Scientific Committee on Health, Environmental and Emerging Risks

SCHEER

Memorandum on Weight of Evidence approach for Risk Assessment

Revision 2024

The SCHEER adopted this Opinion during plenary meeting on 22 October 2024

About the Scientific Committees

Two independent non-food Scientific Committees provide the Commission with the scientific advice it needs when preparing policy and proposals relating to consumer safety, public health and the environment. The Committees also draw the Commission's attention to new or emerging problems that may pose an actual or potential threat.

These committees are the Scientific Committee on Consumer Safety (SCCS) and the Scientific Committee on Health, Environmental and Emerging Risks (SCHEER). The Scientific Committees review and evaluate relevant scientific data and assess potential risks. Each Committee has top independent scientists from all over the world who are committed to working in the public interest.

In addition, the Commission relies upon the work of other Union bodies, such as the European Food Safety Authority (EFSA), the European Medicines Agency (EMA), the European Centre for Disease prevention and Control (ECDC) and the European Chemicals Agency (ECHA).

SCHEER

This Committee, on request of Commission services, provides Opinions on questions concerning health, environmental and emerging risks. The Committee addresses questions on:

- health and ecological risks related to pollutants in the environmental media and other biological and physical factors in relation to air quality, water, waste and soils.

- complex or multidisciplinary issues requiring a comprehensive assessment of risks to consumer safety or public health, for example antimicrobial resistance, nanotechnologies, medical devices and physical hazards such as noise and electromagnetic fields.

Scientific Committee members

Thomas Backhaus, Teresa Borges, Wim de Jong, Pim de Voogt, Raquel Duarte-Davidson, Peter Hoet, Rodica Mariana Ion, Renate Krätke, Demosthenes Panagiotakos, Ana Proykova, Marian Scott, Emanuela Testai, Marco Vighi, Sergej Zacharov

Contact

European Commission
DG Health and Food Safety
Directorate B: Public Health, Cancer and Health security
Unit B3: Health monitoring and cooperation, Health networks
L-2920 Luxembourg
SANTE-SCHEER@ec.europa.eu

ISSN ISBN
doi ND

The Opinions of the Scientific Committees present the views of the independent scientists who are members of the committees. They do not necessarily reflect the views of the European Commission. The Opinions are published by the European Commission in their original language only.

https://health.ec.europa.eu/scientific-committees/scientific-committee-health-environmental-and-emerging-risks-scheer/scheer-opinions_en

TABLE OF CONTENT

ACKNOWLEDGMENTS

Members of the Working Group are acknowledged for their valuable contribution to this Opinion. The members of the Working Group are:

The SCHEER members
Thomas Backhaus
Teresa Borges
Renate Krätke
Demosthenes Panagiotakos (Chair of the WG)
Ana Proykova
Theodoros Samaras
Marian Scott (Rapporteur)

The SCCS member
Corrado Lodovico Galli

All Declarations of Working Group members and supporting experts are available at the following webpage:

Register of Commission expert groups and other similar entities (europa.eu)

ABSTRACT

This Memorandum is focussed on how to use the weight of evidence approach (WoE) to conduct a risk assessment for stressors to which humans and/or the environment may be exposed. It is intended to update the previous SCHEER memorandum (2018). The aim of this document is to support the use of the WoE, wherever appropriate, for the risk assessment activities of the Scientific Committee on Health, Environmental and Emerging Risks (SCHEER). In addition, it should support the consistency in the work of different EU bodies performing risk assessments.

Scientific evidence consists of observations, experimental and model results and expert judgements that serve to support, refute, or modify a scientific hypothesis or theory. The search for relevant information and data for the SCHEER comprises of identifying, collecting and selecting possible sources in order to perform a risk assessment and/or to answer the specific questions being asked. Depending on the issue being addressed, the SCHEER may utilise data provided by the DG tasking the SCHEER or provided by a third party (e.g. stakeholder reports, submissions such as confidential data provided by companies or applicants); reports and Opinions of other scientific, governmental or international bodies; scientific (peer-reviewed) publications; meta-analysis and systematic reviews or personal communications.

The WoE is an iterative process involving:

- Problem formulation of the risk assessment requested,

- Identification of sources, collection and selection of data and information as well as gaps in relation to the aim of the assessment,

- Screening of the data and information to identify those that are relevant to address the question(s) posed by the Commission Services,

- Assessment of the quality and consistency of the content and weighing of the data and information to establish individual lines of evidence,

- Integration of the relevant lines of evidence,

- Conclusions and reporting.

For each piece of information, their validity, reliability and relevance are considered, and the overall quality assessed. Several tools for the analysis and description of uncertainties are also presented. In the integration of the different sources of information or lines of evidence, the strength of the overall evidence depends on the consistency and the quality of the results. Quality and consistency have been defined and graded into categories A-C and I-III, respectively. The weighing of the total evidence is presented in a standardised format.

An overall WoE system is proposed that classifies the assessment in terms of:

- very strong evidence: quality A and consistency I of the lines of evidence
- strong evidence: quality A and consistency II of the lines of evidence
- strong to moderate evidence: quality A and consistency III of the lines of evidence
- moderate evidence: quality B and consistency I of the lines of evidence
- moderate to weak evidence: quality B and consistency II/III of the lines of evidence
- weak evidence: quality C and consistency I of the lines of evidence
- uncertain evidence: quality C and consistency II/III of the lines of evidence.

Keywords (for literature search): human health risk assessment, ecological risk assessment, scientific literature, risk analysis, uncertainty and variability analysis, weight-of-evidence/weighing of evidence, data integration, lines of evidence/line of evidence, levels of evidence/level of evidence, strength of evidence/strengths of evidence, quality of evidence, quality criteria, evidence integration/integration of evidence.

Memorandum to be cited as: SCHEER (Scientific Committee on Health, Environmental and Emerging Risks), Memorandum on Weight of Evidence approach for Risk Assessment - Revision 2024; Date of adoption 22 October 2024

1. BACKGROUND

According to the Commission Decision EU 2024/1514, the mission of the Scientific Committees is to provide the Commission services with scientific advice and risk assessments in the areas of public health, consumer safety and environmental risks, including, when relevant, identification of research needs to address critical information gaps and the assessment of proposed future research actions and research results. The scientific risk assessments carried out by the Scientific Committees should always be based on scientifically accepted standards of best practice and be transparent regarding the data, methods and assumptions that are used in the assessment process. In addition, they should identify uncertainties and use harmonised terminology, where possible, based on internationally accepted terms.

Addressing health, environmental and emerging risks to support decision-making bodies and to identify gaps in knowledge that require further scientific research requires evaluating the gathered evidence. Several definitions have given about the approach of Weighing of Evidence.

The World Health Organization (WHO, 2009) defines Weight-of-Evidence as "*A process in which all of the evidence considered relevant for a risk assessment is evaluated and weighted*".

ECHA (2010) defines it as "*the process of considering the strengths and weaknesses of various pieces of information in reaching and supporting a conclusion concerning a property of the substance*", and "*An evidence-based approach involves an assessment of the relative values/weights of different pieces of the available information that have been retrieved and gathered in previous steps. To this end, a value needs to be assigned to each piece of information. These weights/values can be assigned either in an objective way by using a formalised procedure or by using expert judgement. The weight given to the available evidence will be influenced by factors such as the quality of the data, consistency of results, nature and severity of effects, relevance of the information for the given regulatory endpoint*. The weight of evidence approach requires the use of scientific judgement and therefore it is essential to provide adequate and reliable documentation."[1]

EFSA (2017) defines weight of evidence approach as "*A process in which evidence is integrated to determine the relative support for possible answers to a question*".

The *Memorandum on Weight of Evidence and Uncertainties* by Scientific Committees was proposed by SCENIHR in 2012 to provide greater transparency in the risk assessments carried out by this Scientific Committee, and to provide greater consistency between Opinions and to be helpful to stakeholders. In 2018, the SCHEER revised the *Memorandum* and published a modified version (SCHEER, 2018). The SCHEER approach is in line with the previous definitions, since they described *Weight of Evidence* as "A process of weighted integration of lines of evidence to determine the relative support for hypotheses or answers to a question". A line of evidence is considered as a set of evidence of similar type (EFSA, 2017).

However, human and environmental research is a continuous process that faces several challenges, novel aspects and sources of information that are necessary to incorporate and integrate within *Weight of Evidence and Uncertainties* approaches, before formulating scientific Opinions.

[1] https://echa.europa.eu/support/registration/assessing-hazard-and-risk

2. TERMS OF REFERENCE

On 26 June 2018, the SCHEER adopted its "Memorandum on weight of evidence and uncertainties - Revision 2018"[1] with the aim of providing greater transparency about its risk assessment process, ensuring consistency between Opinions, assisting stakeholders and being in line with the provisions stipulated in the Rules of Procedure.

The previously published Memorandum (2018) focussed on how to use the weight of evidence approach (WoE) to conduct a risk assessment for stressors to which humans and/or the environment may be exposed. The aim of that document was to support the use of the WoE, wherever appropriate, for the risk assessment activities of the Scientific Committee on Health, Environmental and Emerging Risks (SCHEER). In addition, it aimed to support the consistency in the work of different EU bodies performing risk assessments.

The SCHEER proposed that when sufficient practical experience had been gained from the application of the WoE approach to the different mandates, the SCHEER would consider the need for any updates and modifications of the Memorandum. In order to evaluate the implementation of the WoE approach, the SCHEER established a separate Working Group, tasked with collecting practical experience with the approach. Members of the group were assigned to a WoE task in the different mandates.

The SCHEER is requested to prepare a revised version of the 'Memorandum on weight of evidence and uncertainties', adopted in 2018, making explicit the approach to be used by the SCHEER for determining the weight of evidence and the uncertainties involved in the development of their Opinions.

The approach should take into consideration the newest available, evidence-based methodology that has been developed by various national and international risk assessment bodies, including European Union bodies and agencies, as appropriate, and the practical experiences gained from application of the WoE approach to SCHEER mandates (2018-2023).

The Revised Memorandum should be applicable for human health, environmental and ecological risk assessments.

3. METHODOLOGY

3.1 Introduction

This memorandum is intended to make explicit the approach used by the SCHEER for determining the weight of evidence (WoE) and the uncertainties involved in developing Opinions based on risk assessment. The Memorandum draws on the methodology sections of previous Opinions of the SCENIHR, SCHER and SCHEER, identifying the best common practices in the different domains. It involves a staged approach and several additional elements that are considered to improve the transparency and consistency of human health and ecological risk assessments carried out by the Scientific Committee. The approach draws on several schemes that have been developed by various national and international bodies. Particular attention has been paid to ensuring that the format of the schemes can be applied to a wide range of lines of evidence and types of publication.

