment calculation, single values – either default or specific or a mixture of both – are used to represent quantities that are uncertain, and in many cases also variable. For each default or specific value, there is in reality a single true value that would allow for the uncertainty and variability that is being addressed. However, this true value is unknown. The degree to which each default or specific value is conservative depends on the probability that the true value would lead to a higher estimate of risk, and how much higher it could be. Figures B.15.1, B.15.2 and B.15.3 illustrate this for the case of parameters that are positively related to risk; for parameters that are negatively related to risk, the grey areas would be on the left side of the distribution instead of the right.

There are two main ways by which default and specific values can be established. Where suitable data are available to estimate distributions quantifying the uncertainty and variability they are intended to address, it is preferable to do this by statistical analysis and then choose an appropriately conservative value from the distribution. Where this is not possible or such data are not available, it is necessary to use expert judgement. In the latter case, the distribution should be elicited by formal or informal EKE, depending on the importance of the choice and the time and resources available (see sections B.8 and B.9). Alternatively, if the required degree of conservatism were known in advance, that percentile of the distribution could be elicited directly, without eliciting the full distribution.

It is especially important to ensure the appropriateness of default factors, assumptions and decision criteria, as they are intended for repeated use in many assessments. The context for which they are appropriate must be defined, that is, for what types of assessment problem, with which types and quality of data. When using them in a particular assessment, users must check whether the problem and data are consistent with the context for which the defaults are valid. If the assessment in hand differs, e.g. if the data available differ from those for which the defaults were designed, then the assessor needs to consider adjusting the defaults or adding specific factors to adjust the assessment appropriately (e.g. an additional factor allowing for non-standard data). The need to ensure default procedures for screening assessments are appropriately conservative, and to adjust them for non-standard cases, was recognised previously in the Scientific Committee's guidance on uncertainty in exposure assessment (EFSA, 2006).

<u>Overall conservatism of deterministic calculations</u>

Most deterministic assessments involve a combination of default and specific values, each of which may be more or less conservative in themselves. Ultimately, it is the *overall conservatism* of the assessment as a whole that matters for decision-making, not the conservatism of individual elements within it. This is why assessors often combine some conservative elements with others that are less conservative, aiming to arrive at an appropriate degree of conservatism overall.

Conservative is a relative term, and can only be assessed relative to a specified objective or target value. Overall conservatism needs to be assessed relative to the quantity the assessment output is intended to estimate, i.e. the measure of risk or outcome that is of interest to decision-makers. When the measure of interest is a variable quantity (e.g. exposure), the percentile of interest must also be defined. The overall conservatism of a point estimate produced by deterministic assessment can then be quantified in relation to that target value, as illustrated in Figure B.15.4.

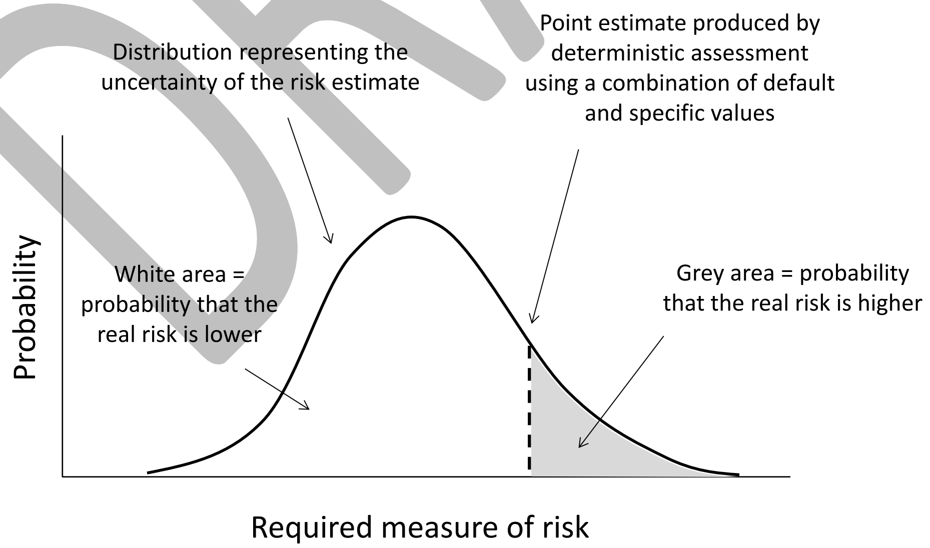


Figure B.15.4: Graphical illustration of assessing the overall conservatism of the output of a deterministic assessment, relative to a specified measure of risk. The distribution is not quantified by the deterministic assessment, so conservatism of the point estimate has to be assessed either by expert judgement, by probabilistic modelling, or by comparison with measured data on risk.

Assessing overall conservatism is very hard to do by expert judgement. Although assessors may not think in terms of distributions, judgement of overall conservatism implies considering first what distribution would represent each element, then how those distributions would combine if they were propagated through the assessment – taking account of any dependencies between them – and then what value should be taken from the combined distribution to achieve an appropriate degree of conservatism overall. Finally, the assessors have to choose values for all the individual elements such that, when used together, they produce a result equal to the appropriately conservative point in the combined distribution.

It is much more reliable to assess overall conservatism using probabilistic calculations, when time and resources permit. If it is done by expert judgement this will introduce additional uncertainty, which the assessors should try to take into account by increasing one or more of the factors involved (in a manner resembling the concept depicted in Figure B.15.2), or by adding an additional uncertainty factor at the end.

It is important that the overall degree of conservatism is appropriate: high enough to provide adequate protection against risk, but not so high that the assessment uses clearly impossible values or scenarios or leads to excessively precautionary decision-making. In terms of Figure B.15.4, the vertical dashed line should be placed neither too far to the left, nor too far to the right. Achieving this for the overall assessment output requires using appropriate values for each default and specific value in the assessment, as explained in the preceding section.

Quantifying the degree of conservativism requires scientific assessment, but deciding *what degree of conservatism is required or acceptable* is a value judgement which should be made by decision-makers (see Section 3 of main document). In terms of Figure B.15.4,characterising the distribution requires scientific consideration, while placing the dashed line requires a value judgement: what probability of conservative outcomes is required? If decision-makers were able to specify this in advance, assessors could then place the dashed line in Figure B.15.4 accordingly. Otherwise, assessors will have to choose what level of conservatism to apply when conducting the assessment, and seek confirmation from decision-makers at the end. In order for decision-makers to understand the choice they are making, they need information on the probability that the true risk exceeds the estimate produced by the assessment, and on how much higher the true risk might be. In other words, they need information on the uncertainty of the assessment. One of the benefits of establishing defaults is that once approved by decision-makers, they can be used repeatedly in multiple assessments without requiring confirmation on each occasion.

In refined assessments, default factors or values may be replaced by specific values. This often changes the overall conservatism of the assessment, because that depends on the combined effect of all elements of the assessment (as explained above). Therefore, whenever a default value is replaced by a specific value, the conservatism of the overall assessment needs to be reviewed to confirm it is still appropriate. This issue was recognised previously in EFSA's guidance on risk assessment for birds and mammals (EFSA, 2009).

Applicability in areas relevant for EFSA

<u>Human risk assessment</u>

Default factors, assumptions and decision criteria are, together with descriptive expression, the most common approaches to addressing uncertainty in EFSA and other regulatory agencies, and are used in many areas of EFSA's work. A comprehensive review is outside the scope of this document, but the following examples illustrate the range of applications involved.

Default assessment factors (AFs) and chemical-specific adjustment factors for inter- and intra-species extrapolation of chemical toxicity are described earlier in this section, and are key tools in setting health-based guidance values for human health (e.g. TDI and ADI). In recent years, efforts have been made to evaluate the conservatism of the default factors based on analysis, for suitable datasets, of inter-chemical variability for particular extrapolation steps (e.g. Dourson and Stara 1983, Vermeire et al. 1999). More recently, it has been proposed (e.g. Cooke 2010) to do a fully probabilistic analysis of uncertainty about such variability in order to derive default assessment factors. WHO/IPCS (2014) have developed a probabilistic approach to inter- and intra-species extrapolation that quantifies the conservatism of the default factors, and includes options for chemical-specific adjustments. The Scientific Committee has recommended that probabilistic approaches to assessment factors for toxicity are further investigated before harmonisation is proposed within EFSA (EFSA, 2012).

Factors and assumptions for other aspects of human health assessment, including exposure, are reviewed by EFSA (2012). Topics considered include body weight, food and liquid intake, conversion of concentrations in food or water in animal experiments to daily doses, deficiencies in data and study design, extrapolation for duration of exposure, absence of a NOAEL, the severity and nature of observed effects, and the interpretation of Margins of Exposure for genotoxic carcinogens. EFSA (2012) recommends the use of defaults for some of these issues, and case-by-case assignment of specific factors for others.

An example of an exposure assessment where the overall conservatism of case-specific assumptions was explicitly assessed is provided by the 2015 opinion on bisphenol A. Deterministic calculations were aimed at estimating an approximate 95th percentile for each source of exposure by combining conservative estimates for some parameters with average estimates for others. The uncertainty of these, and their combined impact on the overall conservatism of the resulting estimate, was assessed by expert judgement using uncertainty tables (EFSA, 2015a).

An example of probabilistic analysis being used to evaluate the conservatism of default assumptions in human exposure assessment is provided by EFSA (2007). This used probabilistic exposure estimates for multiple pesticides and commodities to evaluate what proportion of the population are protected by the deterministic 'IESTI' equations used in routine exposure assessment.

<u>Environmental risk assessment</u>

Default factors for inter-species differences, similar to those used for human risk, have been used for some time in setting environmental standards for ecosystems such as the predicted no effect concentration (PNEC). In some guidance documents for environmental risk assessment, a reference point from toxicity testing is divided by a default assessment factor and the result compared to the predicted exposure by computing their ratio, which is known as the *risk quotient (RQ)* (EC, 2003*).* In others the reference point is first divided by the predicted exposure to find the *toxicity-exposure ratio (TER)* and the result is then compared to a decision criterion, which is equivalent to an assessment factor (91/414/EWG). Although the calculations appear different, they lead to the same result and it is clear from the reasoning in the respective guidance documents that the assessment factors are intended to address variability and uncertainties relating to toxicity.

Most environmental exposure assessments are deterministic, using a combination of conservative factors and assumptions, some of which are defaults and some specific. Examples of these include the Tier 1 procedures for assessing acute and reproductive risks from pesticides to birds and mammals, which define different combinations of default assumptions to be used for different species that may be exposed, depending on the type of pesticide use involved. The guidance includes the option to replace the defaults with specific assumptions in refined assessment, where justified (EFSA, 2009). In assessing exposure of aquatic organisms to pesticides, a range of 'FOCUS' scenarios with differing defaults are used, representing different combinations of environmental conditions found in different parts of the EU (FOCUS, 2001).

