COMMITTEE ON TOXICITY OF CHEMICALS IN FOOD, CONSUMER PRODUCTS AND THE ENVIRONMENT

FSA funded project on evaluation and expression of uncertainty in risk assessment

1. As part of the FSA-funded research contract T01056, in February 3rd 2010 the COT held a Workshop at which COT Members and invited guests participated in discussions exploring the evaluation and expression of uncertainties in risk assessment. Participants looked at examples of risk assessments previously considered by the Committee and used a draft framework to consider whether this could make the steps of the risk assessment process easier, and the risk assessment process more transparent. The potential usefulness of the framework for COT work was also considered.

2. A draft final report from this project is contained at Annex A.

3. Annex B, contains a revised draft of the framework for evaluating uncertainty, which Members provided feedback on at the April meeting.

4. Members are asked:

    i. To comment on the content and quality of the project report.

    ii. To consider how the Committee may wish to trial this adapted framework for COT work.

    iii. To suggest additional changes to tailor the revised draft framework for COT use.

Secretariat
June 2010

COMMITTEE ON TOXICITY OF CHEMICALS IN FOOD, CONSUMER PRODUCTS AND THE ENVIRONMENT

FSA funded project on evaluation and expression of uncertainty in risk assessment

Report of the FSA funded research project on evaluation and expression of uncertainty in risk assessment.

Secretariat
June 2010

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

Summary

In its report on variability and uncertainty, the Committee on Toxicology (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 quantitative question (e.g. determine a reference dose or estimate exposure) or a qualitative question (e.g. is this chemical a mutagen?). 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 relevant authorities at national, European (SANCO, EFSA 2006, 2009, ECHA 2008) and international levels (WHO/IPCS 2008, Codex 2009), 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'.

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,
- 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 are 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 qualitative questions (e.g. is this chemical a carcinogen?).

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.

Wiedemann & Schutz (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.

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 WHO/IPCS (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 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. Methods are needed for evaluating evidence and uncertainty in two types of assessment: those addressing quantitative questions (e.g. estimating exposure) and those addressing qualitative questions (e.g. carcinogenicity). 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 qualitative 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 quantitative questions, such estimating exposure or determining a threshold dose, and those addressing qualitative ('yes/no') questions, such as whether a chemical is carcinogenic[2]. Some of the existing approaches were designed specifically for quantitative questions (e.g. EFSA 2006 was designed for exposure assessment), and some specifically for qualitative 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 qualitative 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.

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

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

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 qualitative questions, as explained in the previous section, so different approaches will be needed to develop a mathematical basis for them.

For quantitative questions, a satisfactory 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). Briefly, the symbols are considered to represent how the answer produced by the assessment (assessment endpoint) might change if each uncertainty were resolved. The user is asked to define the ranges represented by one, two or three pluses or minuses on a suitable numerical scale. A mathematical framework can then be defined for 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 illustrated graphically in Figure 1, and explained in more detail in the Annex 3. This work demonstrates that the approach for quantitative questions proposed in this report can be expressed mathematically in a theoretically coherent way. It is not necessary when using the plus/minus symbols to understand or use any aspect of the mathematics: 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 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).