This chapter provides a brief description of the framework and steps required to complete a WoE analysis to be used for risk assessments by the SCHEER. It updates the approach developed previously (SCENIHR, 2012) and SCHEER (2018). Several organisations have established their own frameworks for assessing/evaluating evidence (e.g. EFSA 2017, 2018). These have been considered wherever appropriate in the development of this memorandum.

3.2 General framework

The SCHEER has identified the following key steps in implementing a WoE approach in risk assessment, which will be addressed in detail in the subsequent chapters (Figure 1):

- Problem formulation of the risk assessment requested (see section 4.2.1),
- Identification of sources, collection and selection of data and information as well as gaps in relation to the aim of the assessment (see section 4),
- Screening of the data and information to identify those that are relevant to address the question(s) posed by the Commission Services (see section 4),
- Assessment of the quality and consistency of the content and weighing of the data and information to establish individual lines of evidence (see section 5),
- Integration of the relevant lines of evidence (see section 6),
- Conclusions and reporting (see section 8).

At each step in the WoE approach, consideration is given to the associated uncertainties including the data variability, any data gaps and model uncertainties (see section 6).

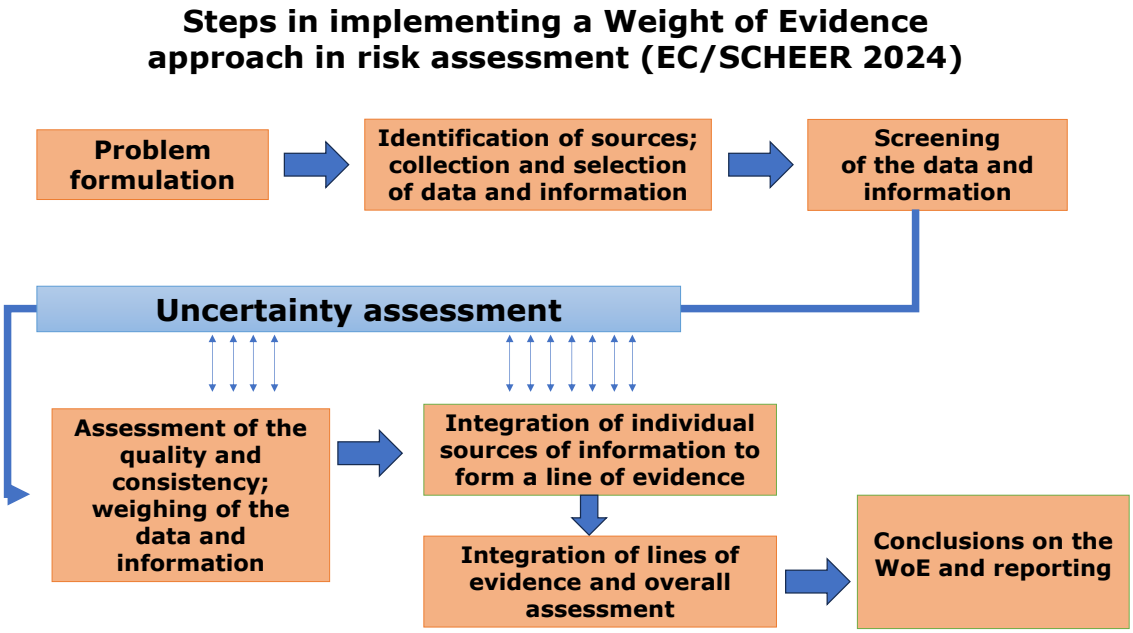

Figure 1: The Weight of Evidence process in risk assessment

3.2.1 Problem formulation

A critical aspect of the risk assessment process is the framing of the questions asked. The problem formulation must clearly state the purpose of the risk assessment and should include the protection goals of concern, the stressors of interest, their sources, routes of exposures, target populations and critical endpoints. The Terms of Reference (ToR) provided to the SCHEER usually contains the problem formulation.

3.2.2 Identification of sources, collection and selection of data and information

Depending on the issue being addressed, the SCHEER uses different sources for information. These may include systematic reviews and meta-analysis, publications of observational studies and clinical trials, computational analysis and experimental data, as well as expert judgements that serve to support, refute or modify a scientific hypothesis or theory. The SCHEER primarily relies on peer-reviewed publications. Non-peer-reviewed reports are considered and weighted on a case-by-case basis based on the expert judgement of the SCHEER.

3.2.2.1 Real World Data

The increasing volume of available data, including but not limited to product (e.g., medical devices) and disease registries, e-health services, as well as industry, hospital, and pharmacy records, together with rising costs and recognised limitations of traditional epidemiological studies and clinical trials, has spurred great interest in the use of Real-World Data (RWD)[2].

[2] RWD is, structured or unstructured, data derived from a number of sources of a heterogeneous population in real-world settings, i.e., registries, databanks, repositories, industry, official statistics, insurance and health claims, etc. As no universal definition of RWD exists, it is typically understood as distinct from data sourced from typical designed studies (Sherman RE, Anderson SA,

Based on RWD, real-world studies (RWS) have been conducted to improve the efficiency of research and to bridge the gap between clinical, animal and environmental research, (bio)monitoring networks and daily practice, i.e., the real world. However, it should be underlined that RWD may have several drawbacks, including heterogeneity, different types of measurement errors/biases, and quality issues. Key concerns include missing or incomplete data, errors in data entry or coding, lack of uniformity across data sources that needs thorough data harmonization, as well as selection/reporting/observer bias. Thus, due to the nature of RWD, their use to draw Real World Evidence (RWE) for decision making and causal inference needs special attention.

3.2.2.2 Presentation of Sources of Information and Data

The presentation of the identified pieces of information and data gathered during the search procedure to inform a line of evidence should be clear and the search strategy used should be transparent. Databases used, timeframe of searches and keywords used should be presented in the methodology section of each Opinion.

Several guidelines have been proposed on how to report the results of a study involving human data (e.g., a meta-analysis, a clinical trial, an epidemiological study or a case-report). It is strongly recommended to follow such guidelines, like the QUORUM statement (Improving the quality of reports of meta-analyses of randomised controlled trials: the QUORUM statement, Lancet 1999; http://dx.doi.org/10.1016/S0140-6736(99)04149-5), the PRISMA statement (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), an evidence-based set of items for reporting systematic reviews and meta-analyses, mainly of randomised trials, or the Meta-analysis of Observational Studies in Epidemiology (MOOSE) checklist (Brooke, 2021). For randomised clinical trials the study's main characteristics should be reported, according to the CONSORT statement (http://www.consort-statement.org/), and for epidemiological studies, the STROBE statement (https://www.strobe-statement.org/index.php?id=strobe-home), or other similar guidelines, could be followed. Moreover, the STaRT-RWE template serves as a guiding tool for designing and conducting reproducible RWE studies. In the case that Opinions are using RWE studies, it is suggested using this or similar templates for the presentation of RWS (Wang, 2021).

Animal studies are widely performed according to the OECD test guidelines, which are internationally agreed testing methods that include the relevant information to allow reproducibility and comparability of results. The OECD test guidelines for animal experiments play an important role in evaluating the chemical hazards. Animal tests performed using OECD guidelines, especially when the good laboratory practice (GLP) principle is applied, reduce the duplication of toxicity testing and ensure the best mutual acceptance of data by the OECD's Mutual Acceptance of Data (MAD) (*Animals* 2022, *12*(23), 3305; https://doi.org/10.3390/ani12233305). In regulatory testing, the principle of Replacement, Reduction, and Refinement (3Rs) should be applied according to REACH Article 4 of Directive 2010/63/EU, which recommends that the number of animals used in projects should be reduced to a minimum without compromising the objectives of the project.

Regarding environmental studies the CIEM (2017) guidelines could be followed, providing description of each survey method used, description of constraints/limitations on the methodology and a clear statement of any assumptions that have been made.

Dal Pan GJ, Gray GW, Gross T, Hunter NL, LaVange L, Marinac-Dabic D, Marks PW, Robb MA, Shuren J, Temple R, Woodcock J, Yue LQ, Califf RM. Real-World Evidence - What Is It and What Can It Tell Us? N Engl J Med. 2016 Dec 8;375(23):2293-2297).

For any study, it is recommended to follow general guidelines concerning the statistical analysis (see Annex).

3.2.3 Assessment of the quality of content and consistency; weighing of the data and information to establish lines of evidence

A major task of the SCHEER while conducting a risk assessment is to identify and assess information sources in terms of quality, and to judge their validity, reliability and relevance (Klimisch *et al.,* 1997; Nendza *et al.,* 2010; ECHA, 2010, Moermond, 2016) before evaluating their contribution to a line of evidence. EFSA has recently published an updated guidance on fit-for-purpose protocols for the development of generic scientific assessments based on a harmonised and flexible framework (EFSA, 2023) which has been considered in this section.

- Relevance: This defines whether a set of data is appropriate for a particular hazard identification or risk characterisation and therefore has the potential to contribute to addressing the ToR. Any source of information that is not considered relevant is not considered further.
- Validity: Validation is the process of evaluating the methods used for data generation or for the development and application of a model and of the data so generated. Accepted standards or guidelines (e.g. OECD, EN, ISO) are preferred, but do not always exist. If that is the case, the underlying scientific principles should be of known validity. For more recent developments in NAMs (New Approach Methodologies, including in-vitro studies) and validation, protocols are still under development (OECD, 2018).
- Reliability: Evaluating an individual source of information with regard to the inherent quality of a test report or publication relating to a validated methodology and the way that the procedure and results are described to demonstrate the robustness and plausibility of the findings.

Key issues to be evaluated when considering relevance, validity and consistency include:

- An unambiguous characterisation of the stressor
- Soundness and appropriateness of the methodology and models
- Extent to which the full details of methodology and experiments are provided
- Reproducibility of findings between experiments/observations
- Relevance of a set of data for a particular endpoint
- Causality and plausibility of the findings
- Availability of the underlying experimental, observational or real-world data.

3.2.4 Integration of the selected sources of information to provide a line of evidence

Integrative assessment means that the results from all relevant individual sources of information are combined into an overall assessment, taking into account their reliability, validity, relevance and uncertainty. The integration of the different sources of information to create a line of evidence requires expert judgement. The overall assessment depends on the consistency and the quality of the individual sources of information. Consistency is defined as the coherence in the findings between the relevant sources of information (SCENIHR, 2012), but also as the extent to which contributions of different sources of information relevant to the specified question are compatible (EFSA, 2017). Quality is defined as the combined result of the judgement on reliability and validity (having judged them relevant).

