FSA Project Number T01056

DEVELOPMENT OF A FRAMEWORK FOR EVALUATION AND EXPRESSION OF UNCERTAINTIES IN HAZARD AND RISK ASSESSMENT

Andy Hart and John Paul Gosling, The Food and Environment Protection Agency[1]

Alan Boobis, Imperial College London

David Coggon, University of Southampton

Peter Craig, University of Durham

David Jones, University of Leicester

December 2010

Summary

In its report on variability and uncertainty, the Committee on Toxicity of Chemicals in Food, Consumer Products and the Environment (COT) concluded that the development of a framework for transparent expression of uncertainty in hazard characterisation would enable COT and other committees that perform toxicological evaluations to improve communication of the sources of variability and uncertainty in their risk assessments. This project reviewed existing approaches for qualitative evaluation and expression of uncertainties and assessed their suitability for routine use by committees like the COT. The theoretical basis for different ways of expressing and combining uncertainties was also considered. It was concluded that different approaches would be required for evaluating uncertainty, depending on whether the hazard or risk assessment addressed a categorical question (e.g. is this chemical an allergen?) or a quantitative question (e.g. determination of a reference dose or estimation of exposure). Promising approaches were combined and adapted to form a draft framework for assessing uncertainty, which was then evaluated in a workshop with members of COT and other potential users, by applying it retrospectively to four assessments previously published by COT. Feedback from the workshop and subsequent COT meetings was used to adapt and refine the framework. Further work is required to evaluate application of the framework to different types of assessments, to develop effective approaches for communicating uncertainty to decision-makers and other stakeholders, and to develop further the mathematical underpinning for the framework. This report summarises the work of the project, and includes the final draft of the framework and a brief worked example relating to caffeine. The annexes include a 2-page summary of the framework, the report of the COT workshop and a draft paper on a mathematical framework for evaluating uncertainties in assessments of quantitative questions.

[1] Contact: Dr Andy Hart, Fera, Sand Hutton, York, YO41 1LH, UK; andy.hart@fera.gsi.gov.uk; 01904 462053

Contents

Introduction and objectives

In its report on variability and uncertainty, the COT (2007) recommended that "hazard identification and characterisation should take into account variability and uncertainty, using a systematic approach that will facilitate transparency and confidence". Similar goals have been expressed by other authorities at national, European (Madelin 2004, EFSA 2006, 2009, ECHA 2008) and international levels (IPCS/WHO 2008, Codex 2010), and in the scientific aims of the FSA's research program on risk assessment of food chemicals (T01). The COT concluded that "the development of a framework for transparent expression of uncertainty in hazard characterisation would enable COT and other committees that perform toxicological evaluations to improve communication of the sources of variability and uncertainty in their risk assessments".

The basic motivation for addressing uncertainty is twofold. First, it is fundamentally necessary for risk management. Information on the relative likelihood of alternative outcomes is necessary to enable decision-makers to choose policies that increase the probability of favourable outcomes, and reduce the probability of unfavourable ones: in other words, to manage risk. As Cicero wrote, 'Probabilities direct the conduct of the wise man'[2].

Second, addressing uncertainty is important for the credibility of science-based policy. As illustrated by the BSE crisis, 'To establish credibility it is necessary to generate trust. Trust can only be generated by openness. Openness requires recognition of uncertainty, where it exists.' (Phillips 2000).

This project aimed to address the COT's requirement by developing and testing a framework for expression of uncertainties that is suitable for use by COT and other committees such as COC and COM, and by FSA itself. The project focussed on qualitative or semi-quantitative approaches, because there is already an extensive literature on quantitative methods for evaluating uncertainty (e.g. sensitivity analysis and probabilistic modelling), and it is expected that qualitative evaluation of uncertainties will be sufficient and more practical for the majority of COT assessments. A qualitative framework will also be useful for targeting quantitative analysis on the most important sources of uncertainty, when required.

Existing approaches to evaluating uncertainty

As a first step, a review was conducted to identify existing approaches that might be suitable for use by the COT, or as a starting point for developing an improved framework. Potentially relevant approaches were identified via searches of the literature (Web of Science) and internet, contacts with relevant EU, US and Canadian authorities, and the authors' previous experience.

Criteria for assessing the suitability of different approaches for the COT were drafted in consultation with FSA. It was considered that suitable approaches should be:

- practical for use by the COT and other FSA committees, and adapted to their work,
- systematic & comprehensive, helping the user address all relevant uncertainties,

[2] Marcus Tullius Cicero (106-43 BC): De Natura Deorum, Book I, Chapter 5, Sec.12.

- efficient, using a tiered approach to minimise the effort required,
- helpful for developing conclusions and in subsequent decision-making, by evaluating uncertainties in terms of their impact on the key issues,
- conceptually compatible with mathematical approaches to uncertainty.

None of the existing approaches that were identified by the review fully met all the criteria, but several contain potentially useful features and these are summarised below.

Similar approaches that meet most of the review criteria have been published by EFSA (2006) and ECHA (2008). All parts of the assessment are examined for uncertainties, which are listed in a table. The individual and combined impact of the uncertainties is evaluated, with a scoring system using plus and minus symbols to indicate the direction and magnitude of their potential effects on the assessment outcome. In practice, however, users rarely define a quantitative scale for these symbols, which reduces their interpretability. The approach was originally developed for use in exposure assessment and has also been used for toxicity and risk, but there is no guidance on how to apply it to assessments of categorical questions (e.g. does this chemical cause allergic sensitisation?).

Draft guidance documents currently under development by Health Canada Contaminated Sites Division (Hans Yu, personal communication) state that sensitivity analysis for a deterministic risk assessment should consist, at a minimum, of a qualitative summary of the uncertainties and variability associated with each input variable and a prediction of how these uncertainties are expected to affect the risk estimates. Example assessment extracts provided by Health Canada variously included narrative text sections discussing uncertainties in each part of the assessment (exposure, toxicity, risk characterisation), tables listing and justifying key assumptions and describing their potential impact on the assessment (e.g. neutral, underestimate, overestimate, unknown) and, in one example, a summary statement on the reasonableness of the assessment and its degree of conservatism.

The US Nuclear Regulatory Commission has published two guidance documents on treatment of uncertainties associated with probabilistic risk assessment (NUREG-1855, 2009 and EPRI, 2009). Both concentrate mainly on quantitative approaches but include qualitative screening steps, that are used to identify uncertainties that may require sensitivity analysis or probabilistic modelling. NUREG-1855 states that the final output should include a qualitative statement of confidence in the conclusion of the assessment and how it has been reached, supported by identification of key uncertainties that were addressed. EPRI (2009) includes a tabular format for listing sources of model uncertainty and narrative evaluations of their impact on the model and risk characterisation/degree of conservatism.

The International Agency for Research on Cancer has established systematic procedures for evaluating carcinogenicity of chemicals (IARC, 2006). Human and animal studies are evaluated and classified as to the level of evidence of carcinogenicity they provide (sufficient, limited or inadequate evidence, or evidence suggesting lack of carcinogenicity). These are then considered together with other relevant data, including evidence from studies on the mechanism of effects, to reach an overall conclusion classifying the chemical in one of 5 categories: 1, Carcinogenic to humans; 2A, Probably carcinogenic to humans; 2B, Possibly carcinogenic to humans; 3, Not classifiable as to its carcinogenicity to humans; 4, Probably not carcinogenic to humans.

Schutz et al. (2008) propose the use of 'evidence maps' as a simple graphical format for summarising the arguments for and against a given hazard or risk, factors that attenuate those arguments (e.g. weaknesses of the underpinning studies), the overall evidence base (number of relevant studies), the current conclusion and remaining uncertainties. Strengths and weaknesses of the evidence and the overall conclusion are expressed by narrative statements. However, such maps will rapidly become unwieldy for assessments with many different lines of evidence or uncertainties. No guidance is given on how to combine the different lines of evidence and uncertainties, or on how to apply the approach to quantitative conclusions (e.g. exposure or reference dose).

The Intergovernmental Panel on Climate Change published very concise guidance (4 pages, IPCC 2005) for lead authors of IPCC 4th Assessment Report on climate change. This offers three approaches for characterising uncertainty: (a) express evidence and level of consensus on two scales of low to high, (b) scale of terms for confidence that an analysis or statement is correct, from very low (less than 1 out of 10 chance) to very high (at least 9 out of 10 chance), (c) scale of terms for likelihood of some well defined outcome, from exceptionally unlikely (<1% probability) to virtually certain (>99%). Little specific guidance is given on how to determine which level of confidence, likelihood etc. applies.

The 'California EMF Risk Assessment' (Neutra et al. 2002) also used a defined scale of terms to express different levels of probability. Three experts first expressed in numbers their individual professional judgments that added personal risks suggested by epidemiological studies of electric and magnetic fields were 'real', as a numerical 'degree of certainty' on a scale of 0 to 100. For the conditions with the most suggestive evidence of EMF risk, they produced a graph depicting their best judgments with a cross and their margin of uncertainty around this with a shaded bar (i.e., their 'uncertainty about their degree of certainty'). Finally, the numbers were used to assign narrative conclusions using the defined scale of terms, e.g. for adult leukemia, two of the scientists were 'close to the dividing line between believing or not believing' and one was 'prone to believe' that EMFs cause some degree of increased risk.

Pedigree analysis is a qualitative approach within the NUSAP (numeral, unit, spread, assessment, pedigree) system for evaluating uncertainty (Van der Sluijs et al. 2005). Pedigree analysis is intended especially for characterising 'deep' uncertainties that cannot be easily quantified, including qualitative issues such as problem framing, choice of methods, level of knowledge or consensus, and value-ladenness. Users define the issues that are relevant for their problem, and rate them on a scale from low to high, which they also define. A number of publications summarise the scores using "kite" or radar diagrams, but more recently these have been considered misleading and replaced with a type of bar chart (Wardekker et al. 2008). Some publications that use Pedigree analysis assign quantitative scores to the ratings and combine them by averaging across issues and across different parts of the assessment: no theoretical basis is claimed for this. Pedigree analysis can be used to characterise uncertainties affecting an assessment, but does not attempt to characterise how much they might change the conclusion.

The IPCS/WHO (2008) guidance document on uncertainty and data quality in exposure assessment includes a qualitative approach similar to Pedigree analysis (see above), but prescribing 11 issues to be evaluated and scales for evaluating them. The 11 issues are grouped in 3 dimensions: level of uncertainty (1 scale), appraisal of the knowledge base (scales for accuracy, reliability, plausibility,

scientific backing, robustness) and subjectivity of choices (scales for choice space, inter-subjectivity, influence of limitations, sensitivity of choices, influence of choices). Scores on the 11 scales are displayed in a separate 'evaluation matrix' for each source of uncertainty. These are then summarised into a single table, accompanied by a brief explanation of the weights used in reaching an overall conclusion, but no guidance is given on the method for doing this. The approach expresses the nature and severity of uncertainty rather than its impact on the estimated exposure, and may become cumbersome as the number of uncertainties increases.

Other approaches reviewed include a number related to weight of evidence assessment, including GRADE (www.gradeworkinggroup.org), a system for evidence assessment that takes explicit account of uncertainties but is strongly focussed on evaluation of clinical trials and procedures. UK Climate Impacts Programme guidance on risk and uncertainty assessment (Willows and Connell 2003) is closely based on Pedigree analysis and the IPCC approach (see above). The US EPA (2000) Risk Characterization Handbook states the assessor should identify residual uncertainties and their impact on the range of plausible risk estimates, but also states that no single recognised guidance currently exists for uncertainty analysis.

In summary, several existing approaches contain elements that are potentially useful in constructing a framework suitable for use by the COT. Key elements of this are likely to include systematic identification and listing of relevant uncertainties, identification of which of these are covered by the standard uncertainty factors, a method to evaluate uncertainties in terms of their impact on the assessment outcome and a tiered approach to identify when additional evaluation is required: features represented in the approaches of EFSA (2006) and ECHA (2008). It is also important to characterise uncertainties whose impact cannot be evaluated (as emphasised by Van der Sluijs et al. 2005), though we suggest that a simple narrative description of these may be more effective for COT than the more complex approach of Pedigree analysis. Methods are needed for evaluating evidence and uncertainty in two types of assessment: those addressing quantitative questions (e.g. estimating exposure) and those addressing categorical questions (e.g. allergenicity). For quantitative questions, tabular approaches similar to EFSA (2006) and ECHA (2008) appear suitable, whereas for categorical questions we suggest a different tabular approach incorporating elements of the 'evidence maps' of Schutz et al. (2008). In both types of assessment, simple graphics in the form of number lines may be helpful when expressing uncertainties numerically (similar to those of Neutra et al. 2002). A key challenge is devising methods for expressing the combined impact of multiple uncertainties that are both compatible with mathematical principles, and practical and intuitive for use by the COT.

Distinction between quantitative and categorical questions

It became clear during the review that, when considering how to evaluate uncertainty, there is a need to distinguish between two types of assessments: those addressing categorical ('yes/no') questions, such as whether a chemical causes allergic sensitisation[3] and those addressing quantitative questions, such estimating exposure or determining a threshold dose. Some of the existing approaches were designed specifically for quantitative questions (e.g. EFSA 2006 was

[3] Note that one may also have qualitative questions about quantitative issues, e.g. does exposure exceed the threshold dose.

designed for exposure assessment), and some specifically for categorical questions (e.g. IARC 2006, designed for assessing carcinogenicity).

It is necessary to distinguish these two types of question because different scales are needed to express their uncertainty. For quantitative questions, uncertainty is naturally expressed on the same scale as the answer to the question, i.e. the assessment 'endpoint'. For example, if an exposure assessment produces an estimate of exposure in units of mg/kg body weight/day, then it is logical to express uncertainty about that estimate on the same scale, indicating how different the answer might be. An obvious example of this is when (part of) the uncertainty of the estimate is expressed as a confidence interval or credibility interval, with the same units as the estimate. For categorical questions, however, the answer is either 'yes' or 'no', and probability (either the probability of yes, or the probability of no) is the natural scale for expressing uncertainty about which answer is true.

Some categorical questions in toxicology may have more than two possible answers. In such cases, multiple probabilities will be required to express the relative likelihood of the different categories.

Theoretical basis for qualitative evaluation of uncertainty

This project focussed on qualitative or, at most, semi-quantitative approaches, as these were expected to be more practical than quantitative (e.g. probabilistic) approaches and sufficient for the majority of COT assessments. However, to provide useful information for risk management, qualitative approaches should, as far as possible, be conceptually compatible with valid mathematical approaches to uncertainty. Therefore, as part of the project, part of the team investigated ways of providing a mathematical underpinning for the representation and combination of uncertainties in qualitative approaches. Uncertainty needs to be expressed on different scales for quantitative and categorical questions, as explained in the previous section, so different approaches will be needed to develop a mathematical basis for them.

For quantitative questions, a basic mathematical framework has been developed, and is described in detail in a draft paper that is included as Annex 3 to this report. The mathematical framework provides a theoretical underpinning for expressing and combining uncertainties using plus and minus symbols, as in the approaches of EFSA (2006) and ECHA (2008). This is done by converting the ranges defined by the symbols to probability distributions, and then combining the distributions for different sources of uncertainty to produce a distribution for the overall shift in the assessment endpoint if all those uncertainties were resolved. This process is explained in more detail in Annex 3. This work demonstrates that the approach for quantitative questions proposed in this report can be expressed mathematically in a theoretically coherent way.

Currently, the mathematical framework presented in Annex 3 treats the users' estimates of uncertainty as precise and assumes that the effects of different uncertainties on the assessment endpoint are independent and additive on the chosen scale. In reality, users' estimates of uncertainties are often imprecise and the effects of the uncertainties on the endpoint may be inter-dependent: both these facts are accommodated in the qualitative tabular approach of the draft framework (assessments of uncertainty can be imprecise, and users are advised to consider possible interdependencies when evaluating the overall uncertainty). In principle, the mathematical framework could be developed further to take account of deviations from additive independence, which would require user to make explicit judgments about dependencies. This could be combined

with more formalised methods for eliciting judgments (e.g. O'Hagan et al. 2006) and different mathematical approaches to take account of the subjectivity and imprecision of those judgments (e.g. Dubois 2010[4]). However, these refinements would result in more complex procedures that are less likely to be considered practical by expert committees such as the COT, at least as a first step. Instead, we suggest the basic theoretical framework is sufficient to justify the proposed qualitative approach for quantitative questions, provided that users take account subjectively of the uncertainty of their judgments and of deviations from additive independence, as is emphasised in the description of the proposed framework later in this report and in the concise guidance developed for users (Annex 1). Although less rigorous than a more quantitative approach, this is more likely to be implemented and will deliver substantial improvements in the recognition and characterisation of uncertainties compared to current practice.

Note that it is not necessary, when constructing uncertainty tables using the plus/minus symbols, to understand or use any aspect of the mathematical framework. It is only necessary to define a scale for the symbols that is convenient in the sense that it helps the user to make subjective judgments about how different uncertainties combine. It would be possible to implement the current mathematical framework as a user-friendly software tool, to assist (but not replace) users' judgment in assessing how uncertainties combine (including the potential for interactions between them).

Developing a mathematical framework for categorical questions is more challenging, for several reasons. Firstly, it is not clear whether it is best to work on the scale of probability or log-odds: the former is more familiar and intuitive for most potential users, but log-odds are routinely used in medical statistics and seem likely to be better suited for constructing the mathematical framework required. Second, it became clear during the project (especially the workshop) that the logical reasoning used by the COT and others for answering categorical toxicological questions can have a complex structure, involving different lines of evidence, each of which has associated strengths and weaknesses, and each of which may have a different weight in contributing to the overall answer. Furthermore, the structure of the reasoning varies between assessments. Thirdly, it became clear that many people are reluctant to give a point estimate for probability because they feel it implies excessive precision, although there are mathematical approaches that can accommodate this (e.g. imprecise probability theory).

It was therefore decided to develop a practical heuristic approach for categorical questions pending further research on a theoretical framework. The heuristic approach should be designed to facilitate current approaches to expert judgment, and express them more explicitly, not to replace or change them. Such an approach would list the lines of evidence and their strengths and weaknesses (uncertainties) in a way which reflects the structure of the logical reasoning used to answer the question; express the direction and magnitude of influence each line of evidence has on the overall answer, taking into account their strengths and weaknesses; express the overall answer either as a numerical probability, as a range of probabilities (to accommodate concern about over-precision), or as a verbal expression of how likely a yes (or no) answer is. These principles were followed in developing the approach proposed below, which combines elements from the evidence maps of Schutz et al. (2008) and the probability scales of IPPC (2005) and Neutra et al. (2002).

[4] See also accompanying papers in the same journal issue for a discussion of the probabilistic representation of subjective uncertainties.

Draft framework for evaluating uncertainty

This section sets out the approaches proposed by the project team for meeting the needs identified by the COT's report on variability and uncertainty (COT 2007), taking account of the review of existing approaches, the need to distinguish between quantitative and categorical assessments and the investigation of a mathematical basis for evaluating uncertainty in each type of assessment. It also takes account of feedback from the project workshop (see Annex 2) and from the COT on two earlier drafts of the framework.

The rationale for each element of the framework is explained below. A more concise version, confined to 2 pages and omitting the explanations, is provided in Annex 1. It is envisaged that Annex 1 could be printed separately to provide a convenient reference sheet for day-to-day use, and that this section of the main report could form the basis (with minor editing) for a longer document providing a more detailed explanation for introducing the approach to first-time users.

The following sections present the recommendations of the project team. The COT and other users may of course consider which parts of the draft recommendations they consider appropriate for their purposes, and either edit the documents or develop new documents accordingly. To assist with this, the draft framework indicates aspects where the project team see opportunity for a range of alternative choices.

When is evaluation of uncertainty needed?

In principle, uncertainty should be considered in every risk assessment (Codex Working Principles for Risk Analysis, Codex 2010). However, in some areas of exposure and risk assessment, standard screening procedures have been established that are considered to provide appropriate allowance for uncertainty. Assessments conducted according to these procedures logically do not require a new evaluation of uncertainty on every occasion, provided that it is clear that they do indeed include appropriate allowance for uncertainty. To demonstrate that an assessment procedure meets this requirement requires an uncertainty analysis, but this does not need to be repeated on every occasion the procedure is used. This is why EFSA (2006) concluded that, in the context of exposure assessment, 'screening assessments do not require an analysis of uncertainty on every occasion, provided that they include appropriate conservative assumptions and default values to take account of uncertainty'. This situation is more common in routine regulatory assessments with highly standardised data than in the work of expert committees like the COT, which tends to focus on non-standard questions with more uncertainties.

In any assessment that deviates from or goes beyond standard procedure, e.g. by use of non-standard data or scenarios, or modified assumptions or assessment factors, general provisions for uncertainty may no longer apply and therefore a specific evaluation of uncertainty becomes necessary. Case by case evaluation of uncertainty is of course required in all other assessments, for which there is no established procedure to allow for uncertainty.

Specification of assessment questions

The importance of good problem formulation – precisely defining the question(s) to be addressed by an assessment – is widely recognised, and was reiterated in conclusions of the project workshop (Annex 1).