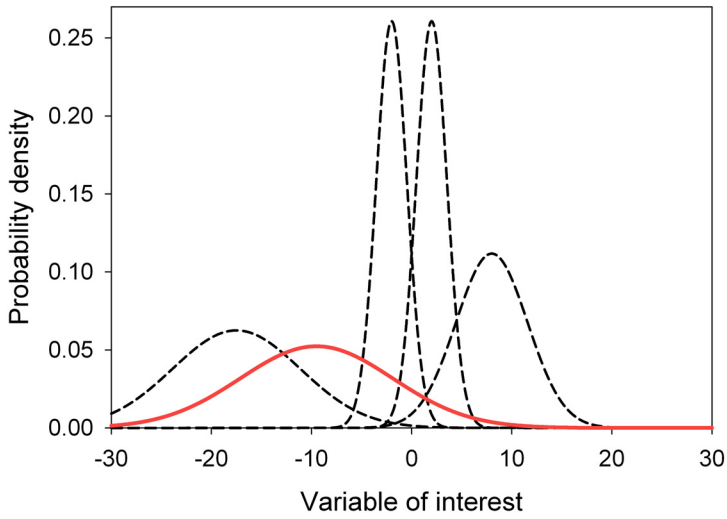


Figure 1. Graphical illustration of a mathematical basis for the qualitative procedure proposed for expression and combination of uncertainties affecting quantitative assessments. The impact of each uncertainty on the assessment endpoint is expressed qualitatively using plus and minus symbols (see Table 1, later). To represent this mathematically, the symbols are converted to probability distributions (dashed curves). Each distribution represents the degree by which the assessment endpoint might shift if the corresponding uncertainty was resolved. The distributions can then be combined mathematically, to estimate the degree to which the assessment endpoint would shift if all the uncertainties were resolved (red curve). For details see Annex 3.

Developing a mathematical framework for qualitative 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 qualitative 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, making it difficult to develop a framework which is both generally applicable and also enables explicit expression of logical structure. Thirdly, it became clear that although, in theory, a point estimate for probability can express all one's uncertainty about a question, many people are reluctant to give a point estimate because it seems to imply excessive precision. There are mathematical approaches that can accommodate this (e.g. imprecise probability theory), which could be considered.

It was concluded that it would not be feasible to develop a mathematical underpinning for qualitative questions within the scope of this project, but that a practical heuristic approach could be developed pending further research. This 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. If using verbal expressions it will be desirable to standardise and define them, to aid consistent interpretation both within advisory committees and by those they communicate with. In addition, it is likely that users would benefit from some simple examples to help in thinking about how to express uncertainty using probability (e.g. the chance of rain tomorrow) and how probabilities combine (including the potential for interactions).

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 qualitative 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 represents the draft recommendations of the project team. These will be revised to take account of any feedback received during finalisation of the project report. Subsequently, 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?

Uncertainty evaluation is not required for assessments conducted to an established procedure with standard, unequivocal studies, data, assumptions or assessment factors, which have been determined by appropriate procedures to provide adequate allowance for uncertainty (EFSA 2006).

In many 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.

Whenever an assessment deviates from or goes beyond standard procedure, e.g. by use of non-standard data or scenarios, or modified assumptions or assessment factors, the general provision for uncertainty may no longer apply and therefore a specific evaluation of uncertainty becomes necessary. Case by case evaluation of uncertainty is also required in new areas of exposure or risk assessment, where there is no established procedure to allow for uncertainty.

For assessments that require a specific evaluation of uncertainty, proceed as follows.

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:

- **Quantitative questions** with numerical answers, e.g. threshold dose or estimate of exposure.
- **Qualitative questions** with yes/no answers e.g. assessment of carcinogenicity.

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 qualitative (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 qualitative.

Some questions actually comprise several subquestions, for example, an estimate of risk is based on separate estimates of hazard and exposure. 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 to evaluate uncertainty separately for each subquestion, before considering 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 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 them (or refer to a checklist). 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 standard 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 1). It may be helpful to group the uncertainties according to which *component* (e.g. study, model input etc.) they affect, as illustrated in Table 1.

   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.

Table 1. Tabular format recommended for evaluating uncertainties affecting assessments of quantitative questions. See text for details.

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

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. 95%[3]) sure the adjusted endpoint would lie. Record your evaluations for all the identified uncertainties in a table (see example).

   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

[3] 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.

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. 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 using numbers or symbols, in the same way as the individual uncertainties (e.g. 0.5× – 2× or –/++, see step 4 above). 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.

If symbols or words are used to evaluate the uncertainties, it is recommended to define their meaning using a quantitative scale. 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 (e.g. within a committee).

When defining a quantitative scale, adjust it to the magnitude of the largest uncertainties. 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.

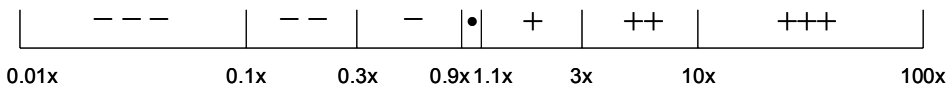


Figure 2. Example of a quantitative scale to define the meaning of symbols used to express uncertainties affecting assessment of quantitative questions (see Table 1).

