A Section 508–conformant HTML version of this article is available at https://doi.org/10.1289/EHP3067.

Weight of Evidence for Hazard Identification: A Critical Review of the Literature

Pierre Martin,[1,2] Claire Bladier,[3] Bette Meek,[4] Olivier Bruyere,[5] Eve Feinblatt,[3] Mathilde Touvier,[6] Laurence Watier,[7] and David Makowski[8]

[1]French Agricultural Research Centre for International Development (CIRAD), Agroecology and sustainable intensification of annual crops (UPR AIDA), Montpellier, France
[2]AIDA, CIRAD, Montpellier University, Montpellier, France
[3]French Agency for Food, Environmental and Occupational Health and Safety (ANSES), Maisons-Alfort, France
[4]McLaughlin Center for Risk Science, University of Ottawa, Ottawa, Canada
[5]WHO Collaborating Center for Public Health Aspects of Musculo-Skeletal Health and Aging, Department of Public Health, Epidemiology, and Health Economics, University of Liège, Liège, Belgium
[6]Nutritional Epidemiology Research Team (EREN), Center of Research in Epidemiology and Statistics, Sorbonne Paris Cité (CRESS), Institute for Health and Medical Research (INSERM, U1153), French National Institute of Research for Agriculture (INRA, U1125), National Conservatory of Arts and Crafts (CNAM), Paris University, Bobigny, France
[7]Biostatistics, Biomathematics, Pharmacoepidemiology and Infectious Diseases (B2PHI), INSERM, UVSQ, Pasteur Institute, University of Paris-Saclay, Paris, France
[8]UMR Agronomy, INRA, AgroParisTech, University of Paris-Saclay, Thiverval-Grignon, France

Background: Transparency when documenting and assessing weight of evidence (WOE) has been an area of increasing focus for national and international health agencies.

Objective: The objective of this work was to conduct a critical review of WOE analysis methods as a basis for developing a practical framework for considering and assessing WOE in hazard identification in areas of application at the French Agency for Food, Environmental and Occupational Health and Safety (ANSES).

Methods: Based on a review of the literature and directed requests to 63 international and national agencies, 116 relevant articles and guidance documents were selected. The WOE approaches were assessed based on three aspects: the extent of their prescriptive nature, their purpose-specific relevance, and their ease of implementation.

Results: Twenty-four approaches meeting the specified criteria were identified from selected reviewed documents. Most approaches satisfied one or two of the assessed considerations, but not all three. The approaches were grouped within a practical framework comprising the following four stages: (1) planning the assessment, including scoping, formulating the question, and developing the assessment method; (2) establishing lines of evidence (LOEs), including identifying and selecting studies, assessing their quality, and integrating with studies of similar type; (3) integrating the LOEs to evaluate WOE; and (4) presenting conclusions.

Discussion: Based on the review, considerations for selecting methods for a wide range of applications are proposed. Priority areas for further development are identified. https://doi.org/10.1289/EHP3067

Introduction

Risk assessment is usually characterized by four components: hazard identification, hazard characterization (including dose–response analysis), exposure assessment, and risk characterization. Identifying relevant hazards for subsequent consideration in dose–response analysis and risk characterization requires the assimilation and assessment of a wide range of different types of data (NRC 2014; OECD 2014; U.S. EPA 2014). Variations in conclusions drawn by different organizations on the potential of specific substances to cause hazards in such assessments have highlighted the need for greater consistency in the analysis of such data. Examples include variations among the conclusions of the European Food Safety Agency (EFSA), the U.S. Environmental Protection Agency (U.S. EPA 2017), and the International Agency for Research on Cancer (IARC) regarding the carcinogenicity of glyphosate (EFSA 2017) and among those of the EFSA, the ANSES, and the U.S. National Toxicology Program on the reproductive/developmental hazards of bisphenol A (ANSES 2015; U.S. NTP 2008). These variations have led to an increasing focus of national and international agencies on the robustness and transparency of expert-informed assessments (Hardy et al. 2015; OHAT 2015) as a basis for increasing the understanding and confidence of the relevant scientific community, stakeholders, and the public. Although the term "weight of evidence" (WOE) appears frequently in the scientific literature, it is often poorly and inconsistently defined, with limited documentation of the supporting expert-informed process and methodology (NRC 2014; Weed 2005).

For example, WOE has long been referenced in a range of disciplines, including the medical sector, where it was introduced principally as a clinical decision-support tool for prioritizing knowledge of medical research, focusing on a critical review of the literature (Sackett et al. 1996). WOE assessment has also been widely referenced and applied in environmental health (Mandrioli and Silbergeld 2016; Krimsky 2005). In various disciplines, approaches to the assessment of WOE have evolved beyond a review of the literature to include expert-informed reviews and the integration of different types of information in a transparent and systematic manner (e.g., meta-analysis). The reviews of Linkov et al. (2009) and Rhomberg et al. (2013) described a wide range of approaches, ranging from those that are largely qualitative in nature (e.g., Guyatt et al. 2011a) to fully quantitative techniques (e.g., Gosling et al. 2013). The inclusiveness and organization of different approaches vary, with some including references to establishing lines of evidence

These authors contributed equally to this work.

Address correspondence to C. Bladier, Agence nationale de sécurité sanitaire, Direction Évaluation des Risques, 14 rue Pierre et Marie Curie 94701, Maisons-Alfort Cedex, France. Telephone: (+33) 1 49 77 26 06. Email: claire.bladier@anses.fr

Supplemental Material is available online (https://doi.org/10.1289/EHP3067).

The authors declare that they have no actual or potential competing financial interests.

Received 7 November 2017; Revised 22 May 2018; Accepted 25 May 2018; Published 17 July 2018.

Note to readers with disabilities: *EHP* strives to ensure that all journal content is accessible to all readers. However, some figures and Supplemental Material published in *EHP* articles may not conform to 508 standards due to the complexity of the information being presented. If you need assistance accessing journal content, please contact ehponline@niehs.nih.gov. Our staff will work with you to access and meet your accessibility needs within 3 working days.