As for human risk, some quantitative analyses have been conducted to justify or calibrate the defaults used in environmental risk. When developing the current guidance on pesticide risk assessment for birds and mammals, the procedure for acute risk to birds was calibrated by comparison with data on bird mortality in field experiments and history of use, as well as assessing its conservatism by expert judgement. For acute risk to mammals and reproductive risks, field data were lacking and it was necessary to rely on expert judgement alone (EFSA, 2008). For aquatic organisms, factors for extrapolating from laboratory toxicity studies with individual species to effects on communities of multiple species have been calibrated by comparing results from single species tests with semi-field experiments (Maltby et al 2009, Wijngaarden et al, 2014). As for human risk, it has been proposed that, in future, default factors used in environmental risk assessment should be derived from a fully probabilistic analysis taking both variability and uncertainty into account (EFSA 2015b).

Potential contribution to the main steps of uncertainty analysis

Steps in uncertainty analysis	Potential contribution of this approach
Identifying uncertainties	Not applicable. However, by discussing the need for assessment factor(s) you also identify some uncertainties.
Describing uncertainties	Not applicable.
Assessing the magnitude of individual uncertainties	Yes. Some assessment factors and assumptions are used to address individual uncertainties.
Assessing the combined impact of multiple uncertainties on the assessment output, taking account of dependencies	Yes. Decision criteria, and some assessment factors, address the combined effect of multiple uncertainties. The way they are used implies that they account for dependencies, though this is rarely explicit.
Assessing the contribution of individual uncertainties to overall uncertainty	In assessments that include multiple assessment factors, their magnitudes should reflect the assessors' evaluation of their relative importance.

Melamine example

In this guidance, the case study of melamine as described in EFSA (2008b) is used to illustrate the different approaches to assessing uncertainty. In EFSA (2008b) a TDI set by the SCF (EC, 1986) was used. Since that document does not describe the RP and the AFs used for deriving the TDI, an example of the use of assessment factors for toxicity is taken from an assessment made by the US-FDA (FDA, 2007), which is also referenced by EFSA(2008b). The following quote from FDA (2007) explains how the TDI was derived from combining a point of departure based on a detailed evaluation of toxicity studies with default assessment factors for inter- and intra-species extrapolation:

"The NOAEL for stone formation of melamine toxicity is 63 mg/kg bw/day in a 13-week rat study. This value is the lowest NOAEL noted in the published literature and is used with human exposure assessments below to provide an estimate of human safety/ risk… This POD was then divided by two 10-fold safety/uncertainty factors (SF/UF) to account for inter- and intra-species sensitivity, for a total SF/UF of 100. The resulting Tolerable Daily Intake (TDI) is 0.63 mg/kg bw/day. The TDI is defined as the estimated maximum amount of an agent to which individuals in a population may be exposed daily over their lifetimes without an appreciable health risk with respect to the endpoint from which the NOAEL is calculated."

The exposure assessment in the EFSA (2008b) statement addressed variability and uncertainty by estimating exposure for a range of scenarios using different combinations of assumptions, with varying degrees of conservatism. The factors that were varied included age and body weight (60kg adult or 20kg child), diet (plain biscuit, filled biscuit, quality filled biscuit, milk toffee, chocolate; plus two combinations of biscuit and chocolate), assumptions regarding the proportion of milk powder used in producing each food, and the concentration of melamine in milk powder (median or maximum of reported values). An estimate of exposure was calculated for each scenario, and expressed as a percentage of the TDI of 0.5 mg/kg taken from the SCF assessment (EC 1986). The results are reproduced in Table B.15.1.

Table B.15.1. Exposure estimates for different combinations of assumptions, expressed as a percentage of the TDI of 0.5 mg/kg (reproduced from EFSA, 2008b).

Melamine concentration	**Dietary exposure in proportion of TDI**			
	60 kg adult		**20 kg child**	
	Mean	95th percentile	Mean	95th percentile
Plain biscuit (2%)				
Median	0.0%	0.1%	0.1%	0.3%
High	4%	8%	11%	23%
Filled biscuit (3.5%)				
Median	0.1%	0.1%	0.2%	0.4%
High	7%	13%	20%	40%
Quality filled biscuit (16%)				
Median	0.3%	0.7%	1%	2%
High	30%	60%	90%	180%
Milk toffee (10%)				
Median	0.1%	0.4%	0.4%	1.2%
High	12%	36%	36%	108%
Chocolate (25%)				
Median	0.3%	1%	1%	3%
High	30%	90%	90%	269%
Combined consumption				
Biscuit	30%		90%	
Chocolate		90%		269%
Combined	120%		359%	
Biscuit		60%		180%
Chocolate	30%		90%	
Combined	90%		270%	

The estimates in Table B.15.1 involve additional assumptions and uncertainties, some of which are likely to be conservative. For example, EFSA (2008b) notes that the calculation involving quality filled biscuits might be a gross overestimation since there was no indication that China exported such products to Europe at that time, though it could not be completely excluded. The chocolate scenario was considered more realistic.

For adults, EFSA (2008b) concluded that:

"Based on these scenarios, estimated exposure does not raise concerns for the health of adults in Europe should they consume chocolates and biscuits containing contaminated milk powder."

This implies a judgement by the assessors that, although the estimated adult exposures exceeded the TDI in one scenario (mean consumption of biscuit combined with high level consumption of chocolate), overall – considering the likelihood of this scenario, the combined conservatism of the assumptions made, and the impact of other uncertainties identified in the text – the likelihood of adverse effects was sufficiently low not to 'raise concerns'. This could be made more transparent by specifying the assessors' judgement of level of likelihood.

For children, EFSA (2008) concluded that:

"Children with a mean consumption of biscuits, milk toffee and chocolate made with such milk powder would not exceed the tolerable daily intake (TDI). However, in worst case

scenarios with the highest level of contamination, children with high daily consumption of milk toffee, chocolate or biscuits containing high levels of milk powder would exceed the TDI. Children who consume both such biscuits and chocolate could potentially exceed the TDI by more than threefold. However, EFSA noted that it is presently unknown whether such high level exposure scenarios may occur in Europe."

The conclusion for children is more uncertain than for adults. The assessors state that the exposure could 'potentially' exceed the TDI by more than threefold in one scenario, but do not express a judgement on how likely that is to occur.

Strengths

1. Conservative assessment factors, assumptions and decision criteria address uncertainty using a one-sided approach that aims to be conservative but not over-conservative.
2. The methodology is widely adopted, well accepted by authorities, and easy to communicate.
3. It can be used in any type of quantitative assessment.
4. Once established, default factors are straightforward to apply and do not require any special mathematical or statistical skills.
5. Some default factors and criteria are supported by quantitative analysis of data that supports their appropriateness for their intended use. Similar analyses could be attempted for others, where suitable data exist.

Weaknesses and possible approaches to reduce them

1. While some default assessment factors are generally well-accepted and research has provided quantitative support, the use of other default factors and most specific factors is based mainly on expert judgment without quantitative detail and it can be difficult to establish either the reasoning that led to a particular value or exactly what sources of uncertainty are included.
2. Generation of specific factors, and providing quantitative support for default factors where this is currently lacking, require relevant expertise to evaluate the available evidence and statistical expertise for analysis.
3. Assessment factors which are based on analysis of data without quantification of uncertainty about variability may be less conservative than intended (as illustrated in Figure B.15.2).
4. It is often unclear how conservative the result is intended to be. This could be addressed by defining more precisely what extrapolation or adjustment is being made and what level of confidence is required, in consultation with decision-makers.
5. There is little theoretical basis for assuming that assessment factors should be multiplied together, as is often done. However such multiplication tends to contribute to the conservatism of the approach (Gaylor and Kodell, 2000). Section B.13 of this annex on *probability bounds* provides a rationale for multiplication if a probability is attached to each individual AF.
6. Division of AFs into subfactors could lead to reduced conservatism if, for example, a CSAF greater than the default subfactor is needed to cover a particular source of variability. The reduction of conservatism could be quantified by a probabilistic analysis.
7. AFs do not provide a range for the outcome, based on the propagation of the uncertainty around the various input factors, but only a conservative estimate of the outcome.

8. Risk management decisions, about the level of conservatism required, are embedded in the AF. For the process to be transparent, such decisions need to be made explicit.
9. Assessment factors do not generally provide a mechanism to assess the relative contribution of different sources of uncertainty to overall uncertainty or to distinguish contributions of variability and uncertainty. A probabilistic analysis can provide a general indication of relative contributions for the selected group of chemicals.

Assessment against evaluation criteria

This method is assessed against the criteria in Table B.15.2.

Conclusions

Assessment factors, conservative assumptions and decision criteria are widely used to account for uncertainty, variability and extrapolation in many areas of EFSA assessment. Some are defaults that can be used in many assessments, while others are specific to particular assessments. They are simple to use and communicate. When well specified and justified they are a valuable tool, providing an appropriate degree of conservatism for the issues they address. They are more reliable when it is possible to calibrate them by statistical analysis of relevant data.

Most assessments involve a combination of multiple factors and assumptions, some default and some specific. Conservatism needs to be evaluated for the assessment as a whole, taking account of all the elements involved. This is much more reliable when done by probabilistic analysis than by expert judgement.

References

Cooke R (2010), Conundrums with Uncertainty Factor (with discussion), Risk Analysis, 30(3), pp 330–338

Dourson ML, Stara JF (1983), Regulatory History and Experimental Support of Uncertainty (Safety) Factors, Regulatory Toxicology and Pharmacology, 3, pp 224–238.

EC (European Commission), 1986. Report of the Scientific Committee for Food on certain monomers of other starting substances to be used in the manufacture of plastic materials and articles intended to come into contact with foodstuffs. Seventeenth series. Opinion expressed on 14 December 1984. Available at URL http://ec.europa.eu/food/fs/sc/scf/reports/scf_reports_17.pdf.

EC (European Commission), 2003. Technical Guidance Document on Risk Assessment. In support of Commission Directive 93/67/EEC on Risk Assessment for new notified substances Commission Regulation (EC) No 1488/94 on Risk Assessment for existing substances Directive 98/8/EC of the European Parliament and of the Council concerning the placing of biocidal products on the market. Part II, Chapter 3 Environmental Risk Assessment.

EFSA, 2005. Scientific opinion on a request from EFSA related to a harmonised approach for risk assessment of substances which are both genotoxic and carcinogenic. The EFSA Journal ; 282, 1-31.

EFSA 2008. Scientific Opinion of the Panel on Plant protection products and their Residues (PPR) on the Science behind the Guidance Document on Risk Assessment for birds and mammals. The EFSA Journal (2008) 734: 1-181.

EFSA, 2009. Guidance Document on Risk Assessment for Birds & Mammals. EFSA Journal 2009; 7(12):1438.