The method for uncertainty evaluation depends on the type of question:

- **Categorical questions** with categorical answers e.g. assessment of allergenicity, or human relevance of a toxic effect.
- **Quantitative questions** with numerical answers, e.g. reference dose or estimate of exposure.

The primary reason for distinguishing the two types of question is that they require different scales for expressing uncertainty, which in turn require different approaches to evaluating uncertainty. For quantitative questions, uncertainty is expressed on the same scale as the answer to the question, e.g. as a range or distribution around a best estimate. For categorical (yes/no) questions, uncertainty is expressed as a probability of the answer being yes (or no).

Therefore it is good practice to write down in precise terms the question(s) addressed in the assessment and identify whether each question is quantitative or categorical.

Most questions comprise several subquestions, for example, an estimate of risk is based on separate estimates of hazard and exposure, which themselves often involve multiple subquestions. In such cases, it may be difficult to judge directly how uncertainties affecting the different subquestions impact on the answer to the overall question. Therefore, it is recommended that the user should disaggregate the overall question into as many subquestions as they find helpful, evaluate uncertainty separately for each subquestion, and finally consider the uncertainty of the overall question (see later).

Systematic identification of uncertainties

It is recommended to systematically examine all parts of the assessment for potential sources of uncertainty. This may include (but is not limited to) limitations in the amount, quality or relevance of data; assumptions, extrapolations, dependencies, confounding, expert judgments; applicability of standard factors or assumptions; inconsistent results; alternative models or mechanisms; and gaps in knowledge. In principle, every individual study, data input or assumption in the assessment may be subject to one or more of these uncertainties.

A systematic approach, examining each part of the assessment for each type of uncertainty, is recommended by EFSA (2006) in order to maximise the likelihood that significant uncertainties that might influence the assessment outcome are identified. This is also consistent with the Codex (2009) Working Principles for Risk Analysis, 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* and documented in a transparent manner'(italics added).

It may be helpful to develop a generic checklist of types of uncertainties to be considered for a given type of assessment. For example, EFSA (2008) have published a general list of the types of uncertainties to be expected in cumulative risk assessments for pesticides. This helps assessors to identify quickly which uncertainties to consider for particular assessments, and also documents for stakeholders that these uncertainties are being considered.

Evaluation of uncertainties for categorical questions

This section describes the approach recommended for evaluating uncertainty in assessments of categorical questions involving only two categories[5], e.g. is this chemical an allergen? A modification of this for questions with multiple questions is considered in the next section. Again, we refer to the final answer to the question as the 'endpoint' of the assessment.

The answer to a categorical question is often generated by logical reasoning, in which various lines of evidence are weighed to reach a conclusion, taking account of their individual strengths and weaknesses (or uncertainties). It is necessary to take account of this logical structure, in order to evaluate how the uncertainties impact on the overall endpoint. The procedures of IARC (2006) and the 'evidence maps' of Wiedemann & Schutz (2008) offer contrasting approaches to articulating the logical structure of reasoning in assessment of categorical questions.

Several alternatives were explored in this project. The first was found not to represent well the structure of reasoning in case studies conducted at the project workshop (Annex 2). In response to this, a second approach was developed which provided for separate evaluation of uncertainties relating to the quality of evidence and its relevance to the assessment question. However, this required a significantly more complex tabular representation, and it was considered that while it might be sufficient for some problems, there might in other cases be types of uncertainty that do not relate clearly to either relevance or quality. Therefore, a third approach was developed, as presented below, which allows the listing of strengths and weaknesses affecting different lines of evidence, and does not attempt to evaluate and propagate them individually, but rather allows the user to consider them together when evaluating the influence of each line of evidence on the overall question. It was considered by the project team that this better captures the structure of reasoning used by the COT, in a way that can be applied flexibly to the variety of questions assessed by the COT.

So far this approach has been applied only to a single case study (see later), so it should be subjected to further evaluation after the current project.

1. The first step is to identify the studies or lines of evidence that contribute to answering the question. List them in a table (Table 1) together with their main strengths and weaknesses (uncertainties).

   Some lines of evidence might comprise individual experimental or observational studies. If there are large numbers of studies, it may be helpful to consider them in groups, e.g. consider all animal studies as one line of evidence and all human studies as another (see the example for caffeine, Table 5 below).

   Other lines of evidence may include theoretical or experimental evidence on the mechanisms of effects, or evidence from historical information, e.g. about the use of a product and the absence or frequency of reports of effects.

2. Evaluate the *influence* of each study or line of evidence on the overall outcome of the question, taking account of its strengths and weaknesses. Use up arrows for lines of evidence which push

[5] Some categorical questions involve more than two categories. These could be addressed using the same approach and tabular structure as proposed here for two-category questions, but adding extra columns on the right side of the table so that separate influences and probabilities can be assigned for each category.

the outcome towards 'yes' and down arrows for those pushing towards 'no'. The meaning of the symbols must be defined. One option is to define them on an ordinal scale, e.g. as representing weak, medium or strong lines of evidence, giving a relative indication of their impact on the overall conclusion. Alternatively, it may be helpful to use definitions that give a more explicit indication of the effect of each line of evidence on the overall conclusion, e.g.:

↑↑↑ or ↓↓↓: line of evidence could be sufficient on its own to be confident of yes or no

↑↑ or ↓↓: contributes importantly towards yes or no

↑ or ↓: minor contribution towards yes or no

• : negligible influence on outcome in either direction.

Combinations of symbols (e.g. ↑/↑↑) could be used if it is desired to express more uncertainty about the influence of a line of evidence, or to indicate when the influence of one line of evidence depends on uncertainty about another line of evidence. In both cases, the meaning of the combined symbols should be explained within the table or accompanying text. If the influence of a line of evidence cannot be evaluated, this could be indicated in the table using a question mark in place of a symbol, or the uncertainty could be discussed separately outside the table[6].

3. Make a judgment about the overall answer to the question, taking careful account of all the studies or lines of evidence and their associated strengths and weaknesses: this should not be done by any simplistic aggregation such as counting the numbers of symbols. Express your uncertainty about the overall outcome as a probability, i.e. your degree of belief that the categorical condition is true (e.g. your estimate of the probability of rain tomorrow). Express the probability in a suitable way: as a number, a range, or using a defined scale of verbal expressions (e.g. Table 2). Drawing a line may help in thinking about probability, as illustrated in Figure 1.

[6] It might be argued that information does not qualify as 'evidence' if its influence on the conclusion cannot be evaluated. However, it may be useful to include it in the table for transparency, so that others can see that the information was considered.

Table 1. Tabular format proposed for evaluating uncertainties in assessments of categorical questions involving two categories. The arrows in the right hand column express the influence of each line evidence on the overall conclusion, taking account of their strengths and weaknesses (uncertainties). Symbols and terms used must be defined in the table legend, or in accompanying text or tables (e.g. Table 2).

Overall question: insert question text here	**Influence on conclusion**
Study/line of evidence 1 – insert text description of the line of evidence including the direction of its influence on the conclusion (e.g. 'Four of five studies in animals showed a clear dose-response')	↑↑
• *Strength: text describing strength 1*	
• *Weakness: text describing weakness 1*	
Study/line of evidence 2 - insert text description including the direction of its influence on the conclusion	↓
Add more rows as needed	
Overall conclusion: **Insert verbal description of likelihood of 'true' answer being yes (or no)**	**Optional expression of likelihood as a probability, range of probabilities, or standard phrase**

Table 2. Table of standard terms established by the Intergovernmental Panel on Climate Change for expressing different degrees of likelihood (IPCC 2005), which could be adapted for expressing uncertainty in the assessment of qualitative questions. Care should be taken to ensure that the spacing of intervals and the terms used to express them are appropriate for the content and context of each assessment, while also avoiding using the same term with different meanings in different assessments.

Virtually certain	*> 99% probability*
Very likely	*90-99% probability*
Likely	*66-90% probability*
About as likely as not	*33 to 66% probability*
Unlikely	*10-33% probability*
Very unlikely	*1-10% probability*
Exceptionally unlikely	*< 1% probability*

After IPCC (2005)

Figure 1. Example of a probability line that may be helpful when evaluating the probability that the 'true' answer to a categorical question is yes. A range instead of an arrow could be used, if the user prefers not to specify a precise probability.

A range of options are available for expressing the uncertainty of the overall answer, including a point estimate of probability, a range of probabilities, a standard phrase corresponding to a defined range of probabilities (e.g. Table 2), a standard phrase with a verbal definition, or free text description. The use of standard phrases with verbal definitions is recommended as a minimum for consistency and transparency. Feedback from both the project workshop and the COT indicated a widespread preference to avoid using numerical expressions of probability, partly because they found it difficult to express uncertainties about categorical questions in this way, partly to avoid implying excessive precision, and partly due to concerns that numerical estimates would be misinterpreted by decision-makers, stakeholders and the public.

However, using words without quantitative definitions may make it more difficult for decision-makers to understand fully the assessors' level of certainty, which may in turn make it difficult for them to judge the appropriate degree of precaution or opportunism in their decisions. Using words defined by ranges of probabilities (as in Table 2) should in principle help to avoid implying excessive precision, and aid consistency of interpretation between different assessors (e.g. within a committee). In addition, verbal terms (with or without a numerical definition) have the potential disadvantage that they tend to imply (or be interpreted as implying) a judgment about the seriousness of the consequences as well as their probability. For example, saying something is 'virtually certain' tends to be interpreted as meaning that the probability of a different outcome can be ignored for practical purposes, whereas in principle this depends on the seriousness as well as the probability of the alternative outcome.

Given the potential advantages and disadvantages of the different options for expressing uncertainty in categorical questions, ranging from numerical probabilities through verbal terms with numerical definitions to standardised terms with verbal definitions, it is recommended that they should be evaluated further in practice before making a decision.

Evaluation of uncertainties for quantitative questions

This section describes the approach recommended for evaluating uncertainty in quantitative assessments, such as estimation of exposure or determination of a threshold dose. For convenience, we refer to the quantity that is being estimated as the 'endpoint' of the assessment.

The first two steps are screening steps that seek to identify cases where a very simple consideration of uncertainty may be sufficient for decision-making.

1. If it is obvious that all the identified uncertainties are negligible or covered by default uncertainty factors, then it is sufficient to state this and list the uncertainties (or refer to a checklist). Similarly, if all the identified uncertainties are covered by uncertainty factors established in a previous assessment of the same issue, it will be sufficient to refer to the earlier assessment. This step seeks to cater for those cases where, having listed all the identifiable uncertainties, it is immediately clear to the assessors that they are all either covered by established factors included in the assessment, or too small to affect the assessment endpoint by enough to make a difference to decision-making (note that this implies either that there are explicit criteria for decision-making, or that the risk assessor has some other means of knowing how large a change in the assessment endpoint will impact decision-making).

2. If the uncertainties not covered by default factors all affect the assessment endpoint in a conservative way, then it may be sufficient to state this and – for transparency – either list the uncertainties, or refer to a checklist.

   Note that this step assumes that decision-makers require only to be assured that the assessment is conservative, and are not concerned about the degree of conservatism (e.g. by how much is the risk over-estimated). This will not be appropriate in cases where decision-makers are concerned not only to avoid excessive risk but also to avoid being excessively precautionary, e.g. if this imposes disproportionate costs. In such cases, step 2 is not applicable and it will be necessary to proceed to steps 3-5, in order to characterise the *degree* of conservatism.

3. If neither (1) or (2) apply, it will be necessary to evaluate the uncertainties in more detail. The approach recommended for this is closely based on that of EFSA (2006) and ECHA (2008), and similar to approaches developed in the US and Canada (EPRI 2009 and Hans Yu, personal communication). It involves using a table to list and evaluate the uncertainties (Table 3). It may be helpful to group the uncertainties according to which *component* (e.g. study, model input etc.) they affect, as illustrated in Table 3.

   Such tables can rapidly become large if all identified uncertainties are included. One practical solution to this is to list the less important uncertainties separately, e.g. EFSA 2007 tabulated major uncertainties in the main text of the opinion, and smaller uncertainties in an annex. Another option is to restrict the detailed evaluation to non-negligible uncertainties, and refer to a standard checklist where other uncertainties that were considered are listed.

4. Consider each tabulated source of uncertainty in turn, and evaluate how much the overall endpoint of the assessment might change if that uncertainty was resolved, i.e. its contribution to how different the 'true' endpoint might be. Express your judgment about this by using pairs of numbers (e.g. 0.5× – 2×), symbols (e.g. –/++), or words to cover the *range* in which you are reasonably (e.g. 90%[7]) sure the adjustment for each uncertainty would lie. Record your evaluations for all the identified uncertainties, as illustrated in Table 3.

   It is emphasised that even though the uncertainties may relate to inputs to a calculation, their impact should be evaluated in terms of their effect on the calculation output, the assessment endpoint, because this is what matters for decision-making. This means that when assessing an individual uncertainty, the assessor needs to consider how that uncertainty 'propagates' through the assessment to influence the endpoint. This requires thinking about the structure of the calculation and the relative importance of its different elements. For example, concentration data for a particular food may be highly uncertain, but have little impact on estimates of dietary exposure if the majority of exposure comes from other foods.

   The approach described here is intended for expressing uncertainty about a point estimate. If the endpoint of the assessment is in fact a variable, e.g. the distribution of exposures in a population, then it is recommended to evaluate the impact of uncertainty on the point estimate for a specified percentile of the distribution, e.g. a percentile that is known to be of interest for risk management.

5. Review the evaluated uncertainties and form a judgment about their overall, combined impact, i.e. how different the 'true' endpoint might be, if all the uncertainties were resolved. This should

[7] It is desirable, though not essential, to define the probability interval for the ranges that are used, to aid consistency of interpretation between different assessors (e.g. within a committee) and also between assessors and decision-makers. It will be necessary to define the interval if it is desired to convert the ranges to probability distributions for a quantitative evaluation, as in Annex 3.

not be done by any simplistic aggregation, such as counting the numbers of symbols. Consider carefully how the different uncertainties combine, including how they combine in calculating the endpoint from the inputs, and any dependencies between different uncertainties (e.g. if new data showed that one input was an underestimate, this might alter the evaluation of uncertainty around other inputs). Express the outcome in the same way as the individual uncertainties, using pairs of numbers (e.g. 0.5× – 2×), symbols (e.g. –/++), or words to cover the range in which you are reasonably (e.g. 90%) sure the adjustment for each uncertainty would lie. Also express the outcome in words as a short narrative (e.g. 'the true exposure is unlikely to be greater than the estimate and may be as much as tenfold lower'): this will assist readers in interpreting the symbols, and may also be useful to include in the conclusion or summary of the assessment. Terms used to express likelihood in this narrative (e.g. 'unlikely') should have a defined meaning (e.g. Table 2, above).

Table 3. Tabular format recommended for evaluating uncertainties affecting assessments of quantitative questions. The symbols used must be defined in the table legend, in accompanying text or in a diagram (e.g. Figure 2).

Question: *precise statement of quantity to be estimated*	**Evaluation of uncertainty**
Assessment component 1: *(e.g. study, model input, etc.) - brief text description*	
• Uncertainty 1: *brief text description*	-/+
• Uncertainty 2	-/++
Assessment component 2: *brief text description*	
• Uncertainty 1	-/•
• *more rows as needed*	
Overall assessment: *verbal description of overall uncertainty in assessment endpoint*	**-/++** (or numeric range)

If symbols or words are used to evaluate the uncertainties, it is recommended to define their meaning using a quantitative scale. It has been demonstrated experimentally that different individuals give varying interpretations of the quantitative meaning of verbal expressions (e.g. Theil 2002). Therefore, providing quantitative definitions is very important to enable the evaluated uncertainty to be given appropriate consideration in decision-making. If uncertainty is expressed only with symbols (e.g. - -/+) or words then it is very difficult for decision-makers to understand how much lower or higher the true exposure or risk might be, which makes it difficult for them to judge the appropriate degree of precaution or opportunism in their decisions. A secondary benefit of using quantitative definitions is to aid consistency of interpretation between different assessors, or between different members of an expert committee.

When defining a quantitative scale, make it wide enough to accommodate the largest uncertainties in the assessment[8]. Set the intervals for different symbols in a way that seems effective to express the variation in the magnitudes of the uncertainties, and that helps in thinking about how the uncertainties combine. Sometimes it may be convenient to use a natural scale, on other occasions a logarithmic scale. An example of a logarithmic scale is shown in Figure 2.

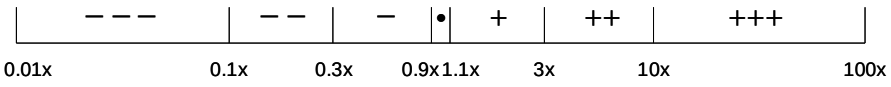


Figure 2. Example of a quantitative scale to define the meaning of symbols used to express uncertainties affecting assessment of quantitative questions (see Table 3).

[8] If there is a defined threshold for decision-making, then it may be helpful to make the scale wide enough to enclose it, as this will make explicit whether the uncertainties are judged large enough to change the decision.

Where users are able to express the effect of uncertainties on a quantitative scale, they may prefer to use numbers in the table, as this expresses their judgments directly, and dispense with symbols. Other users may prefer to use only words, or symbols with verbal, ordinal definitions (e.g. small, medium, large). There are differing views about the advantages and disadvantages of these choices (see Discussion).

Expression of uncertainties whose impact cannot be evaluated

Van der Sluijs et al. (2005) have emphasised that some uncertainties are too 'deep' for their impact on assessment conclusions to be evaluated, and proposed a systematic approach to characterising such uncertainties (Pedigree analysis). It is clearly essential to identify such uncertainties and communicate them to decision-makers, since if they really defy evaluation completely, this implies the assessment outcome could differ to any degree, in any direction and no indication of the probabilities of different outcomes can be given. If true, this would imply the current state of science cannot inform the decision in question, which should therefore be based on other considerations (legal, economic, precautionary, opportunistic, etc.).

In cases where such uncertainties exist, the user could consider using an approach such as Pedigree analysis to characterise them, but it may be as effective to do this by means of a narrative description. In either case, the consequences for the assessment outcome (that it is totally uncertain) need to be explained very clearly and prominently. If a tabular analysis of other uncertainties is presented, it may be helpful to include the 'deep' uncertainties, mark their impact as unknown ('?'), and consequently mark the overall conclusion as unknown.

Given the consequences of uncertainties whose impact cannot be evaluated, it will be worthwhile to examine potential cases of this very carefully to determine whether an approximate or tentative evaluation of their impact can be given. This can then be characterised in an appropriate way by the approaches presented in the preceding sections. For example, in some cases it may be possible to accommodate very wide uncertainty by drawing a conservative conclusion. This implies that the uncertainty in question can be bounded, at least on one side. If that is the case, then this uncertainty is not in fact 'deep' and can be evaluated within the main framework by assigning symbols accordingly, e.g. - - -/• would indicate a source of wide uncertainty for which a worst-case upper bound has been assumed.

Assessments comprising multiple subquestions

As explained earlier (under 'specification of assessment question'), most questions actually comprise two or more sub-questions and, in such cases, it is recommended to evaluate uncertainties affecting the sub-questions before considering how they combine to affect the overall question.

It is suggested that after evaluating each subquestion separately, using the approach for quantitative or categorical questions as appropriate, the outcome of those evaluations should be summarised in a table (e.g. Table 4). These should then be used to make a judgment about the uncertainty of the overall conclusion, taking full account of the structure of the calculations or logic by which the different subquestions combine in the overall answer. The structure of the calculation or logic should be clearly explained in accompanying text, to help others understand the evaluation.

Express the uncertainty of the overall answer using appropriate symbols, words, probabilities or numbers, using a format appropriate to whether the overall question is quantitative or categorical (see above).

Table 4. Tabular format suggested for combining the evaluation of uncertainty for multiple sub-questions implied by an overall assessment question. See text for details.

	Endpoint or Outcome	**Uncertainty**
Subquestion 1: *text*	*Subquestion answer*	*score/numbers*
Subquestion 2: *text*		
Etc.		
Overall question: *text*	*Overall outcome (and) uncertainty*	

Further refinement of the assessment

In some assessments, the degree of uncertainty indicated by the evaluation may be too great for the decision-maker to choose between the available decision options with the desired level of confidence. In such cases, one option is to evaluate one or more of the uncertainties affecting the assessment using quantitative methods (e.g. by sensitivity analysis or probabilistic modelling). Another option is to obtain further data with the aim of reducing uncertainty. Both options may usefully be targeted on the most important uncertainties, as identified by the preceding evaluations.

Communication of results

The importance of effective risk communication is well-recognised, and special care is required when communicating information about uncertainty in risk assessment. Because understanding uncertainty is fundamental for risk management and decision-making, it is important to include information about uncertainty in the headline conclusions of the assessment, for example in the executive summary as well as in the conclusions section. It is recommended that this should include: one sentence summarising the overall impact of uncertainties on the assessment outcome; 1-2 sentences outlining the major sources of uncertainty. If there are any uncertainties whose impact on the outcome could not be evaluated, these should be described and their consequences for the assessment should be clearly explained (see earlier).