(i.e., groupings of evidence of similar types to assess a hypothesis) and integrating evidence of different types (e.g., toxicological, epidemiological, and mechanistic data). Others address only integrating different types of evidence without reference to the prerequisite stages, such as identifying and selecting relevant evidence and establishing "lines of evidence" (LOEs). A framework has been proposed here, then, to support the selection of WOE methodologies, depending on the objectives and focus of assessments.

The specific objective of this work was to propose harmonized approaches to assessing and communicating WOE in environmental, occupational, and food safety, as well as plant and animal health, for the French Agency for Food, Environmental and Occupational Health and Safety (ANSES). The review was limited to considering documented approaches to WOE assessment (interpreted here as the structured synthesis of evidence) and did not address issues related to the selection of experts and conflicts of interest.

The scope and basis of the current review are broader than those of earlier reviews by, for example, Rhomberg et al. (2013), which was confined to chemical hazards to human health. The review addresses approaches relevant to a wide range of applications within the purview of ANSES, including, for example, microbiological quality. It includes not only an extensive review of the literature through PubMed and Scopus but also a focused consultation of 63 public health and environmental agencies worldwide and characterizes identified approaches in component stages of the proposed practical framework for WOE analysis, relevant to this broader range of assessments. Each approach is also rated according to three criteria assessing their prescriptive nature, relevance, and feasibility for screening of their potential for application within ANSES and possibly within other food and environmental safety agencies.

Methods

Both peer-reviewed journal articles and guidance developed by health and environmental agencies were considered in the review. To limit the search to WOE assessment in risk analysis, the query of the review was composed of the combination of two sets of terms using the AND operator, with the first one related to WOE and the second one related to risk analysis (Figure 1). Two databases were queried on 16 March 2015 (i.e., Scopus and PubMed), and the title, summary, and keyword sections were searched. Papers were excluded if they were published before 2010 and after March 2015 and in languages other than English or French, as were case studies, editorials, or papers without identifiable content related to WOE approaches.

Sixty-three national and international agencies or organizations performing risk assessment were also consulted to identify relevant guidance (Table S1). Additional documents identified from the lists of references in the selected articles and relevant reports were also reviewed. Titles and abstracts were screened by at least two people. Descriptions of the approaches based on extraction and assessment of relevant information were completed for each selected article by individual authors within their area of expertise and were reviewed collectively by all the authors of this manuscript. Critical aspects included the domain and scope of the study, the definition of terms (e.g., WOE) and the approach and methodology for WOE assessment, including the nature and number of considerations taken into account for the stage or stages of assessment addressed by the approach (Figure S1).

For application in assessment planning at ANSES, these descriptions were also considered collectively by the authors to characterize and relatively rank the following aspects (Table 1):

- The "extent of their prescriptive nature" which contributes to transparency and reproducibility. This consideration addressed the degree of prescription of the factors assessed in considering the quality and subsequent weighting of studies and bodies of evidence and often derives from the extent of expert-informed experience in developing and applying the approach (i.e., approaches based on extensive application experience are often more prescriptive). Relative ranking was based on the extent to which considerations for implementation were precisely delineated and defined in the various approaches and ranged from "no explicit rules provided" (rank = 1) to "implementation rules defined in significant detail, facilitating use by non-experts" (rank = 4).
- "Relevance" was related to the extent to which the approaches could be broadly applied within the types of assessments conducted within ANSES. For example, were they specific to specialized components or aspects of WOE consideration (e.g., mechanistic data), or were they more broadly applicable to all aspects of assessments of hazard commonly conducted within ANSES? Rankings ranged from "the specificity of the methodology restricts its use to a relatively narrow application for which it was developed" (rank = 1) to "the methodology is sufficiently generic to be broadly applicable to most aspects of a broad range of assessments of hazard within ANSES" (rank = 4).

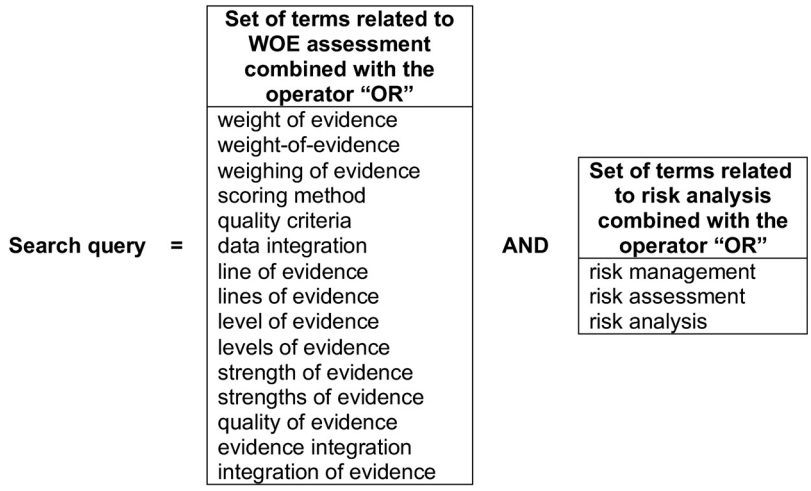

Figure 1. Keywords combined to produce the Scopus and PubMed search query.

Table 1. Ranking considerations for the prescriptive nature, relevance, and feasibility of WOE approaches for ANSES evaluations.

Consideration	Rank	Ranking
Prescriptive nature	1	No explicit rules
	2	Some methodological elements for assessment and weighting defined but insufficiently detailed for non-expert users
	3	Implementation rules are well defined for most aspects of the WOE assessment
	4	Implementation rules are defined in sufficient detail to permit application by non-expert users
Relevance	1	The specificity of the methodology restricts its use to specialized aspects or applications of WOE assessment for which it was developed
	2	The methodology can be applied for a limited range of aspects or applications in hazard assessment within ANSES
	3	The methodology is applicable to most aspects or applications of a broader range of assessments of hazard within ANSES
	4	The methodology is sufficiently generic to be applicable to most aspects of a broad range of assessments of hazard within ANSES
Feasibility	1	Implementation of the method is resource intensive (complexity high) and requires considerable specialized expertise and/or material resources
	2	Implementation of the method impacts moderately on resources (moderate complexity), requiring some specific training
	3	Implementation of the method impacts minimally on resources and does not require specialized training, expertise and/or material resources
	4	Implementation of the method not anticipated to impact significantly on timeframe and resources for assessment

- "Ease of implementation" (feasibility) in terms of time and material/human resources, including the requirement for specific and often advanced methodological skills (modeling, statistics, etc.). Relative ranking of the ease of implementation was based on the extent of complexity of the approach and the associated nature and extent of required experience, skills, time, and material resources for application. Scores ranged from "resource intensive, requiring considerable resources and expertise" (rank = 1) to "limited requirement for specialized expertise, material resources and/or time" (rank = 4).