In the final weight of evidence assignment involving the integration of the different lines of evidence, the basis for the judgement must be provided in the SCHEER Opinion. Relevant information gaps should be clearly identified and their impact on the overall conclusions should be assessed.

3.2.5 Uncertainty evaluation in WoE assessments

A useful generic definition of "uncertainty" is provided by EFSA (2017), which states that "uncertainty refers to all types of limitations in the knowledge available to assessors at the time an assessment is conducted and within the time and resources agreed for the assessment." The EFSA (2023) guidance emphasises that uncertainty is inherent in each step of the scientific assessment process (from problem formulation to drawing conclusions). Therefore, the approach to addressing uncertainty should be carefully planned and implemented *for each step of the scientific assessment*. The extent to which uncertainty can be defined at the formulation phase will vary between assessments based on a variety of factors including data availability, established practices in the field and/or the level of detail required by the problem formulation at hand.

The SCHEER follows a similar approach to EFSA (Chapter 7).

The strength of the evidence is inversely related to the overall degree of uncertainty. Characterisation of the uncertainties in WoE assessments is important for transparency and is also a valuable aid to help risk managers formulate risk advice. In addition, it points to priorities for further work to improve the robustness of hazard identification and risk assessments. However, if not clearly and suitably described, expressing uncertainty may raise unwarranted concerns and/or provoke unwarranted actions, as the degree to which characterisation of uncertainty is needed will depend on the risk assessment and risk management contexts as determined by the questions asked, *i.e.* the problem formulation. Uncertainty analysis should be incorporated during the assessment, during the weighing of individual sources of information and when integrating lines of evidence, rather than after this process is completed.

3.2.6 Conclusions

Clear and transparent documentation and argumentation is essential for allowing stakeholders and policymakers to understand how the different sources of information were selected, assessed and integrated into lines of evidence and in the final WoE assessment used in the SCHEER Scientific Opinions.

More specifically, what is needed is an explicit and transparent documentation of the assumptions, defaults, data sources, decision criteria, applications of expert judgment and other descriptive information used to reach the conclusions of the assessment. The conclusions/reporting should include any uncertainties and gaps identified. SCHEER Opinions include, whenever appropriate, a separate file containing the sources of information considered, as well as the scoring for relevance, reliability and validity.

4. IDENTIFICATION, COLLECTION AND SELECTION OF THE POSSIBLE SOURCES OF INFORMATION

The search for relevant information and data aims to identify and collect the potential sources of information to answer the specific questions being asked in the ToR. Accordingly, the SCHEER may utilise one or more of the following:

- Data provided by the DG tasking the SCHEER
- Data provided by a third party (e.g. stakeholder submissions, which may also include confidential data)
- Real world data
- New approach methodology (NAMs) data
- Reports and Opinions of other scientific bodies
- Reports of various governmental and EU or international bodies (e.g. WHO, FAO, EMA, EFSA, JECFA, IARC, OECD, WMO, NIEHS)
- Reports of stakeholder bodies (e.g. ILSI, ECETOC, WWF)
- Scientific (peer reviewed) publications, including meta-analyses and systematic or narrative reviews.

In this memorandum, the SCHEER has updated the sections relating to newer forms of data (real world data and NAMs data). Full discussion of more traditional forms of data can be found in SCHEER (2018).

4.1 Use of confidential data

For the purposes of carrying out its work, the SCHEER may make use of confidential data. This is the case when the data will become publicly available in the near future or when data and/or information are not publicly available because they are provided by e.g. an applicant and/or by different stakeholders after a call for information is published by the Commission Services.

Confidential data will only be considered by the SCHEER if the provider agrees that the summary of the evaluation of the data can be incorporated in the text of the Opinion and thereby made publicly available.

The Commission Services retain the confidential files if they have been used to generate an Opinion.

4.2 Screening of data sources

Relevant peer-review papers are identified from different literature databases including PubMed, Scopus, Web of Science, Toxline, Chemical and Biological Abstracts or Google Scholar. In each Opinion, the search engines used and the period covered in searching relevant documents should be identified, along with the search terms used. Both keywords and index/subject search terms (e.g., MeSH, Medical Subjects Headings, the NLM-controlled vocabulary thesaurus used for indexing articles for PubMed) are useful and should be used in the search procedure. Keywords should be tagged to search all texts in the documents. Index/subject terms help to focus the screening appropriately, enabling researchers to look for items using the specific terms applied by SCHEER members. As the issues that need to be addressed differ significantly between Opinions, the specific data sources will also differ.

The search strategy (including inclusion and exclusion criteria) will always be clearly stated in the resulting SCHEER Opinion.

Relevant data that become available during the preparation of an Opinion and from the public consultation, are assessed in the same fashion as the data from the literature search. Every Opinion is based on relevant data, published prior to and available within the period of its drafting.

The acceptability of a publication may be based on the criteria proposed by Klimisch *et al.* (1997) and described in the OECD Manual for the investigation of high production volume (HPV) chemicals and further elaborated by Nendza *et al.* (2010) and ECHA (2017). These criteria may be modified to be applied to data sources pertinent to various stressors, including physical agents. Ecotoxicological data can and should be evaluated according to the CRED criteria (Moermond *et al*., 2016). The CRED evaluation method uses 4 reliability categories, similar to the Klimisch *et al.* scores: reliable without restrictions (R1), reliable with restrictions (R2), not reliable (R3), and not assignable (R4).

Prior to the screening of data sources, it is important to consider the problem formulation in the ToR and all aspects of the risk(s) to be evaluated since incomplete identification of the risk(s) and/or risk factors may lead to an inappropriate literature search.

The process of incorporating systematic review methodology into literature-based evaluations has been published by OHAT (2015). However, this process is very resource-intensive and the SCHEER may therefore not be able to follow it.

In preparing an Opinion, the SCHEER should cross-check all references that are intended to be cited from reports and Opinions of other scientific bodies included in the reference list. Personal communications can only be used if supported by raw data and details of the methodology used.

Conflicts of interest, in particular those of a financial nature, will be considered by the SCHEER for each source of data.

Only publications that meet the criteria of relevance, reliability and validity for the development of the Opinion appear in the reference list. In an Appendix, all other relevant publications from the literature search should be mentioned as "publications not considered suitable for the purposes of developing the Opinion".

4.3 Assessment of Real-World Evidence

A wide range of methodologies have been proposed to make appropriate and effective use of RWD as a source of evidence. However, work is still ongoing to determine how to improve data quality and how to properly use RWD to generate unbiased RWE. Although RWS can offer substantial information on the line of evidence, the SCHEER, before using RWD as a source of evidence, should:

1. describe the source(s) of RWD
2. evaluate the
    - completeness / missingness
    - heterogeneity
    - variability
    - validity
    - reliability
3. compare RWD against the available conventional data sources in terms of relevance (Makady *et al*., 2017; Liu and Panagiotakos, 2022).

4.4 Assessment of data from New Approach Methodologies

In order to reduce animal testing in research and regulatory toxicity, numerous national and international initiatives and programs have been launched to develop and validate alternative *in vitro, in chemico* or *in silico* methods, the so-called NAMs. It can be assumed that in the near future, safety assessments will increasingly become based on data generated by NAMs. One task of the EU Reference Laboratory for alternatives to animal testing (EURL ECVAM) is the validation of alternative methods for the safety assessment of chemicals to enhance the quality and reproducibility of these types of data. In this regard, also specific OECD guidelines have been developed for *in vitro* studies.

Data generated by NAMs may be used for purposes of safety assessments, even if the methods have not undergone an official validation process or are part of an OECD guideline. In these cases, quality and reliability of the data have to be judged by the safety assessor. For both conducting and assessing i*n vitro* studies, several tools are available including:

- the OECD Guidance Document on Good In Vitro Method Practices (GIVIMP, 2018) for the development and implementation of *in vitro* methods for regulatory use in human safety assessment,
- the ToxRTool (Toxicological Data Reliability Assessment Tool), developed by the European Commission's Joint Research Centre (Schneider *et al*., 2009),
- the Oral Health Assessment Tool (OHAT) from the National Health and Medical Research Council or
- the Science in Risk Assessment and Policy (SciRAP) tool for evaluation of reliability and relevance of *in vitro* studies (Roth *et al.,* 2021).

In vitro study designs are increasingly prevalent in the toxicological literature driven by efforts to develop experimental models. Such designs are more suitable for detecting adverse effects of environmental exposures, providing mechanistic information, and minimising the use of animals in research. Specifically for *in vitro* studies, other tools for quality evaluation have also been published (Svendsen *et al.*, 2023).

In order to evaluate ecotoxicological data, the CRED criteria (Moermond et al., 2016) are an established approach which can also be applied to data generated by NAMs.

5. ASSESSMENT AND WEIGHING OF INDIVIDUAL PIECES OF INFORMATION AND DATA TO FORM A LINE OF EVIDENCE

Several organisations have proposed grading systems for assessing pieces of data or information. In this chapter, a guide to the assessment procedures used to evaluate a piece of information and data gathered from the literature search, as well as their integration to a line of evidence and its quality and consistency is presented. First, each piece of information is judged based on its relevance to the ToR, and then on the validity and reliability of the methodology applied and the results obtained. Then an overall quality score for that piece is assigned. Finally, the individual pieces of information and data are integrated to form a line of evidence. Quality and consistency of that line of evidence is then assessed, which provides its weight of evidence.

5.1 Quality assessment of the individual piece of information and data

To assess its potential contribution to a line of evidence, each piece of information is reviewed to determine its relevance (binary, yes or no), validity and reliability (classified

into high, medium, or low, see Figure 2). The individual pieces of information are then assessed for consistency before being integrated into a line of evidence, which is also assessed for quality (for the definitions of relevance, reliability and validity, see page 13).

In subsequent sections, specific aspects are discussed in order to judge the quality of the different sources of information. Important aspects to be considered are:

- Differentiation between non-adverse and adverse/toxic effects
- Reversibility of toxic effects
- Presence of dose- or time-effect relationship
- Evidence that the adverse/toxic effect is caused by the exposure to the stressor of interest
- Knowledge of the mode of action (MoA) and/or the resulting Adverse Outcome Pathway (AOP)
- Biological plausibility, not mere statistical significance. A biologically relevant effect can be defined as a definite and durable effect considered by expert judgement as important impairment of human, animal, plant or environmental health. It therefore implies a change that may influence how decisions for a specific problem are taken (EFSA, 2017)
- Information on normal variation in the incidence of the disease/effect of interest (e.g. consideration of historical controls).