Information on uncertainty needs very careful expression to avoid implying excessive precision and to minimise the risk of misinterpretation. In particular, it should be stated clearly, as part of the conclusions and summary, that the evaluation of uncertainty is approximate.

Detailed information including the uncertainty evaluation tables should be provided for transparency, and to enable peer review, but may be presented either in the main body of the assessment report or in annexes (or partly in both).

Case study - caffeine

This section explores application of the proposed methodology to an example assessment. The case study is based on the 2008 COT Statement on the Reproductive Effects of Caffeine, which was also used for one of the case studies in the project workshop (Annex 2). It includes both categorical and quantitative questions, but does not illustrate all aspects of the proposed approach.

The case study group at the project workshop identified the following specific questions within the Statement on Caffeine:

1. Is caffeine intake during pregnancy associated with increased risk of FGR?
2. What is the likelihood the association is causal?
3. What is the lowest level of intake above which the risk of FGR is increased?
4. If the relationship is causal what is the maximum increase in incidence of FGR (above the residual) from intakes of 200 mg/day?

Questions 1 and 2 are categorical, while questions 3 and 4 are quantitative. Questions 1 and 2 were combined by the case study group into a single categorical question for the purpose of their evaluation: Is caffeine a cause of fetal growth restriction in humans?

The evaluation developed at the project workshop (Annex 2) was revised by the project team to illustrate more closely the revised version of the framework proposed in this report. The resulting evaluation for the categorical question is shown in Table 5. This illustrates the proposed approach for summarising the lines of evidence and their strengths and weaknesses, and for indicating their influences on the overall conclusion using symbols. The overall conclusion is expressed only in narrative form in Table 5, using the same words as in the original COT statement (COT 2008). The COT did not define the terminology used in this conclusion. The phrase 'it is still not possible to be confident that the association is causal' appears to rule out high levels of probability but provides no further indication of the likelihood that caffeine is a cause of FGR. If one were to interpret the conclusion using the IPCC scale, for example, it might correspond to a range of probabilities from 0-90% (i.e. to all the IPCC terms between 'exceptionally unlikely' and 'likely'), or to other ranges of probability. A defined scale of terms would make it explicit what range of probability was intended.

Table 5. An evaluation of uncertainty for a categorical question, based on the COT (2008) Statement on the Reproductive Effects of Caffeine. For key to symbols see earlier section. *S* = strength, *W* = weakness.

Overall question: Is caffeine a cause of fetal growth restriction in humans?	**Influence on conclusion**
Animal experiments - reproductive effects have not been found in animal studies at doses below those causing maternal toxicity. *W - The numbers of animals per group were relatively small and hence limited the power of the studies to detect a significant effect.* *W – limitations in the design of studies with monkeys and differences in their metabolism of caffeine make results uninformative for assessing risks to humans* *W - The relevance of developmental findings in experimental animals to humans is uncertain.*	↓
New human study - a new study in humans shows caffeine intake in pregnancy is associated with increased risk of FGR, which might indicate a causal relationship *S – This was a prospective study, which is an inherently more reliable design than retrospective analysis, with better subject control* *S - The study was well designed, with a high rate of completion, detailed assessment of exposure and outcome, FGR being a robust endpoint* *W – While some important confounding factors were controlled, residual confounding is always possible, and caffeine intake may have been a surrogate for some other lifestyle factor such as some other component of tea* *W – Results may have been affected by reductions of caffeine intake by some women in pregnancy, or by other lifestyle changes associated with this* *W – Although exposure assessment was thorough, there were still potential errors as it relied on subject recall, particularly during the first trimester. However, this would be more likely to obscure a relationship than create a spurious one.*	↑/↑↑
Previous human studies - earlier studies showed varying degrees of association between caffeine intake and FGR which might indicate a causal relationship *W – The confidence in the results of these studies was not strong because of limitations in their design, and lack of consistency in the findings* *W – Most of the studies were retrospective, and hence relied upon recall and were subject to potential bias* *W – It is possible effects on FGR were caused by factors other than caffeine*	↑
Experimental evidence for biological mechanism - no plausible biological mechanism has been identified for an effect of caffeine on FGR *W – potential biological mechanisms for an effect of caffeine on FGR have been investigated only to a very limited extent.*	↓/•
Overall conclusion: **Caffeine intake during pregnancy is associated with an increased risk of FGR but it is still not possible to be confident that the association is causal rather than a result of residual confounding.**	

The case study group at the project workshop produced a single evaluation for a quantitative question (Annex 2). However, it was not stated explicitly in their presentation, which of the two quantitative questions it addressed: 'what is the lowest level of intake above which the risk of FGR is increased?' or 'if the relationship is causal what is the maximum increase in incidence of FGR (above the baseline) from intakes of 200 mg/day?'. The project partners developed an evaluation for the first of those questions, taking account of the uncertainties identified at the workshop. The resulting evaluation is shown in Table 6.

The proposed approach for quantitative questions expresses the impact of uncertainties in terms of how much the overall answer would change if each uncertainty was resolved. For this purpose, the project team took 200 mg caffeine per day as the nominal answer, as this was the value mentioned in the COT conclusion. Also, a table in the COT (2008) statement indicates a significant increase in risk of FGR above 200 mg per day but not at 100-199 mg per day. However, a modelled dose-response shown graphically in the COT statement indicates more uncertainty about the level of intake at which risk of FGR rises above the background risk for the population. This and other uncertainties were considered in constructing the evaluation in Table 6.

The evaluation at the workshop was constructed without specifying quantitative definitions for the symbols, using + and – symbols only to indicate the direction of influence and not magnitude. Table 6 was initially constructed in the same way, but using multiple symbols (- -) to indicate larger influences. However, as stated earlier, it is recommended to define a quantitative scale for the symbols in order to facilitate consistency in interpretation and communication. This was done retrospectively by the project members who finalised Table 6 (Boobis and Hart). They considered that, on the scale implied by the evaluations in Table 6, a single minus symbol represents up to 50% decrease in the estimated value, and two minuses up to 75% decrease (to 100 and 50 mg per day respectively), whereas a single plus symbol represents up to a 50% increase (up to 300 mg per day). In future, it would be preferable to define the scale for the symbols earlier in the process.

Table 6. Example of evaluation of uncertainty using the approach proposed for quantitative questions, based on the COT (2008) Statement on the Reproductive Effects of Caffeine.Key: • indicates negligible impact on assessment outcome, plus symbols indicate resolving uncertainty would lead to increased estimates, minus symbols indicate resolving uncertainty would lead to reduced estimates. See text for further discussion of symbol interpretation.

Question: What is the lowest level of caffeine intake above which the risk of FGR is increased?	Evaluation of uncertainty
New human study	
Fitted curve could be compatible with higher or lower thresholds, although effects at much lower levels could be due to residual confounding	- -/+
Alternative model shapes are plausible and might either increase or decrease the level at which risk appears elevated	-/+
Differences between fast and slow metabolisers might have an effect on the dose response by altering the internal dose. The data showed a non-significant trend (P=0.06) but in the opposite direction to expected.	-/+
Study sample might be unrepresentative of the population by chance	-/+
Random error in estimated caffeine intake might flatten the dose-response and make it harder to detect a threshold, or overestimate it.	–/•
Errors in measurement of FGR – were considered to be minor	•
Consumers with very high intakes might distort the shape of the fitted dose-response but excluding consumers >300 mg made no difference	•
Smoking and energy intake are potential confounders but analysis of the data gave no indication of an important effect. Cotinine measurement confirmed the accuracy with which smoking habits were reported.	•
Previous human studies Range of results, could be consistent with thresholds below 200 mg/day or between 200 and 300 mg/day	-/+
Animal studies Range of results, but given the uncertainty of extrapolation to humans they do not alter the assessment based on human studies.	•
Overall Assessment of Uncertainty: The evidence that is now available does not make it possible to identify a threshold level below which there is no elevation of risk. It seems likely that risk is increased in association with intakes in the order of 200mg per day and perhaps even lower.	

Conclusions

The criteria defined at the start of the project were that methods for evaluating uncertainty should be:

- practical for use by the COT and other FSA committees, and adapted to their work,
- systematic & comprehensive, helping the user address all relevant uncertainties,
- efficient, using a tiered approach to minimise the effort required,
- helpful for developing conclusions and in subsequent decision-making, by evaluating uncertainties in terms of their impact on the key issues,
- conceptually compatible with mathematical approaches to uncertainty.

Progress has been made towards all of these. Particular effort has been made to develop approaches adapted to the work of the COT, taking account of feedback from the project workshop and from the COT itself. However further case studies would be beneficial to evaluate the methods more fully, and the project partners welcome the COT's indication that it intends to try out the methodology in a suitable agenda item in the future. The methodology encourages a systematic and comprehensive approach, and has been tiered to the extent possible, to avoid uncertainty evaluation in basic assessments if it is unnecessary, and to allow early exit from evaluation of quantitative questions when the uncertainties are negligible or conservative. The methodologies for both categorical and quantitative questions evaluate uncertainties in terms of their impact on the assessment endpoint, which is the relevant information for decision-making. And finally, a basic mathematical underpinning has been developed for the method for quantitative questions, but further work is needed on this for categorical questions.

In feedback from discussion of an earlier draft of the framework at its meeting in May 2010, the COT stated that 'The challenges for the Committee in expressing uncertainty are not easily addressed by a simple mathematical approach. Members reiterated their reluctance to put a numerical value on uncertainty as this could easily be misinterpreted. It would be helpful to develop a scale of terms describing different levels of uncertainty with advice from the FSA Social Science Research Committee (SSRC). The COT secretariat would pursue this with the SSRC secretariat.' This course of action was supported again at the subsequent COT meeting in June 2010. The project team agree that communicating uncertainty raises issues for risk communication, which go beyond the technical scope of this project, and welcomes the initiative to explore these issues from a social science perspective. The project team accept that there are different views about the appropriateness of expressing subjective judgments about uncertainty in quantitative terms, and that it can be difficult to do in practice. We believe the difficulty of the task could be reduced with more experience and with the help of simple training examples such as are often used when preparing for formal expert elicitation studies (e.g. asking participants to express judgments about everyday questions such as the weather or road distances).

In detailed comment from the FSA, it was suggested that difficulties in expressing uncertainty quantitatively are due to limitations of the available data. A similar limitation is suggested in the IPCC guidance on the treatment of uncertainty (2005), which advises use of quantitative expressions

"where the level of confidence is '*high agreement much evidence*'", though they add "or where otherwise appropriate". A firmer view was expressed by the Inter-Academy Council (IAC) review of the IPCC, which included a recommendation that "Quantitative probabilities (as in the likelihood scale) should be used to describe the probability of well-defined outcomes only when there is sufficient evidence", although in the following sentence they also endorse the subjective estimation of probabilities by expert judgment (IAC, 2010). A draft revision of the IPCC's guidance on the treatment of uncertainty accepts this recommendation and explicitly restricts the use of the IPCC's quantitative 'likelihood' scale to findings with high agreement or robust evidence, or both (IPCC, 2010).

However, a contrasting view is expressed in a major report from the US Climate Change Science Program, which states that "so long as one carefully specifies the question to be addressed" (a requirement also emphasised by the IPCC and IAC), "our judgment is that all four boxes in Figure 1.1" (i.e. all levels of evidence, and all degrees of agreement) "can be appropriately handled through the use of subjective probability, allowing a wide range or a multiple set of plausible distributions to represent the high levels of uncertainty, and retaining the axioms of probability" (Morgan et al. 2009). This is not an academic view uninformed by practical experience – a footnote (p. 149) comments that collectively the 8 authors have roughly 200 person-years of experience in addressing these issues both theoretically and practically in the context of climate and other similar areas.

The participants of the current project are divided on the question of quantitative expression of uncertainty. Some agree with the IPCC view (above), while others agree with Morgan et al. that uncertainty can be expressed quantitatively even when evidence is limited.

Views on this issue can be expected to vary also amongst potential users of the framework produced by this project, in the COT and elsewhere. Therefore, to maximise the usefulness of the framework produced by the project, it includes a range of options for expressing uncertainties, including using words alone, standard phrases with quantitative definitions, ranges of numbers or probabilities, or point estimates.

We understand from earlier feedback that the COT envisages at least developing standardised phrases. Given the potential advantages for transparency, consistency of interpretation, and improved information for decision-makers (outlined in preceding sections), we would encourage the COT to explore all the options in the framework when trialling it, and when consulting the FSA Social Science Research Committee. As part of this, the COT may wish to give some consideration to using standardised phrases with quantitative definitions, possibly similar to those used by the IPCC (2005, 2010) but adapted and refined for the purposes of the COT.

In summary, the following recommendations are made for next steps:

- Further investigation to refine the theoretical basis for the evaluation of uncertainties in quantitative questions, and to establish a theoretical basis for categorical questions.
- Further evaluation and, if appropriate, refinement of the proposed approaches by applying them to a range of practical examples.
- Exploration of challenges in standardising terminology for communicating uncertainties associated with toxicological risk assessments.

Acknowledgements

The authors are grateful to the Food Standards Agency for funding this work, to the participants of the COT workshop, and to FSA staff and the COT for feedback on earlier drafts.

References

Codex, 2010. Codex Alimentarius Commisssion – procedural manual - 19th Edition http://www.codexalimentarius.net/web/procedural_manual.jsp

COT, 2007. Variability and Uncertainty in Toxicology of Chemicals in Food, Consumer Products and the Environment, Committee on Toxicity of Chemicals in Food, Consumer Products and the Environment. Food Standards Agency, London.

Dubois D, 2010. Representation, propagation, and decision issues in risk analysis under incomplete probabilistic information. Risk Analysis, 30, 361-368.

ECHA, 2008. Guidance for implementation of REACH. Chapter 19, Uncertainty analysis.

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. Opinion of the Scientific Panel on Plant Protection products and their Residues to evaluate the suitability of existing methodologies and, if appropriate, the identification of new approaches to assess cumulative and synergistic risks from pesticides to human health with a view to set MRLs for those pesticides in the frame of Regulation (EC) 396/2005. EFSA Journal, 704, 1-84.

EFSA, 2009. Transparency in Risk Assessment – Scientific Aspects 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, 1051, 1-22.

EPRI, 2009. Treatment of Parameter and Model Uncertainty for Probabilistic Risk Assessments. Download at www.epri.com.

IAC, 2010. Climate change assessments. Review of the processes and procedures of the IPCC. InterAcademy Council. Available on the IPCC website.

IPCC, 2005. Guidance notes for lead authors of the IPCC Fourth Assessment Report on addressing uncertainties. International Panel on Climate Change, WMO/UNEP, July 2005, 4 pp.

IPCC, 2010. Draft guidance notes for lead authors of the IPCC Fifth Assessment Report on consistent treatment of uncertainties. Appendix 4 in: Decisions taken by the Panel at its 32nd Session. International Panel on Climate Change, WMO/UNEP, Geneva.

IPCS/WHO 2008. Uncertainty and data quality in exposure assessment. Harmonisation Project Document No. 6. WHO, Geneva.

Neutra R, DelPizzo V, and Lee GM, 2002. An evaluation of the possible risks from electric and magnetic fields (EMFs) from power lines, internal wiring, electrical occupations and appliances, California Dept. of Health Services, Oakland, CA. [http://www.ehib.org/emf/RiskEvaluation/riskeval.html]

O' Hagan A, Buck CE, Daneshkhah A, Eiser JE, Garthwaite PH, Jenkinson DJ, Oakley JE and Rakow T, 2006. *Uncertain Judgements: Eliciting Expert Probabilities*. John Wiley and Sons.

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. http://ec.europa.eu/health/ph_risk/documents/ev_20040907_co01_en.pdf

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.

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

Schutz H, Wiedemann PM and Spangenberg A, 2008. Evidence maps – a tool for summarizing and communicating evidence in risk assessment. In: PM Wiedemann and H Schutz (eds.), The role of evidence in risk characterization: making sense of conflicting data. Wiley-VCH, Weinheim.

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

USEPA, 2000. Risk characterization handbook. EPA 100-B-00-002. Office of Science Policy, US EPA, Washington DC.

US NRC, 2009. Guidance on the Treatment of Uncertainties Associated with PRAs in Risk-Informed Decision Making. NUREG-1855. US Nuclear Regulatory Commission.

Van der Sluijs JP, Craye M, Funtowicz S, Kloprogge P, Ravetz J, Risbey J. 2005 Combining Quantitative and Qualitative Measures of Uncertainty in Model based Environmental Assessment: the NUSAP System, Risk Analysis, 25, 481-492.

Wardekker JA, van der Sluijs JP, Janssen PHM, Kloprogge P and Petersen AC, 2008. Uncertainty communication in environmental assessments: views from the Dutch science-policy interface. Envir. Sci. & Policy, 11, 627-641.

Willows R and Connell R. 2003. Climate adaptation: risk, uncertainty and decision-making. UKCIP Technical Report. UKCIP, Oxford.

Annex 1: Summary of framework

Annex 2: Report of COT Workshop

Annex 3: Draft paper on mathematical framework for uncertainties in quantitative questions

REVISED DRAFT FRAMEWORK FOR EVALUATING UNCERTAINTY[1] v. 13, 27/7/10

Every risk assessment should include an evaluation of uncertainty *unless* it follows a standard procedure including provisions for uncertainty, such as uncertainty factors or conservative assumptions, which make appropriate allowance for the uncertainties involved.

Specification of assessment questions

The method for uncertainty evaluation depends on the type of question:

- **Categorical questions**, e.g. assessment of allergenicity.
- **Quantitative questions** with numerical answers, e.g. reference dose or exposure estimate.

Write down in precise terms the question(s) addressed in the assessment. If the assessment addresses several subquestions, evaluate uncertainty separately for each subquestion before considering the uncertainty of the overall question (see later).

Systematic identification of uncertainties

Systematically examine all parts of the assessment for potential sources of uncertainty including limitations in the amount, quality or relevance of data; assumptions, extrapolations, dependencies, confounding, expert judgments; applicability of standard factors or assumptions; inconsistent results; alternative models or mechanisms; and gaps in knowledge. It may be helpful to develop a checklist of types of uncertainties to be considered for a given type of assessment.

Evaluation of uncertainties for *categorical questions* (e.g. is this chemical an allergen?)

1. Identify the studies or *lines of evidence* that contribute to answering the question. List them in a table (as shown) together with their main *strengths* and *weaknesses* (i.e. uncertainties).

Overall question: insert text here	**Influence on conclusion**
Line of evidence 1 text description	↑↑
• *S - strength 1*	
• *W - weakness 1*	
Line of evidence 2 text description	↓
Add more rows as needed	
Overall conclusion: text statement	**Likely (66-90%)**

2. Evaluate the *influence* of each study or line of evidence on the overall conclusion, taking account of its strengths and weaknesses. Use up arrows for lines of evidence which push the outcome towards 'yes' and down arrows for those pushing towards 'no', for example:

   ↑↑↑ or ↓↓↓: line of evidence could be sufficient on its own to be confident of yes or no
   ↑↑ or ↓↓: contributes importantly towards yes or no
   ↑ or ↓: minor contribution towards yes or no
   • : no influence on the overall conclusion.

3. Make a *judgment* about the overall conclusion to the question, taking into account all the studies or lines of evidence and their strengths and weaknesses. Do not simply add symbols. Express your uncertainty about the answer as a probability, i.e. your degree of belief that the categorical condition is true (e.g. your estimate of the probability of rain tomorrow). Express the probability in a suitable way: as a number, a range, or using a defined scale of verbal expressions (see IPCC example: adapt as appropriate).

Virtually certain	*> 99% probability*
Very likely	*90-99% probability*
Likely	*66-90% probability*
About as likely as not	*33 to 66% probability*
Unlikely	*10-33% probability*
Very unlikely	*1-10% probability*
Exceptionally unlikely	*< 1% probability*

After IPCC (2005)

   Drawing a line may help in thinking about probability, e.g.:

[1] This is a concise version. For more details on each part, see the corresponding sections in Hart et al., (ref. to be added). Contact for questions or feedback on this draft*: andy.hart@fera.gsi.gov.uk*

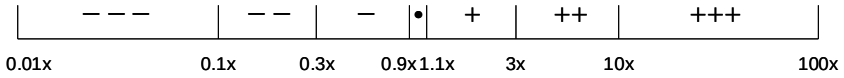
Evaluation of uncertainties for *quantitative questions* (e.g. a reference dose or exposure estimate)

1. If it is obvious that all the identified uncertainties are negligible or covered by default uncertainty factors, then it is sufficient to state this and list them (or refer to a checklist).
2. If the uncertainties not covered by default factors all affect the assessment endpoint in a conservative way, then it may be sufficient to state this and list or refer to them[2].
3. If neither (1) or (2) apply, construct a table to evaluate the uncertainties, as shown below. It may be helpful to group the uncertainties according to which *component* (e.g. study, model input etc.) they affect. Negligible uncertainties may be listed separately (or refer to a checklist).
4. Consider each tabulated source of uncertainty in turn, and evaluate how much the *overall endpoint*[3] of the assessment might change if that uncertainty was resolved, i.e. its contribution to how different the 'true' endpoint might be. Express your judgment about this by using pairs of numbers (e.g. 0.5× – 2×), symbols (e.g. –/++), or words to cover the *range* in which you are reasonably (e.g. 90%) sure the adjustment for each uncertainty would lie. Record your evaluations in the table (as shown).