If users feel comfortable expressing the effect of uncertainties on a quantitative scale, it may be preferable to use the numbers directly in the table, and dispense with symbols. On the other hand, users who are uncomfortable implying any quantitative judgment may prefer to use only words, or symbols with verbal definitions, and accept that this provides less information for decision-making.

Evaluation of uncertainties for qualitative (yes/no) questions

This section describes the approach recommended for evaluating uncertainty in assessments of qualitative questions, e.g. is this chemical a carcinogen? Again, we refer to the final answer to the question as the 'endpoint' of the assessment.

The answer to a qualitative 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 qualitative 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 only briefly applied to a single case study (see later), so further feedback and comment will be very welcome and may be used to adjust the approach in the final report.

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 3) together with their main strengths and weaknesses (uncertainties).

   Lines of evidence might include individual experimental or observational studies, or groups of studies that can conveniently be considered together in the uncertainty evaluation. 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 the outcome towards 'yes' and down arrows for those pushing towards 'no':

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

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

   ↑ or ↓: minor contribution towards yes or no

- : negligible influence on outcome in either direction.

Table 3. Tabular format proposed for evaluating uncertainties in assessments of qualitative questions. See text for details.

Question: *verbal description*	Influence on outcome
Study/line of evidence 1 *text description*	↑↑
• S *text describing strength 1*	
• W *text describing weakness 1*	
Study/line of evidence 2 *text description*	↓
Add more rows as needed	
Overall outcome *verbal description of likelihood of 'true' answer being yes (or no)*	**Optional expression of likelihood as a probability, range of probabilities, or standard phrase**

3. Make a judgment about the overall answer to the question, taking into account all the studies or lines of evidence. Express your uncertainty about the answer as a probability, i.e. your degree of belief that the qualitative 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 4).

Table 4. Table of standard terms established by the Intergovernmental Panel on Climate Change for expressing different degrees of likelihood (IPCC 2005), which could be used or adapted for expressing uncertainty in the assessment of qualitative questions.

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)

4. Drawing a line may help in thinking about probability, as illustrated in Figure 3.

Figure 3. Example of a probability line that may be helpful when evaluating the probability that the 'true' answer to a qualitative question is yes.

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 4), 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 qualitative 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 makes it more difficult for decision-makers to understand the assessors' level of certainty, which makes 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 4) should in principle help to avoid implying excessive precision, and aid consistency of interpretation between different assessors (e.g. within a committee). Given the potential advantages and disadvantages of the different options, it is recommended that they should be evaluated further in practice before making a decision.

Assessments comprising of multiple subquestions

As explained earlier in this section (under 'identification of assessment question'), some 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 qualitative questions as appropriate, the outcome of those evaluations should be summarised in a table (e.g. Table 5). These should then be used to make a judgment about the uncertainty of the overall conclusion, taking 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 qualitative (see above).

Table 5. 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; plus a narrative description of any 'deep' uncertainties whose impact on the outcome could not be evaluated.

However, this information 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).

Examples (caffeine)

This section offers some examples of applying the proposed methodology. The examples are 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). This case study provides examples of both qualitative and quantitative questions.

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 qualitative, while questions 3 and 4 are quantitative.

Questions 1 and 2 were combined by the case study group into a single qualitative 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) has been used by the project team to develop an evaluation following the updated proposals for evaluating uncertainty in qualitative questions. This was done rather quickly, but serves to illustrate the approach.

Table 6. Example of evaluation of uncertainty using the approach proposed for qualitative questions, based on the COT (2008) Statement on the Reproductive Effects of Caffeine. For key to symbols see earlier section.