Results

The study selection process is described in Figure 2. The study selection process is described in Figure 2, using PRISMA (Moher et al. 2009). In all, 643 articles identified in the Scopus and PubMed searches, 25 relevant reports from agencies, and 67 documents from the screening of the associated lists of references were retrieved. This corresponded to 663 documents after the removal of duplicates. We reviewed the titles and abstracts of the 663 documents and excluded 538 due to the lack of reference to WOE in the title or abstract or the lack of a description of a WOE approach. We reviewed the remaining 125 full-text articles for eligibility and excluded 9 because the reported data were not relevant to the objective of the paper. The remaining 116 documents formed the principal basis for the current review/analysis. Twenty-four relevant approaches were identified in the 116 selected documents (Table 2). A previous review of WOE frameworks was also identified in the selected documents, i.e., Rhomberg et al. (2013).

Each of the methods/frameworks cited in Table 2 has been applied in one or more fields (Table 3). A wide range of approaches has been adopted in environmental health, food safety and nutrition, and medical applications. The most commonly adopted approaches based on the numbers of examples of applications in different domains are IARC classifications, followed by Bradford Hill considerations in the assessment of

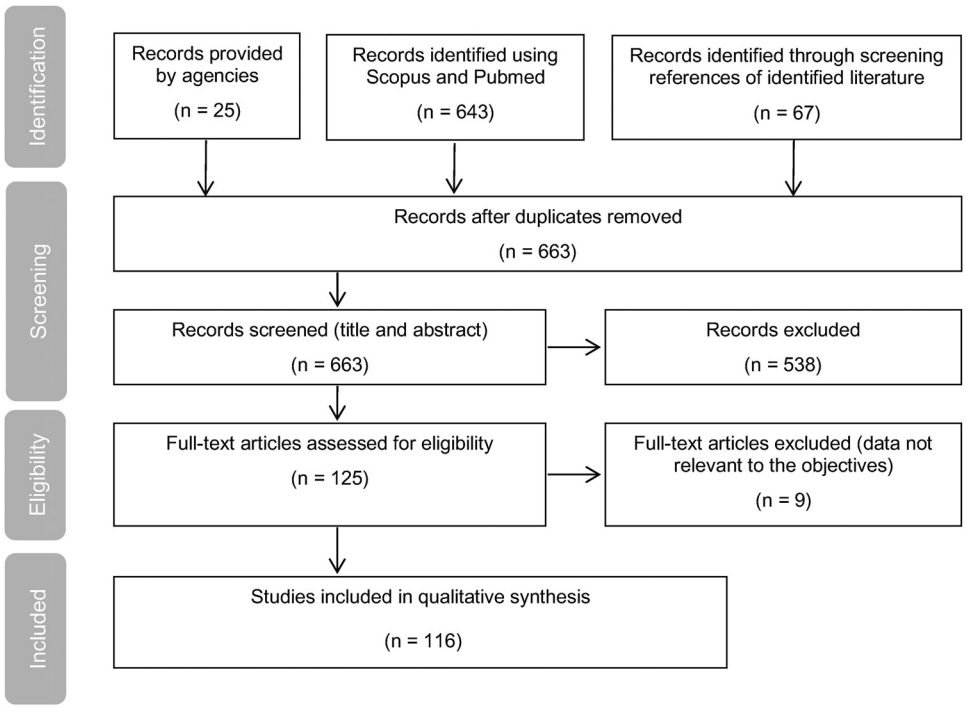

Figure 2. Flow diagram of the study selection process using Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA, Moher et al. 2009).

Table 2. WOE Approaches identified in the literature.