Question: *precise statement of quantity to be estimated*	**Evaluation of uncertainty**
Assessment component 1	
• Uncertainty 1	-/+
• Uncertainty 2	-/++
Assessment component 2	
• Uncertainty 1	-/•
• *more rows as needed*	
Overall assessment: *verbal description of overall uncertainty in endpoint*	-/++ (or numeric range)

5. Review the evaluated uncertainties and form a judgment about their combined impact, i.e. how different the 'true' endpoint might be, if all the uncertainties were resolved. Do not simply add symbols. Consider how the uncertainties combine, taking account of any dependencies between them. Express the outcome using numbers or symbols and also as a narrative.

If you use symbols or words to evaluate the uncertainties, define their meaning using a convenient scale, adjusted to the magnitude of the largest uncertainties. For example:



Assessments comprising of multiple subquestions

1. Evaluate each subquestion separately, then summarise the evaluations in a table (as shown).
2. Make a judgment about the uncertainty of the overall conclusion. Express this using appropriate symbols, words or numbers (format depending whether the overall question is quantitative or qualitative, see above). Explain clearly, in accompanying text, the logic of how the assessments combine.

	Endpoint or Outcome	**Uncertainty**
Subquestion 1: *text*	*Subquestion answer*	*score/numbers*
Subquestion 2: *text*		
Etc.		
Overall question: *text*	*Overall outcome (and) uncertainty*	

Further refinement of the assessment

If further refinement is required to support decision-making, one option is to evaluate uncertainty quantitatively (e.g. by sensitivity analysis or probabilistic modelling). Another is to obtain further data to reduce uncertainty. Both may be targeted on key uncertainties, identified by the evaluation.

Communication of results

- *In the assessment conclusion*: one sentence summarising the overall impact of uncertainties on the assessment outcome; 1-2 sentences outlining the major sources of uncertainty; *plus* a description of any uncertainties whose impact on the outcome could not be evaluated.
- State clearly that the evaluation of uncertainty is approximate, to avoid over-interpretation. Communicate with care to facilitate proper understanding by decision-makers and others.
- *In the main assessment report or as an annex*: lists/tables plus supporting text as appropriate.

[2] If it is desired to evaluate the degree of conservatism, proceed to the next step.

[3] If the endpoint of the assessment is a variable, e.g. the distribution of exposures in a population, then identify the percentile of interest for risk management and evaluate uncertainties in terms of their impact on that percentile.

FOOD STANDARDS AGENCY

WORKSHOP ON THE EVALUATION AND EXPRESSION OF UNCERTAINTIES IN RISK ASSESSMENT

DRAFT REPORT

2nd & 3rd FEBRUARY 2010

OULTON HALL, LEEDS

INDEX

INTRODUCTION

In March 2007, the Committe on Toxicity of Chemical in Food Consumer Products and the Environment (COT) published its report on Variability and Uncertainty in Toxicology of Chemicals in Food, Consumer Products and the Environment (the report summary is enclosed in the following section of this document). The Committee concluded that the development of a framework for transparent expression of uncertainty in hazard characterisation would enable the COT and other committees that perform toxicological evaluations to improve communication of the sources of variability and uncertainty in their risk assessments.

The Food Standards Agency commissioned a research project (led by Dr Andy Hart, FERA) to review the existing approaches for qualitative evaluation and expression of uncertainties and assess their suitability for routine use by the COT and other committees (a summary copy of this review is enclosed). Based on this review, the project team has developed a draft framework for evaluating uncertainty (also enclosed).

The purpose of the workshop was to evaluate the draft framework using four case studies, based on previous assessments by the COT. These were:

- COT Statement on the tolerable daily intake for perfluorooctanoic acid (2006). COT Statement 2006/10, October 2006.
- Statement on the review of the 1998 COT recommendations on peanut avoidance. COT Statement 2008/07, December 2008.
- COT Statement on the reproductive effects of caffeine. COT statement 2008/04, September 2008.
- COT Statement on the 2006 UK total diet study of metals and other elements. COT statement 2008/08, December 2008.

Each case study was considered by a separate breakout group. Participants were provided with copies of the COT assessment in question, extracts of COT minutes relating to the assessment and, where relevant, previous COT assessments on the same issue.

Each breakout group prepared a presentation summarising their evaluation of the framework, which were discussed in a plenary session. The overall conclusions of the workshop were summarised by the COT chair, Dr David Coggon, in his closing remarks. The breakout group presentations and overall conclusions are enclosed in this report.

These findings, together with any additional feedback from the COT, will be taken into account by the project team when revising the review and draft framework for the final report of this project.

COT REPORT ON VARIABILITY AND UNCERTAINTY IN TOXICOLOGY – CONCLUSIONS AND RECOMMENDATIONS

Conclusions

13.1 The conclusions are set out below, under the respective terms of reference:

To review the evidence of the bases and range of variability in response to toxic chemicals

13.2 Variability in response to chemicals is determined by the fate of the chemical within the body (toxicokinetics) and the toxicity of the chemical and its metabolites (toxicodynamics). Variability in both toxicokinetics and toxicodynamics arises from a combination of factors that are inherent to the individual (or organism), and other factors that relate to the physiology and environment of the individual and which change over time. Inherent characteristics include species, sex and genotype. The modulating factors include age, stage of development and functional maturation of organs and systems, co-exposure to other agents and compounds (e.g. nutrients), lifestyle, and environmental factors, and disease. In principle, variability is measurable, and any lack of knowledge of variability is a source of uncertainty.

13.3 Recent scientific advances have led to greater insight into genetic factors responsible for susceptibility. It is now becoming possible to explain some variability in terms of genetic polymorphisms and postgenomic molecular biology. However, genetic heterogeneity and gene expression are not new forms of variation. They have always existed as part of the underlying differences between individuals, and as such would have been part of the variability that has informed the development of current methods that are applied to deal with variability in risk assessment.

13.4 The range of human variability in response to chemicals cannot be measured directly in all sectors of the population. It is inferred from studies in different animal species, and from knowledge of the differences between humans and animal models in toxicokinetics and toxicodynamics.

13.5 Each source of variability results in a distribution of activity or functionality amongst individuals of a population, leading to either increased or decreased susceptibility to the toxic effect. The overall response to a toxic chemical is determined by the combination of the many different sources of variability. The factors contributing to variability are, unless linked, unlikely to all act in the same direction.

To consider sources of uncertainty in hazard identification and characterisation

13.6 Uncertainty in hazard identification or characterisation relates to incomplete knowledge of the relevance of the results of studies conducted in animals and experimental models, or of human populations, and of possible effects that have not been adequately investigated or recognised.

13.7 The few available direct data relating to human variability are largely derived from studies in young men. Thus the extent to which these data reflect the susceptibilities of women, older people, the conceptus, or children is uncertain.

13.8 Often there is uncertainty about the association of early exposure with particular health effects later in life. Improved understanding of the pathogenesis of such effects might enable identification of early predictive markers of significant adverse effects that would reduce this uncertainty.

13.9 There is more uncertainty in the hazard identification and characterisation of contaminants and natural constituents in foods because, unlike food additives and pesticides, they are not subject to a formal approval process requiring systematic studies to support safety assessment.

13.10 Another source of uncertainty relates to interpretation of studies giving apparently contradictory results with no obvious explanation. There is a need for an agreed robust mechanism for assessing the results of studies that give contradictory results, and demonstrating clearly how the hazard characterisation resolves such problems.

To consider the appropriateness of uncertainty factors customarily used to extrapolate toxicological data from animals to humans

13.11 Differences between the animal species used in laboratory experiments and humans derive from anatomical and physiological differences, as well as the variation in genetic factors that occurs within a species. Data from the available research in which compounds have been studied in both animals and man suggest that the default uncertainty factor of 10 allows adequately for interspecies differences.

13.12 The question of special vulnerability of the developing nervous system to neurotoxicity is addressed by current regulatory testing with specific consideration of neurobehavioural and neurodevelopment outcomes. Similarly, data derived from developmental and reproductive toxicity studies in animals can be extrapolated to humans in considering other effects on the fetus and infant. Results suggest that the current approaches and uncertainty factors are adequate. However, it is recommended that this area be kept under review.

To consider the appropriateness of uncertainty factors customarily used to allow for variation within the human population, including subgroups such as children

13.13 Inter-individual differences in the activity of xenobiotic metabolising enzymes are often characterised in well defined populations of subjects, focusing only on the pathway of interest. Whilst 10-fold or greater differences have been demonstrated between groups, there are frequently no comparable differences in the overall kinetics of the parent chemical, because of compensation by alternative pathways.

13.14 The default uncertainty factor for interindividual variability has been explored empirically on a number of occasions. This has usually been performed with

pharmaceutical agents, but these studies can be related to other chemicals, and they suggest that the default uncertainty factor is generally appropriate.

13.15 With a few exceptions, particularly susceptible subgroups cannot be identified by genotype. Some subgroups are potentially vulnerable due to physiological, dietary or environmental factors. With respect to infants and children, it is recognised that the young can be either more suceptible or less susceptible than adults to the toxicity of particular substances. Since more information on newly introduced human pharmaceutical agents will be expected to be derived directly from observations of treated children, it should in future be possible to test the adequacy of current uncertainty factors in protecting young children.

13.16 The possibly increased susceptibilty of the elderly and the consequences of a lifetime of exposure are representatively investigated in chronic toxicity studies. Even so, there is a need for better characterisation of the uncertainties related to possible altered susceptibility arising from environmental, physiological and metabolic changes during the course of life and in older life. An additional uncertainty factor for this is probably unnecessary in most cases but should be considered and decided during hazard characterisation on a case-by-case basis.

To consider other methods that might be used in setting acceptable or tolerable intakes for chemicals in food, consumer products and the environment

13.17 The COT uses current internationally-accepted methods in its risk assessments. These make good use of state-of-the-art knowledge and of the methodologies available to take account of variability and sub-group vulnerability in toxicological data. Given the wide range in primary data quality and the frequent occurrence of critical data gaps, it is not possible to propose the use of any single approach to risk assessment, but rather it is necessary to continue with the present flexible use of the assessment methodology best suited to the specific data set available. As a continuing process, the COT will consider improving and refining the methods and approaches it uses.

13.18 *In vitro* studies have important roles to play in hazard characterisation and investigations of toxicological mechanisms. However, there remain uncertainties with regard to the extrapolation of the results of *in vitro* studies to humans. Complete replacement of animal tests in toxicology is not possible at present.

13.19 Application of the default 100-fold uncertainty factor, which allows for 10-fold factors each for inter- and intra-species variation continues to be a reasonable approach, in the absence of better information. In some instances, e.g. if there is good evidence that humans are not more sensitive than animals, application of the full 100-fold factor is not necessary. Subdivision of the default uncertainty factors to incorporate chemical-specific toxicokinetic and toxicodynamic adjustment factors should be used, whenever data allow. If chemical-specific adjustment factors are used, the adequacy of the remaining default factors should be explicitly considered.

13.20 Statistical and modelling approaches, including physiologically-based pharmaco- or toxicokinetic models, have been used to refine the risk assessment process. Greater use of such methods, when suitable data are available, would support a more systematic

approach to risk assessment. Probabilistic models could be used to explore and quantify uncertainty.

13.21 Description of assumptions and uncertainties in the evaluation is important for transparency of the risk assessment:

- Systematic reviews of the relevant toxicology and epidemiology literature are important tools in hazard identification and hazard characterisation and for the presentation of data

- There should be a description of the criteria for inclusion or exclusion of studies in a review and details of the uncertainties and variabilities in parameters of interest in both the test subjects (eg. Laboratory animals or human cohorts) and the human population of interest (eg. consumers or exposed workers)

- The choice of critical event used to set guidance values should be justified

- The validity and robustness of biomarkers of exposure, intake, susceptibility and outcomes should be discussed, along with environmental and lifestyle factors that might impinge on these factors

- Vulnerable groups of people should be identified

To consider how to express the level of confidence that one can have in the risk assessment

13.22 The degrees of variability and uncertainty at each stage of a particular risk assessment should be clearly described and communicated to those involved in risk management. This should include identification of whether all relevant responses were investigated. Particular attention should be given to stating assumptions and subjective elements in the risk assessment, justification of the choices of uncertainty factors used, and of the selection of the adverse health effects used as the basis for risk assessment. Transparency in these factors aids an informed assessment of uncertainty and enables risk managers to communicate this to stakeholders. Furthermore, such transparency is particularly important in reconciling differences in risk assessments reached by different expert groups.

Recommendations

13.23 There is a need to introduce methods to increase the transparency and reproducibility of hazard identification and characterisation. Several recommendations are made for future areas of research and changes in policy to ensure such transparency and reproducibility.

13.24 Research needs relate to the following areas:

Addressing the best use of existing data:

- Exploration of methods for assessing the quality of the toxicological evidence and the sources of uncertainty and variability

- Development of a framework for transparent expression of uncertainty in hazard characterisation, such as addressing and identifying critical data gaps.

Vulnerable sub-groups:

- Improved understanding of the relevance to susceptibility of the genetic polymorphisms that have been identified in human populations

- Evaluating whether there are specific subgroups not protected by the default uncertainty factors, due to genetic, physiological (e.g. early and older life) or environmental sources of variability

- Developing valid mechanism-based biomarkers of uptake, effect and susceptibility that would help to identify subgroups at risk

- Better characterisation of hazards to older people to determine whether current uncertainty factors are appropriate

Mixtures of substances:-

- Improve understanding of the combined effects of chemicals occurring in food.

13.25 In relation to policy and practice, it is recommended that:

Hazard characterisation:

- Hazard identification and characterisation should take into account variability and uncertainty, using a systematic approach that will facilitate transparency and confidence

- Greater use should be made of statistical and modelling approaches, including probabilistic and physiologically-based pharmaco- and toxicokinetic models. Use of such methods, when suitable data are available, would support a more systematic approach to risk assessment allowing for variability within the human population

- Subdivision of the default uncertainty factors to incorporate chemical-specific toxicokinetic and toxicodynamic adjusted factors should be used, whenever data allow

Risk communication:

- The development of a framework for transparent expression of uncertainty in hazard characterisation would enable COT and other committees that perform toxicological evaluations to improve communication of the sources of variability and uncertainty in their risk assessments. Particular attention should be given to describing assumptions and subjective elements of the risk assessment, clearly

describing where contradictions in information occur and how the resultant uncertainty is resolved.

SUMMARY OF REVIEW OF APPROACHES FOR EVALUATING UNCERTAINTY – JANUARY 2010

In its 2007 report on variability and uncertainty, the COT recommended 'development of a transparent expression of uncertainty in hazard characterisation'. This paper summarises the preliminary findings of a review[1] of existing approaches conducted to provide a basis for developing such a framework, suitable for use by the COT. The remit for the review focussed on qualitative approaches, as these were expected to be more suitable than quantitative (e.g. probabilistic) approaches for the majority of COT assessments. However, qualitative approaches should, if possible, be conceptually compatible with valid mathematical approaches to uncertainty.

Potentially relevant approaches were identified via Web of Science, Google, contacts with relevant EU, US and Canadian authorities, and the reviewers' previous experience.

Criteria for assessing suitability for COT were drafted in consultation with FSA (see introduction to draft framework). None of the existing approaches we identified meet all the criteria, but some contain potentially useful features. These are briefly summarised below in alphabetic order.

EFSA and REACH – EFSA Journal (2006) 438, 1-54; REACH guidance, Chapter R.19.

All parts of the assessment are examined for uncertainties, which are listed in a table. The individual and combined impact of the uncertainties is evaluated, with a scoring system using plus and minus symbols to indicate the direction and magnitude of their potential effects on the assessment outcome. However, users rarely define the quantitative scale for these judgements. The approach was originally developed for use in exposure assessment and has also been used for toxicity and risk, but there is no guidance on how to apply it to assessments where the conclusions are qualitative (e.g. is this chemical a carcinogen?).

Evidence maps – Wiedemann & Schutz (2008) The role of evidence in risk characterization.

A simple graphical format for summarising the arguments for and against a given hazard or risk, factors that attenuate those arguments (e.g. weaknesses of the underpinning studies), the overall evidence base (number of relevant studies), the current conclusion and remaining uncertainties. Strengths and weaknesses of the evidence and the overall conclusion are expressed by narrative statements. Map could become unwieldy for assessments with many different lines of evidence or uncertainties. No guidance on how to combine the different lines of evidence and uncertainties, or on how to apply the approach to quantitative conclusions (e.g. exposure or reference dose).

[1] FSA project T01056. Participants: A Hart & JP Gosling (Fera), A Boobis (Imperial College), D Coggon (Southampton Univ.), P Craig (Durham Univ.), D Jones (Leicester Univ.).

Health Canada Contaminated Sites Division (HC CSD, pers. comm.)

Draft guidance documents developed by HC CSD state that sensitivity analysis for a deterministic risk assessment should consist, at a minimum, of a qualitative summary of the uncertainties and variability associated with each input variable and a prediction of how these uncertainties are expected to affect the risk estimates. Example assessment extracts provided by Health Canada variously included narrative text sections discussing uncertainties in each part of the assessment (exposure, toxicity, risk characterisation), tables listing and justifying key assumptions and describing their potential impact on the assessment (e.g. neutral, underestimate, overestimate, unknown) and, in one example, a summary statement on the reasonableness of the assessment and its degree of conservatism.

IPCC – Guidance on addressing uncertainties (Intergovernmental Panel on Climate Change, 2005)

Very concise guidance (4pp.) for lead authors of IPCC 4th Assessment Report on climate change. Offers three approaches for characterising uncertainty: (a) express evidence and level of consensus on two scales of low to high, (b) scale of terms for confidence that an analysis or statement is correct, from very low (less than 1 out of 10 chance) to very high (at least 9 out of 10 chance), (c) scale of terms for likelihood of some well defined outcome, from exceptionally unlikely (<1% probability) to virtually certain (>99%). Little specific guidance on how to determine which level of confidence, likelihood etc. applies. Primarily aimed at achieving consistency in communication.

Pedigree analysis (NUSAP) – www.nusap.net

Pedigree analysis is intended for characterising 'deep' uncertainties that cannot be easily quantified, including qualitative issues such as problem framing, choice of methods, level of knowledge or consensus, and value-ladenness. Users define the issues that are relevant for their problem, and rate them on a scale from low to high, which they also define. A number of publications summarise the scores using "kite" or radar diagrams, but more recently these have been considered misleading and replaced with a type of bar chart. Some publications assign quantitative scores to the ratings and combine them by averaging across issues and across different parts of the assessment. Characterises uncertainties but does not attempt to indicate how much they might change the conclusion.

US NRC – US Nuclear Regulatory Commission guidance on treatment of uncertainties associated with probabilistic risk assessment (NUREG-1855, 2009 and EPRI, 2009)

These two documents concentrate mainly on quantitative approaches but include qualitative screening steps, that are used to identify uncertainties requiring sensitivity analysis or probabilistic modelling. NUREG-1855 states that the final output should include a qualitative statement of confidence in the conclusion of the assessment and how it has been reached, supported by identification of key uncertainties that were addressed. EPRI (2009) includes a tabular format for listing sources of model uncertainty and narrative evaluations of their impact on the model and risk characterisation/degree of conservatism.

WHO/IPCS – (2008) Uncertainty and data quality in exposure assessment.

Includes a qualitative approach similar to Pedigree analysis, but prescribing 11 issues to be evaluated and scales for evaluating them. The 11 issues are grouped in 3 dimensions: level of uncertainty (1 scale), appraisal of the knowledge base (scales for accuracy, reliability, plausibility, scientific backing, robustness) and subjectivity of choices (scales for choice space, inter-subjectivity, influence of limitations, sensitivity of choices, influence of choices). Scores on the 11 scales are displayed in a separate 'evaluation matrix' for each source of uncertainty. These are then summarised into a single table, accompanied by a brief explanation of the weights used in reaching an overall conclusion, but no guidance is given on the method for doing this. The approach expresses the nature and severity of uncertainty rather than its impact on the estimated exposure, and appears liable to become rapidly cumbersome as the number of uncertainties increases.

Other approaches reviewed include a number related to weight of evidence assessment, including GRADE (www.gradeworkinggroup.org), a system for evidence assessment that takes explicit account of uncertainties but is strongly focussed on evaluation of clinical trials and procedures. UK Climate Impacts Programme (UKCIP) guidance on risk and uncertainty assessment is closely based on Pedigree analysis and the IPCC approach (see above). The US EPA Risk Characterization Handbook (2000) states the assessor should identify residual uncertainties and their impact on the range of plausible risk estimates, but also states that no single recognised guidance currently exists for uncertainty analysis.