Lines of evidence and their main strengths (S) and weaknesses (W)	Influence on conclusion
Animal experiments S - The available experimental evidence indicates that any effects on FGR occurred only at maternally toxic doses. W - The numbers of animals per group in the studies in experimental animals were relatively small and hence limited the power of the studies to detect a significant effect. W - The relevance of developmental findings in experimental animals to humans is uncertain.	↓
New human study 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 – As in most such studies, possible residual confounding is always possible, for example caffeine intake may have been a surrogate for some other lifestyle factor W – Although exposure assessment was thorough, there were still potential errors as it relied on subject recall, particularly during the first trimester	↑↑
Previous human studies 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, hence relied upon recall and were subject to potential bias	↑
Experimental evidence for biological mechanism W – No plausible biological mechanism has been identified for an effect of caffeine on FGR, but this has been investigated only to a very limited extent. Absence of evidence is not evidence of absence.	↓
Overall question: Is caffeine a cause of fetal growth restriction in humans? Whilst there is good evidence for an association between caffeine intake and FGR, it is not clear whether the association is causal.	*It is likely that caffeine is a cause of FGR*

One observation on this example is that, while the descriptions of strengths and weaknesses help to explain the different levels of influence attributed to the four lines of evidence (the new human study being strongest, with two symbols), it might also be useful to contain more information on the basis for the direction of each line of evidence. For example, the text entries for the new human study do not actually state that the result showed a significant association between caffeine intake during pregnancy and FGR: instead, this is inferred from the direction of the arrows. A possible way of including this information would be to add a sentence under each line of evidence describing what it indicates about the question (e.g. 'Caffeine intake during pregnancy was associated with an increased risk of FGR').

The short description of the overall conclusion in Table 6 is 'It is likely that caffeine is a cause of FGR'. This appears somewhat stronger than the longer text, which ends 'it is not clear whether the association is causal'. The case study group at the project workshop gave an overall probability for this question as 0.65 (Annex 2). The 2008 COT Statement conclusion was: 'we conclude that caffeine intake during pregnancy is associated with an increased risk of FGR. It is still not possible to be confident that the association is causal rather than a consequence of residual confounding, but it would be prudent to assume causation'.

The case study group at the project workshop produced a single evaluation for a quantitative question, but 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?
- If the relationship is causal what is the maximum increase in incidence of FGR (above the residual) from intakes of 200 mg/day?

The table produced by the project workshop is shown in Table 7. The project partners have added only a narrative conclusion in the bottom row on the table.

Table 7. Example of evaluation of uncertainty using the approach proposed for quantitative questions, based on the COT (2008) Statement on the Reproductive Effects of Caffeine. For key to symbols see earlier section.

Question: *enter question addressed here*	Evaluation of uncertainty
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:	+/-
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: *Whilst there evidence for a dose-response relationship between intake of caffeine and FGR, there is some uncertainty about the nature of this relationship, particularly at lower levels of caffeine intake.*	**+/-**

This table would benefit from some expansion of the text for the individual uncertainties, to explain more readily to the reader the nature of the uncertainty and the direction and magnitude of its influence. In addition, the narrative for the overall assessment reads more like the evaluation for a qualitative question (is there a dose response relationship?), and does not explicitly address the quantitative questions identified by the project workshop group. The conclusion on these questions

in the 2008 COT Statement read as follows: 'The evidence that is now available does not make it possible to identify a threshold level of caffeine intake below which there is no elevation of risk, and it seems likely that risk is increased in association with intakes in the order of 200mg per day and perhaps even lower. However, if the relation is indeed causal, then the absolute increase in incidence of FGR from intakes less than 200mg per day is likely to be less than 2% of infants'. The contents of the table could be refined to express more clearly the basis for these conclusions and the conclusions they already convey regarding uncertainty (indicated by the words likely, perhaps, likely to be less). However, separate tables would be required for the two questions.

