Annex 1

Guidance of the Synthesis and Integration of Epidemiological and Toxicological Evidence Subgroup (SETE) of the Committee on Toxicity and the Committee on Carcinogenicity

Report of the Synthesis and Integration of Epidemiological and Toxicological Evidence Subgroup (SETE) of the Committee on Toxicity and the Committee on Carcinogenicity

Table of Contents

Guidance on evidence synthesis

Following the work of the Synthesising Epidemiological Evidence Subgroup (SEES) of the COT and COC, the subgroup identified the following over-arching guidance on the synthesis of epidemiological and toxicological evidence. It was recognised that issues on which advice from the committees is sought vary considerably and, hence, the guidance proposed should be sufficiently flexible to address this. For example, in some situations (e.g. risk from exposure to a relatively new product) studies in experimental animals may provide the most valuable, and perhaps even the only, information, whereas in other situations (e.g. long-term and significant exposure to an environmental contaminant), epidemiological studies may provide the most relevant information. For both epidemiological and toxicological information, a weight of evidence approach is proposed, the details differing, depending on the type of information available.

Problem formulation and literature retrieval

The first step in the process of evidence synthesis is scoping and problem formulation. This ensures that the right questions are asked, helps make the most efficient use of resources and identifies the best approaches to use in the assessment. Problem formulation is developed by the risk manager (e.g. FSA) in discussion with the committee. The following points should be considered.

- Has the issue been addressed previously by the committee?
- Why is a review of the evidence needed now?
- How urgent is the review?
- Which sub-populations are of potential concern?
- Is there any systemic exposure (determines the need for an assessment and if so, are data on systemic o local exposure of most concern)?
- Is a systematic review required?
    - If advice is needed urgently, a formal systematic review will not be possible, so what form will the review take, e.g. a focused literature search or use of a review by another authoritative body or from the published literature? If the issue is of major, long-term significance, a new or updated systematic review may be required.
    - Is qualitative (hazard) or quantitative (risk) advice needed? If the latter, a systematic review is most likely to be necessary, followed by meta-analysis if possible, to ensure all risk estimates are identified and included.
- Has the issue been addressed recently by another authoritative body (e.g. JECFA, EFSA, IARC)?
    - If yes, does this serve the needs of the committee, e.g. is it systematic and of satisfactory quality?
    - Does the review only need updating, or is a new review necessary?

- o Is the starting date for literature retrieval adequate, or could useful older literature be missing?
- o Was the characterisation of risk appropriate to the needs of the committee (e.g. were both acute and chronic risks addressed; were risks in the sub-population of concern assessed)?
- o Is there in an existing meta-analysis, and if so, does it need to be updated? Would this be possible, with the information provided on the existing analysis?
- As information is retrieved and evaluated, this may necessitate some change or refinement of the problem formulation or lead to additional questions being asked. Any such changes should be agreed with the risk manager (e.g. FSA) and clearly recorded.

Overarching principles

- An established system or guidance should be followed where appropriate (e.g. for a systematic review; quality assessment of toxicological studies).
- The evidence synthesis should include an expression of uncertainty to the extent possible.
- Potential conflicts of interest should be identified and considered, including for published papers and reviews.

Information retrieval

- What information is being sought (e.g. potential adverse health effects of substance X in the general population)?
- What are the constraints on the search for information, if any (e.g. within a specific time frame; for a specific geographic region)?
- How extensive will be the search for information (e.g. systematic review, focused review)?
- What are the potential sources of information (e.g. bibliographical databases, proprietary information from food producers)?
- What search strategy will be used for open literature, i.e. search terms?
- Will the grey literature be searched, and if so, how will this be done?
- How will other potential sources of information be searched, if necessary?

Epidemiological information

The Report of the Synthesising Epidemiological Evidence Subgroup (SEES) of the Committee on Toxicity and Committee on Carcinogenicity provides detailed information and guidance for the committees on the evaluation of epidemiological information. The current guidance summarises and updates the recommendations of the SEES report.

Focused literature search

As a minimum, this should include the details described under Information Retrieval, above: i.e. purpose of search; information sources searched (e.g. PubMed); period covered (e.g. < Jan 2010; > June 2019); search terms and their combinations.