Name	Description	Category	SR included	Form of evaluation[a]	PF stages (step of stage)	Reference
AMSTAR	Assessment of syntheses of observational and clinical studies through the scoring (1–4) of 11 aspects	Method	Yes	Scoring	2 (2)	Kung et al. (2010); Pieper et al. (2015); Shea et al. (2007a, 2007b, 2009)
Bayesian inference	Statistical analysis combining expert knowledge (described by a prior probability distribution) with data to estimate a quantity of interest and analyze uncertainty	Method	No	Quantitative	3; 4	BioBayes Group (2015); Gosling et al. (2013); Guha et al. (2013); Schleier et al. (2015); Spiegelhalter et al. (2004); Williams et al. (2011)
Bradford Hill	Qualitative consideration of causality (9 aspects) in epidemiological studies	Method	No	Qualitative	2 (3); 3	ANSES (2012) Bergman et al. (2015); Guzelian et al. (2005); Hill (1965); Rothman and Greenland (2005); Vinken (2013)
Decision tree	Tool based on a tree-like graph describing options for various decision points	Method	No	Qualitative or quantitative	3	ANSES (2013a, 2013b); FAO/WHO (2001); Khosrovyan et al. (2015); Metcalfe (2005)
Epid-Tox	Grid based on a five-step process to evaluate the quality of epidemiological and toxicological studies, and their intersection, to establish causal inference	Method	No	Qualitative	2 (2); 2 (3); 3; 4	Adami et al. (2011); ECETOC (2009)
FDA	Qualitative evaluation of individual studies in humans and of the total scientific evidence based on study type, quantity of evidence, relevance to the target population, replication of study results and overall consistency	Method	No	Qualitative	2 (1); 2 (2); 2 (3)	FDA (2009)
GRADE	Assessment of methodological flaws within the component studies, the consistency of results across different studies, the generalizability of research results to the wider patient base, and the effectiveness of treatments	Method	Yes	Scoring	1; 2 (2); 2 (3); 4	Akl et al. (2007); Andrews et al. (2013a, 2013b); Balshem et al. (2011); Berkman et al. (2012); Guyatt et al. (2011a, 2011b, 2011c, 2011d, 2011e, 2011f, 2011g, 2011h, 2011i); HAS (2013); Kho and Brouwers (2012); WHO (2012)
Hope and Clarkson	Weighting and integration of information relating cause and effect to estimate the probability of an adverse outcome for an ecological assessment endpoint	Framework	No	Scoring	1; 2 (2); 2 (3); 3; 4	Hope and Clarkson (2014)
Hypothesis-based	Fully expert-dependent assessment for various hypotheses for hazard identification of chemical substances	Method	No	Qualitative	3; 4	Bailey et al. (2016); Rhomberg (2015)
IARC	Assessment of the quality of individual studies based on "principles of good practice" without reporting templates. Four categories for classification of combined evidence on toxicology and epidemiology and three for mode of action. Expert dependent	Method	No	Qualitative	2 (2); 2 (3); 3; 4	IARC (2006)
ILSI	Set of qualitative criteria to assess evidence on allergens proposed by the International Life Sciences Institute (ILSI) Europe Food Allergy Task Force	Method	No	Rank ordering	2 (2); 2 (3)	Van Bilsen et al. (2011)
INCa	Criteria to assess evidence on nutritional factors and their associated cancer risk	Method	No	Qualitative	2 (3); 3; 4	INCa (2015)
Klimisch	Scoring of the quality of individual toxicological studies based on limited indicators for reliability, relevance, and adequacy of data	Method	No	Scoring	2 (2)	ECHA (2011); Klimisch et al. (1997); Money et al. (2013); Schneider et al. (2009)
Meta-analysis	Statistical analysis of data collected in separate but similar studies, leading to the estimation of the magnitude of an effect and associated confidence interval	Method	Yes	Quantitative	2 (3)	Chalmers et al. (2002); EFSA (2014); Goodman et al. (2010); Marvier et al. (2007, 2011); Moher et al. (2015); Murad et al. 2014
Modified Bradford Hill	Comparative analysis for alternative mode of action hypotheses based on rank ordering of a subset of Bradford Hill considerations, taking into account epidemiological, toxicological and mechanistic data	Method	No	Rank ordering	3; 4	Boobis et al. (2006, 2008); Meek 2008; Meek et al. (2003, 2014a, 2014b); OCDE (2014)

Note: AMSTAR, Assessing the Methodological Quality of Systematic Reviews; EFSA, European Food Safety Authority; FDA, U.S. Food and Drug Administration; GRADE, Grading of Recommendations Assessment, Development and Evaluation; IARC, International Agency for Research on Cancer; ILSI, International Life Sciences Institute; INCa, Institut National du Cancer/French National Cancer Institute; NRC, U.S. National Research Council; OHAT, Office of Health Assessment and Translation; PF, Practical Framework; SCENIHR, Scientific Committee on Emerging and Newly Identified Health Risks; SR, Systematic Review; WCRF/AICR, World Cancer Research Fund and American Institute for Cancer Research; WHO, World Health Organization.

[a]Semiquantitative refers to approaches that include scoring and rank ordering of various components, without quantitation.

Table 2. (*Continued.*)

Name	Description	Category	SR included	Form of evaluation[a]	PF stages (step of stage)	Reference
Multi-criteria analysis	Expert-based quantitative judgment of quality of studies and their integration, including sensitivity and uncertainty analysis	Method	No	Quantitative or scoring	2 (2); 2 (3); 3; 4	Hristozov et al. (2014a, 2014b); Linkov et al. (2009), 2011; U.S. EPA (1997, 2003)
Navigation Guide	Synthesis of results for the reproductive and developmental hazards of chemical agents in the research context through 4 steps focused principally on systematic review	Method	Yes	Scoring	2 (1); 2 (3); 3	Viswanathan et al. (2012); Woodruff and Sutton (2011, 2014)
NRC	Principal focus on systematic review	Framework	Yes	Scoring	1; 2 (1); 2 (2); 3; 4	NRC 2014
OHAT	Detailed documentation of components for all stages	Framework	Yes	Scoring	1; 2 (1); 2 (2); 2 (3); 3; 4	Howard et al. (2014); OHAT (2015); Rooney et al. (2014); U.S. NTP (2015)
SCENIHR	Considerations to address individual studies in 3 categories for quality and relevance and 3 categories for coherence between studies of similar type with weighting of lines of evidence by utility/coherence	Framework	No	Scoring	2 (1); 2 (2); 2 (3); 3; 4	SCENIHR (2012)
SR-Cochrane	Handbook for Systematic Reviews of Interventions. Planning: PICO for question formulation, detailed specification of search strategy, documentation of bias in study selection and presentation of results, their applicability, quality (in 4 categories) and outcome (EPICOT)	Method	Yes	Qualitative	1; 2 (1); 2 (2); 4	Bilotta et al. (2014); Higgins and Green (2011); Mandrioli and Silbergeld (2016); O'Connor et al. (2011); Schünemann et al. (2011)
SR-EFSA	Detailed planning, process and documentation of systematic review, including PICO, PECO, PIT and PO and selection of studies (modification of SR-Cochrane)	Method	Yes	Qualitative	1; 2 (1); 4	EFSA (2010)
WCRF/AICR	Classification regarding nutrition and cancer risk relationships. Evaluation of individual studies (epidemiological and mechanistic) based on good practice, meta-analysis of epidemiological studies identified through systematic review, and consideration of mechanistic data in relation to the biological plausibility of human data. Classification of WOE for each nutritional factor in 5 classes	Method	Yes	Qualitative	2 (2); 2 (3); 3; 4	WCRF/AICR (2014)
Weighted Bradford Hill	Estimation of the probability of causality in epidemiological studies through expert assessment of the extent of supporting data for each of 9 weighted Bradford Hill considerations	Method	No	Quantitative	2 (3); 3	Swaen and van Amelsvoort (2009)

causality in epidemiological studies, modified Bradford Hill considerations in mode of action analyses, expert rule-based decision trees and systematic reviews proposed by EFSA (SR-EFSA). As expected, the methodologies proposed by the Cochrane Collaboration (SR-Cochrane) have only been adopted within the medical community, whereas Grading of Recommendations, Assessment, Development, and Evaluations (GRADE) guidelines have also been applied in the nutrition and environmental health fields.

