SPECIAL ISSUE

EFSA Journal

doi: 10.2903/j.efsa.2016.s0511

# Weighing evidence and assessing uncertainties

**Jean Lou C. M. Dorne[1], Bernard Bottex[1], Caroline Merten[1], Andrea Germini[1], Nikolaos Georgiadis[1], Elisa Aiassa[1], Laura Martino[1], Lorenz Rhomberg[2], Harvey J. Clewell[3], Matthias Greiner[4], Glenn W. Suter[5], Maurice Whelan[6], Andrew D. M. Hart[7], Derek Knight[8], Prabhat Agarwal[9], Maged Younes[10], Jan Alexander[11] and Anthony R. Hardy[1]**

## Abstract

Methodologies for integrating (weighing) evidence and assessing uncertainties are of utmost importance to ensure that scientific assessments are transparent, robust and fit for purpose to support decision-makers. One of the key challenges remains the development of harmonised methodologies for both weighing scientific evidence and assessing uncertainties in the food safety area mainly because of the multidisciplinary and complex nature of the topics involved. The breakout session 'Weighing evidence and assessing uncertainties' was held at the EFSA 2nd Scientific Conference 'Shaping the Future of Food Safety, Together'. This paper aims at summarising the contributions of this breakout session and formulates recommendations to further support the development of harmonised methodologies and practical applications for weighing evidence and analysing uncertainty in key areas of food safety, including chemical risk assessment, microbiological risk assessment and environmental risk assessment.



**Keywords:** weight of evidence, uncertainty, human risk assessment, ecological risk assessment, microbiological risk assessment, biological relevance, problem formulation

**Correspondence:** editor-in-chief.efsajournal@efsa.europa.eu

**Author affiliations:** [1] European Food Safety Authority (EFSA), Parma, Italy; [2] Gradient, Cambridge, MA, USA; [3] ScitoVation, Research Triangle Park, Durham, NC, USA; [4] BfR, Berlin, Germany; [5] US-EPA, Office and Research and Development, Cincinnati, OH, USA; [6] Systems Toxicology Unit, European Commission Joint Research Centre, Ispra, Italy; [7] FERA, York, UK; [8] Executive Director Office, European Chemicals Agency, Helsinki, Finland; [9] Directorate-General for Communication Networks, Content and Technology Knowledge Sharing, Brussels, Belgium; [10] Mies, Switzerland; [11] Institute of Public Health, Oslo, Norway.

**Acknowledgements:** The European Food Safety Authority (EFSA) thanks the participants of the breakout session 'weighing evidence and assessing uncertainties' held at the EFSA 2nd Scientific Conference 'Shaping the Future of Food Safety, Together' (Milan, Italy, 14–16 October 2015) for their active and valuable contribution to the discussion.

**Disclaimer:** The views expressed in this publication are those from the authors and do not necessarily represent the official views or policies of EFSA or the US-EPA. EFSA assumes no responsibility or liability for any errors or inaccuracies that may appear.

**Suggested citation:** Dorne JLCM, Bottex B, Merten C, Germini A, Georgiadis N, Aiassa E, Martino L, Rhomberg L, Clewell HJ, Greiner M, Suter GW, Whelan M, Hart ADM, Knight D, Agarwal P, Younes M, Alexander J and Hardy AR, 2016. Special issue: weighing evidence and assessing uncertainties. EFSA Journal 2016;14(S1):s0511, 13 pp. doi:10.2903/j.efsa.2016.s0511

**ISSN:** 1831-4732



The EFSA Journal is a publication of the European Food Safety Authority, an agency of the European Union.

# Table of contents

# 1. Introduction

Within the risk analysis framework, risk assessment has the critical role of setting health standards to protect humans, animals and the environment from hazards of a physical, chemical or biological nature. The framing of the scientific question during 'problem formulation' is of utmost importance. It ensures that the issues of decision-makers and stakeholders issues are agreed to and clearly expressed prior to the start of the assessment. This ensures that the risk assessment is fit for purpose in relation to the question and delivered in a timely manner to support risk managers. The steps of risk assessment, namely hazard identification, hazard characterisation, exposure assessment and risk characterisation, are harmonised throughout the world, whether the hazard is of physical, chemical or biological origin. The methods, processes and results of the risk assessment should be communicated in a transparent, open and unambiguous manner to support decision-making and inform the public. In this context, assessing, weighing and integrating scientific evidence and characterising uncertainties are critical components of each step of the risk assessment process. The European Food Safety Authority (EFSA) has recently initiated cross-cutting activities including the development of guidance documents by the EFSA Scientific Committee: (1) characterisation of uncertainties in scientific assessments; (2) use of the weight-of-evidence approach in scientific assessments; and (3) identification of biological relevance of adverse/positive health effects from experimental animal and human studies. In addition, the Prometheus initiative 'promoting methods for evidence use in scientific assessments' supports the coordination and consistency of EFSA projects that aim to develop or refine methodological approaches (EFSA, 2015a–c).

The breakout session 'Weighing evidence and assessing uncertainties' was held at the EFSA 2nd Scientific Conference 'Shaping the Future of Food Safety, Together' (Milan, Italy, 14–16 October 2015)[1] and included presentations on key areas relevant to food safety:

1) 'Weighing evidence and assessing uncertainties: moving forward' set the scene of the session providing global methodological perspectives;
2) 'Weighing evidence of biological relevance: from empirical testing in rats to 21st Century mode of action analysis' illustrated the World Health Organization (WHO) mode of action framework to address weight of evidence and biological relevance in chemical risk assessment;
3) 'Weighing evidence and assessing uncertainty in microbiological risk assessment: approaches for preparing appropriate scientific support for decision-making in complex questions?' discussed applications in microbiological risk assessment;
4) 'Uncertainty, variability and weight of evidence: how well do we know environmental risks?' gave a global perspective in the area of environmental risk assessment of chemicals;
5) 'Coming to grips with unfamiliar uncertainties of a new predictive toxicology paradigm' discussed the contribution of new testing methods for predictive toxicology and related uncertainties to the future of chemical risk assessment;
6) 'Assessing and communicating uncertainties for risk assessment and risk management: recent international developments' brought an overview on international developments in uncertainty analysis for risk assessment and risk management.