The results of the search should be summarised, as follows:

- Numbers of papers identified, and numbers included in the review
- Reasons for exclusion of papers (e.g. not covering health effects of the substance of concern)
- Extraction of key information from relevant literature in narrative, graphical and/or tabular format. It can be particularly useful to determine what information is needed for the committee assessment (e.g. effects of substance X on developmental outcomes) and to tabulate relevant information from each paper on this (e.g. exposure metrics, outcomes, affected population). Guidance such as the Meta-analysis of Observational Studies in Epidemiology (MOOSE) and Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), as adapted in Appendix A1 of the SEES report, should be consulted for the types of information that would be value.

Evaluating an existing systematic review

Details are provided in the SEES report.

- As a minimum, an adapted checklist, such as from MOOSE or PRISMA (see SEES report) should be consulted to assess the completeness of the information available in the literature used in an existing systematic reviews and meta-analyses. This can help provide an indication of whether there are any major gaps in the information reviewed.
- The evidence synthesis method and any scoring system used by the authors should be described, and the potential implications on the conclusions noted, e.g. GRADE gives lower weight to evidence from observational studies, which the Committee may feel is appropriate in the given circumstances.

Conducting a new systematic review

The recommended approach for reviewing the open literature comprises four stages.

- Scoping: Criteria for the search strategy
    - Define the criteria for the search (see above)
    - Identify the information sources to be used
- Relevance
    - Define inclusion and exclusion criteria
    - Select studies relevant to the assessment
- Reliability: Quality of the studies
    - Assess the reliability of the studies

- o Compliance with appropriate guidelines (e.g. Good Epidemiological Practice (GEP))
  - o Assessment of uncertainty and potential bias
  - o Peer review
- Outcomes: Reporting
  - o Collect and interpret the evidence
  - o Evidence synthesis
  - o ARRIVE or GOLD publication checklist

An adapted checklist such as from the MOOSE guidelines (SEES Report Appendix A1) should be used to help summarize the papers. SEES recommends a number of additions to the published MOOSE checklist:

- o Include a flow chart for the identification of papers at the different stages of the systematic review
- o Assess the adequacy of study data presentation
- o Describe how data were extracted
- o Use forest plots to illustrate findings from the studies reviewed
- o Include patterns of association and confidence intervals where possible

Assessment of epidemiological evidence

All available studies, e.g. observational studies, meta-analyses, should be evaluated individually to identify potential sources of confounding, other possible biases, their direction and likely impact on estimated parameters; nature of exposure; outcomes; and conclusions. This should not necessarily lead to individual studies being excluded, since such a study may still be highly informative and it is recommended that all relevant studies should be included in evidence synthesis, using a weight of evidence approach (see below). A description of possible sources of confounding and other bias in epidemiological studies is provided in the SETE report. These include confounding, including effect modifiers, selection bias, effect modifiers and information bias.

- Assess risk of bias. The type, direction and magnitude of potential biases identified across all studies should be considered
  - o When most available studies suffer from the same type of bias, e.g. selection/recall bias in case control studies, the overall body of evidence should still be considered using the different study characteristics, e.g. smoking status, source population, to assess the potential impact of the bias.
  - o Where studies have different types of bias, the type and direction of biases must be assessed in parallel.
  - o Identify the most likely influential sources of bias, classifying each study on how effectively it has addressed each of these potential biases, and determine whether results differ across studies in relation to susceptibility to each potential source of bias