When assessing and weighing individual pieces of information, any potential for biases should be considered. It is also important to consider whether a study has been conducted in accordance with accepted test guidelines, e.g. OECD guidance documents, and whether a study was performed under a quality system e.g. GLP (Good Laboratory Practice).

In ***Figure 2*** the procedures for assessing the relevance, reliability and validity, and ultimately the quality of individual pieces of information and data are illustrated. At first, each piece is evaluated based on its relevance in a dichotomous way, *i.e.* relevant or irrelevant.

Then, an assessment is performed based on the reliability and validity, according to a 3-scale scoring, *i.e.*, low, medium and high. Low validity is considered when the piece of information or data gathered were derived through studies with considerable methodological limitations, medium and high are used when the piece of information and data are based on widely adopted, robust methodologies or on "gold standard" approaches (for high). Regarding reliability, a low grade is given when the piece of information or data cannot be repeated or replicated under the identical conditions, whereas, medium and high are given when the information or data gathered are based on methodologies that are proven to be repeatable and or replicable.

For relevant pieces, the combination of reliability and validity defines the quality of each individual piece of information and data, taking also into account the level of uncertainty. This approach is illustrated in ***Figure 2***.

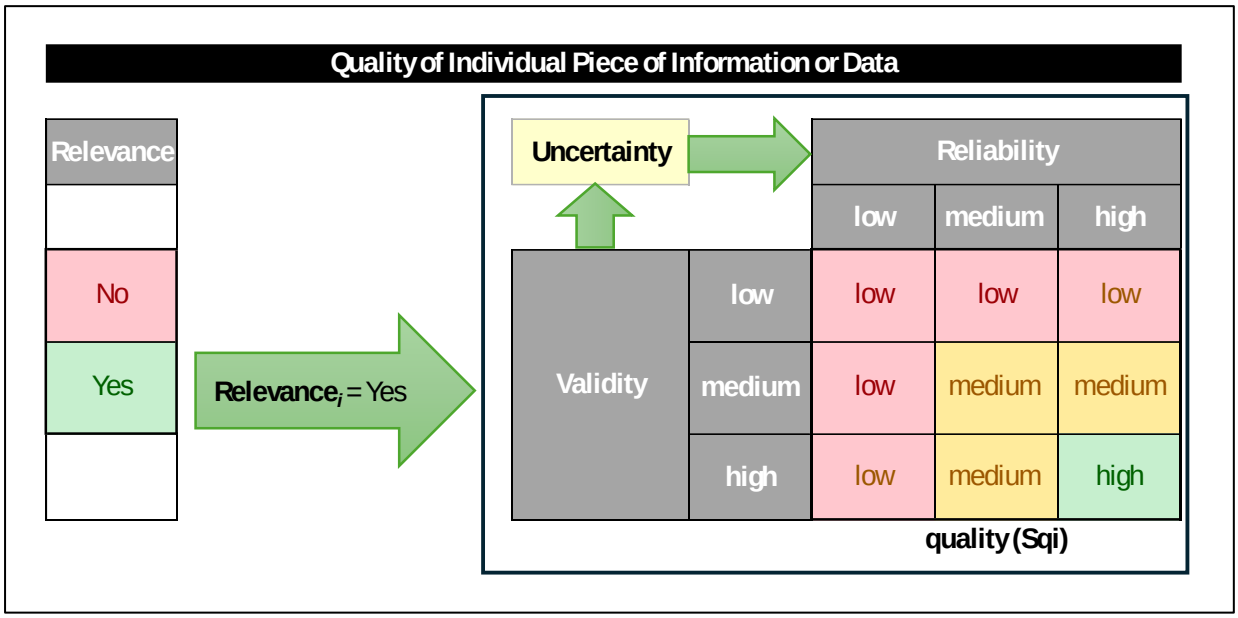

Figure 2. A framework for assessing the relevance (Relevance$_i$) and quality (SQ$_i$) of each individual piece of information and data *i*. Quality (SQ$_i$) is assessed with the help of validity and reliability, taking into account uncertainty.

5.2 Other quality assessment approaches of the individual piece of information and data

At this point it should be noted that a number of organisations have established their own frameworks for assessing and/or evaluating quality of experimental evidence. For example, the preamble to the IARC Monograph Series (IARC, 2006), where a weight is attributed to the study findings as a result of the hazard or risk assessment. In this process, risk assessors need to assess uncertainties in the underlying data as well as in their own interpretations of these data (Levin *et al.,* 2004).

EFSA (2010, 2015, 2023) provides guidance on literature searching and systematic review and ECHA provides practical guidelines on how to apply weight of evidence (ECHA, 2010, 2017).

Concerning mainly human studies, the U.S. Preventive Services Task Force (US PSTF) suggested five different levels, based on type and quality of the study (Level I, Level IIa, IIb, IIc, and Level III). The Oxford Centre of Evidence Based Medicine (CEBM, UK) provided similar levels of evidence for claims about the prognosis, diagnosis, treatment benefits or harms and screening of a treatment.

In 2000, a 4-level systematic approach was proposed by the GRADE (i.e., Grading of Recommendations Assessment, Development and Evaluation) working group that takes into account risk of bias, imprecision, how the results are actually going to be applied as well as inconsistencies and publication bias (Balshem *et al*., 2011).

All these risk assessment approaches are mainly for human studies and focus on the design and interpretation of the results of individual sources of information, giving less emphasis to animal, ecological, environmental sources of information and data.

6. EVIDENCE INTEGRATION PROCESS

6.1 Integration of individual pieces of information to form a line of evidence and its quality and consistency assessment

Following the quality assessment of individual pieces of information and data, pieces of information which are judged to be of low quality are excluded. Low-quality data can lead to significant issues that impact the validity, reliability, and therefore the robustness, and usefulness of research outcomes for decision-making; these data increase uncertainty and can pose significant ethical and legal risks for regulatory organisations and researchers. Implementing strict data quality standards, validation protocols and governance frameworks is crucial for maintaining the integrity of the outcomes of the WoE approach for risk assessment. Data must be integrated to an individual line of evidence and a quality assessment must be performed. This is a two-step procedure involving assessment of overall quality and consistency.

The overall quality of a line of evidence is based on expert judgement from the evaluation of the quality of the individual pieces. A three-level scale is used:

- **A** – the majority (at least 75% of the individual pieces of information or data, with a minimum of 3) are of high quality,
- **B** – the individual pieces of information or data are mixed in terms of quality (high/medium),
- **C** – all individual pieces of information or data are of medium quality.

The second step is the assessment of the consistency of the individual pieces of information and data before they can be integrated to a line of evidence. The classification of consistency of the line of evidence is, according to the 3-scale scoring, I, II, III, as presented below:

- **I** – the majority (at least 75% of individual pieces of information or data, with a minimum of 3) lead to similar conclusions,
- **II** – between 50% and 75% of individual pieces of information or data lead to similar conclusions,
- **III** – fewer than 50% of the individual pieces of information or data lead to similar conclusions.

If there are fewer than 3 pieces of information or data, it is not possible to integrate them into a line of evidence following this approach, since consistency cannot be judged. In such a case, the SCHEER uses the data for the risk assessment and the quality of the individual pieces of information or data and decides whether or not they are in agreement.

In ***Figure 3*** the procedure for assessing the quality and consistency of a line of evidence is illustrated. The assessment of similarity of conclusions is based on expert judgment. Note that a line of evidence can be based on various types of studies, e.g., human studies, animal studies, *in-vitro* studies, or on toxicological studies, or based on exposure, etc.

At this point it should be noted that a thorough analysis of the causes of inconsistencies is recommended. Consistency (or lack) of findings is one measure of uncertainty that is considered in the integration.

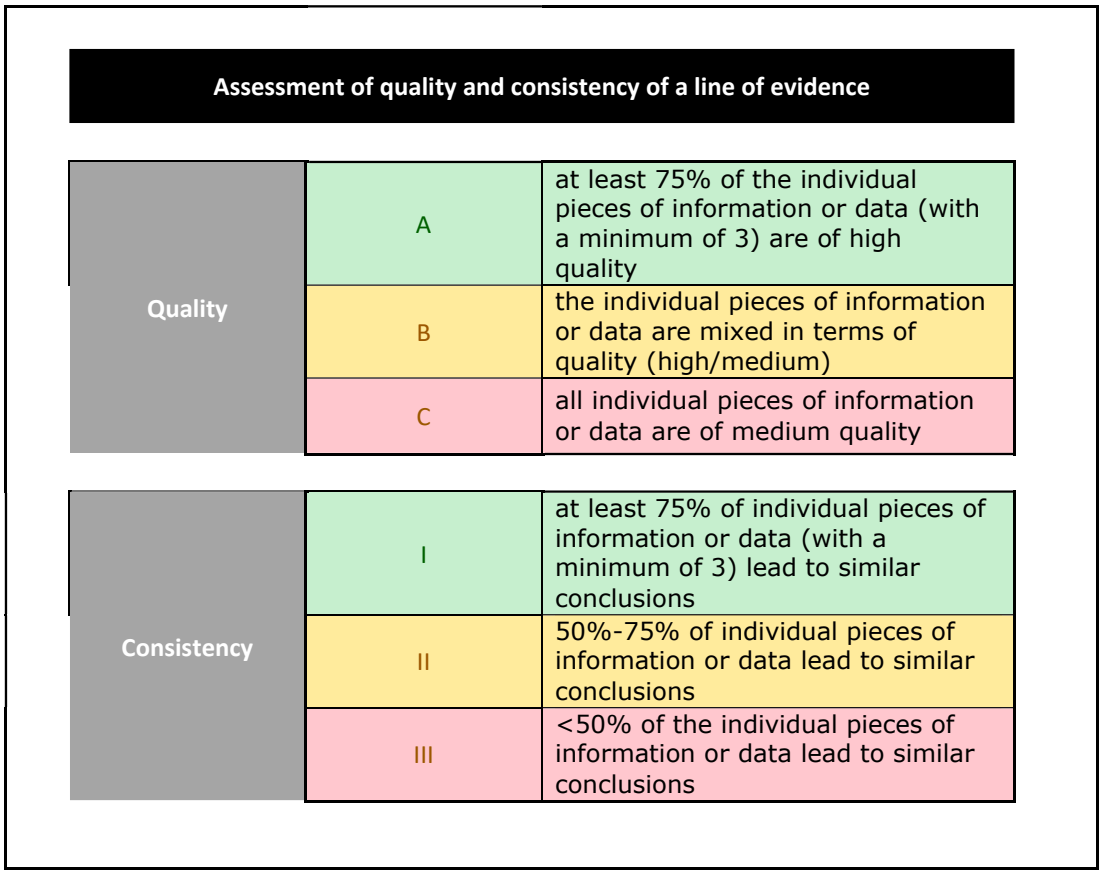

Figure 3: Grading a line of evidence, based on individual sources of information and data, (LQ_i) and consistency of individual sources, (LC_i).