Following these presentations, the audience was given the opportunity to discuss and ask questions in a final panel discussion and the chairs concluded on the session. This paper aims to summarise the contribution of each speaker during this session and formulate recommendations to further support the development of harmonised methodologies and practical applications in key areas of food safety.

# 2. Weighing evidence and assessing uncertainties in scientific assessments: moving forward

We are facing many challenges in the process of weighing scientific evidence regarding chemical toxicity. In applying weight of evidence in risk assessment, a regulatory process needs to follow a set of established procedures. Several procedures for weighing evidence were reviewed at the meeting and questions concerning their adequacy were addressed. In addition, strategies for structuring weight-of-evidence enquiry were discussed. Finally, some approaches that may achieve the twin aims

[1] The scientific programme of the conference is available at http://www.efsaexpo2015.eu/programme/. All the conference material of the breakout session (including briefing notes, presentations and videos) is available at http://www.efsaexpo2015.eu/show-session/?idsession=10

of flexibility in the face of diverse scientific evidence and sufficient structure to ensure that consistency, rigour and justification of conclusions can be documented were compared. There are inevitable limits to what the available data can directly demonstrate concerning exposures in foods, the environment and the workplace (Weed, 2005; Linkov et al., 2009; Rhomberg et al., 2013). These methods range from qualitative methods including listing evidence and best professional judgement; to methods dealing with causal criteria, logic, scoring and indexing; and to full quantitative methods including meta-analysis (Higgins et al., 2009; Linkov et al., 2009).

Scientific judgements about the existence and nature of causal processes of toxicity need to be made while contending with the data gaps, extrapolations, inconsistencies and shortcomings in the available studies. There is need to characterise not only what conclusions can reasonably be drawn but also the degree of confidence in them, noting different interpretations that might also be considered. In pure science, the 'scientific method', comprising an iterative process of hypothesising general explanations and seeking critical tests of them in further experiments, is pursued, with continued scepticism towards and testing of tentative conclusions. In the regulatory context, decisions to take or forgo actions must be made, and the making of judgements about whether the interpretation of evidence is sufficiently robust to support such decisions is delegated to a limited set of assessors who must defend their legitimacy to stakeholders and the public. To ensure consistency in standards of evidence to support conclusions, and to communicate the judgement process and its justifications, a variety of risk assessment frameworks (i.e. procedures for gathering, interpreting and drawing conclusions from available evidence) have been put in place and used by various governmental and international organisations (Rhomberg et al., 2013).

In recent years, the sufficiency of some of these evidence-evaluation frameworks, and their ability to make sound, well-justified and well-communicated judgements, has been questioned. This stems in part from a deeper understanding of underlying modes of toxic action and their diversity among different experimental animal strains and humans, and at different exposure levels. This knowledge exposes the limits of earlier assumptions about toxicity processes being parallel in test systems and humans. In part, it is due to an increasing number of examples in which existing evaluation frameworks seem to miss important scientific considerations that have been revealed by deeper probing of underlying biology. New kinds of test data, in particular, high-throughput *in vitro* testing and gene-expression arrays, have opened new avenues for characterising toxicity pathways and alternative apical endpoints and challenge traditional methods (EFSA, 2014; Benfenati et al., 2016).

Critiques by high-level review panels of several key regulatory assessments have found insufficient explanation of the basis for weight-of-evidence judgements. The advent of evidence-based medicine as a means for evaluating the clinical efficacy of alternative treatments has provided a model of how a more systematic and rigorous process might provide better and more objective justifications for judgements. In consequence, a great deal of recent attention has focused on how the weight-of-evidence process can and should be reformed, and some reform activities have been undertaken by regulatory and scientific bodies at the national and international levels (EFSA, 2015a,c).

Progress has been made in instituting systematic review processes for identifying relevant studies, objectively evaluating strengths and weaknesses, making inclusion/exclusion decisions and tabulating results (EFSA, 2010; Rhomberg et al., 2013; Aiassa et al., 2015). In reviewing some of these and, while noting the benefits, one could also argue that this by itself goes only so far in resolving the challenges. The relevance of studies, interpretations of their interactions and their support for overarching hypotheses about the bases for possible toxicity still need to be considered. A systematic process to do so that weighs diverse evidence is challenging to define.

Insights into the challenges and means to address them can be gained by examining the differing strategies that have been employed in constructing evaluation frameworks. One strategy, a rules-based or algorithmic approach, aims to build a decision-tree process that embodies the interpretive wisdom of the field, such that each decision can be made objectively, and conclusions are justified by how the decision-tree process disposes of the data at hand. The advantage is objectivity, but the shortcoming is that the interpretive wisdom needs to be built into the algorithm, which may be faulty, become out of date or be unable to accommodate novel kinds of evidence. An alternative strategy is to be more unstructured but to rely on expert judgement from a set of appropriately chosen scientists who then explain the basis for their judgements. The advantage is flexibility and, possibly, extra scientific insight, but the shortcoming is that the choice of experts becomes controversial, the justifications are articulated after the fact and are keyed to judgements already made, and the process can lack transparency. The conclusions are justified by asserting the expertise of the judges. A process analogous to evidence-based medicine can be rigorous and transparent, but it does not easily deal

with evidence that is not direct observation of the question of interest itself; that is, it emphasises consistency of repeated observations, but does not handle inference across data sets very well.