- Exposure assessment
    - Assessment method used: direct measurement (e.g. personal monitoring, biomarkers) or indirect methods (e.g. exposure modelling, food consumption pattern)
    - Exposure patterns over time: duration; frequency; continuous or intermittent; critical time windows
    - Relevance of the exposure metric to the exposure patterns: ever/never; duration of exposure; cumulative exposure; shorter-term intermittent exposure (e.g. maximum/average intensity)
    - Have all key sources of exposure, via all possible routes, via all relevant media, been included, to give an estimate of aggregate exposure?
    - Many exposures are part of mixtures and may therefore be highly correlated, making it difficult to evaluate the effects of individual substances. How much uncertainty does this introduce into the conclusions?
    - If possible, determine whether uncertainty in exposure assessment in a study is likely to under- or overestimate the exposure.
- Outcome assessment
    - Nature of adverse health effects, e.g. testicular cancer, decreased birth weight
    - Affected population, e.g. all exposed individuals, young children
    - Any possible differences in sensitivity, which cannot be accounted for by exposure, e.g. atopic individuals
    - Strength of the effect in terms of severity and number of individuals affected, as a fraction of the exposed population
    - Uncertainty associated with effect estimates
- Conclusions of the study or review
    - Risk metric, e.g. relative risk, odds ratio, incremental risk
    - Statistical significance of findings
    - Confidence intervals in risk estimates
    - Likelihood findings were by chance, e.g. confidence intervals, number and types of exposure-effect comparisons
    - Power of the study, e.g. minimum detectable effect given the size of the population studied

Triangulation. Even if individual studies have different uncertainties and biases, the totality of the evidence should be evaluated, to determine whether the combination of individual studies can overcome the different biases and provide suitable evidence.

Assessment of toxicological evidence

Following information retrieval as described above, toxicological data should be evaluated using a weight of evidence approach, analogous to the triangulation approach described above for the assessment of epidemiological data. This should

include assessment of uncertainty, both qualitative e.g. the toxicological significance of an effect observed, and quantitative, e.g. dose without observable effects. It is recommended that a framework for the systematic assessment of data and study quality should be used for this purpose. This should be sufficiently comprehensive that, together with expert judgement, it provides a robust evidence-based approach to risk assessment, whilst being easy to use.

In vivo studies

- Assess the quality of each study, using the criteria proposed by Klimisch et al (1997) for reliability, relevance and adequacy. Published modifications to the scheme proposed by Klimisch et al (1997) may be more appropriate for a given assessment (e.g. Schneider et al, 2009; Kaltenhäuser et al, 2017; Goodman et al, 2020).
    - Does the study comply with Good Laboratory Practice (GLP) or the principles of GLP?
    - Was the study conducted according to an accepted guideline (e.g. OECD, EPA)?
- For each study, consider the following
    - Was the test material clearly identified and defined?
    - Was the experimental system (e.g. test species, strain, husbandry) appropriate?
    - Was the study suitably designed (e.g. route of administration, dose selection)?
    - Was exposure suitably assessed?
    - Was the dose expressed appropriately, i.e. what was the dose metric?
    - Were the results reported adequately (e.g. sufficient detail)?
    - Were the statistical analyses of the results appropriate (e.g. expression of uncertainty (e.g. CIs) where necessary, power calculations, assumptions on data distribution)

There are a number of published schemes that provide details on how to check the quality of scientific studies, e.g. Nature journals' checklist for Life Sciences articles, ARRIVE guidelines.

In vitro studies

In vitro studies should be evaluated using similar principles to those above for *in vivo* studies. However, relatively few *in vitro* methods have been fully validated for use in regulatory toxicity testing (e.g. no OECD guideline). Hence, reliability needs to be assessed using other approaches.

- Was the test material clearly identified and defined?
- Has the method used been formally validated (e.g. EURL ECVAM)?
- Is there a guideline for the method from an authoritative body (e.g. OECD)?
- Was the study conducted according to the OECD (2018) Good *In Vitro* Method Practices (GIVIMP)?

- Is sufficient information provided to assess the relevance of the method?
  - Is the endpoint used being measured reliably (e.g. specificity, variability, metabolic capacity)?
  - Was exposure assessed suitably?
  - Is the endpoint measured biologically/toxicologically relevant (e.g. cell line, culture conditions, duration of exposure)?
  - Is it possible to extrapolate the findings *in vitro* to a mode of action or adverse outcome pathway (AOP) *in vivo* (e.g. known relationship to adversity *in vivo*)*?
    - Qualitatively (e.g. key event in an AOP, known relationship to adversity *in vivo*)?
    - Quantitatively (e.g. suitable PBPK extrapolation available)?
- Were the statistical analyses of the results appropriate (e.g. expression of uncertainty (e.g. CIs) where necessary, power calculations, assumptions on data distribution)?