EFSA, 2012. Guidance on selected default values to be used by the EFSA Scientific Committee, Scientific Panels and Units in the absence of actual measured data. EFSA Journal 2012;10(3):2579.

EFSA, 2015a. Scientific Opinion on the risks to public health related to the presence of bisphenol A (BPA) in foodstuffs – Part I: exposure assessment. The EFSA Journal, 2015;13(1):3978.

EFSA, 2015b. Scientific Opinion addressing the state of the science on risk assessment of plant protection products for non-target arthropods. EFSA Journal 2015;13(2):3996.

FDA (United States Food and Drug Administration) (2007). Interim Safety and Risk Assessment of Melamine and its Analogues in Food for Humans. http://www.fda.gov/food/foodborneillnesscontaminants/chemicalcontaminants/ucm164658.htm, accessed 13-03-2015.

FOCUS (2001). "FOCUS Surface Water Scenarios in the EU Evaluation Process under 91/414/EEC". Report of the FOCUS Working Group on Surface Water Scenarios, EC Document Reference SANCO/4802/2001-rev.2. 245 pp.

Gaylor DW, Kodell R. (2000), Percentiles of the Product of Uncertainty Factors for Establishing Probabilistic Reference Doses, Risk Analysis, 20(2), pp 245–250.

Maltby L, Brock TCM, Brink PJ. Fungicide risk assessment for aquatic ecosystems: importance of interspecific variation, toxic mode of action, and exposure regime. Environ Sci Technol.2009;43:7556–7563.

Vermeire T, Stevenson H, Pieters M., Rennen M, Slob W, Hakkert BC. (1999), Assessment Factors for Human Health Risk Assessment: A Discussion Paper, Critical Reviews in Toxicology, 29(5), pp 439–490.

WHO/IPCS, 2005. Chemical-specific adjustment factors for interspecies differences and human variability: guidance document for use of data in dose/concentration–response assessment. IPCS Harmonization Project Document No. 2. World Health Organisation, Geneva.

WHO/IPCS, 2014. Guidance document on evaluating and expressing uncertainty in hazard characterization. Harmonization Project Document No. 11. International Programme on Chemical Safety. WHO, Geneva.

van Wijngaarden RPA, Maltby L and BrockTCM. 2014. Acute tier-1 and tier-2 effect assessment approaches in the EFSA Aquatic Guidance Document: are they sufficiently protective for insecticides? *In press.* DOI: 10.1002/ps.3937

Table B.15.2: Assessment of Deterministic calculations with conservative assumptions (when applied well) against evaluation criteria.

Criteria	Evidence of current acceptance	Expertise needed to conduct	Time needed	Theoretical basis	Degree/ extent of subjectivity	Method of propagation	Treatment of uncertainty and variability	Meaning of output	Transparency and reproducibility	Ease of understanding for non-specialist
Stronger character-istics	International guidelines or standard scientific method	No specialist knowledge required	Hours	Well established, coherent basis for all aspects	Judgement used only to choose method of analysis	Calculation based on appropriate theory	Different types of uncert. & var. quantified separately	Range and probability of alternative outcomes	All aspects of process and reasoning fully documented	All aspects fully understandable
	EU level guidelines or widespread in practice	Can be used with guidelines or literature	Days	Most but not all aspects supported by theory	Combination of data and expert judgment	Formal expert judgment	Uncertainty and variability quantified separately	Range and relative possibility of outcomes	Most aspects of process and reasoning well documented	Outputs and most of process understandable
	National guidelines, or well established in practice or literature	Training course needed	Weeks	Some aspects supported by theory	Expert judgment on defined quantitative scales	Informal expert judgment	Uncertainty and variability distinguished qualitatively	Range of outcomes but no weighting	Process well documented but limited explanation of reasoning	Outputs and principles of process understandable
	Some publications and/or regulatory practice	Substantial expertise or experience needed	A few months	Limited theoretical basis	Expert judgment on defined ordinal scales	Calculation or matrices without theoretical basis		Quantitative measure of degree of uncertainty	Limited explanation of process and/or basis for conclusions	Outputs understandable but not process
Weaker character-istics	Newly developed	Professional statistician needed	Many months	Pragmatic approach without theoretical basis	Verbal description, no defined scale	No propagation	No distinction between variability and uncertainty	Ordinal scale or narrative description for degree of uncertainty	No explanation of process or basis for conclusions	Process and outputs only understandable for specialists

B.16 Sensitivity and Scenario analysis

Purpose, origin, and principal features

In the context of uncertainty assessment, sensitivity analysis aims to identify both the magnitude of the contributions of individual sources of uncertainty to uncertainty about the assessment output(s) and the relative contributions of different sources. The purpose of doing so is (i) to help prioritise uncertainties for quantification: (ii) to help prioritise uncertainties for collecting additional data; (iii) to investigate sensitivity of final output to assumptions made; (iv) to investigate robustness of final results to assumptions made.

Saltelli et al. (2004) defines sensitivity analysis of a model as '*the study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input*'. A broader definition of Sensitivity Analysis is given in the Oxford business dictionary where it is described as '*Simulation analysis in which key quantitative assumptions and computations (underlying a decision, estimate, or project) are changed systematically to assess their effect on the final outcome. Employed commonly in evaluation of the overall risk or in identification of critical factors, it attempts to predict alternative outcomes of the same course of action*'. According to Saltelli, desirable properties of a sensitivity analysis method for models include the ability to cope with influence of scale and shape; the allowance for multidimensional averaging (all factors should be able to vary at the same time); model independence (i.e. the method should work regardless of additively or linearity of the model); ability to treat grouped factors as if they were single factors.

There is a very large and diverse literature on sensitivity analysis, including a number of reviews (e.g. Clemson et al., 1995; Eschenbach and Gimpel, 1990; Hamby, 1994; Lomas and Eppel, 1992; Rios Insua, 1990; Sobieszczanski-Sobieski, 1990; Tzafestas et al., 1988, Frey & Patil 2002, 2004, Tian 2013) reflecting the fact that historically sensitivity analysis methods have been widely used across various disciplines including engineering systems, economics, physics, social sciences and decision making (e.g., Oh and Yang, 2000; Baniotopoulos, 1991; Helton and Breeding, 1993; Cheng, 1991; Beck et al., 1997; Agro et al., 1997; Kewley et al., 2000; Merz et al., 1992). Most of the literature, however, deals with the use of sensitivity analysis methods in the presence of a model.

Two general approaches to sensitivity analysis have been developed. The first approach looks at the effects on the output of infinitesimal changes to the default values of the inputs (local) while the second one investigates the influence on the output of changes of the inputs over their whole range of values (global). In the following the discussion will focus only on methods for global sensitivity analysis since local analysis is considered of limited relevance in the uncertainty assessment context because it does not provide for an exploration of the whole space of the input factors that is necessary when dealing with uncertainty. Whatever the type and number of input uncertainty factors, it is important that the purpose of sensitivity analysis is clearly defined after consideration and, when needed, prioritization of the inputs to be included in the sensitivity analysis.

One special type of sensitivity analysis is scenario analysis (sometimes named conditional Sensitivity Analysis). It is generally helpful when there is a dependency in the inputs and it is difficult to assess the sensitivity of the output to changes in a single input without fixing some pre-specified values of the other inputs. Scenario analysis express the sensitivity for one input conditional on a set of values of the other factors kept constant at pre-specified values (more likely or of special interest). It is also called 'what-if analysis'. The most common approach in Scenario Analysis is to combine key variables making reference to three possible cases: a. worst-case or conservative scenario; b. most likely or base scenario; c. best-case or optimistic scenario.

Frey and Patil (2002) suggest grouping methodologies for sensitivity analysis in three categories: mathematical methods, statistical methods, graphical methods. These categories could be further classified according to other important aspects such as the kind of input effects that they are able to capture (individual or joint) and the form of the relationship between inputs and output (linear or non-linear). A comparison of the main methodologies and their most appropriate use in relation to the objective of the sensitivity analysis is provided by the same authors. Only those methods that are deemed to be relevant in the framework of uncertainty analysis and applicable to the risk assessment context are described in this section. Therefore the list of methods that follows is not comprehensive. Different methods and sensitivity indexes can provide a range of different factor rankings. Where this happens, the assessor needs to consider the cause of the differences and their implications for interpretation of the results.

A summary of the methods considered in this Guidance for Sensitivity Analysis are provided in Table B.16.1.

Table B.16.1: Summary table of methods to perform sensitivity analysis

Group	Method	Acronym	Characteristics
Graphical	Tornado plot		Input factors sorted by their influence on the output in a decreasing order
	Scatter plot		Highlight relationship between output and each input factor. No interaction among factors
	Spider plot		Plot all the input factors as lines crossing at the nominal value of the ouput. The inputs with the highest slope are those with highest influence on the output
	Box plot		Range of variation of the output with respect to each input
	Pie chart		Split of the pie in slices whose size is proportional to the influence of each input
Mathematical/deterministic	Nominal Range Sensitivity Analysis	NRSA	No interaction among input factors, monotonic relationship
	difference of log odds ratio	ΔLOR	Special case of NRSA when output is a probability
	Breakeven analysis	BEA	Output is a dichotomous variable
Probabilistic	Morris	Morris	Qualitative screening of inputs
	Monte Carlo filtering	MCF	Analogous of BEA with probabilistic approach
	Linear rank regression analysis	SRC, SRRC, PCC, PRCC.	Strong assumptions: normality residuals, uncorrelation among inputs, linear relationship
	Analysis of Variance	ANOVA	Non parametric method
	Fourier Amplitude Sensitivity Test and Extended version	FAST, E-FAST	Variance-base method. No assumptions required.
	Sobol index	S	Widely applicable

Graphical methods

These are normally used to complement mathematical or statistical methodologies especially to represent complex dependency and facilitate their interpretation. They are also used in the early stage to help prioritizing among sources of uncertainty. Graphical methods include: Scatter plot, tornado plots, box plots, spider plots and pie charts (Patil & Fray 2004). In the context of this Guidance they are considered only as supporting methods to help interpretation of the sensitivity analysis results. Examples of graphical methods for sensitivity analysis are provided in Figure B.16.1.