The process as described above may require modification, and expert judgement should be used to consider integration and assessment e.g. when there are very limited sources of information. This would be transparently reported in the Opinion.

6.2 Integration of lines of evidence and overall assessment

In the overall assessment, the individual lines of evidence and their assessment (as outlined in *Section 6.1*) are evaluated before being integrated to arrive at the overall conclusions. For some mandates, based on the problem formulation, this may lead to identification of primary sources of information, i.e. those that would be judged by experts to address the problem, leading to a pre-eminent line of evidence (sometimes called a primary line of evidence). Identification of a primary line is, therefore, dependent on the mandate. For example, a mandate may stipulate to include only human studies. There may (or may not) also be secondary lines of evidence. In such situations, expert judgement is used to prioritise and weight the different lines of evidence when there is no primary line of evidence.

The main goals of the integration procedure are:

- To assess the consistency of different lines of evidence and their uncertainties like key events, concordance of dose-response relationships, temporal association, strength, consistency and specificity of association of toxic response

with key events, biological plausibility and coherence, causality, uncertainties, data gaps, assessment of postulated mode of action (human relevance, human biomonitoring data). Consistency is the extent to which the contributions of different lines of evidence drawing a specific conclusion are compatible (EFSA, 2017; Boobis *et al.,* 2006, 2008; Meek *et al.,* 2014).

- When there are inconsistencies, to try to understand and explain the reasons for them, possibly deciding if more than one answer to the formulated problem is plausible.
- To reject cases of unacceptable or inconsistent lines of evidence, which might be described as outliers to conclude on the WoE based on consistency and quality.

First, it should be checked if a primary line of evidence exists, which will be the sole basis for the overall assessment.

- If yes, then the overall assessment will be based on the quality and consistency of that primary line of evidence.
    - Overall assessment is graded as:
        - very strong evidence: quality A and consistency I of the primary line of evidence
        - strong evidence: quality A and consistency II of the primary line of evidence
        - strong to moderate evidence: quality A and consistency III of the primary line of evidence
        - moderate evidence: quality B and consistency I of the primary line of evidence
        - moderate to weak evidence: quality B and consistency II/III of the primary line of evidence
        - weak evidence: quality C and consistency I of the primary line of evidence
        - uncertain evidence: quality C and consistency II/III of the primary line of evidence

	A	**B**	**C**
I	very strong	moderate	weak
II	strong	moderate to weak	uncertain
III	strong to moderate	moderate to weak	uncertain

Table 6.2.1: Overall weight of evidence (primary line of evidence)

Exceptionally, if the assessment is based on fewer than 3 pieces of information or data, then there would be no categorisation as in Table 6.2.1. Instead, the individual pieces would be described in terms of quality and agreement, and it would be noted that the overall weight of evidence could not be assessed.

Where there are multiple lines of evidence, the relevant weighting based on the overall assessment of quality and consistency of the different lines of evidence is done by expert judgement.

Based on expert judgement, the overall quality of the lines of evidence can be considered as A, B or C, according to:

- **A:** all lines of evidence are of quality A
- **B:** the quality of the lines of evidence is mixed (A/B)
- **C:** all lines of evidence are of quality C

Consistency among the lines of evidence can be considered as:

- **I:** more than 75% of the lines of evidence show similar findings
- **II:** between 50% and 75% of the individual lines of evidence result in similar findings
- **III:** fewer than 50% of the lines of evidence have similar findings.

A thorough analysis of the causes of inconsistencies is recommended.

The final phase of the integration is the assessment on the basis of the complete information available and the critical evaluation of the individual lines of evidence. This process can result in an overall assessment graded as:

	A	B	C
I	very strong	moderate	weak
II	strong	moderate to weak	uncertain
III	strong to moderate	moderate to weak	uncertain

Table 6.2.2: Overall weight of evidence (multiple lines of evidence)

6.3 Remarks

To draw conclusions, integration involves combining conclusions from various lines of evidence while considering their consistency and quality. For example, integrating evidence for a mechanism of action (MoA) to explain the occurrence of a specific effect at biologically plausible doses in experimental animals, but not in humans, or vice versa. This process ultimately aims to lead to a comprehensive understanding, but it requires an element of expert judgement. The severity of the effect/outcome and the likelihood of its occurrence in individuals or in the reference population are other factors to be considered in the integration process. One of the crucial points is the identification of the critical effect both in animal and/or human studies.
In weighing the lines of evidence, the type of question, described carefully in the problem formulation, determines what evidence can be considered as relevant, reliable and valid, and how they should be weighed.

In the traditional risk assessment process, the results from NAMs, *in silico* and validated *in vitro* tests, as well as knowledge of MoA and Absorption, Distribution, Metabolism, Excretion (ADME), generally strengthens the supporting evidence to the overall weighting. In the Next Generation Risk Assessment process (NGRA), the NAMs might provide results as the primary line of evidence in the future.

When there are physical stressors, the methodologies used can be compared against best laboratory practice or, if possible, some of the tools used for evaluating chemical toxicity data can be adapted to the data collected for physical stressors.

The key issues in the evaluation of human health risks are to assess whether the results demonstrate qualitative and quantitative similarities in key events, mode of action, kinetic or dynamic factors between animals and humans, true causal effect, the identification of

the affected population and dthe extent to which the adverse effects of the exposure might be mitigated. This involves:

- Estimating the incidence and severity of adverse effects likely to occur in a population due to exposure to a substance or a mixture
- Addressing several potential toxic effects and human (sub)populations, and considering each (sub)population's exposure by relevant exposure routes
- Focusing on the most critical effect(s) (with consideration of population, exposure route, and time scale)
- Characterising the sources and magnitude of uncertainties

When there are fewer than three individual sources of information or data, special attention should be given when integrating them into a line of evidence. In such cases, expert judgement becomes especially important.

For ecological risk assessments, the WoE based procedures should overcome the traditional approach of risk characterisation based on the comparison of two numerical values (PEC and PNEC) obtained with univocal and deterministic procedures. As mentioned above, ideally, a risk characterisation should be based on probabilistic values for exposure and assessment with statistically determined uncertainty (van de Meent *et al*., 2023).

If a complete data set would be available for exposure (emissions, modelling and monitoring data) and effects (laboratory data, species sensitivity distribution (SSD), higher tier data), the WoE should y provide the information necessary for a more detailed characterisation of hazards and risks, capable of better describing the type of effect likely to occur (losses of biodiversity, reduction of ecosystem services, etc.), as well as the probability of its occurrence evaluated as a function of the variability of exposed system (vulnerability, ecological value, etc.) and of the uncertainty of the results.

7. DESCRIPTION OF UNCERTAINTY IN WoE

Definitions of uncertainty vary depending on the context. In general, uncertainty is defined as "*lack of knowledge*" or "*limitations in knowledge*".

- Uncertainty can be defined as "*imperfect knowledge concerning the present or future state of an organism, system or (sub) population under consideration*" (WHO/IPCS, 2014)
- According to *EFSA Guidance on Uncertainty in Scientific Assessments*: "*uncertainty refers to all types of limitations in the knowledge available to assessors at the time an assessment is conducted and within the time and resources agreed for the assessmen*t".

In SCHEER Opinions, uncertainty should be expressed in relation to the question asked by risk managers and decision makers and should be appropriate regarding the quality and quantity of information or data available to the Committee. In its uncertainty analysis, the SCHEER rely on methods described by EFSA (EFSA, 2017, 2018 and EFSA toolbox) and ECHA (2021a). Uncertainty can be assigned both to individual pieces of information and to a line of evidence, though the uncertainty in either case may be quantified differently.

Uncertainty is evaluated taking into account the outcome of the assessment of the pieces of information (individually) or the line of evidence (collectively) using the criteria specified in the corresponding steps of the WoE approach (such as validity, relevance, reliability for individual sources of information), and consistency (for the line of evidence assessment).

Characterisation of uncertainties is useful for identifying the priorities for generating more or better data.

This means that not all uncertainties are quantified using the most sophisticated scientific methods available (e.g. a fully probabilistic analysis); this would be inefficient when simpler methods of quantification would provide sufficient information on uncertainty for decision making.

Thus, the SCHEER frequently uses a qualitative approach based on expert judgement. In this update to the WoE approach, the SCHEER has provided details concerning the consistency evaluation within a line of evidence and between lines of evidence.

7.1 Identification of sources of uncertainty

An important objective of any uncertainty analysis is the description of the critical sources of uncertainty and characterisation of their impact on the Opinion for the formulated problem. The uncertainties in the WoE are largely based on the evaluation of the relevance, reliability, validity, quality and consistency of a piece of information or data (Section 4.2.3). Uncertainties in a line of evidence are evaluated based on their quality and consistency (sections 6.1, 6.2). In addition to considering the uncertainties involved in the evidence and its synthesis, it is important to identify any other uncertainties in the judgement used (Meek *et al.*, 2014) in the overall risk assessment.

In the WoE approach, the uncertainty attributed to individual lines of evidence must be considered and also accounted for in the final integration. It should be communicated in a proper and transparent way.

Uncertainties in risk assessment (for human health and the environment) are discussed below.

In summary, the SCHEER predominantly adopt a qualitative approach to uncertainty assessment since this is often the only approach viable given the availability of data and information. The SCHEER explicitly recognises the existence of uncertainty in all the stages of the WoE approach and relies on expert judgement with respect to the assessment of uncertainty. In its assessment, the SCHEER recognise that uncertainty and variability are related but not identical. Variability in systems can be better characterised but not reduced, whereas uncertainty referring to a lack of data or incomplete understanding can be reduced. The uncertainty evaluations made by the SCHEER take into account both accuracy and precision. Lack of data and knowledge gaps form part of the uncertainty assessment in the WoE approach.

7.2 Uncertainties in risk assessment

Uncertainty associated to the different steps of risk assessment may reflect different types of knowledge limitations like incompleteness in the dataset and/or knowledge gaps. Uncertainty may be related to measurement errors or to any bias when selecting, aggregating and interpreting data, as well as inherent variability.