Assessment of mode of action

Information on mode of action (MOA) can be invaluable for evidence integration by enabling the qualitative and quantitative bridging between experimental data and observations in humans. MOA underpins weight of evidence considerations by providing the mechanistic link between empirical observation and biological plausibility. The WHO IPCS has developed a well-established framework for assessing MOA and its implications of human health risk assessment (Boobis et al, 2006, 2008; Meek et al, 2014).

The key elements in assessing a MOA are as follows:

- Is there a substance related adverse effect (adverse outcome) in an experimental system? This requires considerations of study quality, consistency and weight of evidence as described above.
- Is there sufficient evidence in experimental studies to establish a MOA for this adverse effect? This requires assessment of weight of evidence using considerations modified from those proposed by Bradford Hill (1965).
- If so, is it possible that the MOA may occur in humans? This requires qualitative consideration of the biology underlying the key events. For example, does a key event depend on a biological process operating only in the experimental species, with no functional equivalent in humans?
- If it is considered possible that a MOA would be operative in humans, considering kinetic and dynamic differences, how probable is it that the MOA would be operative in humans? This requires a quantitative concordance analysis of the key events in the experimental animals and in humans (or human-derived systems, such as isolated cells).

- If it is not possible to dismiss human relevance of a MOA, how can qualitative and quantitative information on the key events be used to inform the risk assessment?

AOPs are in many ways conceptually analogous to modes of action. However, there is a greater focus on forward prediction from assays for key events, usually *in vitro*. Hence, AOPs provide an important link between non-animal methods and assessing possible adverse health effects in humans. The OECD has a major programme on AOPs and their website should be consulted for details. The use of AOPs in risk assessment is still at an early stage and hence their current application is largely case-by-case. However, the Committees are developing separate guidance on this (as of June 2021).

Evidence integration

All lines of evidence should be considered, with no specific hierarchy *a priori*. However, assessment of the strength of evidence from a particular approach, as described above, will provide an indication of how reliable a line of evidence is. For example, it may be that the epidemiological evidence for a given compound is considered extremely robust, whereas the evidence from *in vivo* toxicological studies is considered very weak. This should be reflected in how the respective lines of evidence are weighted. This is different from consideration of the nature of the evidence.

The guidance provided here has been developed from published approaches, such as the "Epid-Tox" process developed by Adami et al (2011). For each question some upper and lower estimate of uncertainty should be made.

- Epidemiological evidence
    - How strong is the evidence that exposure to the substance of concern causes an adverse health effect in humans?
    - Are the exposures at which effects are reasonably anticipated to occur in humans realistically achievable in the population(s) of concern?
    - Is the same adverse health effect observed in toxicological studies, recognising that some effects are not produced in toxicological studies?
    - Are there any modifying factors in sub-populations that increase or decrease susceptibility, consistent with the MOA (see below) (e.g. genetic polymorphisms in molecular targets for the AOP, differences in life-stage sensitivity)?
- Experimental evidence
    - How strong is the evidence that the substance of concern causes an adverse outcome on administration to experimental animals?
    - Is the adverse outcome observed relevant to humans (e.g. known species or strain specific sensitivity to a class of compounds)?

- o Is the same adverse outcome observed in exposed human populations?
- Mechanistic data/MOA
  - o Is there sufficient information to establish a MOA?
  - o Is there evidence that the key events (precursor events) observed experimentally occur is exposed humans?
  - o Is there evidence from other information (e.g. pathophysiology) that should a key event occur in humans it will lead to the adverse outcome?
- Exposure
  - o Is the exposure in experimental models (laboratory species, *in vitro*) at which adverse effects are observed achieved in the subjects of an epidemiological study? If not, it may be difficult to draw conclusions on causation, as no effects would be expected at this exposure level.
  - o Is the predicted (e.g. using PBPK modelling) or measured internal exposure at which adverse effects are observed in humans consistent with that at which adverse outcomes are observed in experimental animals?
  - o Is the predicted (e.g. using PBPK modelling) target site concentration at which adverse effects are observed in humans consistent with the predicted concentration at which adverse outcomes are observed in experimental animals?
  - o Is the predicted (e.g. using PBPK modelling) target site concentration at which adverse effects are observed in humans consistent with the predicted concentration at which adverse effects are observed *in vitro*?
  - o If the relative sensitivity of the molecular target in humans and experimental models (e.g. laboratory species, cell line *in vitro*) is known, is the dose/concentration-effect relationship in humans consistent with the experimental observations?