Frameworks for Assessing the Body of Evidence

Of the 24 WOE approaches identified in Table 2, Hope and Clarkson (2014), U.S. National Research Council (NRC 2014), Office of Health Assessment and Translation (OHAT 2015), Scientific Committee on Emerging and Newly Identified Health Risks (SCENIHR 2012), and Rhomberg et al. (2013) addressed more than one stage of systematic data compilation, assessment and integration. These five approaches are subsequently referred to here as frameworks, and the other 19 are described as methods.

Based on the definitions identified in the literature review (Table S2), LOE and WOE are defined in the practical framework proposed here as follows:

- An LOE is a set of relevant items of information of similar type grouped to assess a hypothesis; and
- WOE is the structured synthesis of lines of evidence, possibly of varying quality, to determine the degree of support for hypotheses.

The term "strength of evidence," although appearing in some of the selected documents, was defined differently in varying contexts related to WOE by different authors, e.g., as a constitutive element of WOE (Suter and Cormier 2011; Linkov et al. 2009) or as a distinct entity (EFSA 2010). Subsequently, no definition is elaborated here.

The five identified frameworks differ in terms of the number of stages and level of detail (Figure 3). Three frameworks – namely, those of the NRC, Hope and Clarkson, and OHAT – address planning, scoping, problem formulation, and protocol development. Rhomberg et al. (2013) define causal questions and identify criteria for study selection. All five of the frameworks distinguish additional steps in establishing LOEs, namely, identification and selection of studies and an evaluation of their quality (based on specific criteria), and an assessment of LOEs. For all five frameworks, weighting and/or integrating one or more LOEs to assess WOE is addressed, and, lastly, conclusions are drawn. To support conclusions, SCENIHR adds an expression of

Table 3. Areas of application for the 24 WOE approaches screened by the literature review.

Approaches[a]	Safety at work	Food microbiology	Food chemistry	Nutrition	Animal feed and health	Environmental health	Crop protection products, biocides and fertilizers	Medical	Ecology-environment
AMSTAR								X	
Bayesian inference						X	X		
Bradford Hill	X			X		X		X	
Decision tree		X		X		X			X
Epid-Tox			X			X		X	
FDA			X	X					
GRADE				X		X		X	
Hope and Clarkson									X
Hypothesis based						X			
IARC	X		X	X		X	X	X	
ILSI				X					
INCa				X				X	
Klimisch			X			X	X		
Meta-analysis				X		X		X	
Modified Bradford Hill	X		X	X		X			
Multi-criteria analysis	X					X			X
Navigation Guide						X			
NRC			X			X			
OHAT						X			
SCENIHR						X			
SR-Cochrane								X	
SR-EFSA		X	X	X	X				
WCRF/AICR				X				X	
Weighted Bradford Hill						X			

Note: AMSTAR, Assessing the Methodological Quality of Systematic Reviews; EFSA, European Food Safety Authority; FDA, U.S. Food and Drug Administration; GRADE, Grading of Recommendations Assessment, Development and Evaluation; IARC, International Agency for Research on Cancer; ILSI, International Life Sciences Institute; INCa, Institut National du Cancer/French National Cancer Institute; NRC, U.S. National Research Council; OHAT, Office of Health Assessment and Translation; PF, Practical Framework; SCENIHR, Scientific Committee on Emerging and Newly Identified Health Risks; SR, Systematic Review; WCRF/AICR, World Cancer Research Fund and American Institute for Cancer Research.

[a]Descriptions of approaches and associated references are included in Table 2.

uncertainty, and Hope and Clarkson estimate ecological risk based on WOE.

Based on this analysis, and in view of the broad scope of ANSES expert-informed evaluations, a practical framework including four main stages is proposed here (Figure 4). The four stages are as follows: planning the assessment, establishing LOEs, integrating LOEs, and expressing WOE conclusions. For each stage of this framework, the identified methods were considered according to the three aspects introduced above, namely, the extent of their prescriptive nature, relevance, and feasibility.

The aim of formally documenting assessment planning (stage 1) is to increase transparency in the focus and methodology selected for the assessment. This first stage has three operational steps:

- Scoping (i.e., determining the appropriate focus, based on the objectives and preliminary consideration of available data),
- Formulating the question(s) to be assessed, and
- Developing the protocol for WOE assessment

Establishing LOEs (stage 2) also has three operational steps:

- Identifying and selecting studies
- Assessing the quality of the studies; and
- Analyzing a set of studies of similar type (epidemiological, toxicological, etc.) to establish LOEs.

Stage 3 addresses the integration of data from available LOEs to establish WOE in order to determine the degree of support for hypotheses or to estimate quantities of interest.

The objective of Stage 4 (i.e., the formal expression of conclusions) is an explicit presentation of WOE in a form that maximally supports decision-making.

Stages of WOE Addressed in the Identified Methods

Each method/framework identified in the literature addresses one or more key stages and steps of the WOE practical framework presented in Figure 4. Most address steps 2 (assessing the quality of the studies) and 3 (analyzing studies of similar type to establish LOEs) of stage 2, stage 3 (integration to establish WOE), and stage 4 (formal expression of conclusions). Few of them consider stage 1 (assessment planning) and step 1 of stage 2 (systematic identification and selection of studies) (see column on stages/steps addressed in Table 2).

Stage 1. Assessment planning. Stage 1 is addressed in six approaches: Hope and Clarkson (2014), NRC (2014), OHAT (2015), GRADE (Guyatt et al. 2011a), SR-Cochrane (Higgins and Green 2011), and SR-EFSA (EFSA 2010). Hope and Clarkson, NRC, and OHAT differentiate between the three operational steps as shown in Figure 4.