7.2.1 Uncertainty in the sources of information and data

Each individual source of information or data is assessed in terms of its reliability, relevance and validity. Assessing only relevant sources, their reliability and validity are

integrated to provide an overall measure of quality. This is also the first stage of uncertainty evaluation. Reliability considers both the methodology and how the experimental procedures and results are described, analysed and presented. This is related to repeatability, which concerns whether the experiment can be repeated and return similar results and is intrinsically linked to variation, which may be natural variability in the test subjects or variation due to the measurement process and which informs an assessment of uncertainty.

The assessment of both reliability and validity, while not explicitly mentioning uncertainty, do require the assessor to assess uncertainty. As an example, in an ecotoxicological study, the precision of the measurements as well as the numbers of samples will all be assessed under reliability. Under validity, accuracy would also be considered. Thus, in the SCHEER WoE scheme, the quality and consistency grade of the source of information or data includes uncertainty assessment.

7.2.1.1 Integration of individual sources of information to form a line of evidence

Assessment of the sources of information or data when forming a line of evidence is based around quality and consistency. The uncertainty in the line of evidence is, thus, similarly embedded in these attributes, and importantly at this stage, the lack of data and evidence as a source of uncertainty is also considered.

The quality in a line of evidence is assessed based on the quality of the individual sources of information, so that the precision and accuracy considerations in these individual sources is tracked through to the final WoE.

Determining consistency, in this case for a given endpoint, involves assessing whether or not the findings of the individual sources of information or data are the 'same'. Implicitly this assessment considers the variability in the findings (and also their precision). Expert judgement is the basis of the consistency assessment.

It should be noted that a thorough analysis of the causes of inconsistencies is recommended. Consistency (or lack of) in the findings is one measure of uncertainty that is considered in the integration.

Part of the uncertainty may be attributed to lack of sources of information or data on a specific endpoint, and this should be noted.

7.2.1.2 Integration to the overall weight of evidence

Quality and consistency assessments of the lines of evidence are discussed in section 6, where a standardised set of phrases that incorporate the expert judgement of uncertainty was presented.

Human health risk assessments and ecological risk assessments are both based on largely independent hazard and exposure evaluations. Depending on whether the assessment is of a predominantly prospective or retrospective nature, the exposure evaluation can be based on predictive models and/or empirical data. Hazard assessments typically use, for both assessment types, empirical (experimental) data. *In silico* data, such as Quantitative

Structure Activity Relationships (QSARs) (OECD, 2007), are usually used as supporting information only.

Human health and environmental hazard assessment are both typically based on realistic worst-case scenarios, implemented by using appropriately scaled assessment factors, also called safety or reduction factors, that account for unquantified uncertainties.

7.2.2 Uncertainties in human health risk assessment

Human health risk assessment is based on information about the hazardous properties of the stressor to be assessed (e.g. a toxic chemical) and the information about the exposure of humans to the stressor (e.g. consumer, when using a product with the chemical under investigation). Uncertainties have to be considered for both, the hazard assessment as well as for the exposure assessment when conducting a risk assessment.

The exposure assessment may rely on:

- direct measures (experimental personal monitoring or biomarkers of exposure)
- predictive modelling (e.g. using environmental levels, personal behaviour patterns, dispersion modelling and emission scenarios)
- deterministic modelling.

In most cases the data base for direct measures or predictive models is limited. In the absence of adequate data, standard default values are used in the modelling. Default values are often conservative to ensure that exposure is not underestimated (worst-case exposure scenarios)

When deriving numerical values, uncertainty factors are introduced (see, for e.g. IPCS, 2009 for a comprehensive review of uncertainties in exposure assessment).

Historically, the hazard assessment is often based on data from experimental animals and these data have to be extrapolated to describe potential hazards in humans. Therefore, so called scaling factors or (un)certainty factors are used for the extrapolation from animals to humans and from high doses (usually in experimental animals) to low doses expected in humans. The data coming from NAMs, which is often used as supporting evidence nowadays, might provide this kind of information in the future.

Further factors are used to address the biological variability among humans. Variability describes a property of the population due to existing differences between individuals. Factors may also be used to extrapolate hazards for susceptible groups like children.

The evaluation of such factors in a human health risk assessment should be included in the uncertainty analysis.

Any lack of knowledge (i.e. lack of data on exposure or on health hazards) needs to be expressed by a respectively high uncertainty of the risk assessment conducted.

7.2.3 Uncertainties in ecological risk assessment

Similar to human health risk assessment, ecological risk assessment integrates the results of two separate evaluations, exposure and hazard assessment, using the predicted environmental concentration (PEC) of a chemical and its predicted no effect concentration (PNEC) by calculating the so-called risk quotient (RQ=PEC/PNEC). Both PEC and PNEC are stochastic variables that follow probability distributions. Often, the information necessary to sufficiently quantify these distributions is missing, but for PNEC, the species sensitivity distribution approach (Posthuma *et al*. 2002) was developed two decades ago to overcome

this problem. In those cases where information about such distributions is missing, chemical assessments are based on appropriate conservative assumptions. Recently, an alternative for the RQ method was proposed that quantifies the probability that the PEC exceeds the PNEC, to account for the stochastic nature of PEC (Van Straalen *et al*. 2022).

The major source of uncertainty in the hazard assessment process stems from the need to extrapolate from simple biotests that are typically performed with individuals or populations of laboratory-cultivated organisms to the protection goal, i.e. ecological structures and functions. To account for the uncertainties resulting from these extrapolations (short laboratory exposures to real-world long-term exposure, isolated populations of laboratory organisms to real-world ecological communities) so-called assessment factors are used. These factors vary between 1 and 10 000, depending on the amount of data available and the applied regulatory framework, and the implemented "realistic worst case" scenarios for the aforementioned extrapolations. Low assessment factors (usually between 1 and 5, but to be evaluated case by case) may be applied if data obtained with higher tier testing approaches (microcosms, mesocosms) are available. In theory, field data should reduce uncertainties. However, considering the multiplicity of stressors potentially present in the natural environment, a precise cause-effect relationship is difficult to determine.

Such assessment factors are not used in the context of predictive environmental exposure assessments. Instead, these assessments are based on worst-case assumptions that are implemented case by case. PECs are usually calculated using (multimedia) exposure models. Exposure concentrations are uncertain due to spatial variability and uncertainty in the model parameters used. Hence, for each chemical the resulting PEC will have a different distribution. The spatial variability and uncertainty can be accounted for and described by spatially resolved environmental models, although this remains a challenging task due to several difficulties related to model setup, computational cost, and obtaining high-resolution input data (Falakdin *et al*., 2022).

By using these distributions mathematically, the overlap with species sensitivity distributions can be calculated to establish a so-called Expected Risk (Van de Meent *et al*. 2023).

Finally, risk assessments are usually based on single substances in isolation. The notion that humans and ecosystems are exposed to mixtures of chemicals that result in cocktail effects that can be triggered even when each chemical is present at low concentrations, should not be overlooked (Rudén *et al*., 2019).

Retrospective exposure assessments can be based on empirical data, i.e. chemical monitoring campaigns. Here, uncertainties can often be better quantified and uncertainties from both the sampling and the analytical phase should be accounted for. A recent technical workshop under the auspice of SETAC has developed a set of detailed criteria for evaluating reliability and relevance of empirical exposure data for estimating exposures in environmental assessments (CREED - Criteria for Reporting and Evaluating Exposure Datasets, see also Merrington *et al*., Hladik *et al*. and Peters *et al*. (all 2024)).

7.3 Expressing and communicating uncertainties

The SCHEER use a set of standardised phrases in the overall weight of evidence, uncertainty being implicitly (and expertly judged) and in tabular form. This follows the approach presented in ECHA (2017).

The SCHEER has adopted the use of standardised terms and has created a series of tables that communicate the strength of the weight of evidence, and ultimately its relationship to the original sources of information and their uncertainties. It is also common SCHEER

practice in any Opinion to identify (and prioritise) where future work as well as additional data are needed, which can be a source of uncertainty.

The risk assessment serves two purposes (1) to provide information about risk management options, prioritisation and decisions by a regulatory agency and (2) to provide regulatory officials with information for communicating with stakeholders and the public about current risks and the expected risks and benefits after taking a particular action (based on WHO, 2014; IPCS, 2014).

Risk assessors need to assess uncertainties in the underlying data as well as in their own interpretations of these data (Levin *et al.,* 2004). Characterisation of the uncertainties in a risk assessment is important for transparency and should also be a valuable aid to risk managers in determining how to respond to risk management advice. In addition, it is a useful way of indicating priorities for further work to improve the robustness of risk assessments. However, if not clearly and suitably described, the expression of uncertainty may result in inappropriate concerns and/or actions. The degree to which characterisation of uncertainty (and variability) is needed will depend on the risk assessment and risk management contexts as determined in the questions formulated, i.e. problem formulation.

Despite the fact that expert judgement is used in the weighing of evidence, there is still a degree of subjectivity. This needs to be recognised and made explicit as far as possible. It may introduce a further degree of concern that is difficult to quantify but which can impact the uncertainties in the risk assessment and consequently the advice to risk managers (Van der Sluijs, 2003 and 2005). Expert judgement in the SCHEER working groups is not based on a single individual's judgement but on that of the whole working group.

8. CONCLUDING REMARKS

Conducting scientifically robust assessments based on a weight of evidence approach requires a structured and clearly documented process if the outcome of the scientific assessment is to be communicated unambiguously to decision makers, the wider scientific community and stakeholders. This process will help to clearly focus on key issues and allow reproducibility of the assessments between expert groups (Hardy *et al.,* 2015). The SCHEER approach is summarised in this document and provides an update to its 2018 WoE Opinion.

The weight of evidence assessment should report the method used for the selection of the individual lines of evidence to ensure that the iterative process leading to the conclusions is fully comprehensible and reproducible. To better achieve this goal, it is advisable that the weight of evidence assessment is reported in a standardised way concerning the choice of methods and all assumptions used, including expert judgement (EFSA, 2017). Biological relevance and associated uncertainty should also be addressed and reported as part of the weight of evidence assessment (EFSA, 2017). Where the methods used are already described in other documents, it is sufficient to cross-reference.

Integration draws together the conclusions drawn from different lines of evidence, (e.g. evidence for a MoA, explaining the occurrence of a certain effect in experimental animals and not in humans or vice-versa). This demands an element of expert judgement. According to this Memorandum, reporting should be performed at each step of the iterative process of building the WoE.