Combining the evidence

- Integration of the lines of evidence
  - o A graphical approach similar to that of Adami et al (2011) is recommended. However, the two axes should be "Epidemiological evidence" (x-axis) and "Experimental evidence" (y-axis).
  - o Start with a clear hypothesis relating exposure to the substance of concern to adverse health effect(s) in humans (e.g. caffeine during pregnancy causes low birthweight). This forms the initial estimate of causal inference and should be placed centrally in the grid.
- Assess the impact of each line of evidence on confidence in the initial estimate, using expert judgment to position the estimate along an axis.
- Positioning of the graph should reflect the Committee's agreed conclusion on the weight of evidence on the likelihood of causation. The axes should not be

considered numerical, but rather they reflect increasing or decreasing weights of evidence based on expert judgement.

- Where possible, include an estimate of uncertainty to provide a range (likely, upper and lower bound of impact)
- Epidemiological Evidence
    - o Consider how the answer to each question would affect confidence in the initial estimate and move the estimate accordingly leftwards or rightwards along the x-axis, as appropriate.
- Experimental Evidence
    - o Include all other lines of evidence under this heading
    - o Consider how the answer to each question would affect confidence in the initial estimate and move the estimate accordingly upwards or downwards along the y-axis, as appropriate.
- Conclusion on the evidence
    - o Based on where the estimate of causal inference appears on the graph, after taking account of all lines of evidence, one of several conclusions is possible:
        - A causal relationship in humans is likely
        - A causal relationship in humans is unlikely
        - A causal relationship in humans is possible, but lacks strong experimental support
        - A causal relationship in humans is possible, but lacks strong epidemiological support
        - There is insufficient information to reach a conclusion on the possibility of a causal relationship
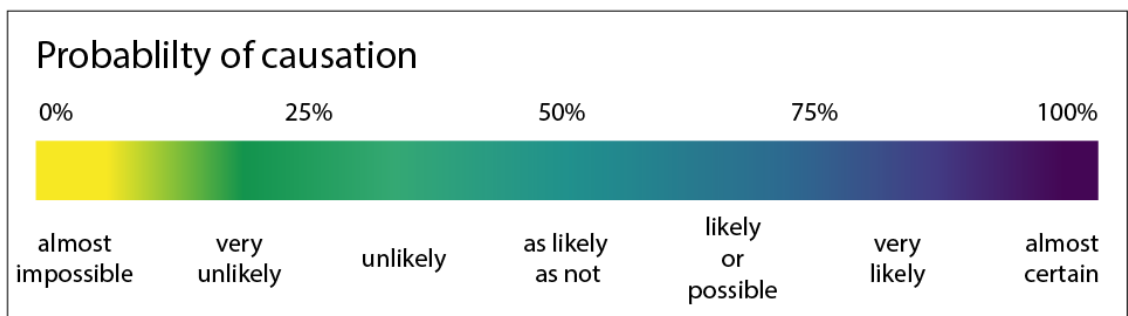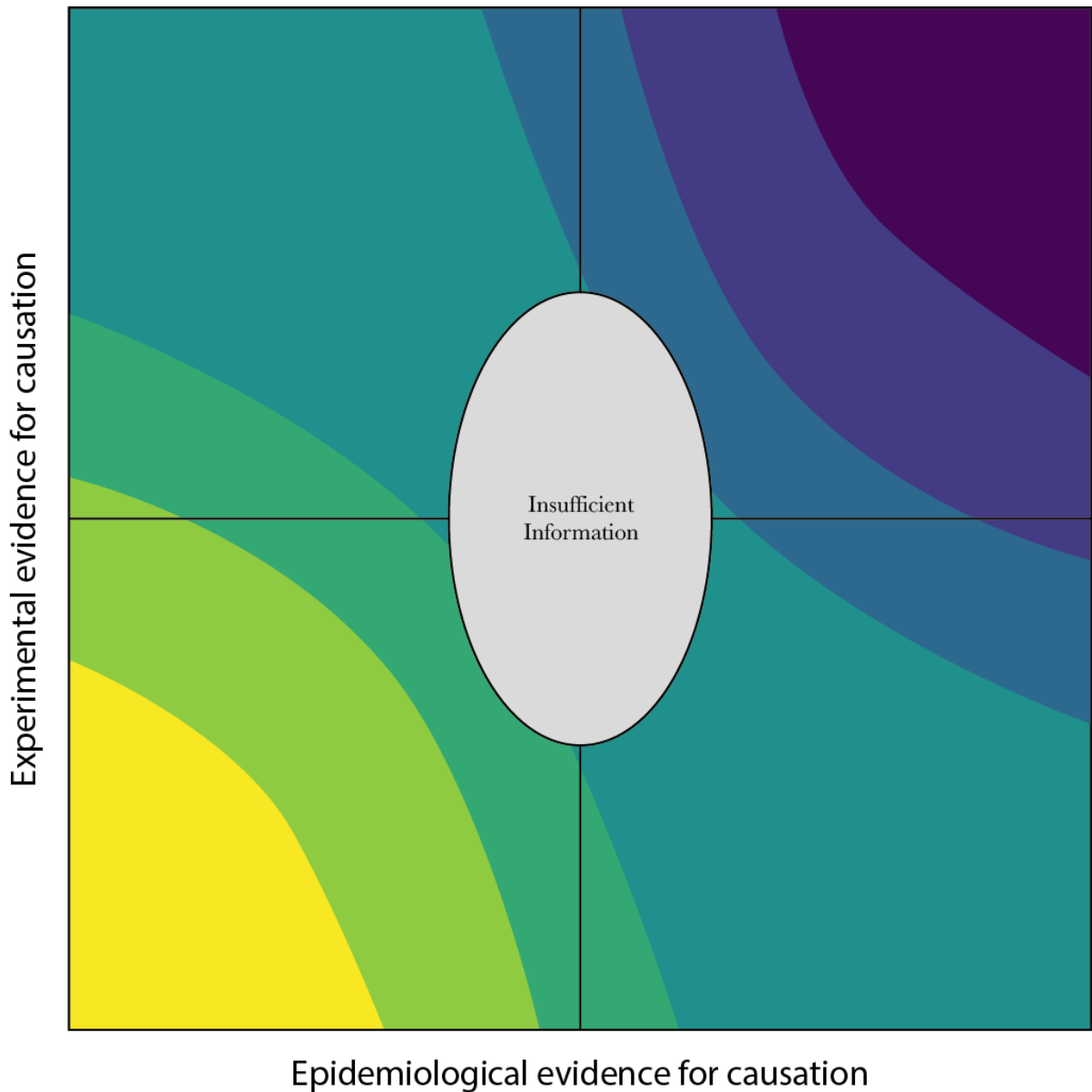
Visual representation of the likelihood of a causal relationship, considering different lines of evidence and guide to interpretation of the conclusion.

Reporting

- The problem being assessed should be clearly stated, together with why it is being reviewed by the committee

- Each step of the procedure should be clearly described
- Information sources should be documented, including the databases searched, details of the search terms used, criteria for selection of papers and the papers identified
- All lines of evidence should be described, together with their identified uncertainties.
- A clear conclusion on how each line of evidence affects the estimate of causal inference should be provided, together with the associated uncertainty. If the Committee cannot agree on a single 'probability' for a line of evidence, the range of suggested 'probabilities' should be reported, e.g. "as likely as not" to "very likely".
- Tabulation of this information may be of value
- A graphical presentation of evidence integration should be provided
- The conclusion of the assessment should be stated, with an estimate of the overall uncertainty and, where appropriate, guidance on how data gaps could be filled