Hypothesis-based weight of evidence seeks to gain the advantages of others while avoiding the disadvantages. It stresses constructing competing sets of tentative explanations for all of the relevant study outcomes, where explanations invoking a common causal toxicity process can be compared for plausibility and dependence on assumptions with an alternative set of possible explanations that denies the tested agent's toxicity and explains outcomes by alternative means, such as chance, confounding and variable operation of non-agent-related causes in different test systems (Rhomberg, 2015). Hypothesis-based weight of evidence has been applied to a number of compounds and toxicological endpoints, including naphthalene carcinogenesis and neurodevelopmental effects of chlorpyrifos (Prueitt et al., 2011; Bailey et al., 2016).

## 3. Moving from empirical testing to mechanistic thinking: mode of action analysis, weight of evidence and biological relevance

In the early days of chemical risk assessment, the focus was essentially qualitative: describing pathological changes observed after the exposure of laboratory animals to relatively high doses of a chemical. Despite the accumulation of a large volume of animal data, there has been growing scepticism regarding its usefulness due to the perceived difficulties of interpreting the relevance of the animal results for humans (WHO/FAO 2009b) . A concept that has proved useful for weighing evidence on the toxicity of a chemical is the 'mode of action'. The International Programme on Chemical Safety (IPCS) (Sonich-Mullin et al., 2001; Meek et al., 2014) guidelines provide a discussion of the desired elements of a mode of action and a description of the kinds of data that can inform its development, using a conceptual framework for mode-of-action evaluation. The IPCS mode-of-action (MoA) evaluation framework is an extension of the considerations of causation originally presented by Bradford Hill, which has been recently modified to include the evaluation of experimental animal data and aid in the interpretation of epidemiological data (Meek et al., 2014). Weighing the evidence for the likely human relevance of an animal outcome is particularly problematic and has frequently been a source of controversy. In order to promote transparent, harmonised approaches for such evaluations, the IPCS extended its MoA framework to address consideration of human relevance for both cancer and noncancer effects observed in animal studies (Boobis et al., 2006, 2008). In 2007, the US National Research Council report on 'Toxicity testing in the 21st Century: a vision and a strategy' argued for a transformative shift away from *in vivo* animal toxicity testing and towards the use of mechanistic *in vitro* assays, typically using human cells in a high-throughput context. The shift to *in vitro* tests for assessing risks of chemicals entails new questions about weighing evidence: (1) how will we define adversity from *in vitro* tests; (2) how will the *in vitro* test results be used to predict expected outcomes in animals and people; and (3) how will regulatory agencies set exposure standards for human populations based on *in vitro* test results. These questions pose the challenges to the development of a 21st-century toxicology that both collects toxicity testing information and weighs evidence for human health risk assessment (Thomas et al., 2013; EFSA, 2014).

## 4. Weighing evidence and assessing uncertainty in microbiological risk assessment: the future approaches for preparing appropriate scientific support for decision-making in complex questions?

The complexity of questions in food safety is addressed in formal exposure or risk assessments by deconstructing the problem into elements at different levels, typically involving a risk question, a scenario, a model, model parameters and data to support parameter estimation. Specific uncertainties at these levels need to be assessed to derive an understanding of the overall uncertainty of the science-based risk assessment. On the other hand, the concept of 'evidence' is mainly applicable to support the choice of a scenario and to characterise the empirical knowledge about key model parameters.

The weight-of-evidence concept has been promoted in several areas in microbiological food safety, including inferences about causal associations (WHO/FAO, 2003). It has been proposed to derive weights from sample sizes, expert beliefs and uncertainty (WHO/FAO, 2008). The WHO has suggested that weight of evidence will become increasingly important in risk assessments of microbiological pathogens in food (WHO/FAO, 2009a). A weight-of-evidence approach has also been proposed for problem formulation, intervention measure and outcome evaluation to evaluate food safety interventions

(e.g. Fazil et al., 2008). Similarly, aspects related to population, agent, vehicle, source and adverse effect have been evaluated using weight of evidence in support of foodborne outbreak management (Health Canada, 2011; Vik et al., 2014). In microbiological risk assessment, weight of evidence has also been used for expressing the empirical support for the choice of a dose–response function and its parameters (Moon et al., 2005). Applications of weight of evidence are also reported in microbiological water management (Olivieri et al., 2014) and risk ranking (EFSA BIOHAZ Panel, 2015). Systematic review methodology (EFSA, 2010; Aiassa et al., 2015) and meta-analysis (Borenstein et al., 2009) are applicable in the area of microbiological food safety and provide powerful tools for exploring and accounting for parameter heterogeneity, which is highly relevant in situations of primary studies with conflicting results. However, meta-analysis requires that parameter estimates from primary studies are comparable in their metric and relate to the same type of study and research question. In contrast, evidence synthesis and evidence integration aim at collating and combining, respectively, information from primary studies with different types of observations (e.g. *in vivo*, *in vitro*, *in silico*, epidemiological) and different study organisms (e.g. human and animal), which are also referred to as lines or streams of evidence. The two methodologies, weighing of evidence and uncertainty assessment, are complementary and have a common overall goal, which is to provide the best possible basis for science-based decision-making.

