SUMMARY OF THE SCENIHR WEIGHT OF EVIDENCE METHODOLOGY

The aim is to develop guidelines on risk assessment of stressors in particular chemicals) to which humans may be exposed. This report is intended to complement the on going activities on the expression of uncertainty.

Currently the assessment of data relies on expert judgement and, although this approach is well established, how the expert judgement is used and its consistency is not clear to many stakeholders.

The report covers:

- the identification and selection of relevant publications for analysis,
- weighing the data, and
- Its application for risk assessment purposes.

The aim is to use it, wherever appropriate, for risk assessment activities

It is intended to apply to both human health and environmental risk assessments.

A risk assessment requires the evaluation of the evidence across all relevant domains/lines of evidence. It is proposed that the acceptability of each publication considered to be relevant should be assessed.

Selection of papers for consideration

It is important to identify how the data for consideration was found and the criteria used. This may involve just the title or the abstract as well, whether only full papers were selected and any other criteria were included.

Individual papers /data sets

For individual papers and/or data sets the quality and relevance need to be assessed independently using the following criteria:

Quality

- *Good Scientific quality.* Study is considered to be appropriately designed, conducted and reported, and using valid methodology.
- *Adequate/utilizable scientific quality but with significant limitations*. Scientifically acceptable but some important deficiencies in the design and/or conduct and /or the reporting of the experimental findings
- *Inadequate scientific quality*. Serious concerns about the design or conduct of the study
- *Not assignable*. Insufficient detail to make an evaluation

Relevance

- *Direct relevance,* i.e. addressing the agent (stressor), model and outcome of interest
- *Indirect relevance*, i.e. addressing a related agent (stressor), model or outcome of interest
- *Insufficient relevance*

Table 1 Matrix to assess individual publications

	Good Scientific quality	**Adequate/utilizable scientific quality**	**Inadequate scientific quality**	**Not assignable**
Direct relevance	X	X		
Indirect relevance	X	X		
Insufficient relevance				

Typically, the data search methods used will identify many papers that could be used. A preliminary screening is then needed in order to focus on those relevant for the specific purposes of the development of the opinion. Papers that are identified initially but on preliminary examination do not meet the criteria of quality and/or relevance for the purposes of the development of the opinion in the reference list or additional document for the report on which the opinion is based as:

- ''Publications noted but not considered suitable for the purposes of developing the opinion'.

Individual lines of evidence

The next stage is the assessment of the weight of evidence for each line of evidence.

For most risk assessments a number of lines of evidence need to be considered these may involve several or all of the following:

- Exposure- sampling and analysis
- toxicokinetics
- animal studies/data in environmentally relevant species
- in vitro studies
- mathematical modelling
- mechanistic/mode of action studies.
- epidemiology studies
- human volunteer studies
- other human data
- Studies in selected environmental species

Table 2 Matrix to weigh individual lines of evidence (scoring indicated by crosses)

		Consistency		
		High	***Medium***	***low***
Utility	***high***	X	X	
	medium	X	X	
	low			

Key issues in the assessment of individual lines of evidence are consistency of findings and their utility.

Studies should be classified into those that:

- indicate the presence of an effect
- indicate the absence of an effect
- are consistent with either the presence or absence of an effect.

Two criteria are used, consistency and utility.

Consistency is defined as the agreement on the outcome between different studies for each line of evidence. The following categories may be identified:

- high – most studies show findings in the same direction;
- medium – the majority of studies show findings either in the same direction or are consistent with either outcome;
- low – little agreement between studies.

Utility is defined as the usability for the purposes of developing the risk assessment. A matrix is proposed that integrates to consideration of both relevance and quality. The following categories may be identified:

- high overall relevance
- moderate overall relevance
- low overall relevance

This analysis leads to the assignment of individual papers to one of the following categories

- 'Publications that are relevant and of sufficient/suitable quality and were important for the development of the opinion'
- 'Publications that are relevant and of sufficient/suitable quality but were not judged to be necessary for the development of the opinion'

Stage 4. Integration of all lines of evidence

Integration of the various lines of evidence (Integrative risk assessment) is the final stage It involves several steps, details of which are set out in the text:

- Describe the nature of the data, including endpoints considered
- Evaluation of exposure. Combining modelling and monitoring data

- Evaluation of hazard data. Combining in vivo, in vitro and in silico data also combining animal\and human data. This stage also includes dose-response and internal exposure modelling and extrapolation based on the critical study or studies.
- Mode(s) of action. The plausibility of the observed or hypothetical mode(s) of action and its validity for extrapolation purposes particularly between species.
- Quantifying the risks (including the statistical analysis) using the hazard and exposure data. Overall impact on man and on the environment**.**

At each stage, a narrative justification should be provided for the final conclusions. It should highlight possible knowledge gaps and other uncertainties.

Dimensions of risk that may need to be expressed include the severity of the effect/outcome (nature of the adverse effect) and the likelihood of its occurrence. Both of these aspects should be addressed in depth in risk characterisation.

The weighing of the total evidence should be presented for the purposes of clarity and consistency in a standard format. A tabulated form is proposed. Although all lines of evidence are considered for human risk assessment human, animal and mechanistic studies comprise the primary line of evidence along with exposure. The result of the tabulation and its analysis should be expressed as follows:

Strong overall weight of evidence: Coherent evidence from human and one or more other lines of evidence (animal or mechanistic studies) in the absence of conflicting evidence from one of the other lines of evidence (no important data gaps)

Moderate overall weight of evidence: good evidence from a primary line of evidence but evidence from several other lines is missing (important data gaps)

Weak overall weight of evidence: weak or conflicting evidence from the primary lines of evidence (severe data gaps)

In each case free text is required to explain the assignment.

In addition to three categories, based on available data, it might be concluded that there is a lack of data to scientifically weigh the evidence, which might be due to either a general lack of studies or mostly studies are available that were classified as being inadequate for the risk assessment

Table 2 Contribution of the different lines of evidence to the opinion

Factor	Strong	Moderate	Weak
A. Weight of evidence from the following lines of evidence: Exposure measurement Exposure modelling Epidemiologic studies Human volunteer studies Other human data sources Animal studies In vitro studies Mathematical models, structure activity and other in silico data Studies on Mechanisms			
Conclusion from the totality of evidence (short description)			
B. Comprehensiveness of evidence base (i.e. absence of critical knowledge gaps)			
C. Uncertainty (cross-reference to activities on uncertainty)			
Overall evaluation (including short description)			

January 21st 2011.