In summary, several existing approaches contain elements that are potentially useful in constructing a framework suitable for use by the COT. Key elements of this are likely to include: systematic identification and listing of relevant uncertainties, identification of which of these are covered by the standard uncertainty factors, a tiered approach to identify when additional evaluation is required, a method to evaluate uncertainties in terms of their impact on the assessment outcome, and narrative description of uncertainties whose impact cannot be evaluated. The key challenge will be devising meaningful scales for expressing the impact of uncertainties that are both compatible with mathematical principles, and practical and intuitive for use by the COT.

DRAFT FRAMEWORK FOR EVALUATING UNCERTAINTY

In its 2007 report on variability and uncertainty, the COT recommended that 'hazard identification and characterisation should take into account variability and uncertainty, using a systematic approach that will facilitate transparency and confidence'.

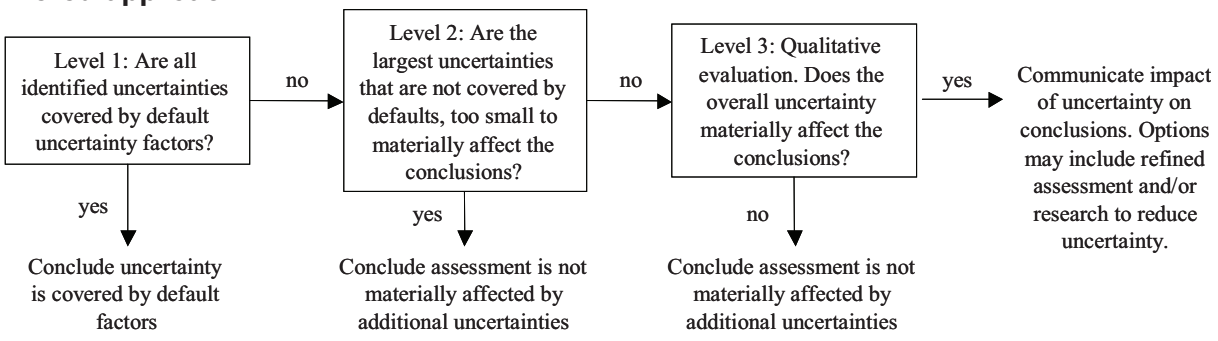
This draft framework is intended to meet that need. It is designed to be:

- Practical for use by the COT and other FSA committees, and adapted to their work
- Systematic & comprehensive, helping the user address all relevant uncertainties
- Efficient, using a tiered approach to minimise the effort required
- Helpful for developing conclusions and in subsequent decision-making, by evaluating uncertainties in terms of their impact on the key issues
- Conceptually compatible with mathematical approaches to uncertainty, where possible

Integration with Committee procedure

Secretariat draft review paper	Secretariat identify and list potentially relevant sources of uncertainty, assisted by checklist (*Annex 1*), and identify which are covered by default uncertainty factors. It may help to make separate lists for different aspects of the assessment (e.g. exposure, toxicity, etc.).
Committee discussion, leading to development of conclusions	Secretariat add additional uncertainties identified in discussion. Committee evaluate identified uncertainties using tiered approach (see below).
Committee report/statement	Include summary of uncertainty evaluation

Tiered approach[2]



Evaluation methods

Level 1: Review the list of uncertainties and judge whether all are covered by default factors.

Level 2: Review those uncertainties that are not covered by default uncertainty factors, and identify any that might materially affect the conclusions.

[2] The three levels for uncertainty evaluation are not tied to corresponding tiers of hazard, exposure or risk assessment.

Level 3: Qualitatively evaluate individual and combined uncertainties using methods from *Annex 2*. Refined assessment (>Level 3): consider analysing key uncertainties quantitatively, starting with simple methods such as what-if calculations.

Communication of results

- Include a brief summary of the uncertainty evaluation in Committee conclusion: 1-2 sentences indicating the nature of the key uncertainties and summarising your overall evaluation.
- Briefly describe any additional uncertainties whose impact could not be evaluated.
- Include list or table of identified uncertainties (and Level 3 evaluation if done) in the discussion section or as an annex.

Annex 1. Common types of uncertainty in food safety risk assessments

This checklist is intended as a prompt. It is not exhaustive and should be updated when appropriate.

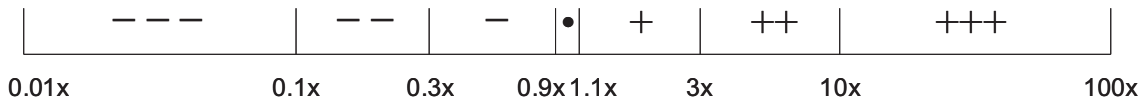
- Measurement uncertainty including accuracy, precision and detection/reporting limits
- Sampling uncertainty (variability & bias)
- Other study quality/design issues including ambiguity and inadequate reporting
- Inconsistency of results across multiple studies
- Extrapolation from animals to humans, between age and sex classes, etc.
- Variability between individuals in the population under assessment
- Relevance of data to assessment scenario, and use of surrogate data
- Uncertainty of expert judgements, including differences between experts
- Applicability of default assumptions or uncertainty factors
- Uncertainty about which factors/mechanisms to include
- Uncertainty about structure of conceptual or quantitative models; residual confounding in clinical studies
- Dependencies between different elements of assessment or model
- Gaps in knowledge, especially (but not only) for novel problems

Annex 2. Methods for evaluating uncertainties at Level 3

Use available evidence and expert judgement to evaluate uncertainties in terms of their potential impact on the assessment outcome, expressed using the following common scale: +++, ++, +, •, –, – –, – – –. The interpretation of the scale differs, depending whether the issue under assessment is quantitative or qualitative. If the assessment addresses multiple issues, evaluate each separately. If two or more uncertainties or lines of evidence are strongly related, assess their effect jointly.

Evaluation for quantitative conclusions (e.g. a threshold dose or exposure estimate)

1. Define a convenient scale for the evaluation symbols, e.g. by adjusting +++ or – – – to the magnitude of the largest identified uncertainties. An example of a potentially useful scale is:



2. Consider each identified source of uncertainty in turn, and use the chosen scale to express your judgement of how much the final endpoint[3] of the quantitative assessment might change if that uncertainty was resolved. Express your uncertainty about this by using pairs of symbols (e.g. –/++) to cover the range in which you are 95% sure the adjusted estimate would lie. Record your evaluations for all the identified uncertainties in a table.
3. Review the table of evaluated uncertainties and form a judgement about their overall, combined impact, i.e. how much the final endpoint might change if all the uncertainties were resolved. Express this using the same symbols, and as a short narrative for use in the conclusions.

Evaluation for qualitative conclusions (e.g. is this chemical a carcinogen?)

1. In this case, the appropriate scale is a probability or percentage, expressing your degree of belief that the qualitative condition is true (0% = certainly false, 100% = certainly true). A potentially useful scale is as follows:

Symbols	– – –	– –	–	•	+	++	+++
Probability	<1%	1-10%	10-33%	33-66%	66-90%	90-99%	>99%

(IPCC, 2005)

2. Consider each relevant piece of evidence in turn, and use the scale to express the probability that the condition is true *if based on that evidence alone*[4]. Express your uncertainty about this probability by using one or more symbols, e.g. +/++ means based on this evidence alone, you are 66-99% sure the condition is true. Repeat this for all relevant lines of evidence.

3. Review your evaluations for all lines of evidence and use them to help you form an *overall judgement* about the probability that the condition is true. Express this using the same symbols, and also as a short narrative statement for use in the assessment conclusions.

[3] If the endpoint of the assessment is a variable, e.g. the distribution of exposures in a population, then identify the percentile of interest for risk management and evaluate the uncertainties in terms of their impact on that percentile.

[4] It may be helpful to first express your initial judgement of the probability, i.e. before considering any of the evidence.

WORKSHOP AGENDA – 2nd February

	Details
15:30 – 16:00	Registration and refreshments
16:00 – 16:10	Welcome and introduction – Alan Boobis, COT (Workshop Chair)
16:10 – 16:35	Presentation: Review of approaches for addressing uncertainty – Andy Hart, FERA
16:35 – 16:50	Presentation: Draft recommendations and framework – Alan Boobis
16:50 – 17:50	Plenary discussion
17:50 – 18:00	Summary – Alan Boobis
18:00	Close
19:30	Drinks reception followed by dinner

WORKSHOP AGENDA – 3rd February

	Evaluation of Draft Framework
09:00 – 09:15	Introduction and briefing for breakout discussions – Alan Boobis
09:15 – 10:45	Breakout discussions – Evaluation of the draft framework using four case studies
10:45 – 11:15	Refreshments
11:15 – 12:30	Breakout discussions continue – Evaluation of the draft framework using four case studies
12:30 – 13:30	Lunch
13:30 – 15:00	Plenary – Breakout conclusions from each group (10 mins presentation & 10 mins discussion)
15:00 – 15:30	Refreshments
15:30 – 16:45	Plenary discussion
16:45 – 17:00	Closing remarks – David Coggon, COT Chair
17:00	Close

BREAKOUT GROUP DETAILS

BREAKOUT GROUP TASKS

1. Apply the draft framework to the case study
2. Evaluate the clarity of the framework, it's applicability to COT work, usability, value added, suggested changes, potential concerns and possible solutions
3. Raise any additional questions arising from the discussions on day 1
4. Prepare a 10 minute presentation on points 1 and 2 above

PFOA CASE STUDY

Facilitator Andy Hart

Invited Experts
Ian Morris
John Foster
David Ray
Alison Ward
Andy Renwick
Jon Ayres
Frances Pollitt

Secretariat
David Gott
Britta Gadeberg

PEANUT CASE STUDY

Facilitator Peter Craig

Invited Experts
Rebecca Dearman
David Tuthill
Anna Hansell
Robert Smith
Peter Aggett
Ian McManus

Secretariat
Frances Hill
Cath Mulholland
Dave Parker

CAFFEINE CASE STUDY

Facilitator Alan Boobis

Invited Experts David Coggon
Justin Konje
Alma Williams
Nick Plant
Joyce Tait
Phil Carthew
Donald Davies

Secretariat Natalie Thatcher
Gary Welsh

TOTAL DIET STUDY (Al & Pd) CASE STUDY

Facilitator David Jones

Invited Experts David Harrison
Cliff Elcombe
Brian Lake
Derek Bodey
Lesley Stanley
Andy Smith
JP Gosling

Secretariat Rosalind Harrison
Joseph Shavila

BREAKOUT GROUP REPORTS

This section contains the reports prepared by the 4 breakout groups during day 2 of the workshop and presented in the afternoon session on that day.

Breakout Report - PFOA

PFOA

Process

- Considered available data and determined going straight to Level 2 assessment
- Chemical Information
  - Uncertainty over what has been tested
  - What are people exposed to – it's in environment and found in human serum

Genotoxicity and Carcinogenicity

- Is it a mutagen?
- *No* with some uncertainties which don't affect conclusion
- Animal carcinogen
- *Yes* 100% certain
- Animal data relevant to humans?
- *Could be*

Genotoxicity and Carcinogencity

- Want COM and COC to provide measures of uncertainty on conclusions they reach to advise COT
- Selection of threshold either cancer or other toxicity as relevant

Other Toxicity data

- Considered Developmental and Reproductive data
- Questioned whether neurotoxicity or immunotoxicity data were available
- Consider how comprehensive database is
- Confidence in lowest NOAEL depends on
  - How complete is the dataset
  - Is there evidence in the studies of other effects (e.g. neuro/immunotoxicity)

$BMDL_{10}$ selection

- Compared BMDL10's from various relevant studies
- Need to think about whether it is likely that there is another effect with a lower NOAEL
- Also think about uncertainty over the dose(s) administered

Breakout Report - PFOA

Toxicokinetics

- Sex and species differences
- Protein binding – consider importance of free concentration – big uncertainty
- Differences in approach taken by US EPA compared to COT and EFSA (update statement)

Certainty around final assessment

- "We consider that on the basis of available information this provisional TDI is adequate to protect against the range of identified toxic effect" (COT 2009 update statement)
- Consideration of TDI
  - Adequate, overconservative or underconservative
- Consideration of exposure data
  - Upper bound concentrations in food
  - Good MOE even if concerns over adequacy of the TDI

Comments on framework approach

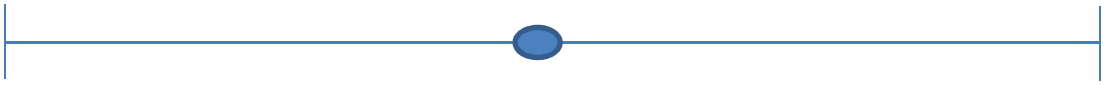
- *Until you make a list difficult to determine if Level 1 assessment appropriate*
- *Need to separate Analysis of Evidence from Decision*
- *There's a need to document what evidence has been gathered by the Secretariat to identify data gaps*
- *Labour intensive*
- *Difficult to communicate*

Flip charts from PFOA breakout group: 1

Mutagenic?

no yes

Comments:
-Would like COM to provide this judgment
- this is key in deciding whether to apply threshold



Evidence of carcinogenicity in human studies?

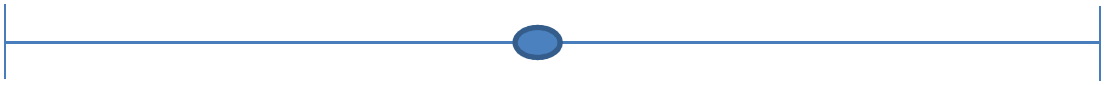
no yes

Evidence of carcinogenicity in animal studies?

no yes

Mode of action?

Unknown Known



Relevant to humans?

no yes

Conclusion: proceed on basis it is relevant

Threshold dose-response?

no yes

Comment: would be higher if had a MoA

Flip charts from PFOA breakout group: 2

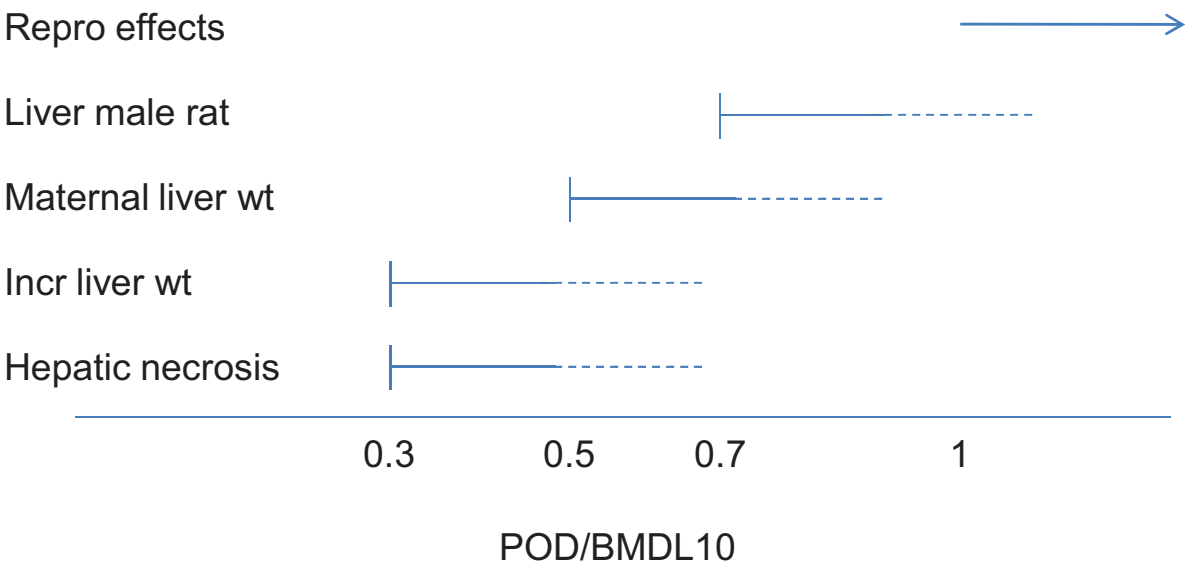
Developmental and reproductive effects

Teratogenic?

no yes

Reproductive effects in absence of maternal tox?

no yes



Pancreatic effects?

no yes

Any other unstudied effects with lower thresholds?

No Yes

Flip charts from PFOA breakout group: 3

Kinetics

Confident of half-life in animals?

no yes

Confident of half-life in humans?

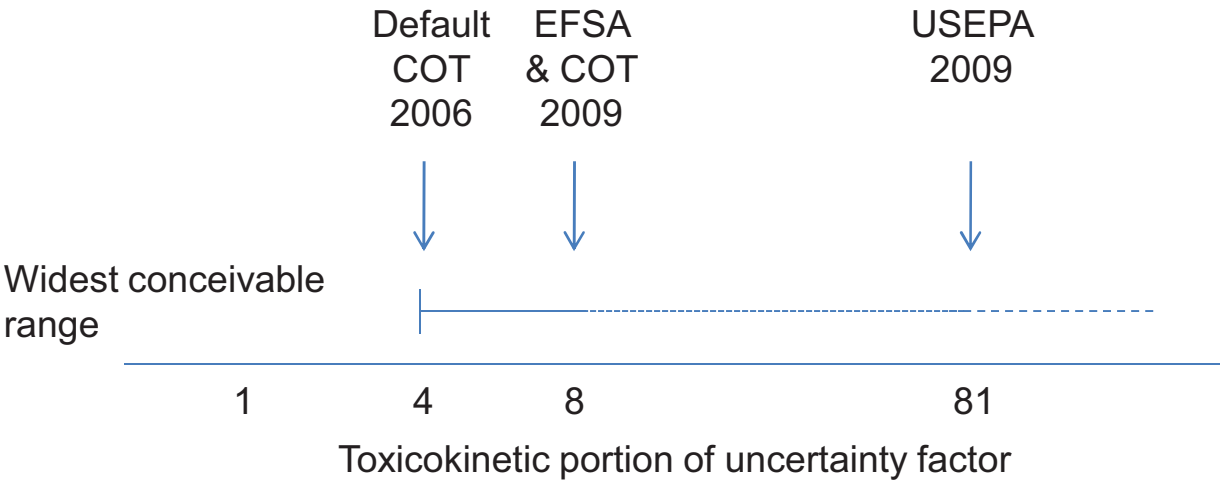
no yes

Comments:
- Post-occupational exposure assumed zero
- no mechanistic explanation for large difference from animals

Confidence that half life reflects toxicity

no yes

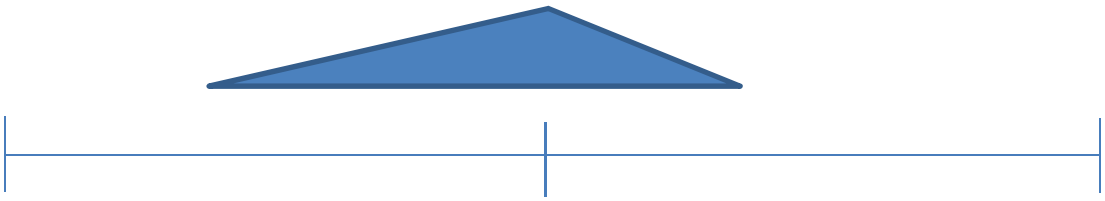
Comment: based on rat data



Nb: USEPA value based on mouse-human clearance ratio

Flip charts from PFOA breakout group: 4

Conclusions



Under-protective | Adequate | Over-protective

Is provisional ADI adequate to protect against the range of identified effects?

Is current exposure of concern?

no yes

Comments:
- Taking account that exposure estimates are conservative

Breakout Report - Peanuts

Would a mother eating peanuts during pregnancy increase the risk of an infant developing peanut allergy?

Areas of Uncertainty

Route of exposure Trans-placental blood transfer of allergens Does trans-uterinal diffusion occur? breast milk maternal trans-cutaneous maternal inhalation	**Timing of exposure** when during gestation
Interaction Different peanut allergens Cross-reactivity (e.g. egg and lupines)	**Vulnerable groups** Do familial links to atopy matter?
Dose External maternal exposure Internal foetal dose Is response(s\ threshold? what is the threshold?	**Compound** What peanut allergen(s) causes infant allergy? Does cooking preparation matter?
Analytical variability Sampling Measurement variability	**Available results** Do available data and analyses agree?

Areas of uncertainty

Exposure	Degree of certainty	Impact on question	Effect on risk
Dose	0	++	↔
Vulnerable groups	++	++	↑
Outcome (infant allergy)	+++	+++	n/a
Cord blood (surrogate)	+	+	↔

KEY to scales

Degree of certainty	Impact on question	Effect on risk
0 = uncertain	0 = no importance	↑ = raises
+ = some uncertainty	+ = some importance	↓ = lowers
++ = quite uncertain	++ = quite import	↔ = equivocal
+++ = certain	+++ = important	

Current COT Advice

- "There are insufficient data in humans to be able to conclude whether maternal consumption or avoidance of dietary allergens during pregnancy or lactation has an impact on the incidence of allergy, sensitisation or tolerance to food allergens in the child, and there is no consistent evidence on whether, and if so in what way, the timing of exposure to allergens is critical."
- "The advice for atopic mothers to avoid peanuts during pregnancy was considered to be no longer appropriate."