Figure 1 presents the steps needed in a WoE assessment:

- Problem formulation of the risk assessment requested (see section 3.2.1),
- Identification of sources, collection and selection of data and information as well as gaps in relation to the aim of the assessment (see section 4),

- Screening of the data and information to identify those that are relevant to address the question(s) posed by the Commission Services (see section 4),
- Assessment of the quality and consistency of the content and weighing of the data and information to establish individual lines of evidence (see section 5),
- Integration of the relevant lines of evidence (see section 6).
- Description of uncertainty in WoE (see section 7).

As a concluding remark, this memorandum is a living document and is intended to make explicit the approach used by the SCHEER for determining the weight of evidence and the uncertainties involved in the development of its Opinions. It involves a staged approach. The approach draws on a number of schemes that have been developed by various national and international bodies. However, it introduces a number of additional elements that are considered to benefit both transparency and consistency.

Particular attention has been paid to ensuring that the format can be applied to a wide range of lines of evidence and assessment.

9. ABBREVIATIONS AND GLOSSARY

ABBREVIATIONS

ADME	Absorption, distribution, metabolism, excretion
AOPs	Adverse outcome pathways
CEBM	Centre of Evidence Based Medicine
CREED	Criteria for Reporting and Evaluating Exposure Datasets
ECDC	European Centre for Disease prevention and Control
ECETOC	European Centre for Ecotoxicology and Toxicology of Chemicals
ECHA	European Chemicals Agency
EFSA	European Food Safety Authority
EMA	European Medicines Agency
ERA	Ecological Risk Assessment
ESA	Ecological Society of America
FAO	Food and Agriculture Organization of the United Nations
GHS	Globally Harmonised System of Classification and Labelling of Chemicals
GLP	Good Laboratory Practices
GRADE	Grading of Recommendations Assessment, Development and Evaluation
IARC	International Agency for Research on Cancer
ILSI	International Life Science Institute Europe
ITS	Integrating Testing Strategy
JECFA	The Joint FAO/WHO Expert Committee on Food Additives
MeSH	Medical Subject Headings
MoA	Mode of Action
NAMs	New Approach Methodologies
NGRA	Next Generation Risk Assessment
NIEHS	National Institute of Environmental Health Sciences
NLM	US National library of medicine
OECD	The Organisation for Economic Co-operation and Development
PECs	Predicted environmental concentrations
PNEC	Predicted No Effect Concentration
QSAR	Quantitative Structure–Activity Relationship models
RCT	Randomised Clinical Trials
REACH	Registration, Evaluation, Authorisation, and Restriction of Chemicals

RWD	Real World Data
RWE	Real World Evidence
RWS	Real World Studies
SSD	Species Sensitivity Distribution
TFEEA	OECD Task Force on Environmental Exposure Assessment
TGD	European Technical Guidance Document on risk assessment
ToR	Terms of Reference
WHO	World Health Organisation
WMO	World Meteorological Organization
WoE	Weight of Evidence
WWF	World Wildlife Fund

GLOSSARY

Term	Definition
Confidence	A measure of certainty or assurance in a finding
Consistency	The degree of agreement in the findings of multiple studies or lines of evidence
Line of evidence	Set of evidence of similar type. EFSA (2017)
NAMs	New Approach Methodologies can be defined as any in vitro, in chemico or computational (in silico) method that when used alone, or in concert with others, enables improved chemical safety assessment through more protective and/or relevant models and as a result, contributes to the replacement of animals.
Quality	Quality is the combined result of the judgement on relevance, reliability, and validity.
Relevance	Relevance/potential importance. This defines whether a set of data (e.g. from a publication) is appropriate for a particular hazard identification or risk characterisation. Klimisch *et al.* (1997); Nendza *et al.* (2010)
Reliability	Evaluating an individual result with regard to the inherent quality of a test report or publication relating to a, preferably standardised, methodology and the way that the experimental procedure and results are described to give evidence of the clarity and plausibility of the findings. Klimisch *et al.* (1997); Nendza *et al.* (2010); ECHA (2010)
Validity	Evaluating the method used for the generation of data for a specific endpoint relative to accepted guidelines. Or: Evaluating the model used for the generation of data against validation principles such as the OECD validation principles. Klimisch *et al.* (1997); Nendza *et al.* (2010); ECHA (2010)
Uncertainty	All types of limitations in the knowledge available, including imperfect or lacking knowledge

10. REFERENCES

Balshem H., Helfand M. and Schünemann H.J. (2011). GRADE guidelines: 3. Rating the quality of evidence". J Clin Epidemiol. 2011, 64 (4): 401–06)

Boobis A.R., Cohen S.M., Dellarco V., McGregor D., Meek M.E., Vickers C., Willcocks D. and Farland W. (2006). IPCS framework for analysing the relevance of a cancer mode of action for humans Crit Rev Toxicol 36 781-792

Boobis A.R., Doe J.E., Heinrich-Hirsch B., Meek M.E., Munn S., Ruchirawat M., Schlatter J., Seed J. and Vickers C. (2008). IPCS framework for analysing the relevance of a noncancer mode of action for humans Crit Rev Toxicol 38, 87-96

Brooke B.S., Schwartz T.A. and Pawlik T.M. (2021). MOOSE Reporting Guidelines for Meta-analyses of Observational Studies. *JAMA Surg.* 2021;156(8): 787–788. doi: 10.1001/jamasurg.2021.0522

CIEM (2017) Guidelines for ecological report writing
https://cieem.net/wp-content/uploads/2019/02/Ecological-Report-Writing-Dec2017.pdf

ECHA (2008). Guidance on information requirements and chemical safety assessment Chapter R.10: Characterisation of dose [concentration]-response for environment. Helsinki, Finland. www.echa.europa.eu

ECHA (2010). Practical guide 2: How to report weight of evidence. Helsinki, Finland. ISBN-13: 978-92-9217-028-8.

ECHA (2017). Weight of Evidence/Uncertainty. Helsinki, Finland. Available at: https://echa.europa.eu/support/guidance-on-reach-and-clp-implementation/formats)

EFSA (2017). Guidance on the use of the weight of evidence approach in scientific assessments. EFSA Journal 2017;15(8):4971 [69 pp.], DOI: 10.2903/j.efsa.2017.4971

EFSA (2018). The principles and methods behind EFSA's Guidance on Uncertainty Analysis in Scientific Assessment. EFSA Journal 2018. https://doi.org/10.2903/j.efsa.2018.5122

EFSA (2010). Systematic review methodology and food and feed safety risk assessment. EFSA Journal 2010; 8(6):1637, DOI: 10.2903/j.efsa.2010.1637

EFSA (2015). Principles and process for dealing with data and evidence. EFSA Journal 2015;13(5):4121 [36 pp.], DOI: 10.2903/j.efsa.2015.4121

EFSA (2023). EFSA Scientific Committee (SC): Guidance on protocol development for EFSA generic scientific assessments. e08312 First Published: 30 October 2023. http://onlinelibrary.wiley.com/doi/10.2903/sp.efsa.2022.EN-7349/full

Falakdin P., Terzaghi E., Di Guardo A., 2022. Spatially resolved environmental fate models: a review. Chemosphere 133394. https://doi.org/10.1016/j.chemosphere.2021.133394

Hardy A., Dorne JLCM, Aiassa E., Alexander J., Bottex B., Chaudhry Q., Germini A., Nørrung B., Schlatter J., Verloo D., Robinson T. (2015). Editorial: Increasing robustness, transparency and openness of scientific assessments. EFSA Journal 2015;13(3):e13031. 3 pp. doi:10.2903/j.efsa.2015.e13031

Hladik, M.L., Markus, A., Helsel, D., Nowell, L.H., Polesello, S., Rüdel, H., Szabo, D. and Wilson, I., 2024. Evaluating the reliability of environmental concentration data to characterize exposure in environmental risk assessments. Integrated Environmental Assessment and Management.

IPCS (2008). Uncertainty and data quality in exposure assessment. Part 2: Hallmarks of data quality in chemical exposure assessment. WHO/IPCS, Geneva, Switzerland. IPCS Harmonization Project Document No. 6. www.who.int/ipcs

IPCS (2009) Risk assessment of combined exposures to multiple chemicals: a WHO/IPCS framework. WHO/IPCS harmonization draft document.

IPCS (2014). Guidance document on evaluating and expressing uncertainty in hazard characterization. ISBN 978 92 4 150761 5

Klimisch H.J., Andreae M. and Tilmann U. (1997). A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data Regul Tox and Pharmacl 25 1-5

Levin R., Jansson S.O. and Rudén C. (2004). Indicators of uncertainty in chemical risk assessments. Regul Toxicol Pharm 39:33-43

Makady A., de Boer A., Hillege H., Klungel O. and Goettsch W (on behalf of GetReal Work Package 1) (2017). What Is Real-World Data? A Review of Definitions Based on Literature and Stakeholder Interviews. Value Health. Jul-Aug;20(7):858-865. doi: 10.1016/j.jval.2017.03.008

Meek M.E., Boobis A., Cote I., Dellarco V., Fotakis G., Munn S. and Vickers C. (2014). New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. J. Appl. Toxicol. 2014; 34: 1–18. doi: 10.1002/jat.2949. Epub 2013 Oct 25

Merrington, G., Nowell, L.H. and Peck, C. (2024). An introduction to Criteria for Reporting and Evaluating Exposure Datasets (CREED) for use in environmental assessments. Integrated Environmental Assessment and Management.

Moermond C.T.A., Kase R., Korkaric M. and Ågerstrand M. (2016). CRED: Criteria for reporting and evaluating ecotoxicity data. Environmental Toxicology and Chemistry. Doi: 10.1002/etc.3259 https://doi.org/10.1002/etc.3259

National Health and Medical Research Council. Assessing risk of bias. Available from: https://www.nhmrc.gov.au/guidelinesforguidelines/develop/assessing-risk-bias

Nendza M., Aldenberg T., Benfenati E., Begnini R., Cronin M.T.D., Escher S., Fernandez A., Gabbert S., Giralt F., Hewitt M., Hrovat M., Jeram S., Kroese D., Madden J.C., Mangelsdorf I., Rallo R., Roncaglioni A., Rorije E., Segner H., Simon-Hettich B. and Vermeire T. (2010). Data quality assessment for in silico methods: a survey of approaches and needs. Chapter 4 in: Cronin, M.T.D., Madden, J.C. (eds.) In silico toxicology, principles and applications. RSC Publishing, Cambridge, UK. ISBN 978-1-84973-004-4

OHAT (2015). Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. P 1-94

OECD (2007). Guidance document on the validation of (Q)SARs. Organisation of Economic Cooperation and Development, Paris, France. Series of Testing and Assessment

OECD (2018). Guidance Document on Good In Vitro Method Practices (GIVIMP), OECD Series on Testing and Assessment, No. 286, OECD Publishing, Paris, https://doi.org/10.1787/9789264304796-en.