Finally, aspects of internal validity and relevance for the question (external validity) may be part of an uncertainty assessment and at the same time can be used to derive weights in a weight-of-evidence approach. Therefore, weight of evidence and uncertainty assessment are complementary and have a common overall goal, which is to provide a robust basis for science-based decision-making. In this context, relevant recommendations were formulated during the special session of this issue on microbiological risk assessment (Cassini et al., 2015) and included the need to: (1) determine the uncertainties in the risk estimates from risk ranking studies; (2) effectively communicate these uncertainties to decision-makers; and (3) establish risk assessment, risk management and risk communication responsibilities for national institutions and government in the case of outbreaks of foodborne disease and foster collaboration across institutions.

## 5. Environmental risk assessment: uncertainty, variability and weight of evidence

Risk assessment is technical support for decision-making under uncertainty and so, without uncertainty, there is no inherent risk. However, what decision-makers need from assessors is not uncertainty *per se*, but rather, relevant risk information and an expression of the confidence that they should place in that information. Confidence in an assessment result has two components: scatter and weight.

The first is the conventional statistical issues of uncertainty and variability in the measurements or analytical results which are both expressed as scatter of the data (less scatter, more confidence). Estimating and expressing scatter can be difficult but it is a well-studied problem and the conceptual issues are relatively well recognised (Bayesian subjectivism vs Frequentist objectivism, statistical testing vs estimation, etc.). When multiple studies estimate the same variable (e.g. multiple estimates of a pesticide's half-life in water), combined estimates and associated scatter can be estimated by meta-analysis (Borenstein et al., 2009). If, as is common practice, the estimates are weighted, meta-analysis constitutes a genuinely quantitative weighing of evidence.

The second component of confidence is weight. The weight-of-evidence metaphor is borrowed from jurisprudence (evidence that has more influence on a decision is weightier). Weight of evidence is qualitative and, although data analysis contributes, it is ultimately a matter of judgement. Constituents of weight of evidence are relevance, strength and reliability. Relevance is the degree to which evidence represents the issue and situation being assessed. Strength expresses the degree to which the evidence is distinguished from random variance, commonly expressed as correlation. Reliability expresses factors such as study quality and transparency of presentation that make evidence more convincing. The best method for weighing evidence is a matter of controversy in the environmental assessment community. Potential methods range from conventional academic narrative reviews of the evidence, through to more formal qualitative methods, and finally to quasi-quantitative methods (Weed, 2005; Linkov et al., 2009; Suter and Cormier, 2011).

Qualitative assessment results such as causation (e.g. carbofuran has caused bird kills) or condition (e.g. the stream is biologically impaired) have weight but not scatter. However, quantitative information results, such as a lethal threshold concentration or a cancer slope factor, should be

accompanied by expressions having both scatter and weight and both are essential information. For example, we may have little uncertainty expressed as scatter (e.g. a small standard deviation) but, if the weight for either relevance or reliability of the evidence is low, we have low confidence in the risk estimate. We must learn to convey the overall appropriate level of confidence in our results in a useful form to decision-makers, stakeholders and the public.

A parallel session in the EFSA EXPO conference addressed 'advancing environmental risk assessment of regulated products under EFSA's remit' and formulated a number of recommendations that are relevant to weighting evidence and assessing uncertainties in an ERA context: (1) use the ecosystem services approach to make protection goals operational for regulated products under EFSA's remit; (2) rely on problem formulation to improve the relevance of ERA studies; (3) comply with quality standards to warrant the reliability of ERA studies; (4) make ERA more contextual by accounting for multiple stressors and environmental benefits; and (5) acknowledge strengths and limitations of environmental monitoring as a tool to resolve scientific uncertainties post-market (Devos et al., 2016).

## 6. Bringing new ways to assess uncertainties within the predictive toxicology paradigm

Key objectives underpinning future food safety policies will never be achievable without a clear paradigm shift in the way that we profile the toxicological properties of chemicals. Regulatory toxicity assessment relies for the most part on laboratory animal tests developed many decades ago. Over the last decade, many research programmes, such as the TOX-21 in the USA and SEURAT in Europe, have examined the application of new methodologies and tools using *in vivo*, *in vitro* and *in silico* approaches to investigate toxicokinetic and toxicodynamic processes of chemicals at the organism, organ, cellular and molecular levels in the light of mode of action and adverse outcome pathways (AOPs) (NRC, 2007; EFSA, 2014). These methodologies and tools address three major scientific, ethical and legislative needs: (1) achieving a mechanistic understanding of toxicity for hazard assessment (e.g. mode of action/AOPs); (2) reducing animal testing under the 3Rs principles (reduce, replace, refine) due to the ban on testing chemical ingredients in animals as under the EU Cosmetics regulation (EU/1223/2009); and (3) assessing thousands of chemicals (particularly under the REACH (Registration, Evaluation, Authorisation and Restriction of Chemicals) regulation) (SCHER, SCENHIR, SCCS, 2012; EFSA, 2014). Recent reviews that discuss such options include the joint report of the three non-food committees of the European Commission 'New challenges in Risk Assessment', and the report of the US-EPA on 'Next Generation (NexGen) Risk Assessment: Incorporation of Recent Advances in Molecular, Computational, and Systems Biology' (SCHER, SCENIHR, and SCCS, 2013; US-EPA, 2013; Goodman et al., 2014).