Step 1. Scoping. Hope and Clarkson describe the objective of scoping as defining environmental management objectives with stakeholders in neutral, precise, and measurable terms. NRC outlines these objectives as understanding the needs of clients in evaluating chemical products or processes. For OHAT, the aims are presented as identifying participants, evaluating the impact of conducting an evaluation and identifying the ongoing and related components of the assessment to be developed. For GRADE, scoping is not considered, as this method is devoted to the examination of alternative clinical management strategies or interventions.

For problem formulation in the consideration of environmental risk (including exposure and hazard), Hope and Clarkson suggest developing a conceptual model for each question and subquestion. These authors describe a conceptual model as a diagram that illustrates the succession of risk hypotheses based on predicted relationships among sources, stressors, exposures and assessment endpoint responses.

Step 2. Formulating the question(s). NRC formulates the problem based on a matrix outlining the testing strategy (i.e., the nature of the effects to be investigated in specified testing

Hope and Clarkson (2014)	NRC (2014)	OHAT (2015)	Rhomberg et al. (2013)	SCENIHR (2012)	Steps shared by the 5 frameworks
Planning and scoping	Scoping	Problem formulation (scoping included) and protocol development	Define causal question and identify criteria for study selection		Planning the assessment
Problem formulation	Problem formulation				
	Protocol development				
Build lines of evidence	Identify relevant studies	Search for and select studies for inclusion	Review of individual studies	Identify the possible sources of data and data gaps based on the objectives of the assessment	Establishing lines of evidence
				Initial screening of data sources to identify those that are relevant	
		Extract data from studies			
	Evaluate studies	Assess internal validity of individual studies		Assess individual data sources	
		Synthesize evidence and rate confidence in body of evidence			
Weight lines of evidence within each evidence group	Integrate Lines of evidence	Translate confidence ratings into level of evidence for health effect	Integrate and evaluate evidence	Weight the individual lines of evidence	Integrating lines of evidence and expressing weight of evidence conclusions
				Weight the totality of evidence	
Evaluate each evidence group to develop collective weighting of that group to estimate risk estimate		Integrate evidence to develop hazard identification conclusions	Draw conclusion based on inference	Express uncertainty	

Figure 3. Stages addressed in the five WOE frameworks identified in the literature.

protocols (*in vivo*, *in vitro*, etc.). OHAT and GRADE adopt the PECO reporting template (population, exposure, comparator, and outcome). The latter is derived from PICO elements (patient/problem/population, intervention, comparator, and outcome) promoting well-developed clinical questions in evidence-based medicine. In addition, NRC and OHAT propose working with a systematic review (SR) specialist but do not specifically outline the requirements for systematic review. The SR-Cochrane method has adapted the PICO reporting templates, whereas OHAT has developed PECOTS by adding the elements of time (T) and information on the setting of interest (S). SR-EFSA recommends additional templates for assessing the accuracy of a test result and quantifying a scenario of interest (prevalence, for instance) and has developed a method for completing the templates based on the literature.

Step 3. Developing the assessment protocol. For protocol development in assessment planning, Hope and Clarkson and NRC list assessment methods to establish LOEs and key elements of a systematic review, respectively. OHAT has developed a reporting template for a detailed analysis that includes scoping elements, the PECO template, and a description of all methods of analysis, from evidence identification to the development and presentation of conclusions. OHAT also recommends the use of text-mining, e.g., SWIFT (Howard et al. 2014), to characterize the extent and nature of available data. SR-EFSA recommends that the criteria for study inclusion or exclusion, the methodology adopted for each step of stage 2, and the process for the conducting the review (i.e., the composition of the multidisciplinary team, the timetable, and allocated resources) be specified to reduce the risk of bias — thus limiting their later criticism — and to increase the level of repeatability. GRADE consists of distinguishing the importance of outcomes in three steps, i.e., specifying all potential patient-important outcomes in their endeavor,

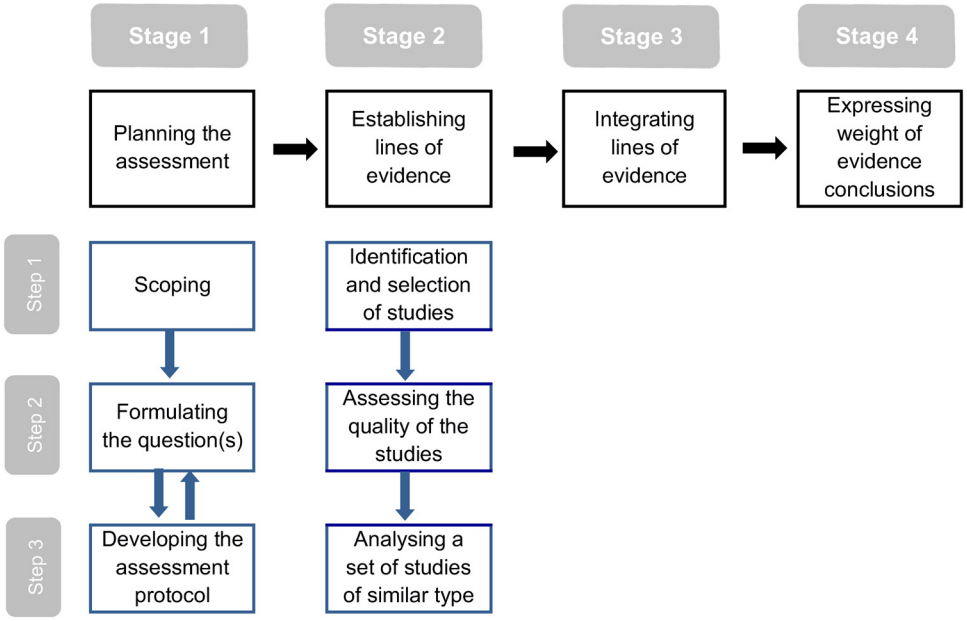

Figure 4. Practical framework for weight of evidence assessment.

distinguishing between critical and important-but-not-critical outcomes and making judgments about the balance between the desirable and undesirable effects of an intervention.