Feedback on the process

- Needs a flexible approach – case by case
- Qualitative rather than quantitative – need to decide early on
- Words may be easier to communicate
- Clarity of definitions – need to ensure consistency for scale, question etc
- Feels "backwards" as we already have the conclusions
- Was hard to fit the framework to this case study
- May make the decision making process more open

Breakout Report - Caffeine

Caffeine Case Study

Main Issues

- Is caffeine intake during pregnancy associated with increased risk of FGR?
- What is the likelihood the association is causal
- What is the lowest level of intake above which the risk of FGR is increased?
- If the relationship is causal what is the maximum increase in incidence of FGR (above the residual) from intakes of 200 mg/day

Uncertainties Identified

- Systematic error in the assessment of caffeine intake
- Random error in the assessment of caffeine intake
- Study sample systematically unrepresentative of wider population
- Study sample unrepresentative by chance

Uncertainties Identified (Continued)

- Model uncertainty
- Systematic error in the assessment of FGR
- Random error in the assessment of FGR
- Is the internal dose of the exposure highly variable between individuals

Quantitative Evaluation

- Systematic error in estimated caffeine intake: •
- Random error in estimated caffeine intake: -
- Population was estimated to be reasonably representative: •
- Study sample was unrepresentative of the population by chance: +/-

Quantitative

Quantitative Evaluation (Continued)

- Model uncertainty: +/-
- Systematic error in FGR: •
- Random error in FGR: •
- Is the internal dose of the exposure highly variable between individuals: No value assigned, as this was considered in the above

Overall Assessment of Uncertainty: +/-

Quantitative

Breakout Report - Caffeine

Qualitative Evaluation

Is caffeine a cause of fetal growth restriction?

Lines of evidence to be considered:

- Animal experiments
- New epidemiology study (including dose response relation)
- Other epidemiology observation studies in humans
- Biological plausible mechanism

Qualitative

Contribution to overall conclusion/Weight of evidence

Line of Evidence - Animal experiments

-Diminishes the likelihood that there is a causal relationship because there isn't a consistent effect on FGR and where there is an effect there is also maternal toxicity.

Contribution to overall conclusion +

Qualitative

Contribution to overall conclusion/Weight of evidence

New epidemiology study (including dose response relation)

– The main uncertainty is about possible residual confounding and the judgement is there is a low probability of that. Contribution to overall conclusion

Contribution to overall conclusion ++

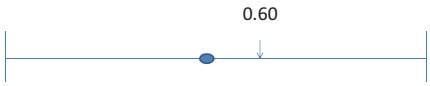
Qualitative

Contribution to overall conclusion/Weight of evidence

Other epidemiology observation studies in humans

– The variability in findings between studies and huge variability in study methodologies and limitations in study design

Contribution to overall conclusion +



Qualitative

Contribution to overall conclusion/Weight of evidence

Biological plausible mechanism

– Nobody has identified a plausible mechanism

Contribution to overall conclusion +

Qualitative

Impact on overall conclusion: ++

(NB. Driven by evidence from new epidemiology study, supported by other observational human studies. Noting lack of evidence of plausible mechanism from human data and lack of evidence of causal relationship from animal studies.)

Qualitative

Breakout Report - Caffeine

Clarity of the framework

- Unclear
- Easily misinterpreted
- Needs wording in clearer language

Summary

Applicability to COT work

- Might be applicable but is not useable
- Is a basis for moving forward

Summary

Usability

- Unusable in current form
- Time consuming

Summary

Value Added

- Increases transparency – but is it possible to communicate this?
- Will help ensure a more rigorous assessment

Summary

Suggested Changes & Possible Solutions

Qualitative

- Use simpler language
- Use IPCC style descriptors – No lines and No numbers
- Provide a statement using standardised IPCC descriptors
- Provide a statement explaining which lines of evidence weighed most strongly in coming to a conclusion

Summary

Suggested Changes & Possible Solutions

Quantitative

- Best estimate for the quantitative parameter
- Describe credible range of values around the best estimate

Summary

Breakout Gp: Total Diet Study (TDS): Aluminium & Palladium

- Group members: David Harrison, Cliff Elcombe, Brian Lake, Derek Bodey, Lesley Stanley, Andy Smith, Jean-Paul Gosling, *Rosalind Harrison, Joseph Shavila,* David Jones
- COT asked '**if the levels of any of the elements in the diet posed a risk to human health**'

Conclusions from TDS report

(COT, 2006)

97. We *note* that whilst the estimates of dietary exposure to **aluminium** are not markedly higher than previous estimates, they present uncertainty with regard to the safety of aluminium in food in light of new data that led to the recent reduction in the Provisional Tolerable Weekly Intake (PTWI), which is exceeded by some population subgroups. There is a need for further information on possible sources and forms of aluminium in the diet and their bioavailability.

103. The toxicological database on **palladium** metal and its compounds is extremely limited. However, we *conclude* that from the available data, there is no reason to believe that current intakes of palladium from the diet pose a risk to health.

Discussion

- Much of the decision making is qualitative, or at best "semi-quantitative"
- Felt we could not apply probability values meaningfully
- Wanted a scheme that was robust, transparent and systematic – and involved all members in decision making

Tiers 1 & 2

FOCUSSED ON ALUMINIUM, PALLADIUM
RATHER RUSHED AT END

- Identified issues such as significant gaps in knowledge as well as uncertainties, and felt that these should be formally recorded
  - for example: absence of longterm studies, formal repro tox studies

Checklist of types of uncertainty

(from AH/AB)

- Measurement uncertainty
- Sampling uncertainty
- Other study/methodology issues
- Variability of studies
- Bioavailability data
- Extrapolation from animals to humans, etc.
- Variability in the population and individuals
- Relevance of data
- Relevant factors/mechanisms
- Speciation
- Residual confounding
- Dependencies
- Total exposure vs food
- Gaps in knowledge
- Applicability of defaults

Tiers 1 & 2

- Left with sufficient residual uncertainties and so proceeded to tier 3.
- We felt that uncertainty factors should be applied late in deliberation, rather than at early stages
- Identified issues such as **gaps in knowledge** as significant, and felt that these should be formally recorded

Tier 3

- **Data lines chosen**
  - Hazard
  - Exposure & bioavailability
  - Individual susceptibility
  - Consensus
- **Two questions**
  - How good is the data in each of these data lines to make call?
  - What is the call?

How good is the data?

A record of each member's degree of certainty

- 1= very weak data
- 5 = very strong data

CONSENSUS

No attempt to group or average this data as it is non linear

- Hazard 2,2,2,2,2,1,2
- Exposure 3,2,2,2,2,2.5(!!),2
- Individual susceptibility 1,1,1,1,1,1,1
- Consensus 5, that is **we agreed that there is considerable uncertainty**

So what is the risk assessment?

- (i) Qualified statement using the uncertainty/ reliability assessment as per previous slide, that is "There is considerable uncertainty…. with significant gaps in data available, however members concluded that….."
- (ii) Conclusion on risk assessment, not scored but using words as descriptors

Conclusions from TDS report

we agree with…….

97. We *note* that whilst the estimates of dietary exposure to aluminium are not markedly higher than previous estimates, they present uncertainty with regard to the safety of aluminium in food in light of new data that led to the recent reduction in the Provisional Tolerable Weekly Intake (PTWI), which is exceeded by some population subgroups. There is a need for further information on possible sources and forms of aluminium in the diet and their bioavailability.

103. The toxicological database on palladium metal and its compounds is extremely limited. However, we *conclude* that from the available data, there is no reason to believe that current intakes of palladium from the diet pose a risk to health.

Benefits

- Systematic - but probably capturing what we already do
- Suggest that we have a quality checklist for all data/literature reviewed
  - Formal gap analysis possible
- Auditable
- Transparent
- Defensible
- Inclusive – for both specialist and non-specialist members

SUMMARY CONCLUSIONS

Closing remarks by Dr David Coggon, COT Chair

1. In deciding whether and how to modify current COT practice regarding the evaluation and expression of uncertainty, it is necessary to weigh expected benefits against costs and risks.

Possible benefits are:
a) Better evaluation of evidence
b) Better communication of conclusions and of the underlying rationale to Government, the scientific community (including the COT in the future) and the public

Costs arise from the additional time and effort that might be required.

The main risk identified was the potential for misinterpretation by others (particularly the media and the public) if descriptions of uncertainty are too complex.

2. The discussion highlighted the need for careful specification of the questions that are addressed by the Committee – what exactly does the user need to know?

3. Currently, the rationale for COT conclusions does not always come across as clearly as we would wish.

4. There was support for a tiered approach to uncertainty, with most detailed consideration where uncertainties could impact critically on the decisions that would follow from a risk assessment.

5. Different approaches are needed for the expression of uncertainty, according to whether questions are qualitative or quantitative.

6. In describing uncertainties in qualitative assessments, it will be best to avoid the use of numbers, which are difficult to specify and liable to misinterpretation. However, it should be clear which lines of evidence weighed most heavily in coming to conclusions, and what, if any, are the major sources of uncertainty. Moreover, the wording that is used to express levels of uncertainty should where possible be standardised. Standardisation of terms will be more effective if it is agreed across FSA, and if possible, more widely.

7. The best way of expressing uncertainty for quantitative estimates of parameters such as risk may vary according to the question that is being addressed. For example, one question might relate to an extensive database, and require a tight answer (perhaps to within +/- 50% or less). In these circumstances, it may be most helpful to specify a numerical credibility interval. For another question, decisions might hinge on accuracy only to within an order of magnitude, and uncertainties might best be expressed in less precise language.

LIST OF PARTICIPANTS

COT MEMBERS

DELEGATE	AFFILIATION
Professor David Coggon	University of Southampton T01056 Team Member (COT Chair)
Mr. Derek Bodey	Consultant on Consumer Affairs (COT)
Professor Alan Boobis	Imperial College London T01056 Team Member (COT)
Dr Rebecca Dearman	University of Manchester (COT)
Dr Cliff Elcombe	CXR Biosciences Ltd (COT)
Dr John Foster	AstraZeneca (COT)
Dr Anna Hansell	Imperial College London (COT)
Professor David Harrison	University of Edinburgh (COT)
Professor Justin Konje	University of Leicester (COT)
Professor Brian Lake	University of Surrey (COT)
Professor Ian Morris	Hull York Medical School (COT)
Dr Nick Plant	University of Surrey (COT)
Dr David Ray	University of Nottingham (COT)
Mr. Robert Smith	Consultant on Consumer Affairs (COT)
Dr David Tuthill	Cardiff and Vale University Health Board (COT)
Ms Alison Ward	Independant Consultant (COT)
Ms Alma Williams	Consultant on Consumer Affairs (COT)

INVITED EXPERTS

DELEGATE	AFFILIATION
Dr Andy Hart	FERA T01056 Project Leader
Dr Peter Craig	Durham University T01056 Team Member
Professor David Jones	University of Leicester T01056 Team Member
Dr John Paul Gosling	FERA T01056 Team Member
Professor Peter Aggett	University of Lancashire
Professor Jon Ayres	University of Birmingham
Dr Philip Carthew	Unilever
Professor Donald Davies	Imperial College London
Mr Ian McManus	Chemicals Regulation Directorate
Ms. Frances Pollitt	Health Protection Agency
Professsor Andrew Renwick	University of Southampton
Dr Andrew Smith	MRC Toxicology Unit University of Leicester
Dr Lesley Stanley	Consultant in Investigative Toxicology (T01 Programme Advisor)
Professor Joyce Tait	University of Edinburgh

FSA STAFF

DELEGATE	AFFILIATION
Mrs. Hattie Lambrou	FSA – Communications Division
Dr. Diane Benford	FSA – Head of Chemical Risk Assessment Unit
Ms. Britta Gadeberg	FSA – Chemical Risk Assessment Unit
Dr. David Gott	FSA – Chemical Risk Assessment Unit
Ms.Toni Gray	FSA – Chemical Risk Assessment Unit
Ms. Rosalind Harrison	FSA – Chemical Risk Assessment Unit
Mrs. Frances Hill	FSA – Chemical Risk Assessment Unit
Ms. Cath Mulholland	FSA – Chemical Risk Assessment Unit
Dr. David Parker	FSA – Chemical Risk Assessment Unit
Dr. Joseph Shavila	FSA – Chemical Risk Assessment Unit
Dr. Natalie Thatcher	FSA – Chemical Risk Assessment Unit
Mr. Gary Welsh	FSA – Chemical Risk Assessment Unit

APPENDIX – PRESENTATIONS ON

REVIEW OF APPROACHES AND DRAFT FRAMEWORK

As presented at the Workshop

Appendix – Workshop presentation - Framework

Review of approaches for evaluating uncertainty

Andy Hart
The Food and Environment Research Agency
York

COT Report on variability and uncertainty, 2007

Research recommendations:

- Exploration of methods for assessing the quality of the toxicological evidence and the sources of uncertainty and variability
- Development of a framework for transparent expression of uncertainty in hazard characterisation, such as addressing and identifying crucial data gaps

Need recognised at national and international levels

- **Phillips report**: 'Advice should identify the nature and extent of any areas of uncertainty'
- **Gov. Office for Science**, 2007: 'Committees should identify the sources and extent of uncertainties in the scientific analysis'
- **DG SANCO**, 2004: risk managers 'need to understand the level of uncertainty in your advice'
- **CODEX Working Principles**: 'Uncertainties should be explicitly considered at each step of the assessment… documented in a transparent manner …quantified to the extent that is scientifically achievable'

FSA Project T01056

- **Development of a framework for evaluation and expression of uncertainties in hazard and risk assessment**
  - Review existing approaches and evaluate suitability for use by FSA and its expert committees
  - Workshop to test most promising approach(es)
- Participants:
  - Andy Hart & John Paul Gosling, Fera
  - Alan Boobis, Imperial College
  - David Coggon, Southampton University
  - Peter Craig, Durham University
  - David Jones, Leicester University

Review of existing approaches

- Focus on qualitative approaches
- Search methods: Web of Science, Google, contacts with relevant EU, US, Canadian authorities, personal knowledge & contacts
- Search terms: combinations of uncertain*, exposure*, toxic*, risk*, framework*, evaluat*, assess*

Evaluation criteria (summary)

- Practical for use by the COT and other FSA committees, and adapted to their work
- Systematic & comprehensive, helping the user address all relevant uncertainties
- Efficient, using a tiered approach to minimise the effort required
- Helpful for developing conclusions and in subsequent decision-making, by evaluating uncertainties in terms of their impact on the key issues
- Conceptually compatible with mathematical approaches to uncertainty, where possible

Appendix – Workshop presentation - Framework

EFSA and REACH

Source of uncertainty	Direction & magnitude
One row for each source of uncertainty	*- -/+++*
	- - -/+
	etc.
Overall uncertainty of assessment outcome	*- - -/++*

- Uncertainties evaluated in terms of impact on outcome
- Users rarely define scale for +/- scores
- No guidance on how to apply to qualitative questions

Health Canada Contaminated Sites Division

- Guidance requires at least a qualitative summary of the uncertainties & their effect on risk estimates
- No specific guidance on methods or format
- Recently introduced, initial draft examples vary
- Summary text on overall uncertainty in at least some cases

Example of tabular format from 2 draft examples:

Factor	Uncertainty	Effect on risk assessment
(which part of RA)	*(description)*	*(qualitative evaluation)*

US Nuclear Regulatory Commission

- Requirement to characterise assumptions & sources of uncertainty in probabilistic assessments in terms of their effect on the quantitative estimates
- Includes example of tabular evaluation
- Results used to target key elements for sensitivity analysis

Topic	Discussion of issue	Part of model affected	Plant-specific approach taken	Assumptions made	Impact on model	Characterization assessment
						e.g. realistic, conservative, candidate for sensitivity analysis

IARC Cancer evaluations

Human studies + 'Hill criteria':
- Sufficient evidence of carcinogenicity
- Limited evidence
- Inadequate evidence
- Evidence suggesting lack of carcinogenicity

Animal experiments:
- Sufficient evidence of carcinogenicity
- Limited evidence
- Inadequate evidence
- Evidence suggesting lack of carcinogenicity

Mechanistic & other relevant data

Overall evaluation:
1: Carcinogenic to humans.
2A: Probably carcinogenic to humans.
2B: Possibly carcinogenic to humans.
3: Not classifiable as to its carcinogenicity to humans.
4: Probably not carcinogenic to humans.

- Well established; specific to cancer evaluations

Evidence maps*

attenuating
- All of these 5 studies demonstrate methodological deficits, above all with determination of exposure. They are therefore only conditionally meaningful.
- Theoretical considerations are missing in the selection of the tumours

Basis of evidence:
•122 studies
•13 selected

Pro-argument:
In 5 studies, noticeable results are found that point to a relationship between exposure to mobile phone communication fields and cancer.

Con-argument:
8 studies find no statistically significant relationship between exposure to mobile phone communication fields and cancer.

Conclusion:
- Vague initial suspicion (Stang/Jöckel)
- On this basis of evidence, an evaluation is not possible (Blettner)

attenuating
7 of these studies are only conditionally meaningful since they possess inadequate exposure determinations, too short exposure durations, too small sample sizes or other methodological deficits

Remaining uncertainties:
Studies are mostly exploratory and single out individual results

- Facilitates & communicates weight-of-evidence evaluation
- Could become unwieldy with more lines of evidence

* Schütz, Wiedemann & Spangenberg (2008)

IPCC - Intergovernmental Panel on Climate Change

- Defined terminology for different levels of uncertainty – suitability for COT assessments?
- Little specific guidance on how to do the evaluation

Chance of a statement being correct

Very High confidence	*At least 9 out of 10 chance*
High confidence	*About 8 out of 10 chance*
Medium confidence	*About 5 out of 10 chance*
Low confidence	*About 2 out of 10 chance*
Very low confidence	*Less than 1 out of 10 chance*

Probability of an outcome occurring

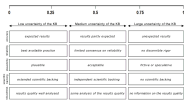
Virtually certain	*> 99% probability*
Very likely	*> 90% probability*
Likely	*> 66% probability*
About as likely as not	*33 to 66% probability*
Unlikely	*< 33% probability*
Very unlikely	*< 10% probability*
Exceptionally unlikely	*< 1% probability*

Appendix – Workshop presentation - Framework

Other approaches

- **Pedigree analysis (NUSAP)**
  - Intended for 'deep' uncertainties whose impact cannot be evaluated
  - May be sufficient to address these more simply

- **WHO/IPCS** – uncertainty in exposure assessment
  - Seems unnecessarily complex
  - Impact of uncertainty on outcome not expressed



Other approaches

Qualitative approaches include:

- GRADE: evaluation of clinical trials
- UK Climate Impacts Programme (NUSAP + IPCC)
- Martin et al. 2007: ratings for incidence trends, association & consequence
- EPPO invasive species RA: uncertainty rated low to high

Quantitative approaches include:

- Many publications on sensitivity analysis & probabilistic modelling
- Literature on mathematical representations of logic/evidence
- Turner et al. 2009: Bias modelling in evidence synthesis
- California EMF assessment methodology (Neutra et al.)

US EPA, 2000: requirement but 'no recognised guidance'

Steps to 'transparent expression of uncertainty in hazard characterisation'

- Combine the best features of existing approaches in a framework that addresses all three steps

Pedigree analysis WHO/IPCS

EFSA/REACH US NRC Canada Evidence Maps IPCC IARC

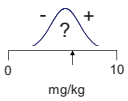
1. Identify uncertainties
2. Evaluate the uncertainties
3. Evaluate combined impact on hazard characterisation

Need to distinguish quantitative & qualitative questions

- **Quantitative questions:**
  - E.g. determine a threshold dose, estimate exposure…
- **Qualitative questions:**
  - E.g. is chemical X a carcinogen, is effect in animals relevant to humans?
- **Some assessments contain both types of question** – may be simpler to treat separately

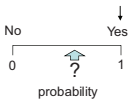
Quantitative questions

- Primary assessment output = threshold dose, estimated exposure, etc.
- Express uncertainty in terms of **how different** the estimate could be, and how likely that is
- Could do this quantitatively or qualitatively (e.g. +/- scores)
- Evaluate effect of each **source of uncertainty** on overall estimate



Qualitative questions

- Is chemical X a carcinogen, is effect E relevant to humans?
- Express uncertainty in terms of the **probability** of the condition being true
- Could do this quantitatively (e.g. probability) or qualitatively (e.g. scores or descriptive terms)
- Evaluate contribution of each **piece of evidence** to overall level of certainty

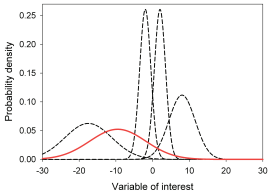


Appendix – Workshop presentation - Framework

Seeking a mathematical underpinning

- **Quantitative questions:**
  - Define quantitative scale
  - Transform scores to distributions
  - Combine → distribution for uncertainty of outcome
  - *Evaluate dependent factors jointly*

  Workable mathematical underpinning emerging



- **Qualitative questions:**
  - What scale to work on – probability or log-odds?
  - Raises questions about imprecise probabilities
  - Dependencies more challenging due to lack of explicit model

  Seek practical heuristic approach pending further research – based on expert judgement as now, but more explicit

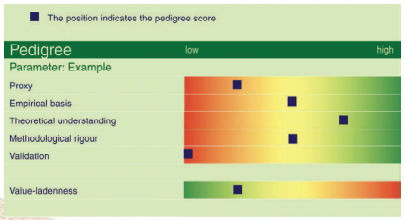
Summary so far…

- Used elements of existing approaches to develop a draft framework addressing all three steps of uncertainty characterisation (see next presentation)
  - included different methods for assessments of quantitative and qualitative questions
  - adapted to COT context (including use of default uncertainty factors) and procedures
- Next steps:
  - evaluate & improve proposed approaches (this workshop)
  - longer-term development of mathematical underpinning

Thank you!