Importantly too, predictive approaches reveal and exploit mechanistic understanding of why a chemical might be toxic to an organism under certain conditions, opening the door to more tailored, substance-specific approaches to hazard assessment. Embracing this new toxicity testing paradigm offers considerable benefits to society including improved assessment of more chemicals, simplified cross-sector legislation based on harmonised assessment approaches, cheaper and faster testing for industry, higher levels of protection for sensitive populations, and incorporation of safety-by-design and green chemistry practices in product lifecycles (Thomas et al., 2013). What exactly a new paradigm will look like and what implications it will have for both risk assessment and risk management remains to be seen, but it is clear that the tipping point is behind us. Change will need to be managed at many levels, and understanding unfamiliar 'non-standard' sources of uncertainty will be an essential step in the process of responsibly yet definitively transitioning to new ways of describing and predicting chemical toxicity. Recently, the focus of current and future initiatives investigating AOPs has been reviewed to increase the acceptance of non-animal approaches and bring about full mechanistic approaches in risk assessment. The key areas where current and future initiatives should be focused will enable the translation of AOPs into routine chemical safety assessment and provide lasting 3Rs benefits (European Commission, 2014; Burden et al., 2015).

These new developments were also discussed in a dedicated breakout session of this special issue 'novel chemical hazard characterisation approaches', which highlighted the concept of mechanistic validation as a way forward to quality-assure new cell-based tests and to incorporate integrated assessment and testing approaches as a means of combining multiple lines of evidence (Hartung et al., 2013; Benfenati et al., 2016).

## 7. Uncertainty analysis in risk assessment and risk management: future needs in methodological development and training

The need to address uncertainties in food safety risk assessment has long been internationally recognised. The Codex Working Principles for Risk Analysis, established in 2003, state that uncertainties should be explicitly considered at each step in risk assessment, documented transparently and quantified to an extent that is scientifically achievable. EFSA Guidance on Transparency, adopted in 2009, states the same. There has been gradual progress towards implementing these principles. Guidance on methods for addressing uncertainty in human exposure assessment were published by EFSA in 2006 and IPCS/WHO in 2008 (EFSA, 2006; IPCS, 2008), whereas the ECHA (2012) guidance also included hazard, risk and environmental assessments (ECHA, 2012). IPCS/WHO published guidance on uncertainty in hazard characterisation, accompanied by a spread sheet calculator (WHO, 2014), and EFSA published, for public consultation, a draft guidance on addressing uncertainty in all areas of its work (EFSA, 2016).

The draft EFSA Guidance (EFSA, 2016) encourages assessors to be systematic in identifying sources of uncertainty, checking each part of their assessment to minimise the risk of overlooking important uncertainties. It also indicates that uncertainties may be expressed qualitatively or quantitatively and that it is not often necessary or possible to quantify separately every individual source of uncertainty. However, the guidance highlights the importance for assessors to express, in quantitative terms, the combined effect of as many as possible of the identified sources of uncertainty, both for transparency and to provide meaningful support to the decision-making process. The draft EFSA guidance offers a framework that is scalable to the needs of each assessment, enabling the assessor to select from a menu of qualitative and quantitative methods while taking account of any limitations in time and resources, including emergency situations. It also recognises that progress is more feasible by evolution than revolution, focusing first on addressing uncertainties within current assessment paradigms. Revising those paradigms to address uncertainty more fully is challenging, as illustrated by the 2014 IPCS/WHO guidance on hazard characterisation, and will require concerted action over a longer period (WHO, 2014).

Implementation of new guidance will require training and support for risk assessors, including provision of user-friendly tools and specialist help with more sophisticated methods. Risk managers will also need training and support in meeting their responsibility, emphasised by Codex, for resolving the impact of uncertainty on decision-making. In addition, risk assessors and managers will need to work together when communicating with stakeholders about how uncertainty has been addressed.

## 8. Conclusions and recommendations

A key challenge in risk assessment is the development of harmonised and robust methodologies for weighing scientific evidence and assessing uncertainties that are applicable, adequate and transparent, and will provide a fit for purpose and timely support to decision-makers. In the food safety area, such challenges arise mainly because of the multidisciplinary nature of the topics, which include microbiology, animal health and welfare, epidemiology, toxicology, ecology, plant health, genetics, nutrition, bioinformatics and statistics. In other words, one size may not fit all and these methods need to achieve both flexibility in the light of diverse scientific evidence, and sufficient structure to ensure consistency, rigour and justification of conclusions. These challenges have been addressed in this session 'weighing evidence and assessing uncertainties' and international developments in critical areas of food safety (i.e. chemical risk assessment including mechanistic and animal-free risk assessment methods and biological and environmental risk assessment) have been reviewed and discussed. In principle, the methods for weighing evidence and assessing uncertainties underlying a risk assessment range from simple qualitative description of the evidence and the uncertainties in a data-poor situation to full probabilistic models integrating uncertainty and variability of the biological processes underpinning the hazard, exposure and risk questions.

Generally speaking, this session has highlighted the need for further development of such qualitative and quantitative methods for weighing evidence and assessing uncertainties using case studies. A key recommendation in this area is the need to further develop approaches and case studies to weigh different lines of evidence, particularly across different levels of biological organisation (e.g. molecular, cellular, organism, population). Ideally, such case studies and approaches would take into account the practical needs of risk assessors and decision-makers, namely problem formulation and the level of knowledge available for the hazard to be assessed, as well as the resources and time

available for the assessment. Two extreme examples in the food safety area include an urgent assessment during a food crisis (time: days to weeks), which may be based on scarce data and/or a previous assessment and expert knowledge, and a full risk assessment based on the combination of a full systematic review of the literature, which may include probabilistic meta-analysis and expert knowledge (time: months to 1 year).