The approaches proposing a reporting template for one or more of the substeps (i.e., SR-EFSA and OHAT) are considered to be the most prescriptive and, as such, they promote transparency in assessment planning (Table 4). Generally, then, for assessment planning, GRADE, NRC, OHAT, SR-Cochrane, and SR-EFSA are equally prescriptive, but are more prescriptive than Hope and Clarkson (Table 4) because the latter authors do not propose a reporting template. NRC, OHAT, GRADE, SR-Cochrane, and SR-EFSA are considered broadly applicable or relevant, whereas the application of Hope and Clarkson is limited to the estimation of ecological risk.

Stage 2. Establishing LOEs. Nineteen methods address the collection and consideration of data to establish LOEs.

Step 1. Identification and selection of studies. Five methods/frameworks consider Step 1: the Navigation Guide (Woodruff and Sutton 2014), OHAT (2015), SR-Cochrane (Higgins and Green 2011), SR-EFSA (EFSA 2010), and Institut National du Cancer/French National Cancer Institute (INCa 2015), primarily through systematic literature review, the objective of which is to limit bias in the assembly, critical appraisal, and synthesis of all relevant studies. The principles adopted by SR-Cochrane, SR-EFSA, IARC, and OHAT are the use of at least two databases, the selection of studies by two independent reviewers, and the identification of the study selection criteria and data extraction format prior to the review. These approaches are considered prescriptive and relevant, but the requirement for considerable human resources makes them less feasible (Table 5). INCa is considered prescriptive and feasible. However, its relevance is limited to consideration of meta-analysis only in the establishment of LOEs.

Table 4. Ranking of the methods for planning the assessment.

Approach	Prescriptive nature[a]	Relevance[a]	Feasibility[a]
GRADE	4	3	3
Hope and Clarkson	2	2	3
NRC	4	3	3
OHAT	4	3	3
SR-Cochrane	4	3	3
SR-EFSA	4	3	3

Note: GRADE, Grading of Recommendations Assessment, Development and Evaluation; EFSA, European Food Safety Authority; NRC, U.S. National Research Council; OHAT, Office of Health Assessment and Translation; SR, Systematic Review.
[a]The rankings were assigned to the methods by the authors collectively and reflect relative consideration of each of the three aspects defined and outlined in the Methods and Table 1: the extent of prescriptive nature contributing to transparency and reproducibility, relevance to be broadly applied within ANSES, and ease of implementation in terms of time and material/human resources (feasibility). Each aspect is ranked from 1 (i.e., the least) to 4 (i.e., the most).

Step 2. Assessing the quality of the studies. The quality of relevant studies considered in the establishment of LOEs is usually assessed according to the degree of transparency in the documentation of the methodology, analysis and results, and the degree to which potential methodological bias, such as information and selection bias, is considered. Alternatively, or in addition, quality is assessed by the extent and nature of the scientific data (e.g., whether supporting data are direct or indirect). Two types of assessment methods are presented in the literature, i.e., those with or without quantitative scoring.

IARC (2006), WCRF/AICR (2014), SR-Cochrane (Higgins and Green 2011), and FDA (2009) are based on a qualitative evaluation of studies, i.e., without scoring. The evaluation criteria relate to good research practices for each area (epidemiology, toxicology, etc.). Epid-Tox (Adami et al. 2011) adopts criteria proposed by the U.S. EPA (2001) for evaluating toxicological studies and those from the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC 2009) for assessing epidemiological studies in three categories: "reliable without restriction" (minimum limitations), "reliable with restrictions" (moderate limitations), and "not reliable" (limitations sufficient to be excluded from WOE assessment). The extent of prescription of the qualitative methods is generally low, a function of their being expert-dependent with a varying degree of transparency in the considerations of the resulting judgments. Their simplicity makes them feasible, and they are broadly applicable or relevant (Table 5).

In multicriteria decision analysis (Linkov et al. 2011), Hope and Clarkson, GRADE, OHAT, International Life Sciences Institute (ILSI) (Van Bilsen et al. 2011), and Klimisch (Klimisch

Table 5. Ranking of the methods for establishing lines of evidence.

Approach	Identifying and selecting studies[a]			Assessing the quality of the studies[a]			Analyzing a set of studies of similar type[a]		
	PN	REL	FEA	PN	REL	FEA	PN	REL	FEA
AMSTAR	NA	NA	NA	4	3	4	NA	NA	NA
Bradford Hill	NA	NA	NA	NA	NA	NA	2	4	4
Epid-Tox	NA	NA	NA	2	4	4	2	4	3
FDA	NA	NA	NA	3	4	4	2	3	3
GRADE	NA	NA	NA	4	3	3	2	3	4
Hope and Clarkson	NA	NA	NA	2	3	3	2	3	3
IARC	NA	NA	NA	2	4	4	2	3	4
ILSI	NA	NA	NA	2	3	3	3	2	3
INCa	3	2	4	3	2	4	NA	NA	NA
Klimisch	NA	NA	NA	2	3	4	NA	NA	NA
Meta-analysis	NA	NA	NA	NA	NA	NA	4	4	1
Modified Bradford Hill	NA	NA	NA	NA	NA	NA	3	3	3
Multi-criteria analysis	NA	NA	NA	2	4	3	2	4	3
Navigation Guide	1	3	2	1	3	4	1	3	3
OHAT	3	3	2	3	3	4	2	3	3
SR-Cochrane	3	3	2	2	4	4	NA	NA	NA
SR-EFSA	3	3	2	NA	NA	NA	NA	NA	NA
SCENIHR	NA	NA	NA	2	3	4	1	3	4
WCRF/AICR	NA	NA	NA	2	4	4	4	4	2
Weighted Bradford Hill	NA	NA	NA	NA	NA	NA	3	3	3

Note: AMSTAR, Assessing the Methodological Quality of Systematic Reviews; EFSA, European Food Safety Authority; FDA, U.S. Food and Drug Administration; FEA, Feasibility; GRADE, Grading of Recommendations Assessment, Development and Evaluation; IARC, International Agency for Research on Cancer; ILSI, International Life Sciences Institute; INCa, Institut National du Cancer/French National Cancer Institute; NA, Not applicable because the corresponding step was not addressed by the approach; NRC, U.S. National Research Council; OHAT, Office of Health Assessment and Translation; PF, Practical Framework; PN, Prescriptive nature; REL, Relevance; SCENIHR, Scientific Committee on Emerging and Newly Identified Health Risks; SR, Systematic Review; WCRF/AICR, World Cancer Research Fund and American Institute for Cancer Research.