Pedigree analysis (NUSAP)

*

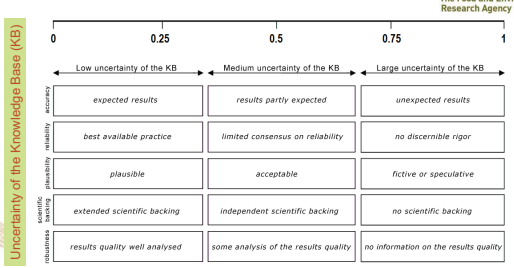


Example of Pedigree criteria for 'Empirical basis':
0 – crude speculation
1 – one expert opinion, rule of thumb estimate
2 – very small sample, model/ indirect data/ structured expert opinion
3 – small sample, direct data, less recent…
4 – large sample, direct recent data, controlled experiments

- Intended for 'deep' unquantifiable uncertainties
- User defines criteria and scales; apply to each parameter
- Can aggregate scores, but relation to outcome not defined

* graphic from Wardekker et al. 2008, Envir. Sci. & Policy, 11, 627-641

WHO/IPCS*



- Similar to Pedigree analysis, 11 defined scales (5 shown here)
- Apply to each source of uncertainty, then aggregate
- Potentially cumbersome; impact on outcome not expressed

* Guidance Document on Characterizing & Communicating Uncertainty in Exposure Asst.

Appendix – Workshop presentation - Review

Imperial College London

Draft framework for evaluating uncertainty

Alan Boobis

COT Report on Variability & Uncertainty, 2007

- 'hazard identification and characterisation should take into account variability and uncertainty, using a systematic approach that will facilitate transparency and confidence'

Draft framework design criteria

- Practical for use by the COT and other FSA committees, and adapted to their work
- Systematic & comprehensive, helping the user address all relevant uncertainties
- Efficient, using a tiered approach to minimise the effort required
- Helpful for developing conclusions and in subsequent decision-making, by evaluating uncertainties in terms of their impact on the key issues
- Conceptually compatible with mathematical approaches to uncertainty, where possible

Integration with COT procedure

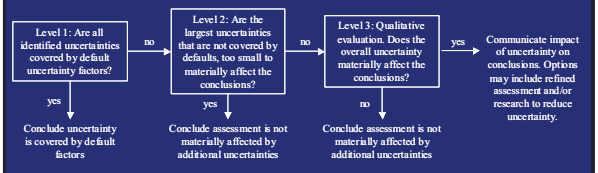
Secretariat draft review paper	• Secretariat identify and list potentially relevant sources of uncertainty, and identify which are covered by default uncertainty factors.
Committee discussion, leading to development of conclusions	• Secretariat add additional uncertainties identified in discussion. • Committee evaluate identified uncertainties using tiered approach.
Committee report/statement	• Include summary of uncertainty evaluation

Checklist of types of uncertainty (Annex 1)

- Measurement uncertainty
- Sampling uncertainty
- Other study issues
- Inconsistency of studies
- Extrapolation from animals to humans, etc.
- Variability in the population
- Relevance of data
- Expert opinion
- Applicability of defaults
- Relevant factors/mechanisms
- Model uncertainty
- Residual confounding
- Dependencies
- Gaps in knowledge

(Update as appropriate)

Tiered approach



N.b. The three levels for uncertainty evaluation are not tied to corresponding tiers of hazard, exposure or risk assessment

Appendix – Workshop presentation - Review

Evaluation methods

Level 1: review the list of uncertainties and judge whether all are covered by default factors.

Level 2: review those uncertainties that are not covered by default uncertainty factors, and identify any that might materially affect the conclusions.

Level 3: qualitatively evaluate individual and combined uncertainties using methods from Annex 2.

Refined assessment (>Level 3): consider analysing key uncertainties quantitatively, starting with simple methods such as what-if calculations.

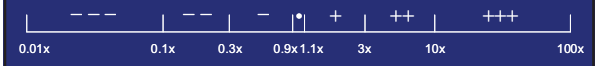
Evaluation methods for Level 3 (when needed)

- Method depends on nature of issue:
  - Quantitative: e.g. determination of exposure or TDI
  - Qualitative: e.g. determination of carcinogenicity (yes/no questions)
- Address multiple issues separately
- If 2 or more uncertainties or lines of evidence for the same issue are interdependent, assess them jointly

Level 3 for **Quantitative** issues

Express uncertainty as range for assessment outcome

- Evaluate how resolving each uncertainty might change the assessment outcome, and express using relative scale (modify if necessary):



- Evaluate how resolving all the uncertainties might change the outcome; express using scale and as short narrative for conclusion

Level 3 for **Qualitative** issues

Express uncertainty as probability that condition is true (e.g. X is carcinogenic)

1. Evaluation based on each piece of evidence considered in isolation (it may help to start by expressing an initial judgement, before considering the evidence)

	– – –	– –	–	•	+	++	+++
Prob.	<1%	1-10%	10-33%	33-66%	66-90%	90-99%	>99%

2. Evaluation based on all lines of evidence considered together; express using scale & narrative statement

Communication of results

- Include a brief summary of the uncertainty evaluation in Committee conclusion
  - 1-2 sentences indicating the nature of the key uncertainties and summarising your overall evaluation.
- Briefly describe any additional uncertainties whose impact could not be evaluated.
- Include list or table of identified uncertainties (and Level 3 evaluation if done) in the discussion section or as an annex.

Example: Glucosamine & hepatotoxicity*

- Popular food supplement taken alone or in combination with chondroitin suplphate usually by sufferers of osteoarthritis
- Small number of case reports linking glucosamine & hepatotoxicity (1 fatality)
- COT asked to consider whether causal association was plausible

* COT statement 2009/1

Appendix – Workshop presentation - Review

Example: Glucosamine & hepatotoxicity

Potential uncertainties:

- Incomplete information on cases (e.g. doses)
- Some cases improved on cessation of treatment
- Non-specific effects, might be caused by other unidentified exposures
- May be too rare to detect in clinical trials
- No plausible mechanism
- …

Level 1 – default factors not applicable

Level 2 – *is it helpful to go to Level 3?*

Example: Glucosamine & hepatotoxicity

LINE OF EVIDENCE:	*Prob. %*	*Symbols*
Widely used, no clear effects	10 – 33	–
Small number of cases, mostly improve on cessation of supplement	33 – 90	• / +
Not specific, other causes possible	10 – 66	– / •
No evidence in human trials, but may be too rare	10 – 66	– / •
No evidence in animals, but limited data	10 – 66	– / •
Glucosamine occurs naturally in human body	10 – 66	– / •
No plausible mechanism for hepatotoxicity	33 – 66	•
OVERALL: Current evidence does not suggest glucosamine is likely to be a cause of hepatitis but causal link cannot be completely excluded	1 – 66	– – / •

Example: Glucosamine & hepatotoxicity

- COT para. 41: 'Current evidence does not suggest that glucosamine is likely to be a cause of hepatitis although a causal link cannot be completely excluded. It should be noted, however, that the likelihood of an individual user experiencing adverse effects is, at most, very low'
  - *Note: uncertainty of frequency could also be evaluated*
- Para. 42: uncertainty unlikely to be resolved by further research – would need extremely large study due to rarity of effect and many potential confounding factors

A simple, structured approach to assessing uncertainties that are not part of a quantitative assessment.

Andy Hart and John Paul Gosling[†]

The Food and Environment Research Agency

Peter Craig

Durham University

June 17, 2010

Abstract

We introduce a tabular approach to acknowledging and accounting for uncertainties that have not been included in a quantitative risk assessment. Uncertainty tables are communication tools that allow risk assessors to highlight the uncertainties that have not been accounted for in a probabilistic risk assessment. The proposed structure is not a replacement for more rigorous methods that account for uncertainty: it is a tool to lay bare the assumptions in a risk assessment and the potential impact that removing those assumptions would have. In this paper, we provide guidance on judging and combining the effects of uncertainties using uncertainty tables.
Keywords: qualitative assessment, subjective judgement, uncertainty characterisation, uncertainty table.

1 Introduction

Recent years have seen a range of crises and controversies concerning food safety, animal health and environmental risks: for example, BSE, foot and mouth, dioxins in seafood, and GM crops. These have led to an increased recognition of the

[†]Address for correspondence: The Food and Environment Research Agency, Sand Hutton, York, YO41 1LZ, UK.
Email: johnpaul.gosling@fera.gsi.gov.uk

need for improvement in several areas of the process of risk analysis. One fundamental need is to improve the handling of uncertainty in risk assessment, so that decision makers and the public are better informed on the limitations of scientific advice. The Codex Alimentarius Commission, which is the international forum for food safety issues, has published a set of working principles for risk analysis that includes the following (see Codex, 2003):

> "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. Expression of uncertainty or variability in risk estimates may be qualitative or quantitative, but should be quantified to the extent that is scientifically achievable."

In practice, very few risk assessments approach this ideal due partly to institutional inertia and partly to a lack of practical methodology.

A guidance document published by the European Food Safety Authority (EFSA) attempts to provide such a methodology (EFSA, 2006). The document is focused on uncertainties in assessments of human exposure to chemical and biological contaminants in food, but, in principle, the approach is equally applicable to any type of quantitative risk assessment. It uses a stepwise approach that starts by qualitatively evaluating all identified uncertainties affecting an assessment, and then proceeds to a quantitative evaluation of selected uncertainties if this is possible. For the qualitative evaluation, the document suggests using the symbols '+', '++', '+ + +', '−', '−−' and '− − −' to represent a subjective assessment of the direction and magnitude of the influence of each uncertainty on the outcome of the exposure assessment. Similar symbols are then assigned to represent a subjective evaluation of the combined influence of all of the identified uncertainties. This approach was proposed as a practical heuristic response to the need to characterise uncertainty, and no theoretical basis was claimed for

it. Nevertheless, it is beginning to be used in some of the assessments published by EFSA's scientific panels (see EFSA, 2007). This suggests that the scientists involved find it an effective means of expressing uncertainty. In Sorvari (2007), a similar scoring approach has been used that lists the environmental and human health risks from use of metal residues in soil at Finnish shooting ranges.

These attempts to be transparent about uncertainties provide some basis for optimism that, if effective procedures for qualitative evaluation of uncertainty can be devised, scientists engaged in regulatory risk assessment will take the methods up. However, this will only improve the management of risk if the procedures evaluate and combine uncertainties in a rational way. In the present paper, we aim to provide a more formal theoretical basis for procedures of this type without detracting from their practicality.

There are several methods being employed that attempt to characterise the uncertainty (and the inadequacy of models and data) in risk assessments. Evidence maps, as introduced in Schütz et al. (2008), graphically and qualitatively outline the role of evidence in risk characterisation. Pedigree analysis, as described in Van der Sluijs et al. (2005), is intended for characterising "deep" uncertainties that cannot be easily quantified, including qualitative issues such as problem framing, choice of methods, level of expert knowledge or consensus, and value-ladenness. The WHO/IPCS (2008) document on uncertainty and data quality in exposure assessment includes a qualitative approach similar to pedigree analysis, but prescribes eleven issues to be evaluated and scales for evaluating them. The issues are grouped into three broad classes: level of uncertainty, appraisal of the knowledge base and subjectivity of choices.

The United States Nuclear Regulatory Commission guidance on treatment of uncertainties associated with probabilistic risk assessments (see EPRI, 2006; Drouin et al., 2009) gives an overview of methods in that area. These two documents concentrate mainly on quantitative approaches, but include qualitative screening steps that are used to identify uncertainties requiring quantitative mod-

elling. Drouin et al. (2009) state that the final output should include a qualitative statement of confidence in the conclusion of the assessment and how it has been reached, supported by identification of key uncertainties that were addressed. EPRI (2006) includes a tabular format for listing sources of model uncertainty and narrative evaluations of their potential impact on the risk model.

There are a great number of mathematical and statistical tools for characterising and accounting for different types of uncertainty that have rigorous theoretical foundations (a recent overview from an environmental standpoint can be found in Beven, 2008). The difficulty is that most mathematical and statistical tools tend to be seen as complicated to use and to require the provision of a lot of detail by the analyst. Therefore, we wish to find out if we can bring mathematics or statistics to bear on the problem in ways other than the standard tools for formal quantification of risk, and, in particular, if we can do so without creating too much of a burden for potential users of new tools.

In the present paper, we introduce a methodology that a user may find simple to employ that is grounded in well-established statistical theory. We have built the process on a principle of transparency with a focus on the effect of the uncertainties on an estimate of interest. For example, our method could be used to evaluate the effects of unquantified uncertainties on an estimate of exposure to some chemical or on an estimate of toxicity. In Section 2, we introduce our uncertainty table method and discuss how it fits into the risk assessment process. In Section 3, we show how the method is underpinned by probability theory and how we can combine simple judgements made by the users of the method. A demonstration of the method concerned with contaminated land is given in Section 4.

2 A structured and transparent approach

An uncertainty table is a tool that aids transparency about the uncertainties that have not been accounted for in a risk assessment. The described structure is not a replacement for more rigorous methods for uncertainty modelling: it is a tool to help lay bare the assumptions in the process and to assess the impact that removing the assumptions would have. Throughout this report, we refer to a female expert for ease of exposition. In practice, a group of experts might be coming to a consensus over the necessary judgements.

The types of uncertainty that we might need to characterise using uncertainty tables include:

- measurement uncertainty including accuracy, precision and detection/reporting limits,
- sampling uncertainty (variability and bias),
- other study quality and design issues including ambiguity and inadequate reporting,
- inconsistency of results across multiple studies,
- extrapolation from animals to humans, between age and sex classes,
- variability between individuals in the population under assessment,
- relevance of the data to assessment scenario, and the use of surrogate data,
- uncertainty of expert judgements, including differences between experts,
- applicability of default assumptions or uncertainty factors,
- uncertainty about which factors and mechanisms to include,
- uncertainty about the structure of conceptual or quantitative models,

- dependencies between different elements of an assessment or model.

This list is not exhaustive, and, for different application areas, other sources of uncertainty might need to be considered. To construct an uncertainty table, we begin by listing all of the uncertainties that were not quantified in the assessment along with a narrative to explain what effect resolving each uncertainty may have on the original estimate.

An uncertainty table consists of just two columns: the first is populated by the list of uncertainties and the second is used to capture the analyst's judgements about the potential effects of formally accounting for the uncertainty in a probabilistic risk assessment. We have seven basic components that an analyst can use to express her beliefs about the impact of the uncertainties on the quantitative end-point of their study. These are the elements of the following set:

$$S = \{---, --, -, \bullet, +, ++, +++\}. \tag{1}$$

We introduce an ordering of these symbols to get

$$--- < -- < - < \bullet < + < ++ < +++. \tag{2}$$

Here $-$ represents a judged effect of a negative move from the current estimate, and $--$ and $---$ represent stronger effects (and *vice versa* for $+$, $++$ and $+++$). The $\bullet$ symbol represents negligible change in the estimate.

Due to the uncertainty about the effects and the fact that the experts have not quantified the effect of the uncertainty during their quantitative study, we believe that it is unreasonable to attach point estimates of the variable to the symbols in S. Therefore, the seven symbols correspond to intervals of possible values rather than point estimates.

Within the uncertainty table framework, we allow the analyst to express her beliefs through ranges of elements. We use s_1/s_2 to denote a range of elements

in S between $s_1 \in S$ and $s_2 \in S$ where $s_1 < s_2$. For example, if she believes that any element between the lower endpoint of $-$ and the upper endpoint of $+++$ is possible, she would write $-/+++$. Given the ordering of elements in S, there are 21 possible ranges of the form s_1/s_2. We define the set of possible ranges as

$$U = \{s_1/s_2 : s_1, s_2 \in S, s_1 \leq s_2\}.$$

The expert can choose between the elements of U when completing the uncertainty table; hence, there are 28 different choices available. This construction does not allow for situations where she believes the element could be disjoint possibilities; for example, elements of U cannot be used to represent beliefs that the effect is either $-$ or $+$, but not $\bullet$.

To complete an uncertainty table, an overall assessment is made of the potential effect of the combined uncertainties. In this final step, an element of U is assigned to the total effect of all the uncertainties. In order to combine the judgements in a rigorous way, we use probability theory in a relatively simple manner. Along with the appropriate symbol, the analyst should provide an explanation of why this representation of total uncertainty is appropriate for her analysis.

Table 1 shows the layout of an uncertainty table. If there are many sources of uncertainty, it could be beneficial to further subdivide the table into separate uncertainties that affect different parts of the assessment; for example, in a risk assessment of exposure to a chemical, we could divide the table into two parts: one for the uncertainties in the exposure and the other for the uncertainties in hazard characterisation.

3 Combining judgements in the table

One of the challenges in completing an uncertainty table is combining the judgements about individual sources of uncertainty into an overall statement of un-

Table 1: The layout of an uncertainty table

Source of uncertainty	Direction & magnitude
First source of uncertainty: description of potential impact.	$-/++$
Second source of uncertainty: description of potential impact.	$-/\bullet$
Third source of uncertainty: description of potential impact.	$-/\bullet$
$\vdots$	$\vdots$
Evaluation of overall effect of identified uncertainties: Description of combined effect of preceding uncertainties.	$-/+$

certainty. We assume that a quantitative study of the variable of interest has been performed. This will have resulted in a point estimate, $\widehat{Y}$ and, possibly, a credible range or probability distribution that captures uncertainty about that estimate.

The first stage of producing an uncertainty table is to list the uncertainties that have not been considered in the quantitative study and briefly explain their potential impacts on the estimate of interest. Next, the analyst should set a scale for subsequent judgements to be made on. The choice of intervals for the elements of S could be informed by the results of the quantitative study. The intervals do not have be regularly spaced or be symmetric about the interval for $\bullet$. The interval for $\bullet$ will contain the estimate for Y. Recall that $\bullet$ represents negligible change in the estimate; therefore, a central portion of the distribution for Y derived in the probabilistic risk analysis could be used (a 50% credible interval for example). The interval for $\bullet$ could also be set as the range over which there would be negligible effect on the subsequent risk management decision.

An important decision here is the scale on which to carry out the uncertainty analysis. For instance, in the example considered later, the quantity of interest is a daily exposure in units of μg kg^{-1}BW day^{-1}, but the uncertainty analysis is actually carried out in terms of the base 10 logarithm of exposure. Underlying our method of combining uncertainties is a model in which uncertainty is expressed

through potential changes (to the assessment output) that are independent and additive.

Once the intervals are specified, the expert should judge which element of U is appropriate for each source of uncertainty in the table. For most uncertainty sources, she will be able to say if she thinks a central value or a value closer to the ends of the interval is more likely. To capture this information, she will be required to choose from a set of distributions that are defined over the corresponding interval. These distributions will cover six different distributional shapes and spreads, so there is no need for her to describe the appropriate probability distribution herself. Figure 1 shows the distributional choices over an arbitrary range. In our research into finding appropriate distributions to use at this stage of the process, we have also considered skew-normal and skew-t distributions. We found that skew-t distributions produce similar results to those of the stable distributions, but the skew-normal can produce narrower 95% intervals as the tails of that type of distribution are thinner. An important consideration is computational time if the method is to be used as part of expert workshops. Combining stable distributions is much less costly than combining skew-normal and skew-t distributions due to the additional level of numerical computation that is required to combine distributions from the latter distributional types.