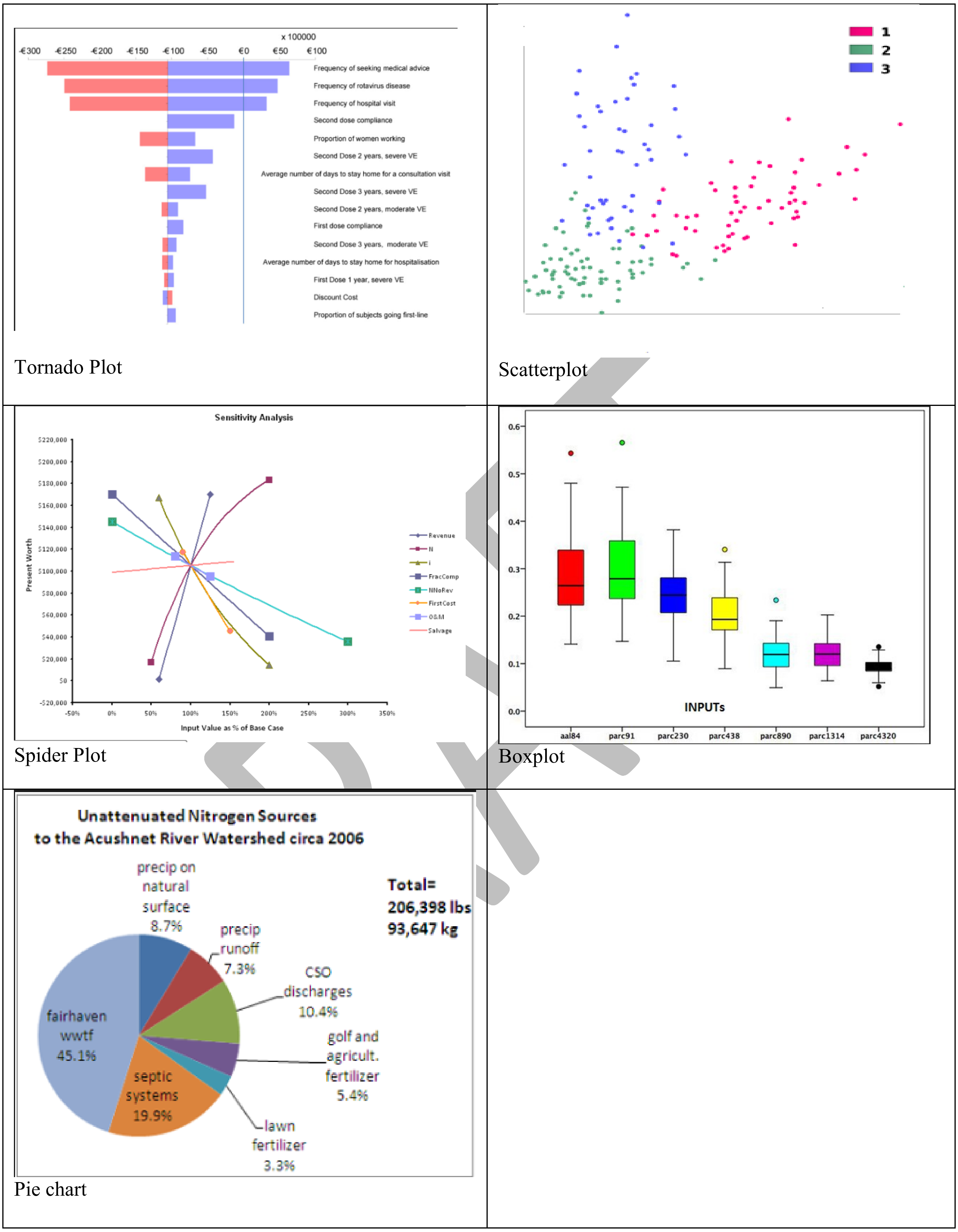


Figure B.16.1: Examples of graphical methods for sensitivity analysis.

Deterministic (named "mathematical" by Patil & Frey) methods

These methods involve evaluating the variability of the output with respect to a range of variation of the input with no further consideration of the probability of occurrence of its values. For this reason and to keep the symmetry with the classification adopted for the uncertainty assessment approaches, they are referred to as 'deterministic' instead of mathematical methods. In case of monotonic relationship these methods can be useful for a first screening of the most influential inputs. Graphical methods and the revised Morris method are suitable alternatives when monotonicity is not met.

1. Nominal Range Sensitivity Analysis (NRSA)

This method is normally applied to deterministic models (Cullen and Frey 1999). It assesses the effect on the output of moving one input from its nominal (often most-likely) value to its upper and lower most extreme plausible values while keeping all the other inputs fixed at their nominal values. The resulting sensitivity measure is the difference in the output variable due to the change in the input (expressed sometimes as percentage).

This approach to sensitivity analysis is closely related to interval analysis (see section B.7).

Interactions among factors are not accounted for by this method which limits its capacity to estimate true sensitivity. Although simple to implement, it fails in case of non monotonic relationships because it does not examine behaviour in for input values between the extremes.

A specific case of the nominal range is the difference of log odds ratio which can be used in case of an output expressed as probability. It is based on the computation of the log-odds or log-odds-ratio of an event.

2. Breakeven analysis (BEA)

The purpose of this method is to identify a set of values of inputs (break-even values) that provide an output for which decision makers would be indifferent among the various risk management options (Patil & Fray 2004). This method is useful to assess the robustness of a decision to change in inputs (i.e. whether a management option still remains optimal or sub-optimal also in case the values of inputs change with respect to the current levels). It is commonly used when the output is expressed as dichotomous variable indicating two possible options such as whether a tolerable daily intake is exceeded or not. It represents a useful tool for evaluating the impact of uncertainty on different possible choices of policy maker (e.g. what level of use to permit for a food additive).

The breakeven analysis has a probabilistic counterpart in Monte Carlo filtering which partitions the outputs in two sets based on compliance/non-compliance with some criterion (see later).

<u>Statistical methods</u>

In statistical methods of sensitivity analysis, the input range of variation is addressed probabilistically so that not only different values of the inputs but also the probability that they occur are considered in the sensitivity analysis. This approach to the sensitivity analysis is naturally linked to the investigation of the uncertainty based on probabilistic methods.

Most of the methods belonging to this group are based on the decomposition of the output variance with respect to the variability of the inputs. They generally allow the assessor to identify the effect of interactions among multiple inputs. Frequently statistical sensitivity analysis is performed using Monte Carlo techniques (sometimes combined with bootstrapping techniques) although this approach is not strictly necessary and sometimes not preferable if it is too computationally intensive.

Identification of the separated influence of variability and uncertainty in the input on the uncertainty in the output is not a trivial issue in sensitivity analysis. Recently Busschaert et al. (2011) proposed an advanced sensitivity analysis to address this issue. This analysis is sometimes referred to as two-dimensional sensitivity analysis. It is not described in detail in this Guidance.

1. Morris method

The Morris method provides a qualitative measure of the importance of each uncertain input factor for the outputs of a model at a very low computational cost, determining factors that have: (i) negligible effects; (ii) linear and additive effects (iii) non-linear and/or non-additive effects (Saltelli et al., 2005). The methods can be used as a qualitative screening procedure to select the most important input factors for computationally more demanding variance-based methods for sensitivity analysis. The Morris method varies one factor at a time across a certain number of levels selected in the space of the input factors. For each variation, the factor's elementary effect is computed, which measures, relative to the size of the change, how much the output changed when the factor value was changed.

The number of computations required is N = T (k+1), where k is the number of model input factors and the number of sampling trajectories T is a number generally ranging between 10 and 20 depending on the required accuracy. Ten trajectories are usually considered sufficient (Saltelli et al., 2004). Different sampling methods are available. Khare et al. (2015) describe a new sampling strategy (sampling for uniformity – SU), which was found to perform better than existing strategies using a number of criteria including: generated input factor distributions' uniformity, time efficiency, trajectory spread, and screening efficiency. We use the SU method in the example that follows on melamine.

The mean of the elementary effects for a factor estimates the factor's overall effect (μ_i). A high value suggests a strong linear effect of that factor, whereas a high value of the standard deviation of the elementary effects (σ_i) indicates a non-linear or non-additive effect. For non-monotonic effects, the mean of the absolute values of the elementary effects can also be computed to avoid cancelling out of opposing signals (Saltelli et al. 2005). When using absolute values the method is known as revised Morris. Visualization is possible by plotting the mean elementary effect for each factor versus the standard deviation. Input factors which have large mean or standard deviation of the elementary effects (or moderately large values of both) are most influential on the model outcome.

2. Monte Carlo Filtering (MCF)

The goal of Monte Carlo filtering is to identify the ranges of these input factors which result in model output which is considered acceptable by decision-makers (Chu-Agor et al, 2012). In MCF, a set of constraints has to be defined that targets the desired characteristics of the model realization (e.g. a threshold value for the risk ratio, set by risk managers or stakeholders). Based on the results of the uncertainty analysis, model results (for example output values of r) are then classified as being "favourable" or "unfavourable". The values of the input factors are then divided into two groups: those which produce favourable output and those which produce unfavourable output. In order to check what drives the difference between a favourable outcome and an unfavourable outcome, a two-sided Smirnov test is performed for each factor to test if the distribution of the factor is different in the favourable output group than in the unfavourable output group. If the null hypothesis is rejected, this indicates that the input factor is a key factor in driving the model towards favourable outcomes, and is a good candidate for risk management intervention. If the null-hypothesis is accepted, this indicates that at any value of the input factor can result in either a favourable or an unfavourable result, and intervening on that factor is not likely to result in changes in the output of the system represented by the model. In addition to the statistical significance, it is important to evaluate the ranges of input factors that produce differential outputs to explore the biological significance of the findings.

3. Linear rank regression analysis

The linear regression analysis can be used as a statistical method for investigating sensitivity when it is reasonable to assume that the relationship between inputs and output is linear (Saltelli, 2008). A variety of indicators can be computed using this broad approach. The magnitude of the regression coefficients, standardized by the ratio of the standard deviations of model independent and dependent variables (SRC: standardized regression coefficient) is commonly used as a measure of sensitivity as well as the rank assigned to the inputs once sorted by their SRC (SRRC: standardized rank regression coefficient)

$$SRC = b_i \cdot \frac{stddev(X_i)}{stddev(Y)}$$

The Partial Correlation Coefficient (PCC) and the Partial Rank Correlation Coefficient (PRCC),), can be used alternatively.

The square of the multiple correlation coefficient (R2) is an indicator of goodness of fit of a linear model. Its incremental change, when performing a multivariate stepwise regression analysis, expresses the additional component of variation of the dependent variable explained by the newly introduced input. In the phase of setting up a model, it can be used as a measure of sensitivity to screen factors most influential on the dependent variables.

Possible drawbacks of this class of indicators are the low robustness of the results of regression analysis when key assumptions are not met (e.g. independence of inputs, normality of residuals). In addition these methods are dependent on the functional form (underlying model) explaining the relationship between output and inputs and the range of variation considered for each input.

4. Analysis of variance

Analysis of Variance (ANOVA) is a sensitivity analysis method that does not require specification of a functional form for the relationship between the output and a set of inputs (non parametric method). The ANOVA aims at investigating whether the variation of the values of the output is significantly associated with the variation of one or more inputs.