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 exclude uncertainty evaluation in basic assessments and to allow early exit from evaluation of quantitative questions when the uncertainties are negligible or conservative. The methodologies for both quantitative and qualitative questions both evaluate uncertainties in terms of their impact on the assessment endpoint, which is the relevant information for decision-making. And finally, a satisfactory mathematical basis has been developed for the method for quantitative questions, but further work is needed on this for qualitative 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.' 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 many find it difficult to express uncertainty in numerical terms, although we believe that this might become at least partly easier with more

experience and with the help of simple examples (such as are often used in training for expert elicitation). To address these twin concerns of difficulty and misinterpretation, the framework produced by the project includes a range of options for expressing uncertainties in both types of question, including using words alone, standard phrases with quantitative definitions, ranges of numbers, or point probabilities. We understand from earlier feedback that the COT envisages at least developing standardised phrases. Given the advantages for transparency, consistency of interpretation, and improved information for decision-makers (outlined in preceding sections), we would encourage the COT, when trialling the approach and consulting the Social Science Research Committee, to give some consideration to using standardised phrases with quantitative definitions, possibly similar to those used by the IPCC (2005) but adapted and refined for the purposes of the COT.

References

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.

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.

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

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

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

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

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 et al. 2008, Envir. Sci. & Policy, 11, 627-641

Wiedemann

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

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

COMMITTEE ON TOXICITY OF CHEMICALS IN FOOD, CONSUMER PRODUCTS AND THE ENVIRONMENT

FSA funded project on evaluation and expression of uncertainty in risk assessment

Revised draft framework for evaluating uncertainty

Secretariat
June 2010

REVISED DRAFT FRAMEWORK FOR EVALUATING UNCERTAINTY v. 11, 17/6/10

Uncertainty evaluation is not required for assessments conducted to an established procedure with standard, unequivocal studies, data, assumptions or assessment factors, that have been determined to provide adequate allowance for uncertainty. For all other assessments, proceed as follows.

Specification of assessment questions

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

- **Quantitative questions** with numerical answers, e.g. threshold dose or estimate of exposure.
- **Qualitative questions** with yes/no answers e.g. assessment of mutagenicity.

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.

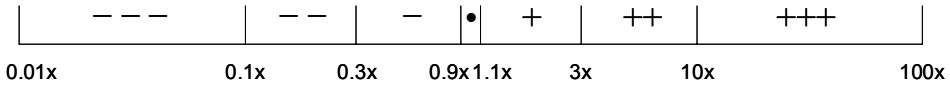
Evaluation of uncertainties for *quantitative questions* (e.g. a threshold dose or exposure estimate)

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[1].
3. If neither (1) or (2) apply, construct a table to evaluate the uncertainties (see example). 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*[2] 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. 95%) sure the adjusted endpoint would lie. Record your evaluations for all the identified uncertainties in a table (see example).

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 assessment endpoint*	**-/++** (or numeric range)

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. Consider carefully how they combine[3]. Express the outcome using numbers or symbols, and as a short narrative for use in the conclusions.

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:



[1] If it is desired to evaluate the degree of conservatism, proceed to the next step.
[2] 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.
[3] Including how they combine in calculating the endpoint, and any dependencies between them.

Evaluation of uncertainties for *qualitative (yes/no) questions* (e.g. is this chemical a mutagen?)

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

	Influence on outcome
Study/line of evidence 1	↑↑
• S *strength 1*	
• W *weakness 1*	
Study/line of evidence 2	↓
Add more rows as needed	
Overall outcome *verbal description*	**Likely (60-90%)**

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 the outcome towards 'yes' and down arrows for those pushing towards 'no':

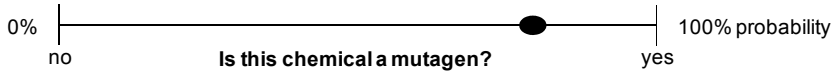
   ↑↑↑ or ↓↓↓: line of evidence could be sufficient on its own to be confident of yes or no
   ↑↑ or ↓↓: contributes significantly towards yes or no
   ↑ or ↓: minor contribution towards yes or no.

3. Make a judgment about the overall answer to the question, taking into account all the studies or lines of evidence. Express your uncertainty about the answer as a probability, i.e. your degree of belief that the qualitative 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 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)

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



Assessments comprising of multiple subquestions

1. Evaluate each subquestion separately, then summarise the evaluations in a table (see example).

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

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.

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 option is to obtain further data with the aim of reducing uncertainty. Both options may be targeted on the most important uncertainties, as identified by the preceding 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 'deep' 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.

Contact for questions or feedback on this draft: andy.hart@fera.gsi.gov.uk