Some of these general recommendations are currently being addressed by the guidance development activities of the EFSA Scientific Committee on weight of evidence and uncertainty analysis (EFSA, 2015a,c, 2016). In addition, a guidance document on biological relevance is also under development and deals with definitions, concepts and criteria to address adverse effects, homeostasis, biological thresholds, nature and size of biological changes, and the relevance of test species. In chemical risk assessment, addressing the biological basis of a toxicological effect is the starting point of hazard assessment (identification and characterisation) prior to a consideration of the evidence available and statistical issues dealing with variability and uncertainty. Hence, the coherent integration of biological relevance with the results of weighing evidence and assessing uncertainty provides the basis of a harmonised framework to further support holistic approaches for risk assessment and decision-making. In food safety, case studies are needed to apply these integrated approaches and to address their applicability in a sound, feasible, timely and fit for purpose manner. Specific recommendations for the chemical, microbiological and environmental areas are discussed below.

The beginning of the 21st century has seen the emergence of new methods for chemical hazard characterisation and tools such as 'omics', systems biology and computational tools (i.e. *in silico* tools). These new methods generate a vast amount of data and evidence (Big Data) that scientists are struggling to integrate into the current risk assessment paradigm. Over the last decade, the MoA and AOP frameworks have provided approaches to bring biological relevance, weight of evidence and uncertainty analysis for the integration of mechanistic data into the risk assessment process. These approaches have contributed very significantly to the 3Rs shifting empirical testing to integrated testing strategies, and progressively reducing animal testing to move towards the mechanistic prediction of toxicity. These include the use of predictive computer models (e.g. quantitative structure activity relationships, physiologically based models, etc.), *in vitro* cell systems and 'omics' technologies. The need for further guidance and the development of case studies applied to these unfamiliar 'non-standard' sources of evidence, uncertainty and variability are recommended for different regulatory applications (e.g. prioritisation under REACH, food safety, etc.).

In the microbiological area, weight of evidence and uncertainty analysis have been applied to risk assessments of microbiological pathogens in food, including microbiological water management, risk ranking and foodborne outbreak management to cite but a few. For all of these examples, further methodological developments applied to microbiological hazards for human health or animal health are recommended, particularly for data-poor/data-rich situations and the development of predictive tools. Examples include the development of weight-of-evidence methods and uncertainty analysis for the integration of wide genome sequence data or the use of systematic review, meta-analysis and modelling approaches based on known pathogens to predict health outcome in 'data-poor emerging pathogens'.

In environmental risk assessment, an example of a key challenge for weighing evidence and assessing uncertainty and variability is the integration of evidence and information from both the laboratory and field across several levels of biological and spatial organisation and across numerous taxa.

As methods and case studies for weighing evidence, assessing uncertainty and biological relevance develop, international cooperation and training for risk assessors and risk managers are crucial, particularly for the development of best practice across disciplines, scientific advisory bodies and regulatory areas. Finally, communication of such a complex and multidisciplinary science has become an integral part of the process and is becoming more and more critical so that risk assessors can provide a reproducible and transparent picture of the assessment to decision-makers, relevant stakeholders and the public.

## References

Aiassa E, Higgins JP, Frampton GK, Greiner M, Afonso A, Amzal B, Deeks J, Dorne JL, Glanville J, Lövei GL and Nienstedt K, 2015. Applicability and feasibility of systematic review for performing evidence-based risk assessment in food and feed safety. Critical Reviews in Food Science and Nutrition, 55, 1026–1034. doi: 10.1080/10408398.2013.769933

Bailey LA, Nascarella MA, Kerper LE and Rhomberg LR, 2016. Hypothesis-based weight-of-evidence evaluation and risk assessment for naphthalene carcinogenesis. Critical Reviews in Toxicology, 46, 1–42. doi: 10.3109/10408444.2015.1061477

Benfenati E, Berggren E, Fritsche E, Hartung T, William Slikker W Jr, Spielmann H, Testai E, Tice RR, Tiramani M and Villenave R, 2016. Novel chemical hazard characterisation approaches. In: EFSA (European Food Safety Authority), 2015. Shaping the Future of Food Safety, Together: Proceedings of the 2nd EFSA Scientific Conference. Milan, Italy, 14–16 October 2015. EFSA Journal 2015;13(10):s1310. 38–41 pp. doi: 10.2903/j.efsa.2015.s1310

Boobis AR, Cohen SM, Dellarco V, McGregor D, Meek ME, Vickers C, Willcocks D and Farland W, 2006. IPCS framework for analyzing the relevance of a cancer mode of action for humans. Critical Reviews in Toxicology, 36, 781–792. doi: 10.1080/10408440600977677

Boobis AR, Doe JE, Heinrich-Hirsch B, Meek ME, Munn S, Ruchirawat M, Schlatter J, Seed J and Vickers C, 2008. IPCS framework for analyzing the relevance of a noncancer mode of action for humans. Critical Reviews in Toxicology, 38, 87–96. doi: 10.1080/10408440701749421

Borenstein M, Hedges LV, Higgins JPT and Rothstein HR, 2009. *Introduction to meta-analysis*. John Wiley & Sons, Hoboken, New Jersey. 452 pp.