5. Fourier Amplitude Sensitivity Test (FAST)

The FAST method belongs to the class of variance-based global sensitivity analysis methods. The effect of the uncertain inputs on the output is computed as the ratio of the conditional variance (variance of the conditional distribution of the output having fixed the value of one input or of a combination of inputs) to the total variance of the output. It takes his name from the multiple Fourier series expansion that is used as a tool for computing the conditional variance. The method has a wide applicability since it does not require any assumptions on the model structure nor on monotonicity. In its original form the FAST method (Cukier et al., 1973) required the assumption of no interaction among inputs. Saltelli et al (1999) developed an extended FAST method that allows accounting for multiple interactions.

Based on Fourier expansion, the total variance of the output can be expressed as the sum of all conditional variances of various orders (from the 1st to the nth):

$$V = \sum_{j=1}^{n} V_j + \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} V_{jk} + ... + V_{12....n}$$

The first order sensitivity index is computed as the ratio of a single input conditional variance and the total variance whereas the multiple effect sensitivity index is a similar ratio obtained using the multiple factors conditional variance in the numerator.

$$S_{j_1 j_2 \ldots j_r} = \frac{V_{j_1 j_2 \ldots j_r}}{V}$$

Higher values of the index indicate a great influence of the factor/s on the output.

6. Sobol Index

Sobol's index (Sobol, 1990) is based on the idea of decomposing the output variance into the contributions associated with each input factor. It expresses the reduction in the output variability that could be achieved if value of an input factor was fixed.

The first-order Sobol index for an input factor is defined as the ratio of the variance of the conditional means of the output (given all possible values of a single input) over the total variance of the output. It indicates the rate of the total output's variance exclusively attributable to a specific input. It does not account for the interaction with other factors.

$$S_j = \frac{V\left[E\left(\mathrm{Y} / \mathrm{X}_j\right)\right]}{V(Y)}$$

In a perfectly additive model the sum of first order sensitivity indices over all the input factors equals 1. Models with a sum greater than 0.6 are considered mostly additive (Saltelli et al., 2004).

The higher order interaction terms express the amount of variance of the output explained by the interaction among factors not already accounted for by lower interaction terms (including first order). It is computed as the ratio of the higher order conditional variance over the total variance of the output.

The total sensitivity index (Homma and Saltelli 1996) of an input is obtained as the sum of the first-order index and all the higher order interaction terms involving that specific input.

Traditionally the computation of the Sobol indexes is performed running simulations with the Monte Carlo algorithm. The computational requirements of the method are $N = M(2k+2)$, with M the Monte Carlo over-sampling rate, $512<M<1024$ and k the number of input factors.

Various software applications have been developed to carry out Sensitivity Analysis. JRC developed a free license tool named SimLab[8] that provides a reference implementation of the most recent global sensitivity analysis techniques. Various packages have been developed to support performance of sensitivity analysis in mathematical and statistical softwares that are commonly used (e.g. R and Matlab). Tools have been included in @Risk and Sensit Excel adds-in allowing computation of some sensitivity indices and their graphical plotting. The EIKOS Simulation Toolbox has been developed by Uppsala University (Ekstrom 2005). A non-comprehensive list of software is given in Table B.16.2.

Table B.16.2: Main software and packages including tools to perform sensitivity analysis

Package	Method
@Risk (Excel adds-in)	Scatter plot, tornado plot multivariate stepwise regression and PRCC
CrystalBall	
ModelRisk	
Simlab software (JRC)	Morris, SRC, SRRC,FAST, E-FAST, Sobol
Matlab	Scatter plot, 3-D plot, PCC, SRC, Morris
EIKOS	SRC, SRRC, PCC, PRCC Sobol, FAST, extended FAST
Sensit (Excel adds-in)	Spider charts, and tornado charts
R packages - Sensitivity	SRC, SRRC, PCC, PRCC, Morris, FAST, Sobol

Applicability in areas relevant for EFSA

The value of sensitivity analysis in the regulatory context and risk assessment is highlighted by Pannell (1997). It opens the possibility for the assessors to provide decision makers with important information related to the robustness of the assessment conclusions with respect to the various sources of uncertainty. This information includes: a. the identification of break-even input values where the conclusions would change; b. the provision of flexible recommendations which depend on circumstances; c. the characterization of a strategy or scenario in terms of riskiness allowing development of priorities for risk mitigations; d. the identification of important sources of uncertainty for prioritizing additional research/data collection.

Despite its informative value, the performance of sensitivity analysis poses some critical challenges in EFSA's assessment models mainly because, when models are used, they are frequently non- linear, contain thresholds and deal with discrete inputs and/or outputs. Non linearity and presence of thresholds generally imply that interactions among input factors cannot be ignored and sensitivity measures accounting for input dependency need to be considered.

A review of the sensitivity analysis methods that deserve consideration in the risk assessment context is provided by Frey and Patil (2002, 2004). An example of the implementation of the global sensitivity analysis developed by Saltelli in the context of contamination assessment of Listeria monocytogenes in smoked salmon is given by Augustin (2011).

[8] http://ipsc.jrc.ec.europa.eu/?id=756

Some examples of applications of sensitivity analysis are available in EFSA risk assessment. The opinion of the AHAW Panel on Framework for EFSA AHAW Risk Assessments (2007) advises to perform a sensitivity analysis 'to determine to what extent various uncertainties affect the conclusions and recommendations'. The PPR Panel Guidance on the Use of Probabilistic Methodology for Modelling Dietary Exposure to Pesticide Residues (2012) suggests the use of sensitivity analysis in probabilistic assessment in order to investigate the impact of model assumptions and other decisions based on expert judgement (e.g. exclusion of extreme values) on the final results. In the EFSA opinion on prevalence of Listeria monocytogenes (2014) the association between the prevalence of Listeria monocytogenes in EU and some potentially associated factors related to fish and meat dishes consumption was investigated using multiple-factor regression models. To get further insight into the stability of the final models, a sensitivity analysis was performed with respect to some methodological changes in the setting up of the model.

Other institutions perform or advise to use sensitivity analysis as part of their assessments. The European Chemical Agency mentions sensitivity analysis in its Guidance on information requirements and chemical safety assessment (ECHA, 2012). The Joint Research Centre of the European Commission has a long history of application of sensitivity analysis in various fields including transport, emission modelling, fish population dynamics, composite indicators, hydrocarbon exploration models, macroeconomic modelling, and radioactive waste management. US Nuclear Regulatory Commission (2013) regularly performs uncertainty and sensitivity analyses in its assessments (http://sesitivity-analysis.ec.europa.eu). The European Safety and Reliability Association (ESRA) has established a Technical Committee on Uncertainty Analysis (http://www.esrahomepage.org/uncertainty.aspx) whose aim is to foster research on new methodologies and innovative applications of Uncertainty and Sensitivity Analysis of simulation models.

Potential contribution to the main steps of uncertainty analysis

Steps in uncertainty analysis	Potential contribution of this approach
Identifying uncertainties	Not applicable. (but: some methods can be used to prioritize among long list of sources of uncertainty)
Describing uncertainties	Not applicable.
Assessing the magnitude of individual uncertainties	Not applicable.
Assessing the combined impact of multiple uncertainties on the assessment output, taking account of dependencies	Not applicable.
Assessing the contribution of individual uncertainties to overall uncertainty	Yes. Sensitivity Analysis methods allow investigating input factors in order to identify those that are more influential on the output. Some methods are not able to quantify the joint effects of all the inputs when evaluating the sensitivity of a single one (i.e. they do not account for higher order interactions among inputs). Sometimes methods are used to screen the inputs in a very preliminary stage in order to prioritize a subsequent more refined analysis of the uncertainty (e.g. scatter plots, mathematical methods)

Melamine example

The melamine risk assessment as published by EFSA (2008) compares calculated exposure to melamine in different scenarios with a previously established tolerable daily intake (TDI) and presents the ratio of exposure to TDI as the decision variable. Calculations are deterministic and based on different point estimates, including medians, means and 95th percentiles.

In this example, different possible approaches for the risk assessment and the uncertainty analysis are considered, in order to present various methods for the sensitivity analysis.

The risk assessment model includes two calculation steps, to calculate exposure (e) and to calculate the risk ratio (r):

$e = c * w * q / bw$ (1)

$r = e / tdi$ (2)

with

c: concentration of melamine in milk powder (mg/kg)

w: weight fraction of milk powder in chocolate (-)

q: consumption of chocolate (kg/day)

bw: body weight of children (kg)

tdi: Tolerable Daily Intake (mg/kg/day)

e: exposure (mg/kg/day)

r: risk ratio (-)

When assessing uncertainty, the computation can be performed using a deterministic or probabilistic approach. The same approaches can be adopted to perform a sensitivity analysis.

For the purpose of uncertainty analysis all types of information and assumptions fed into the assessment could potentially cause variation in the output and therefore should be assessed for their influence. However in this section and the example on melamine, because of the illustrative purpose, we consider as relevant inputs only parameters and variables used in the risk assessment models used to calculate exposure and risk ratio.

Example based on NRSA method

The basis for this example is given by assessment of uncertainty done in section B.7 using interval analysis method. In that section interval values for the uncertain worst case of the input factors were provided as in Table B.16.3

Table B.16.3: Child 1 year old, uncertainty about the worst case (wc) values for parameters.

Parameter/Estimate	Favored value for worst case	Lower bound for wc value	Higher bound for wc value
C_{mel} (mg/kg)	2563	2563	5289
$w_{milk\text{-}powder}$ (-)	0.28	0.28	0.30
$q_{chocolate}$ (kg/d)	0.05	0.05	0.1
bodyweight (kg-bw)	6	5.5	6.5

The Nominal Range Sensitivity Analysis method (Table B.16.4) provides an index to identify input factors that are more influential on the estimated exposure of melamine and on the relative risk (not computed since would provide same results in a different scale).

Table B.16.4: Nominal range sensitivity analysis index for the model input factors.

Parameter/Estimate	$E_{melanine}$ at nominal value of X_i (a)	$E_{melanine}$ at minimum value of X_i and nominal value of the other inputs (b)	$E_{melanine}$ at maximum value of X_i and nominal value of the other inputs (c)	NRSA (c-b)/a
C_{mel} (mg/kg)	6	6	12.34	1.06
$w_{milk\text{-}powder}$ (-)	6	6	6.40	0.07
$q_{chocolate}$ (kg/d)	6	6	12	1

bodyweight (kg-bw)	6	5.52	6.52	0.17

The ranking of the input factors in terms of their influence on the output is as follows: 1. melamine concentration in adulterated milk powder; 2. consumption of chocolate on an extreme day; 3. Body weight; 4. weight fraction of milk powder in chocolate. Consequently the first two variables are those for which a reduction in the uncertainty should be achieved in order to reduce uncertainty in the output.