The details of the distributions used to create the densities in Figure 1 are given in Table 2. We have used stable distributions for all but one of the choices (stable distributions are defined in Appendix A). The family of stable distributions are rich enough to give the different levels of skew and kurtosis we need. The uniform distribution captures the belief that every value in the interval is equally likely. In Table 2, the uniform distribution is defined on an interval that is just slightly longer than (-4,4). An alternative to this is to set 95% of the probability to be in (-4,4) and then uniformly apportion the remaining 5% probability across the rest of the overall interval. More advanced users might want to define their own distributions over the intervals. The framework presented here

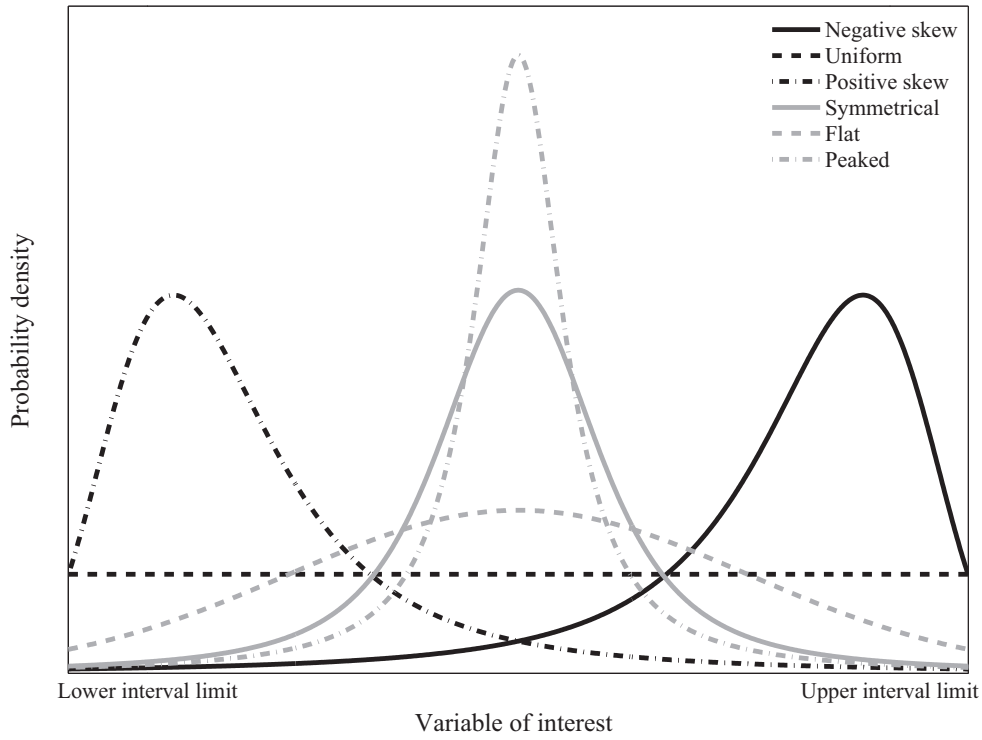


Figure 1: Six densities that can be allocated to intervals specified as part of the uncertainty table procedure.

is flexible enough to accommodate this.

If the analyst is finding it difficult to assign a distribution to the interval she has specified, she can experiment with various choices to investigate the robustness of the overall result from the uncertainty table. As the combination is computationally cheap, this can be done practically instantaneously.

The final task in completing an uncertainty table is to combine all the sources of uncertainty to get an appreciation of the total uncertainty about the variable of interest. Instead of asking the expert to combine all her judgements by selecting an appropriate element of U, we use simple probability theory. The combination of elements of U is based on the addition of probability densities with these pre-specified shapes (addition of random variables is addressed in Appendix B).

Table 2: Details of the six densities displayed in Figure 1.

Descriptor	Distribution
Symmetric	S(1.3, 0, 0.64, 0)
Flat	$N(0, 2.04^2)$
Peaked	S(1.1, 0, 0.41, 0)
Positive skew	S(1.3, 1, 0.61, -1.70)
Negative skew	S(1.3, -1, 0.61, 1.70)
Uniform	U(-4.21, 4.21)

The stable distributions have good properties for variable summation (again see Appendix A).

We need to combine distributions that characterise our beliefs about potential departures from some estimate for Y, $\widehat{Y}$, from our initial studies. For source of uncertainty i, the distribution encapsulates our beliefs about $\widehat{Y}_i$; that is, the altered estimate we would expect to see if we resolved the uncertainty and bias from the i^{th} source of uncertainty. Therefore, when we combine these distributions to capture our beliefs about the combined effect (denoted by $\widetilde{Y}$), we must subtract the initial estimate from each individual distribution:

$$\widetilde{Y} = \widehat{Y} + \sum_i \left(\widehat{Y}_i - \widehat{Y} \right).$$

For each source of uncertainty, the analyst imprecisely specifies a probability distribution for the change to the assessment output that would result from resolving that source of uncertainty. The assumption of independence means that the uncertainty statement made about the potential change due to one source of uncertainty would not be affected by observing the change resulting from resolving one of the other sources of uncertainty. In the example, the analyst is asserting that this holds true on the logarithmic scale, but not for the original units.

Assessment of correlation is an under researched area of expert elicitation, and the limited research that is available is inconsistent in its advice (as described in

Clemen & Reilly, 1999). As

$$Var\left(\sum_{i=1}^{n}\widehat{Y}_i\right)=\sum_{i=1}^{n}\sum_{j=1}^{n}Cov\left(\widehat{Y}_i,\widehat{Y}_j\right),$$

it is clear that any non-zero covariance when $i \neq j$ will have an impact on the spread of the aggregated uncertainty. To try to avoid this, we recommend that correlated sources of uncertainty be combined to form a single source in the uncertainty table. Ignoring the likely small correlations in the analyst's judgements about changes from resolving different sources of uncertainty is likely to be conservative when those correlations are negative and the opposite when those changes are positive. Our intention is that she avoids separating sources of uncertainty when the correlations are not negligible (perhaps less in absolute value than 0.1).

Another important consideration is how we back-transform from the resulting distribution to an element of U. As we assume that the variables are independent, we know that adding them together will cause the variance to grow. Therefore, it does not take many separate sources of uncertainty to have 2.5th and 97.5th percentiles that fall close to the extremes of the scale. This will result in $---/++$ being reported if we use the same 95% argument as we used for the initial distribution specification. If a 95% interval for the combined distribution is much wider than the original scale set by the analyst, then she could revisit the scale and provide a scale that is more appropriate given the judged uncertainty.

By fixing a qualitative scale as intervals and requiring at least 95% credibility for a specified range, the ranges are effectively conservative in nearly all cases as the analyst is not asked to fine-tune the intervals to achieve the appropriate level of credibility. This is one of the key ways in which we seek to make the task of expressing uncertainties manageable.

We anticipate that it may be possible in future work to develop our system to provide formal advice about the effects of correlations and the consequences of the qualitative scale and the small menu of distributional shapes by using

the machinery of imprecise probability. However, we see such developments as secondary to the importance of getting uncertainty tables adopted, and we see our procedure for combining uncertainties as a key part of that process.

4 Example: Exposure to cadmium

In this section, we describe an application of the uncertainty table methodology. As part of the determination of whether a piece of land should be classified as contaminated, a risk assessor must consider the exposure to chemicals of a user of the site. In a recent exposure assessment, a probabilistic assessment was conducted for exposure of a toddler (female aged between 1 and 4 years) to cadmium in a residential setting. The quantitative endpoint of this assessment was the average daily exposure of a randomly-selected toddler over the age range of 1 to 4 measured in μg kg^{-1}BW day^{-1}.

The probabilistic exposure assessment arrived at a best estimate of 0.021 μg kg^{-1}BW day^{-1} (with a standard deviation of 0.0051 μg kg^{-1}BW day^{-1}). The following sources of uncertainty were modelled probabilistically:

- the amount of cadmium at the property,
- the variability in a child's body weight, inhalation rates and soil ingestion rates,
- the amount of time spent at the property,
- properties of air circulation at the property.

To construct an uncertainty table, we begin by listing all of the identified uncertainties that were not considered in the probabilistic assessment; these are displayed in the left-hand column of Table 3. For each identified uncertainty, we provided a brief statement on our beliefs about the effects of accounting for the uncertainty.

Table 3: Uncertainty table identifying the effects of unquantified uncertainty in an exposure assessment

Source of uncertainty	Direction & magnitude
1. The mathematical model of exposure is not a perfect representation of reality. It is not clear in which direction the estimate will move when accounting for the model inadequacy; however, we do not believe this will cause a massive deviation from the original estimate.	$-/+$
2. Many of the parameters in the model were set at either average or conservative values. Due to the cumulative effect of all the conservative values and assumptions, we believe that there could be an overestimate of the true exposure. Also, by using point estimates, some of the uncertainty was ignored in the original assessment.	$--/\bullet$
3. The concentration of cadmium was assumed to be constant over the site and over the time period. There will be great variability on the chemical concentration over the soil at the site. Also, the level of contamination will fall over time as the original source has been remediated.	$-/\bullet$
4. Only exposure to cadmium in the soil is considered. There is potential for considerable exposure to come from sources such as passive smoking and through dietary exposure.	$\bullet/++$
5. The measurements of the building at the site were crudely used to quantify the building's volume and ground floor area. As there is little exposure to cadmium through inhalation, the volume of the building with have little impact on the exposure.	$\bullet$
Qualitative evaluation of overall effect of identified uncertainties: Overall, the estimate of exposure has been produced using overly conservative mechanisms. We feel that if we were to formally model the uncertainty around the parameters the mean estimate would be much lower. The strength of this effect and its balance against the other possible sources of cadmium result in uncertainty about the level of the reduction.	$---/++$

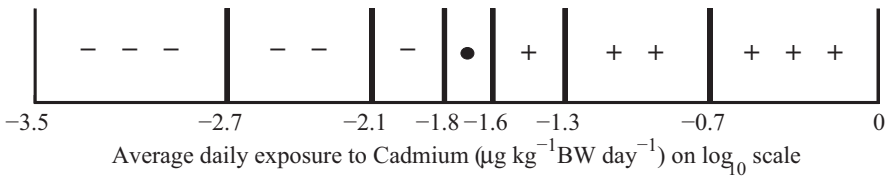


Figure 2: The intervals associated with the elements of S in this example.

The next part of the uncertainty table assessment involves setting up the intervals to map the elements of S onto. Figure 2 displays the intervals we used in this example, which cover changes of approximately two orders of magnitude in both negative and positive directions. Now that we have listed the uncertainties to be considered and we have set a scale to make judgements upon, we can start assigning elements of U to the rows of Table 3.

Our judgements are displayed in the right-hand column of Table 3 and the reasoning for each judgement is given in the associated text. We felt that the conservative assumptions we made during the probabilistic risk assessment could have the greatest impact on the estimated exposure. We also felt that, although we could not rule out a negligible positive effect, there was likely to be a strong negative effect on the result. We were able to capture this when we assigned distributions to each row of the table: in this case, we used the positively skewed option from Figure 1. For the first source of uncertainty listed in Table 3, we believed that the effect could be in either direction, but it was likely to have a negligible effect. Therefore, we selected a symmetric distribution over the corresponding interval. In Table 4, our chosen densities for each of the sources of uncertainty are displayed along with the result of combining those distributions.

The combined distribution, given in Table 4, is calculated using the formulae of Appendix B. Figure 3 plots the densities for the individual uncertainties along with this combined density. As expected, the combined distribution has a greater

Table 4: The chosen distributions for each source of uncertainty in Table 3

Source of uncertainty	Distribution	Description from Table 2
1	S(1.3,0.00,0.064,-1.70)	Symmetric
2	S(1.3,1.00,0.084,-2.38)	Positive skew
3	S(1.3,-1.00,0.038,-1.74)	Negative skew
4	S(1.3,1.00,0.084,-1.48)	Positive skew
5	S(1.3,0.00,0.016,-1.70)	Symmetric
Total	S(1.3,0.56,0.196,-2.29)	—

spread than any of the component distributions. Also, the mode of the combined distribution reflects our belief that the combined effect of accounting for all of the uncertainties would be to reduce the exposure estimate. The combined distribution has a 95% credible interval of (-3.2,-0.7); this allows us to back-transform to $---/++$ as the symbol that represents our assessment of the combined uncertainty.

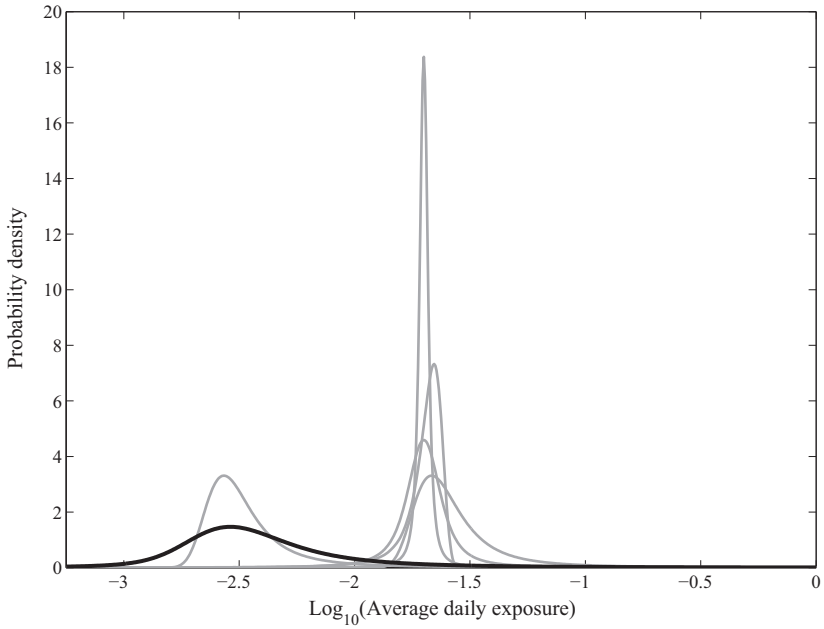


Figure 3: The densities for the individual sources of uncertainty of Table 4 plotted alongside the combined density (black line).

5 Discussion

The uncertainty table procedure offers a simple and transparent methodology to account for uncertainties. The methods for capturing beliefs and the combination of judgements about separate sources of uncertainty within the framework are simplistic. There are many aspects of the method that could be changed to suit the application; for instance, the method for combining the effects and the distributions that represent beliefs about the size of effect. The calculation we propose for the combination of uncertainties is essentially pragmatic and has some features that point towards it being conservative and others that do not clearly point towards or against conservatism. However, we feel they are flexible enough to be useful in many situations and will aid transparency in the risk assessment process. We believe this method should be less likely to give an inappropriate result than direct subjective combination by the experts; however, we would recommend that experts consider the appropriateness of the resulting symbols and adjust them if necessary.

In this paper, we have referred to a single expert. It is generally important to elicit beliefs from a group of experts, rather than a single expert, in order to synthesise the range of knowledge and opinions of the expert community. Kadane (1986) recommends that prior distributions used in medical applications are representative of the community of experts. This may lead to any number of expert's beliefs being combined to form a set of judgements on the intervals and on the corresponding distributions. In our experience, such a group of experts benefit from procedures where they all discuss the problem in hand and try to reach a consensus over the individual judgements.

The structure of the uncertainty table and the uncomplicated nature of the combination of sources of uncertainty mean that the methodology lends itself well to a PC-based application. We are currently developing a tool that leads analysts through the steps required to build an uncertainty table, calculates the

overall assessment of the effects of the uncertainty and outputs an uncertainty table in a format that could be used in reports and publications.

Appendix A: stable distributions

Five of the six distributional choices introduced in Section 3 are *stable* distributions. Stable distributions are used as they have good mathematical properties that makes the combination of distributions through convolution simple, and they cover many distributional shapes (that is, skewed and leptokurtotic distributions). In fact, the family of stable distributions has the normal, Cauchy and Lévy distributions as special cases. Although there is no analytical form for the density of distribution functions of a stable distribution (apart for the aforementioned special cases), the characteristic function is defined, and this is all we need when combining the distributions in the uncertainty table. The characteristic function of a random variable, X, is defined as

$$\phi_X(t) = \mathrm{E}[\exp(itX)],$$

where $i = \sqrt{-1}$ and $t \in \mathbb{R}$.

As with any family of distributions, there are a number of parameterisations available. Throughout this document, we use the following parameterisation of the characteristic function of X with stable distribution, $\mathrm{S}(\alpha, \beta, \gamma, \delta)$:

$$\phi(u) = \begin{cases} \exp\left(-\gamma^\alpha |u|^\alpha \left\{1 - i\beta \left[\tan(\pi\alpha/2)\right] \left[\mathrm{sgn}(u)\right]\right\} + i\delta u\right), & \alpha \neq 1, \\ \exp\left(-\gamma|u| \left\{1 + 2i\beta\pi^{-1} \left[\mathrm{sgn}(u)\right] \left(\log|u|\right)\right\} + i\delta u\right), & \alpha = 1, \end{cases}$$

where $\alpha \in (0, 2]$ is the index of stability, $\beta \in [-1, 1]$ is the skewness parameter, $\gamma > 0$ is the scale parameter, and $\delta \in \mathbb{R}$ is the location parameter. The normal distribution, $\mathrm{N}(0, 2.04^2)$, which is displayed in Figure 1, can be shown to have a

S(2,0,1.44,0) distribution.

All of the distributions listed in Section 3 are scaled to have their 2.5^{th} and 97.5^{th} percentiles at -4 and 4 respectively. In uncertainty table applications, the distributions will need to be defined over different intervals whilst keeping the 2.5^{th} and 97.5^{th} percentiles at the boundaries. Let the lower and upper bounds of the new interval be a and b respectively, and let $c = (a+b)/2$ and $m = (b-a)/8$. For X, we have that

$$mX + c \sim \begin{cases} \mathrm{S}(\alpha, \beta, m\gamma, m\delta + c), & \alpha \neq 1, \\ \mathrm{S}(1, \beta, m\gamma, m\delta + c - 2\beta\pi^{-1}\gamma m \log|m|), & \alpha = 1, \end{cases}$$

where the transformed variable will have its 2.5^{th} and 97.5^{th} percentiles at a and b respectively.

Another useful property of stable distributions in the context of uncertainty tables is that sums of stable distributions with the same value of α result in a stable distribution. Let $X_i \sim \mathrm{S}(\alpha, \beta_i, \gamma_i, \delta_i)$ for $i = 1 \ldots n$, then

$$\sum_{i=1}^{n} X_i \sim S(\alpha, \beta, \gamma, \delta),$$

where

$$\gamma^\alpha = \sum_{i=1}^{n} \gamma_i^\alpha, \quad \beta = \gamma^{-\alpha} \sum_{i=1}^{n} \beta_i \gamma_i^\alpha, \quad \text{and} \quad \delta = \sum_{i=1}^{n} \delta_i.$$

A comprehensive introduction to stable distributions and their properties is given in Nolan (2007).

Appendix B: summing variables

We can recover the density function of X, f_X, from its characteristic function using

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itx)\phi_X(t)dt,$$

provided that ϕ_X is integrable. In practice, we use numerical integration techniques to recover the density function.

Characteristic functions have a property that makes the addition of random variables easy to implement. Assume we have two independent random variables, X and Y, then

$$\phi_{X+Y}(u) = \phi_X(u)\phi_Y(u),$$

due to the linearity of expectation.

In the uncertainty table setting, we need to combine distributions that characterise our beliefs about potential departures from some estimate for Y, $\widehat{Y}$, from our initial studies. For source of uncertainty i, the distribution encapsulates our beliefs about $\widehat{Y}_i$ that is the altered estimate we would expect to see if we accounted for the uncertainty and bias from the i^{th} source of uncertainty. Therefore, when we combine these distributions to capture our beliefs about the combined effect (denoted by $\widetilde{Y}$), we must subtract the initial estimate from each individual distribution:

$$\widetilde{Y} = \widehat{Y} + \sum_i \left(\widehat{Y}_i - \widehat{Y}\right).$$

References

BEVEN, K. (2008). *Environmental Modelling: An Uncertain Future?* Routledge.

Clemen, R. T. & Reilly, T. (1999). Correlations and copulas for decision and risk analysis. *Management Science* **45**, 208–24.

Codex (2003). Working principles for risk analysis for application in the framework of the codex alimentarius. Tech. rep., Codex Alimentarius Commission: ftp://ftp.fao.org/docrep/fao/007/y5817e/y5817e00.pdf.

Drouin, M., Parry, G., Lehner, J., Martinez-Guridi, G., LaChance, J. & Wheeler, T. (2009). Guidance on the treatment of uncertainties associated with PRAs in risk-informed decision making. Tech. rep., US Nuclear Regulatory Commission: http://www.nrc.gov/reading-rm/doc-collections/nuregs/staff/sr1855/v1/sr1855v1.pdf.

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 the risks associated with an increase of the MRL for dieldrin on courgettes. *The EFSA Journal* **554**, 1–48.

EPRI (2006). Guideline for the treatment of uncertainty in risk-informed applications: Applications guide. Tech. rep., Electric Power Research Institute: http://my.epri.com/portal/server.pt?Abstract_id=000000000001013491.

Kadane, J. (1986). Progress toward a more ethical method for clinical trials. *The Journal of Medicine and Philosophy* **11**, 385–404.

Nolan, J. (2007). *Stable Distributions: Models for Heavy-Tailed Data*. Birkhauser.

Schütz, H., Wiedemann, P. & Spangenberg, A. (2008). *The Role of Evidence in Risk Characterization*, chap. Evidence maps — a tool for summarizing and communicating evidence in risk assessment. Wiley, pp. 151–60.

Sorvari, J. (2007). Environmental risks at finnish shooting ranges a case study. *Human and Ecological Risk Assessment* **13**, 1111–46.

Van der Sluijs, J., Craye, M., Funtowicz, S., Kloprogge, P., Ravetz, J. & Risbey, J. (2005). Combining quantitative and qualitative measures of uncertainty in model-based environmental assessment: the NUSAP system. *Risk Analysis* **25**, 481–92.

WHO/IPCS (2008). *Uncertainty and Data Quality in Exposure Assessment*. World Health Organization.