Burden N, Sewell F, Andersen ME, Boobis A, Chipman JK, Cronin MT, Hutchinson TH, Kimber I and Whelan M, 2015. Adverse outcome pathways can drive non-animal approaches for safety assessment. Journal of Applied Toxicology, 35, 971–975. doi: 10.1002/jat.3165

Cassini A, Hathaway S, Havelaar A, Koopmans M, Koutsoumanis K, Messens W, Müller-Seitz G, Nørrung B, Rizzi V and Scheutz F, 2015. Microbiological risk assessment. In: EFSA (European Food Safety Authority), 2015. Shaping the Future of Food Safety, Together: Proceedings of the 2nd EFSA Scientific Conference. Milan, Italy, 14–16 October 2015. EFSA Journal 2015;13(10):s1310. 38–41 pp. doi:10.2903/j.efsa.2015.s1310

Devos Y, Gaugitsch H, Gray AJ , Maltby L, Martin J, Pettis JS, Romeis J, Rortais A, Schoonjans R, Smith J, Streissl F and Suter GW II , 2016. Special issue: advancing environmental risk assessment of regulated products under EFSA's remit. EFSA Journal 2016;14(S1):s0508, 13 pp. doi:10.2903/j.efsa.2016.s0508

ECHA (European Chemicals Agency), 2012. Guidance on information requirements and chemical safety assessment. Chapter R.19: Uncertainty analysis. ECHA, Helsinki. 36 pp. Available online: http://echa.europa.eu/documents/10162/13632/information_requirements_r19_en.pdf

EFSA (European Food Safety Authority), 2006. Guidance of the Scientific Committee on a request from EFSA related to Uncertainties in Dietary Exposure Assessment. EFSA Journal 2006;4(1):438, 1–54. doi:10.2903/j.efsa.2007.438

EFSA (European Food Safety Authority), 2010. Application of systematic review methodology to food and feed safety assessments to support decision making. EFSA Journal 2010;8(6):1637, 90 pp. doi:10.2903/j.efsa.2010.1637

EFSA (European Food Safety Authority), 2014. Modern methodologies and tools for human hazard assessment of chemicals. EFSA Journal 2014;12(4):3638, 87 pp. doi:10.2903/j.efsa.2014.3638

EFSA (European Food Safety Authority), 2015a. Editorial: increasing robustness, transparency and openness of scientific assessments. EFSA Journal 2015;13(3):e13031, 3 pp. doi:10.2903/j.efsa.2015.e13031

EFSA (European Food Safety Authority), 2015b. Scientific report on Principles and process for dealing with data and evidence in scientific assessments. EFSA Journal 2015;13(5):4121, 35 pp. doi:10.2903/j.efsa.2015.4121

EFSA (European Food Safety Authority), 2015c. Outcome of the targeted consultation of the EFSA Journal editorial on increasing openness, robustness and transparency of scientific assessments. EFSA supporting publication 2015:EN-785, 26 pp.

EFSA (European Food Safety Authority), 2016. Draft Guidance document on Uncertainty in Scientific Assessment. Public consultation version. Available online: http://www.efsa.europa.eu/sites/default/?files/consultation/150618.pdf

EFSA BIOHAZ Panel, 2015. Scientific Opinion on the development of a risk ranking toolbox for the EFSA BIOHAZ Panel. EFSA Journal 2015;13:3939, 131 pp. doi:10.2903/j.efsa.2015.3939

European Commission, 2014. EURL ECVAM Status Report on the Development, Validation and Regulatory Acceptance of Alternative Methods and Approaches. Luxembourg: Publications Office of the European Union, 84 pp. doi: 10.2788/86047

Fazil A, Rajic A, Sanchez J and McEwen S, 2008. Choices, choices: the application of multi-criteria decision analysis to a food safety decision-making problem. Journal of Food Protection, 71, 2323–2333.

Goodman JE, Boyce CP, Pizzurro DM and Rhomberg LR, 2014. Strengthening the foundation of next generation risk assessment. Regulatory Toxicology and Pharmacology, 68, 160–170. doi: 10.1016/j.yrtph.2013.12.002

Hartung T, Hoffmann S and Stephens M, 2013. Mechanistic validation. ALTEX, 30, 119–130. doi: 10.14573/altex.2013.2.119

Health Canada, 2011, online. Weight of evidence: factors to consider for appropriate and timely action in a foodborne illness outbreak investigation. Available online: http://www.hc-sc.gc.ca/fn-an/pubs/securit/2011-food-illness-outbreak-eclosion-malad-ailments/index-eng.php [Accessed: 10 October 2015]

Higgins JP, Thompson SG and Spiegelhalter DJ, 2009. A re-evaluation of random-effects meta-analysis. Journal of the Royal Statistical Society. Series A, (Statistics in Society), 172, 137–159. doi: 10.1111/j.1467-985X.2008.00552.x

IPCS (International Programme on Chemical Safety), 2008. Guidance Document on Characterizing and Communicating Uncertainty in Exposure Assessment. IPCS Harmonization Project Document No. 6, Part 1. World Health Organization, Geneva, 138 pp. Available online: http://www.who.int/ipcs/methods/harmonization/areas/uncertainty%20.pdf

IPCS (International Programme on Chemical Safety), 2014. Guidance Document on Evaluating and Expressing Uncertainty in Hazard Assessment. IPCS Harmonization Project Document No. 11. World Health Organization, Geneva, 158 pp. Available online: http://www.who.int/?ipcs/methods/harmonization/uncertainty_in_hazard_characterization.pdf

Linkov I, Long EB, Cormier SM, Satterstrom FK and Bridges T, 2009. Weight-of-evidence evaluation in environmental assessment: review of qualitative and quantitative approaches. Science of The Total Environment, 497, 5199–5205. doi: 10.1016/j.scitotenv.2009.05.004

Meek ME, Boobis A, Cote I, Dellarco V, Fotakis G, Munn S, Seed J and Vickers C, 2014. New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. Journal of Applied Toxicology, 34, 1–18. doi: 10.1002/jat.2949

Moon H, Kim HJ, Chen JJ and Kodell RL, 2005. Model averaging using the Kullback information criterion in estimating effective doses for microbial infection and illness. Risk Analysis, 25, 1147–1159. doi: 10.1111/j.1539-6924.2005.00676.x