Example based on Break-even analysis

The example on the use of a Break-even analysis for sensitivity analysis is based on the uncertainty intervals previously established for the worst case of the concentration of melamine in adulterated milk powder and consumption of chocolate on an extreme day input factors. No uncertainty is assumed for the worst case of the other two factors (weight fraction of milk powder in chocolate and body weight) that are kept at their nominal values due to their reduced influence on the model outcome (Table B.16.5).

Table B.16.5: Child 1 year old, uncertainty about the worst case (wc) values for parameters.

Parameter/Estimate	Favored value for worst case	Lower bound for wc value	Higher bound for wc value
c (mg/kg)	2563	2563	5289
q (kg/d)	0.05	0.05	0.1
bw (kg/bw)	6	6	6
w (-)	0.28	0.28	0.28

Therefore break-even analysis focuses only on the most influential factors previously identified (Table B.16.6).

Table B.16.6: Break-even analysis for *uncertain worst case* chocolate consumption and melamine concentration in milk powder - *Child 1 year old.*

		Chocolate consumption (q)					
		0.05	0.06	0.07	0.08	0.09	0.1
Menalmine Concentration (c)	2563	5.98	7.18	8.37	9.57	10.76	11.96
	3108.2	7.25	8.70	10.15	11.60	13.05	14.50
	3653.4	8.52	10.23	11.93	13.64	15.34	17.05
	4198.6	9.80	11.76	13.72	15.67	17.63	19.59
	4743.8	11.07	13.28	15.50	17.71	19.92	22.14
	5289	12.34	14.81	17.28	19.75	22.21	24.68

The result of the BEA is trivial for this example since clearly in the worst case scenario for chocolate consumption and melamine concentration, the exposure exceeds the TDI by various folds. The results of the analysis would have been informative in case the TDI was, for instance, equal to 10 mg/kg.

In this case, it would be possible to indicate to policy makers which maximum level should be fixed by regulation for melamine concentration to avoid exceeding the TDI given a specific worst case scenario for chocolate consumption. In case, for instance, of a worst case consumption of 0.07 kg/day, a level of 3108 mg/kg melamine should be indicated to regulators as the highest possible level to avoid safety concern in 1 year children eating very high quantity of chocolate. The same approach could be used to identify a possible target of reduction of the amount of chocolate consumed by children with high intake, in case the melamine concentration is kept fixed at the current use level.

This example shows the potential value of sensitivity analysis to inform decisions of risk managers.

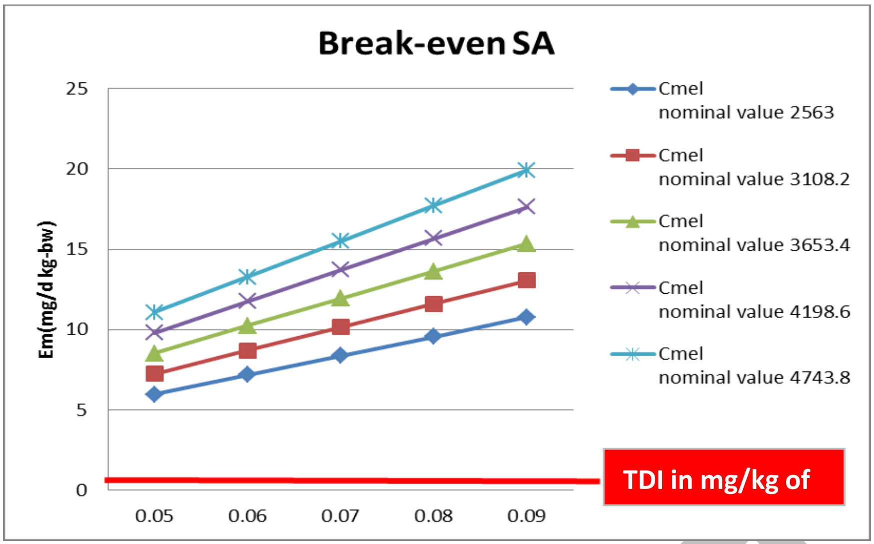


Figure B.16.2: Results of break-even sensitivity analysis

Example based on Morris method for sensitivity analysis

Table B.16.7 presents the input distributions, used for the Morris and Sobol methods. These are based on the outputs of the 2d Monte Carlo simulation, by taking the medians of the uncertainty distributions of the mean and standard deviation of the variability distributions for 1 year old children. These were then converted in parameters for the distributions used in the global sensitivity analysis. As in other examples, uncertainty in the TDI was not considered. For both methods, the distributions were truncated at the 0.1 and 99.9 percentiles to prevent a strong influence of extreme values.

Table B.16.7: Distribution of input factors for computation of exposure distribution.

Input factor	Description	Unit	Mean	Std	Range	Distribution
C	Concentration of melamine in milk powder	mg/kg	232	627	--	LN(4.34, 0.1.46)
W	Weight fraction of milk powder in chocolate	-		--	(0.14,0.30)	U(0.14,0.30)
Q	Consumption of chocolate	kg/day	0.0142	0.0134		Γ(1.12, 79.1, 0]
Bw	Body weight of children	Kg	11.00	1.53	--	LN(2.39, 0.0.138]
Tdi	Tolerable Daily Intake	mg/kg/day	0.50	--	Constant	Constant

Results of the Morris method are given in table B.16.8 and figure B.16.3 below. For this linear model, the mean of the elementary effects *(μ_i)* and the mean of the absolute values of the elementary effects *(μ^*_i)* are the same for all input factors except body weight. All input factors have (almost) linear effects and there are limited interactions among factors (measured by the standard error of the elementary effects - σ_i), as expected from the simple model structure. The risk ratio *r* is most sensitive to variations in *c* and *q* and least sensitive to variations in *bw*. The blue and red lines in the Morris graph (Figure B.16.3) indicate proposed qualitative thresholds where factors' main influence is in the form of direct effects (below the line) or higher order/interactions (above the line). The red line was proposed originally by Morris (1991) for μ_i and the blue line by Muñoz-Carpena et al. (2007) and Chu-Agor et al. (2012)for μ^*_i. The results indicate that there are non-linear effects for all factors.

Table B.16.8. Mean and standard deviation of elementary effects of input factors in the melamine model on the risk ratio *r*, according to the method of Morris (60 samples).

Input factor	μ_i*	μ_i	σ_i
C	0.20	0.20	0.19
W	0.05	0.05	0.08
Q	0.14	0.14	0.17
Bw	0.02	-0.02	0.02

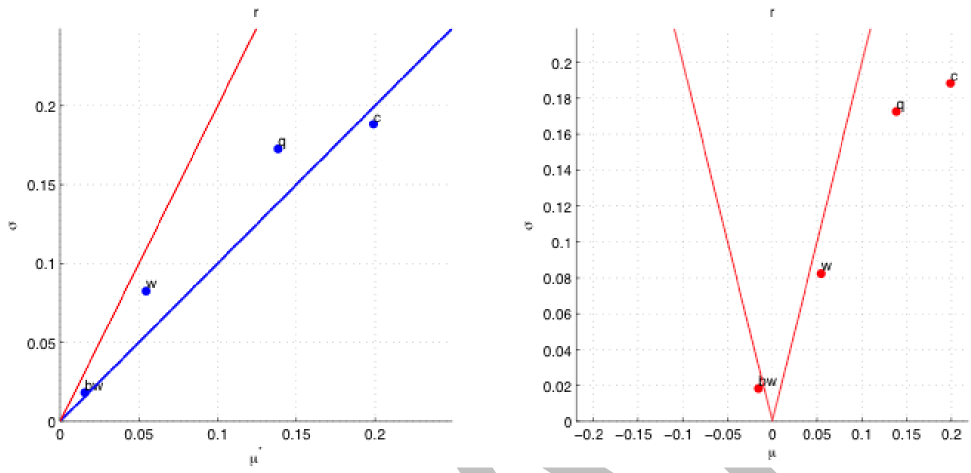


Figure B.16.3: Elementary effects of input factors in the melamine model on the risk ratio *r*, according to the method of Morris (160 samples). See text for explanation of red and blue lines.

Example based on Monte Carlo filtering

For the melamine example, a natural threshold value for the risk ratio, set by risk managers or stakeholders would be *r* = 1 but, since only few realizations of such values were observed, we chose a threshold of *r* = 0.1. Figure B.16.4 shows the MCF results for *q* and *c*, the two input factors with the greatest influence on the model output variance, as identified by the Sobol method. According to the Smirnov test, *c* and *q* distributions are significantly different and the figure demonstrates that the probability density functions (pdfs) of *c* are more separated than those of *q*, indicating that a management intervention to reduce the concentration of melamine in chocolate might be more effective than reducing chocolate consumption. The intersection of the two distributions for *c* is at ~ 100 mg/kg, hence above the median but below the mean of the input distribution. The intersection of the two distributions for q is at 0.009 g/day, somewhat lower than the mean consumption. This implies that an intervention (policy, regulation) to limit values of *c* and *q* at the threshold identified (c < 100mg/kg and q < 0.009 g/day) would result in the reduction of the risk of children being exposed to more that 10% of the TDI. This illustrates the opportunities of this analysis to transfer the results to risk managers. This result must be considered within the ranges specified for these input factors.

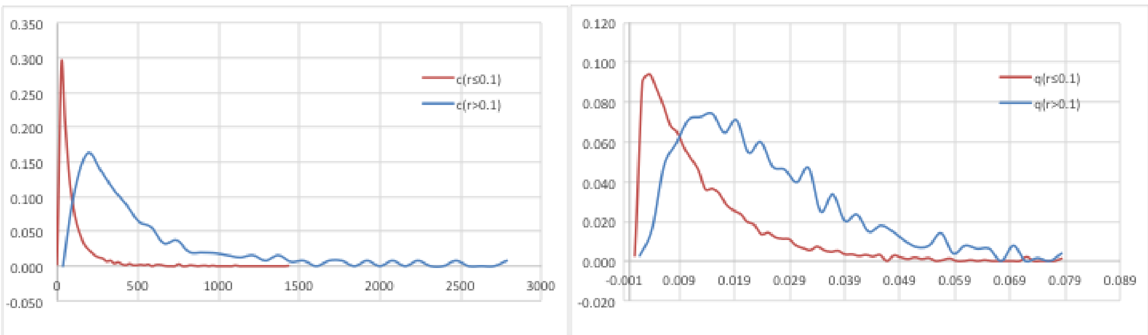


Figure B.16.4: Monte Carlo filtering for melamine example: pdf's of c and q producing favorable (r <= 0.1) or unfavorable (r > 0.1) results.

Example using Sobol Index

For the melamine example, the variance decomposition is shown in Table B.16.9. The sum of the first-order indices is ΣS_i = 0.74 > 0.6,, indicating the model behaves as a mostly additive model for this simple application. Again, the model outputs are most sensitive to variations in *c* (54% of the total model variance) and to a lesser extent to *q* (19%). Variations in *w* and *bw* hardly affect the model results.