Peters, A., Beking, M., Oste, L., Hamer, M., Vuaille, J., Harford, A.J., Backhaus, T., Lofts, S., Svendsen, C. and Peck, C., 2024. Assessing the relevance of environmental exposure data sets. Integrated Environmental Assessment and Management.

Posthuma L., Suter G.W.II, Traas T.P. (2002). Species sensitivity distributions in ecotoxicology. Lewis Publishers, Boca Raton, FL

SCHEER (2018). Memorandum on weight of evidence and uncertainties.

Roth N, Zilliacus J, Beronius A. (2021). Development of the SciRAP approach for evaluating the reliability and relevance of in vitro toxicity data. Front Toxicol. 3:746430.

Rudén, C., Backhaus, T., Bergman, P., Faust, M., Molander, L., Slunge, D., 2019. Future chemical risk management - Accounting for combination effects and assessing chemicals in groups (Swedish Government Official Reports No. SOU 2019:45).

SCENIHR (Scientific Committee on Emerging and Newly Identified Health Risks) (2012) Memorandum on the use of the scientific literature for human health risk assessment purposes – weighing of evidence and expression of uncertainty.

Schneider K, Schwarz M, Burkholder I, Kopp-Schneider A, Edler L, Kinsner-Ovaskainen A, *et al*. (2009). "ToxRTool", a new tool to assess the reliability of toxicological data. Toxicol Lett. 2009;189(2):138–44. https://doi.org/10.1016/j.toxlet.2009.05.013

Sherman R.E., Anderson S.A., Dal Pan G.J., Gray G.W., Gross T., Hunter N.L., LaVange L., Marinac-Dabic D., Marks P.W., Robb M.A., Shuren J., Temple R., Woodcock J., Yue L.Q. and Califf R.M. (2016). Real-World Evidence - What Is It and What Can It Tell Us? N Engl J Med. 2016 Dec 8;375(23):2293-2297

Svendsen C, Whaley P, Vist GE, Husøy T, Beronius A, Consiglio ED, Druwe I, Hartung T, Hatzi VI, Hoffmann S, Hooijmans CR, Machera K, Robinson JF, Roggen E, Rooney AA, Roth N, Spilioti E, Spyropoulou A, Tcheremenskaia O, Testai E, Vinken M, Mathisen GH. Protocol for designing INVITES-IN, a tool for assessing the internal validity of in vitro studies. Evid Based Toxicol. 2023 Aug 31;1(1):1-15. doi: 10.1080/2833373x.2023.2232415.

Van der Sluijs J.P., Risbey J.S., Kloprogge P., Ravetz J.R., Funtowicz S.O., Quintana S.C., Pereira A.G., De Marchi B., Petersen A., Janssen P.H.M., Hoppe R. and Huijs S.W.F. (2003). RIVM/MNP guidance for uncertainty assessment and communication. Detailed Guidance. Utrecht University, Utrecht, The Netherlands, ISBN 90-393-3536-2

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

Van Straalen N.M., den Haan K.H., Hermens J.L.M., van Leeuwen K., van de Meent D., Parsons J.R., de Voogt P., de Zwart, D. (2022). Risk assessment acknowledging variability in both exposure and effect. Environ. Sci. Technol. 56, 14223-14224.

Wang S.V., Pinheiro S., Hua W., Arlett P., Uyama Y., Berlin J.A., Bartels D.B., Kahler K.H., Bessette L.G. and Schneeweiss S. (2021). STaRT-RWE: structured template for planning and reporting on the implementation of real-world evidence studies. BMJ. 2021 Jan 12;372:m4856. doi: 10.1136/bmj.m4856

WHO/IPCS (2014). Guidance document on evaluating and expressing uncertainty in hazard characterization. IPCS harmonization project document; no. 11. World Health Organisation, International Programme on Chemical Safety. ISBN 978 92 4 150761 5

ANNEX

Evaluation of statistical analysis methods in a line of evidence

The statistical analysis section of each research publication gathered as a source of information should be carefully evaluated before the reported findings are considered in an Opinion.

Evaluation should include appropriateness and completeness of the statistical methods followed. Methods used for statistical modelling should be clearly described and appropriate for the nature of the data investigated, together with the necessary testing of any assumptions. Descriptive analytics should present in a clear way the data collected; diagnostic analytics should be used to explore inherent limitations and associations within data, data analytics methods for examining data sets to find trends and draw conclusions about the information they contain, should be transparently presented, whereas predictive analytics through robust methodologies should be used to forecast data series.

Key questions concern the sample size, and so where possible, statistical power analysis should be presented in a clear way, since it is possible that an observed effect or a lack of such an effect could be due to chance. This is a particular problem, especially in human studies with small, inadequate sample sizes or low prevalence of the exposure factors, but it may occur in animal and environmental studies, too. The presence or absence of statistical significance alone should not guide inclusion or exclusion of a study from the Opinion.

Presence of a statistically significant association does not alone constitute sufficient evidence for causality, nor of a practically relevant effect. Bias can produce a spurious association and can also mask existing associations. The effect size of the association, its related statistical uncertainty (e.g., confidence intervals of the effect estimates) and the internal consistency of the results should also be evaluated. The number of statistical tests performed and the reliance on methods correcting for multiple comparisons should also be considered when appropriate.

The effect size measures should be clearly justified, together with their levels of uncertainty, i.e., confidence intervals. In case of meta-analysis, the methodological approach used to derive the combined effect estimates, i.e., fixed or random or mixed effects modelling, should be clearly justified and documented together with measures of heterogeneity. Publication bias should be ascertained in systematic reviews and meta-analyses and considered for the generalization of the results.

Other important characteristics that are taken into consideration are the characteristics of the reference population, the types of controls that have been used in randomised controlled trials, any randomisation procedures and blinding to assure comparability of information, and the degree to which replication studies have been performed.

For ecological studies, unlike human studies, there have been no internationally agreed reporting or statistical methods guidelines. There are criteria which are widely used concerning the reliability and validity of studies (e.g. Klimisch and CRED). This gap in reporting guidelines is being addressed and a number of reports and publications have appeared in recent years. Some of these are briefly mentioned below.

For ecological studies, CIEM (2017) recommend that description of each survey method used is provided in sufficient detail to allow others to validate or repeat the survey. This includes Definition of the study area for each survey; Description of constraints/limitations on the methodology (e.g. time, resources, lack of access, sub-optimal season); A clear statement of any assumptions that have been made; A description of statistical techniques and confidence limits that have been used.

Popovic *et al.* (2024) describe three principles for improved statistical ecology: 1) develop a model that accounts for the distribution and dependence of your data; 2) emphasise effect sizes to replace statistical significance with ecological relevance; and 3) report your methods and findings in sufficient detail so that your research is valid and reproducible.

Popovic *et al.* (2024) argue that a well-defined research question allows researchers to create an efficient study to answer it, and guards against poor research practices that lead to poor estimation of the direction, magnitude, and uncertainty of ecological relationships, and to poor replicability.

ESA (2022) emphasises that the specific statistical procedure must always be described in plain language or using equations. In addition, key components of the computing environment should be cited, and version numbers should be included. The authors should publish their code for analyses. All statistical tests that were performed should be clearly described to avoid problems associated with p-hacking, data dredging, etc., and corrections for multiple testing should be employed when appropriate. Randomization procedures should be clearly described, including the number of randomizations employed.

EFSA (2014) similarly gives guidance on statistical reporting. They summarise the approach as follows:

"The general and specific objectives of the statistical analysis should be stated with scientific background explaining the rationale for the analysis. The sources of information (data) used for the analysis and data quality assurance measures should be reported. These could be pre-existing sources or data specifically collected. The data sources will be dependent on some underlying study design and all measures taken to minimise bias and maximise precision should be detailed. This, together with approaches used to address sample selection, sample size, power, blinding (where relevant) and randomisation (where relevant) should be detailed. Statistical analysis, including data processing (e.g. transformation of data), details of the methodology (e.g. assumptions, models used) and the software used, have to be documented. Deviations and/or noncompliance issues, planned or unplanned, in relation to the a priori protocol (if any)/statistical plan should be described. The reporting of the results should be consistent with the objectives of the study. Descriptive statistics should be presented for relevant data collected for analysis. The point and interval estimates (e.g. confidence) for all results of the statistical analysis should be presented. A statistical interpretation of results to support the biological/scientific interpretation should be given including a discussion about all relevant uncertainties affecting the statistical analysis and its results."

ESA (2022) also provide a set of recommendations on the reporting of statistical analysis. They provide the following guidance:

"The specific statistical procedure must always be described in plain language or using equations (i.e., just referring to a function within a specific statistical program is not sufficient), with the goal that a knowledgeable reader could reproduce the results when provided with the raw data. In addition, key components of the computing environment should be cited, and version numbers should be included. For example, analysis in R using several packages should cite each package based on its requested citation format. Where appropriate, authors should indicate which procedure within a package was used and which methods or options within a procedure were chosen. Relatively novel statistical procedures need to be explained in sufficient detail, including references if appropriate, for the reader to reconstruct the analysis. The authors should publish their code for analyses. All statistical tests that were performed should be clearly described to avoid problems associated with p-hacking, data dredging, etc., and corrections for multiple testing should be employed when appropriate."

References for Annex

CIEM (2017) Guidelines for ecological report writing
https://cieem.net/wp-content/uploads/2019/02/Ecological-Report-Writing-Dec2017.pdf

ESA (2022) Statistical analysis guidelines
https://www.esa.org/wp-content/uploads/2022/05/ESA-Statistical-Analysis-Guidelines.pdf

EFSA (2014) Guidance on statistical reporting
https://www.efsa.europa.eu/en/efsajournal/pub/3908

Popovic, G., Mason, T. J., Drobniak, S. M., Marques, T. A., Potts, J., Joo, R., Altwegg, R., Burns, C. C. I., McCarthy, M. A., Johnston, A., Nakagawa, S., McMillan, L., Devarajan, K., Taggart, P. L., Wunderlich, A., Mair, M. M., Martínez-Lanfranco, J. A., Lagisz, M., & Pottier, P. (2024). Four principles for improved statistical ecology. *Methods in Ecology and Evolution*, 15, 266–281. https://doi.org/10.1111/2041-210X.14270