NRC (National Research Council), 2007. Toxicity testing in the 21st Century: a vision and a strategy. The National Academies Press, Washington, 216 pp. doi: 10.17226/11970

Olivieri AW, Seto E, Cooper RC, Cahn MD, Colford J, Crook J, Debroux JF, Mandrell R, Suslow T, Tchobanoglous G and Hultquist RA, 2014. Risk-based review of California's water-recycling criteria for agricultural irrigation. Journal of Environmental Engineering, 140, 04014015. doi: 10.1061/(ASCE)EE.1943-7870.0000833

Prueitt RL, Goodman JE, Bailey LA and Rhomberg LR, 2011. Hypothesis-based weight-of-evidence evaluation of the neurodevelopmental effects of chlorpyrifos. Critical Reviews in Toxicology, 41, 822–903. doi: 10.3109/10408444.2011.616877

Rhomberg L, 2015. Hypothesis-Based Weight of Evidence: an Approach to Assessing Causation and its Application to Regulatory Toxicology. Risk Analysis, 35, 1114–1124. doi: 10.1111/risa.12206

Rhomberg LR, Goodman JE, Bailey LA, Prueitt RL, Beck NB, Bevan C, Honeycutt M, Kaminski NE, Paoli G, Pottenger LH, Scherer RW, Wise KC and Becker RA, 2013. A survey of frameworks for best practices in weight-of-evidence analyses. Critical Reviews in Toxicology, 43, 753–784. doi: 10.3109/10408444.2013.832727

SCHER, SCENIHR, and SCCS (Scientific Committee on Health and Environmental Risks), (Scientific Committee on Emerging and Newly Identified Health Risks) (Scientific Committee on Consumer Safety), 2013. Addressing the New Challenges for Risk Assessment, March 2013.

Sonich-Mullin C, Fielder R, Wiltse J, Baetcke K, Dempsey J, Fenner-Crisp P, Grant D, Hartley M, Knaap A, Kroese D, Mangelsdorf I, Meek E, Rice J and Younes M, 2001. IPCS conceptual framework for evaluating a mode of action for chemical carcinogenesis. Regulatory Toxicology and Pharmacology, 34, 146–152.

Suter GW II and Cormier SM, 2011. Why and how to combine evidence in environmental assessments: weighing evidence and building cases. Science of the Total Environment, 409, 1406–1417. doi: 10.1016/j.scitotenv.2010.12.029

Thomas RS, Philbert MA, Auerbach SS, Wetmore BA, Devito MJ, Cote I, Rowlands JC, Whelan MP, Hays SM, Andersen ME, Meek ME, Reiter LW, Lambert JC, Clewell HJ, Stephens ML, Zhao QJ, Wesselkamper SC, Flowers L, Carney EW, Pastoor TP, Petersen DD, Yauk CL and Nong A, 2013. Incorporating new technologies into toxicity testing and risk assessment: moving from 21st century vision to a data-driven framework. Toxicological Sciences, 136, 4–18. doi: 10.1093/toxsci/kft178

US-EPA (U.S. Environmental Protection Agency), 2013. Next Generation Risk Assessment: Incorporation of Recent Advances in Molecular, Computational, and Systems Biology. Available online: http://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=259936#Download

Vik J, Hexemer A and Farber J, 2014. Summary: weight of evidence-factors to consider when investigating a food-borne illness outbreak. Canada Communicable Disease Report, 40, 303 pp.

Weed DL, 2005. Weight of evidence: a review of concepts and methods. Risk Analysis, 25, 1545–1557. doi: 10.1111/j.1539-6924.2005.00699.x

WHO (World Health Organization), 2014. Guidance document on evaluating and expressing uncertainty in hazard characterization. Harmonization Project Document 11. WHO, Geneva, 158 pp. available online: http://www.inchem.org/documents/harmproj/harmproj/harmproj11.pdf

WHO/FAO (World Health Organization) (Food and Agriculture Organization of the United Nations), 2003. Hazard characterization for pathogens in food and water: guidelines. Microbiological Risk Assessment Series, No. 3. WHO, Geneva, 76 pp. Available online: http://www.who.int/foodsafety/publications/pathogen/en/

WHO/FAO (World Health Organization) (Food and Agriculture Organization of the United Nations), 2008. Exposure assessment of microbiological hazards in food. Microbiological Risk Assessment Series, No. 7. WHO, Geneva, 102 pp. Available online: http://www.who.int/foodsafety/publications/micro/MRA7.pdf

WHO/FAO (World Health Organization) (Food and Agriculture Organization of the United Nations), 2009a. Risk Characterization of microbiological hazards in food. Microbiological Risk Assessment Series, No. 17 WHO, Geneva, 135 pp.

WHO/FAO (World Health Organization) (Food and Agriculture Organization of the United Nations), 2009b. Annex 1: Glossary of terms. In: Environmental Health Criteria 240: Principles and methods for the risk assessment of chemicals in food. WHO, Geneva, 45 pp. Available online: http://apps.who.int/iris/bitstream/10665/44065/13/WHO_EHC_240_13_eng_Annex1.pdf

## Abbreviations

| | |
|---|---|
| AOP | adverse outcome pathway |
| FAO | Food and Agriculture Organization |
| IPCS | International Programme on Chemical Safety |
| MoA | mode-of-action |
| REACH | Registration, Evaluation, Authorisation and Restriction of CHemicals |
| SCCS | Scientific Committee on Consumer Safety |
| SCENHIR | Scientific Committee on Emerging and Newly Identified Health Risks |
| SCHER | Scientific Committee on Health and Environmental Risks |
| US-EPA | United States Environmental Protection Agency |
| WHO | World Health Organization |