Table B.16.9: Variance decomposition of input factors in the melamine model in relation to the risk ratio *r*, according to the method of Sobol (5120 samples, *M*=512).

Input	First-order index	Total order index	Interaction index
c	0.54	0.82	0.28
w	0.01	0.03	0.02
q	0.19	0.46	0.27
bw	0.00	0.00	0.00

The Sobol method is based on an efficient Monte Carlo sampling algorithm, exploring the joint parameter space instead of the marginal distributions. Therefore, even though the number of samples is limited, the results can directly be used for uncertainty analysis by reading the Cumulative Density Function (CDF) from the samples of the model Y = f(X1, X2, …, Xk). In the melamine example, the uncertainty in *r* is graphically represented as in Figure B.16.5. In this example, the uncertainty should be interpreted as due to variability in the input factors. To include uncertainty in the variability distributions of the input factors, their parameters should be described by probability distributions as in a 2D Monte Carlo simulation. Based on the results of the analysis of variability, parameter uncertainties would only need to be specified for *q* and *c*.

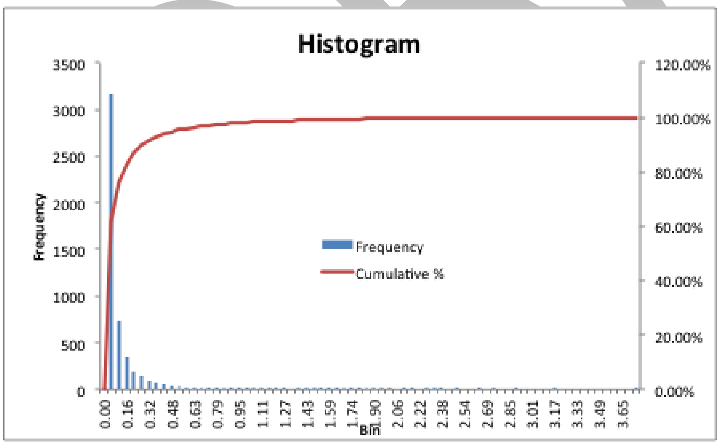


Figure B.16.5. Model output uncertainty pdf for risk ratio *r* (x-axis) (N=5120 samples)

Sensitivity analysis in the melamine example: general considerations

Irrespective of the method used to perform sensitivity analysis, the ranking of the input factors according to their influence on the output of the model is extremely robust. Melamine concentration and chocolate consumption are the variables largely explaining the variability/uncertainty of the exposure and the risk ratio. In a real assessment this result could be communicated to risk managers and support an informed decision about actions to reduce exposure and risk.

The separation of variability and uncertainty in sensitivity analysis methodology is not well established yet. Therefore it has not been proposed in this example. Further research is needed in this direction.

Strengths

1. Provide extremely valuable information for making recommendations to policy makers (e.g. identifying factors on which it is more effective to concentrate resources and actions in order to reduce risk)
2. allows prioritization of parameters for uncertainty analysis and/or further research
3. some methods are very easy to implement and understand (e.g. nominal range methods)

Weaknesses and possible approaches to reduce them

1. When Risk Assessment involves many model parameters, sensitivity analysis can be quite computationally intense. Screening of input factors (e.g. using graphical methods or method of Morris) can be used to reduce dimensionality;
2. Some methodologies rely on assumptions related to relationship between inputs and output (e.g. linearity) and among inputs (e.g. independence). When these assumptions do not hold, conclusions of the SA can be misleading; methods that are able to address non linearity and dependency should be preferred in these cases.
3. It is necessary to clarify prior to start the sensitivity analysis which question it is intended to answer, otherwise its value could be limited and not addressing the informative needs
4. Generally it is not possible to separate influence of each input on the output in terms of variability and uncertainty of the input separately. Only methods recently developed allow so (Busschaert et al. 2011).
5. The sensitivity analysis has been already occasionally applied in EFSA. Still a regular application (especially when models are used as a basis for the assessment) is not in place. The application of scenario analysis (conditional sensitivity analysis) is more frequent but not a common practice.
6. Training should be provided to staff and experts in order to facilitate the performance of sensitivity analysis. This training should include guidance on preferable methods to be included in different domains/scientific assessment types.

Assessment against evaluation criteria

There is a large variability in the nature and complexity of the methods that can be used to perform a sensitivity analysis. Consequently it was decided to have two tables assessing deterministic (Table B.16.10) and probabilistic methods (Table B.16.11) separately against evaluation criteria. The item 'meaning of output' was deliberately not filled in since sensitivity analysis complements uncertainty analysis without providing a direct measure of it.

Conclusions

1. Sensitivity analysis can represent a valuable complement of uncertainty assessment in EFSA. It helps assessors in providing risk managers with information about most influential factors on which to focus actions and further research.
2. It has potential for applicability in any area of work in EFSA.

3. Obstacles to application of the method could be technical complexity and the need to involve an experienced statistician in the computation and interpretation of some specific methods. Training should be provided to staff and experts in order to facilitate the performance of sensitivity analysis.

4. It is necessary to clarify prior to start the sensitivity analysis which question it is intended to reply, otherwise its value could be limited and not addressing the informative needs.

References

Augustin J-C (2011): Global sensitivity analysis applied to food safety risk assessment. Electron. J. App. Stat. Anal., Vol. 4, Issue 2, 255 – 264

Busschaert P., Geeraerd A.H., Uyttendaele M., Van Impe J.F. (2011): sensitivity analysis of a two-dimensional quantitative microbiological risk assessment: keeping variability and uncertainty separated. Risk analysis, 31, 1295-1307

Chu-Agor ML, Munoz-Carpena R, Kiker GA, Aiello-Lammens ME, Akcakaya HR, Convertino M, et al. Simulating the fate of Florida Snowy Plovers with sea-level rise: Exploring research and management priorities with a global uncertainty and sensitivity analysis perspective. Ecol Model. 2012;224(1):33-47.

Committee on Decision Making Under Uncertainty (2013): Environmental decision in the face of uncertainty. Board on Population Health and Public Health Practice; Institute of Medicine. National Academies Press (US)

Crick, M.J.; Hill, M.D. (1987): The role of sensitivity analysis in assessing uncertainty. NEA workshop on uncertainty analysis for performance assessments of radioactive waste disposal systems; Seattle, WA (USA); 24-26 Feb 1987. Report of Nuclear Energy Agency, 75 - Paris (France); 258 p; 1987; p. 116-129.

Cukier R. I., Fortuin C. M., Shuler K. E., Petschek A. G. and Schaibly J. H., Chern J (1973): Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory. The Journal of Chemical Physics, 59(8): 3873-3878.

Cullen AC and Frey HC (1999): Probabilistic techniques in exposure assessment. Plenum Press. New York

ECHA (2012): Guidance on information requirements and chemical safety assessment. Chapter R19: Uncertainty analysis. ECHA-12-G-25-EN. http://echa.europa.eu/documents/10162/13632/information_requirements_r19_en.pdf

EFSA Panel on Plant Protection Products and their Residues (PPR); Guidance on the Use of Probabilistic Methodology for Modelling Dietary Exposure to Pesticide Residues. EFSA Journal 2012;10(10):2839. Available online: www.efsa.europa.eu/efsajournal

EFSA Panel on Animal Health and Welfare (2007): Framework for the AHAW Risk Assessments. EFSA Journal 2007; 550, 1-46.

EFSA (2014): Analysis of the baseline survey on the prevalence of Listeria monocytogenes in certain ready-to-eat foods in the EU, 2010-2011. Part B: analysis of factors related to prevalence and exploring compliance. EFSA Journal 2014;12(8):3810, 73 pp. Available online: www.efsa.europa.eu/efsajournal

Ekstrom P. (2005): EIKOS a simulation toolbox for sensitivity analysis. Degree project 20p. Uppsala University. Faculty of Science and Technology. The UTH_Unit.

Frey H C and Patil S R (2002): Identification and review of sensitivity analysis methods, Risk Analysis, vol. 22(3):553-78

Homma T and Saltelli A (1996). Importance measures in global sensitivity analysis of nonlinear models. Reliability Engineering & System Safety, 52(1):1-17.

Iman, R.L.; Conover, W.J. (1980): Small sample sensitivity analysis techniques for computer models, with an application to risk assessment. Communications in Statistics – Theory and Methods. A9:1749-1842.

Iman, R.L.; Helton, J.C. (1988): An investigation of uncertainty and sensitivity analysis techniques for computer models. Risk Analysis. 8:71-90

Khare YP, Muñoz-Carpena R, Rooney RW, Martinez CJ. A multi-criteria trajectory-based parameter sampling strategy for the screening method of elementary effects. Environmental Modelling & Software. 2015;64(0):230-9.

Le Sage, T. (2013). Scenario based risk assessment. JDiBrief Series. London: UCL Jill Dando Institute of Security and Crime Science. ISSN: 2050-4853.

Morris MD. Factorial sampling plans for preliminary computational experiments. Technometrics. 1991;33:161-74.

Muñoz-Carpena R, Zajac Z, Kuo YM. Evaluation of water quality models through global sensitivity and uncertainty analyses techniques: application to the vegetative filter strip model VFSMOD-W. Trans ASABE. 2007;50:1719-32.

Oakley J.E. & O'Hagan A. (2004): Probabilistic sensitivity analysis of complex models: a Bayesian approach. J. R. Stat. Soc. Serie B, 66, part3, 751-769.

Pannell, D.J. (1997). Sensitivity analysis of normative economic models: Theoretical framework and practical strategies, Agricultural Economics 16: 139-152

Patil S.R. & Fray H.C. (2004): Comparison of sensitivity analysis methods based on applications to a food safety risk assessment model. Risk Analysis, 24, 3, 573-585.

US Nuclear Regulatory Commission (2013). Risk assessment in regulation. http://www .nrc.gov/about-nrc/regulatory/risk-informed.html

Saltelli A., S. Tarantola and K.P.S. Chan (1999). A quantitative model-independent method for global sensitivity analysis of model output. Technometrics, 41, 39–56

Saltelli A, Tarantole S. Campolongo F, Ratto M. (2004): Sensitivity analysis in practice. A guide to assessing scientific models. Wiley-Interscience.

Saltelli A, Ratto M, Tarantola S, Campolongo F. (2005): Sensitivity analysis for chemical models. Chemical reviews. 105(7):2811-28.

Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, et al. (2008): Global sensitivity analysis: the primer. Wiley-Interscience;.

Sobol IM. Sensitivity estimates for nonlinear mathematical models. Translated as I M Sobol 1993 Sensitivity analysis for non-linear mathematical models Math Modeling Comput Experiment, 1: 407-414. 1990;2:112-8.

Stiber NA, Pantazidou M, Small MJ (1999): Expert system methodology for evaluating reductive dechlorination at TCE sites, Environmental Science and Technology, 33(17), 3012-20.

Tian W. (2013): A review of sensitivity analysis methods in building energy analysis. Renewable and sustainable energy reviews 20, 411-419

Table B.16.10: Assessment of Deterministic methods for sensitivity analysis (when applied well) against evaluation criteria.

Criteria	Evidence of current acceptance	Expertise needed to conduct	Time needed	Theoretical basis	Degree/ extent of subjectivity	Method of propagation	Treatment of uncertainty and variability	Meaning of output	Transparency and reproducibility	Ease of understanding for non-specialist
Stronger character-istics	International guidelines or standard scientific method	No specialist knowledge required	Hours	Well established, coherent basis for all aspects	Judgement used only to choose method of analysis	Calculation based on appropriate theory	Different types of uncert. & var. quantified separately	Range and probability of alternative outcomes	All aspects of process and reasoning fully documented	All aspects fully understandable
	EU level guidelines or widespread in practice	Can be used with guidelines or literature	Days	Most but not all aspects supported by theory	Combination of data and expert judgment	Formal expert judgment	Uncertainty and variability quantified separately	Range and relative possibility of outcomes	Most aspects of process and reasoning well documented	Outputs and most of process understandable
	National guidelines, or well established in practice or literature	Training course needed	Weeks	Some aspects supported by theory	Expert judgment on defined quantitative scales	Informal expert judgment	Uncertainty and variability distinguished qualitatively	Range of outcomes but no weighting	Process well documented but limited explanation of reasoning	Outputs and principles of process understandable
	Some publications and/or regulatory practice	Substantial expertise or experience needed	A few months	Limited theoretical basis	Expert judgment on defined ordinal scales	Calculation or matrices without theoretical basis		Quantitative measure of degree of uncertainty	Limited explanation of process and/or basis for conclusions	Outputs understandable but not process
Weaker character-istics	Newly developed	Professional statistician needed	Many months	Pragmatic approach without theoretical basis	Verbal description, no defined scale	No propagation	No distinction between variability and uncertainty	Ordinal scale or narrative description for degree of uncertainty	No explanation of process or basis for conclusions	Process and outputs only understandable for specialists

EFSA Journal

Table B.16.11: Assessment of Probabilistic methods for sensitivity analysis (when applied well) against evaluation criteria.

Criteria	Evidence of current acceptance	Expertise needed to conduct	Time needed	Theoretical basis	Degree/ extent of subjectivity	Method of propagation	Treatment of uncertainty and variability	Meaning of output	Transparency and reproducibility	Ease of understanding for non-specialist
Stronger characteristics	International guidelines or standard scientific method	No specialist knowledge required	Hours	Well established, coherent basis for all aspects	Judgement used only to choose method of analysis	Calculation based on appropriate theory	Different types of uncert. & var. quantified separately	Range and probability of alternative outcomes	All aspects of process and reasoning fully documented	All aspects fully understandable
	EU level guidelines or widespread in practice	Can be used with guidelines or literature	Days	Most but not all aspects supported by theory	Combination of data and expert judgment	Formal expert judgment	Uncertainty and variability quantified separately	Range and relative possibility of outcomes	Most aspects of process and reasoning well documented	Outputs and most of process understandable
	National guidelines, or well established in practice or literature	Training course needed	Weeks	Some aspects supported by theory	Expert judgment on defined quantitative scales	Informal expert judgment	Uncertainty and variability distinguished qualitatively	Range of outcomes but no weighting	Process well documented but limited explanation of reasoning	Outputs and principles of process understandable
	Some publications and/or regulatory practice	Substantial expertise or experience needed	A few months	Limited theoretical basis	Expert judgment on defined ordinal scales	Calculation or matrices without theoretical basis		Quantitative measure of degree of uncertainty	Limited explanation of process and/or basis for conclusions	Outputs understandable but not process
Weaker characteristics	Newly developed	Professional statistician needed	Many months	Pragmatic approach without theoretical basis	Verbal description, no defined scale	No propagation	No distinction between variability and uncertainty	Ordinal scale or narrative description for degree of uncertainty	No explanation of process or basis for conclusions	Process and outputs only understandable for specialists

Annex C – Further details for the melamine case study

C.1 Quantitative model

The basic risk assessment model for the case study includes two calculation steps, to calculate first exposure (e):

$$e = \frac{c \times w \times q}{bw}$$

and then the risk ratio (r): r = e/TDI. The quantities involved in these calculations are:

c	concentration of melamine in milk powder	(mg/kg)	Input variable (dist'n uncertain)
w	weight fraction of milk powder in chocolate	(-)	Input variable (dist'n uncertain)
q	consumption of chocolate	(kg/day)	Input variable (dist'n uncertain)
bw	body weight of children	(kg)	Input variable (dist'n uncertain)
TDI	Tolerable Daily Intake	(mg/kg/day)	Specifed value (but there is uncertainty about whether it is the correct value)
e	exposure	(mg/kg/day)	Output variable (dist'n uncertain)
r	risk ratio	(-)	Output variable (dist'n uncertain)

Two versions of the example are considered: uncertainty about the highest exposure occurring (worst-case) and uncertainty about variability of exposure. For the first version, the issue of variability has been removed by considering the worst case so that there is only uncertainty to be addressed. For the second, both variability and uncertainty need to be addressed.

In the Interval Analysis example (annex B.7.), the worst-case assessment is considered for all children before considering sub-groups to address dependence between body-weight and consumption. In the other quantitative method examples, attention is restricted to children aged from 1 up to 2 years. An advantage of doing so is that very simple statistical models can be used to illustrate the statistical methods of statistical inference.

C.2 Worst-case assessment (uncertainty but no variability)

The worst-case value for the risk-ratio is $r_{max} = e_{max}/\mathrm{TDI}$ where

$$e_{max} = \frac{c_{max} \times w_{max} \times q_{max}}{bw_{min}}$$

The new quantities involved in these calculations are:

r_{max}	Highest occurring value for the risk ratio	(-)	Output parameter (value uncertain)
e_{max}	Highest occurring exposure	(mg/kg/day)	Ouput parameter (value uncertain)
c_{max}	Highest occurring concentration of melamine in milk powder	(mg/kg)	Input parameter (value uncertain)
w_{max}	Highest occurring weight fraction of milk powder in chocolate	(-)	Input parameter (value uncertain)
q_{max}	Highest occurring consumption of chocolate	(kg/day)	Input parameter (value uncertain)
bw_{min}	Lowest occurring body weight of children	(kg)	Input parameter (value uncertain)

C.3 Uncertainty about variability of exposure

Attention was further restricted to children consuming contaminated chocolate from China.

For each of the input variables, a parametric family of distributions was chosen with which to model the variability. In the cases of *q* and *bw*, the choice of distribution family was informed by analysis of the data. For *c* and *w*, the choices were pragmatic ones made for illustrative purposes. Each of the parameters introduced in this table is uncertain and uncertainty about the values of the parameters is the way in which we address uncertainty about the variability for each variable. Details are given in the following table:

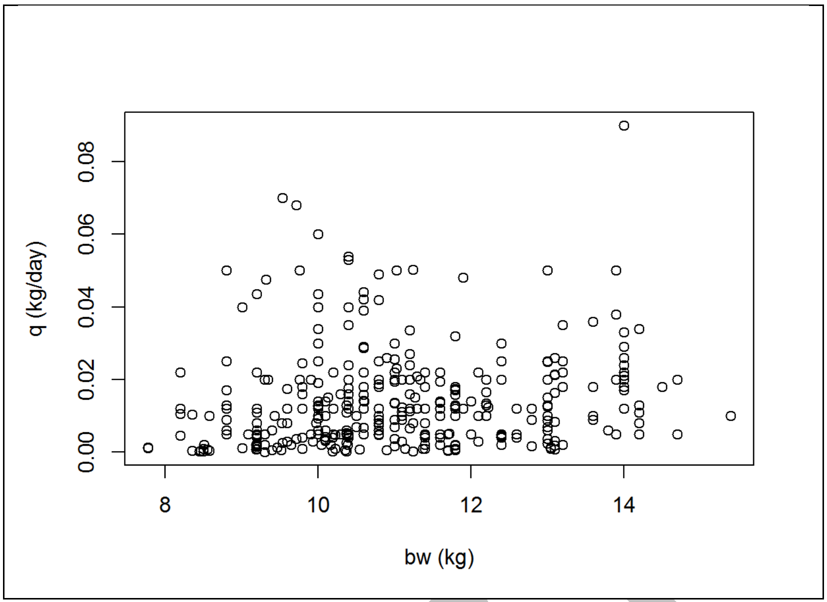
Variable	Distribution family	Parameters (statistical)	Meaning of parameters
c	Log-normal distribution (base 10)	$\mu_{\log c}$ and $\sigma_{\log c}$	Mean and standard deviation of log-concentration
w	Uniform distribution	a_w and b_w	Lower and upper limit for weight-fraction
q	Gamma distribution	α_q and β_q	Shape and rate parameters for gamma distribution for q
bw	Log-normal distribution (base 10)	μ_{logbw} and σ_{logbw}	Mean and standard deviation of log-bod-weight

Data used for modelling variability of body-weight and consumption

For *q* and *bw*, consumption survey data were available, for 1 year old children, from EFSA (http://www.efsa.europa.eu/en/datexfoodcdb/datexfooddb.htm) and which existed in 2008. The data derive from 5 surveys carried out in Finland, Germany, Italy, Poland and Spain. They record daily consumption (weight) of "Chocolate (cocoa) products". Restricting to records with positive consumption, they provide 362 values of q for 171 children and the value of bw for each child.

Standard goodness-of-fit tests show that the log-normal family of distributions is a better fit to the bw data than either the normal or gamma families. The log-normal fit is visually excellent although it does formally fail the tests. For q, the gamma family fits better than normal, log-normal or Weibull and the visual fit is again good.

The plot below shows the relationship between q and bw for the data used. The correlation is statistically significant, with or without logarithmic transformation of variables, but nevertheless small: 0.13 for the raw data and 0.24 after logarithmic transformation of both variables. Since the examples are intended primarily to illustrate the methods and not to be a complete assessment of uncertainty for the melamine case study and incorporating dependence into the examples in annex B would involve considerable extra complexity, variability of b and q is treated as independent in the examples of probability bounds analysis and Monte Carlo.


q (kg/day)
0.00
0.02
0.04
0.06
0.08
8
10
12
14
bw (kg)

Annex D – Case studies in combining methods for the purpose of characterising overall uncertainty

This annex is not yet available but will in due course provide case studies showing how the methods proposed in the guidance may be combined for the purpose of characterising overall uncertainty for an assessment. The case studies will demonstrate a number of approaches of varying complexity and suitable for different situations. Each case study will relate to the melamine example discussed in Annexes A and C and used to provide examples for methods in Annex B.