[a]The rankings were assigned to the methods by the authors collectively and reflect relative consideration of each of the three aspects defined and outlined in the Methods and Table 1: the extent of prescriptive nature contributing to transparency and reproducibility, relevance to be broadly applied within ANSES, and ease of implementation in terms of time and material/human resources (feasibility). Each aspect is ranked from 1 (i.e., the least) to 4 (i.e., the most).

et al. 1997) attribute scores to individual studies, taking into account their quality. The tools proposed by GRADE and OHAT enable the classification of study quality on a qualitative scale based on a set of questions. For instance, the OHAT Bias Risk Tool is composed of eleven questions related to good research practices for various types of studies. A response to a question is expressed in terms of risk of bias (low, probably low, probably high, and high). Multicriteria decision analysis and Hope and Clarkson score individual studies based on a quantitative scale according to specific criteria. For example, Hope and Clarkson address five criteria, i.e., study quality, use of standard methods to design the study, site specificity, spatial representativeness, and temporal representativeness. Each of these criteria is scored in binary fashion for each study, with the value of 1 corresponding to criteria effectiveness for each of the five LOEs, each one addressing a specific aspect, i.e., endpoint/attribute association, exposure/response function, sensitivity to stressor, specificity to stressor, and quantification of response. The weighting for each LOE is then calculated by combining the criteria scores.

None of these methods prescribes a quantitative threshold value for exclusion. Criteria specified by the Klimisch, Hope and Clarkson, and multicriteria decision analysis methods are less specific (i.e., less prescriptive) than those of GRADE and OHAT. Each of the methods considered in this section, i.e., multicriteria decision analysis, Hope and Clarkson, GRADE, OHAT, ILSI, and Klimisch, is considered relevant and feasible (Table 5).

Assessing the Methodological Quality of Systematic Reviews (AMSTAR) and its revised version R-AMSTAR are the only methods considered here that address the quality of a synthesis of studies. The methodology is relatively prescriptive, delineating a questionnaire with eleven items to score, contributing to transparency and reproducibility of reviews. Although relevant for the assessment of syntheses of both clinical trials and observational studies, the method addresses only one component of one stage of the developed practical framework. The method is also considered feasible, requiring limited time to develop the score.

Step 3. Analyzing a set of studies of similar type. Fourteen methods/frameworks include considerations for establishing LOEs of similar type. Meta-analysis and all methods based on meta-analyses of epidemiological studies are considered prescriptive, as specified elements of the considered studies must be sufficiently similar to enable their statistical analysis (Chalmers et al. 2002). These include WCRF/AICR, which systematically performs meta-analyses in all its nutrition–cancer evaluations; IARC, which commissions specific meta-analyses for selected topics, such as asbestos and ovarian cancer; or INCa, which performs systematic reviews of published meta-analyses on nutrition and cancer risk. Prescribed methods are transparent and reproducible and enable a quantitative synthesis of studies of similar type to estimate quantities of interest and to test hypotheses (Table 5). However, these methods are considered less feasible, as implementation is time-consuming and may require specialized computational and/or statistical skills.

Multicriteria decision analysis requires selected experts to define specific considerations and their relative weighting. Although relevant to a broad range of applications, it is not prescriptive nor reproducible, due to its dependence on the judgment of selected experts. The outcome is therefore highly sensitive to the judgment of the participating experts, for whom selection criteria are often not specified or well described.

The other methods/frameworks considered here, namely, Bradford Hill, IARC (for some other topics), Epid-Tox, FDA, GRADE, Hope and Clarkson, OHAT, and SCENIHR, are based on qualitative or semiquantitative approaches to establishing LOEs. Most of these methods/frameworks assign a level of evidence or confidence, utility or consistency, according to predefined scales. Bradford Hill considerations (or modifications thereof) continue to be applied when assessing causality of associations in epidemiological studies (e.g., IARC, Epid-Tox). These approaches

are thus considered relevant and feasible, except GRADE, which is restricted principally to randomized controlled trials and meta-analysis. These approaches are not very prescriptive (or transparent), with results being highly sensitive to expert input; as such, they have a low degree of reproducibility (Table 5).

Stage 3. Integrating LOEs. Nineteen methods/frameworks address the integration of LOEs to establish WOE. One category of methods used to integrate LOEs relies on statistical techniques. Bayesian inference (BioBayes Group 2015) is highly relevant to combining experimental data and expert opinion, but it is rather complex to implement. Thus, although it is prescriptive and relevant, this method is less feasible due to the complexity of accessing expert knowledge through elicitation and statistical methodology to combine experimental data with expert opinion (Table 6).

A second category of approaches for integrating LOEs includes semiquantitative methods, i.e., modified Bradford Hill (Meek et al. 2014b) and Hope and Clarkson (2014), and qualitative approaches, i.e., IARC (2006), WCRF/AICR (2014), OHAT (2015), hypothesis-based (Rhomberg 2015), Epid-Tox (Adami et al. 2011), and INCa (2015), and SCENIHR (2012). These methods are relevant and feasible, with their extent of prescription varying depending on the nature of the expert-informed experience upon which they draw (i.e., for those where there is greater experience, the considerations to be taken into account in integration are often more precisely delineated, drawing on a larger number of documented examples). For example, assessing causality in epidemiological studies based on Bradford Hill considerations is commonly quite subjective, which limits the reproducibility of the evaluation (i.e., the results vary considerably, depending on the experts involved). In the modified Bradford Hill approach, as a basis for increasing the consistency and reproducibility of mode of action analyses, selected considerations have been modified for the specific application and precisely defined and rank ordered (i.e., weighted) by their relative importance, taking into account acquired experience. Examples of the types of datasets (integrating epidemiological, toxicological and mechanistic data) associated with higher or lower confidence are also provided.

Table 6. Ranking of the methods for integrating lines of evidence.

Approach	Prescriptive nature[a]	Relevance[a]	Feasibility[a]
Bayesian inference	3	4	2
Bradford Hill	2	4	4
Decision tree	1	3	3
Epid-Tox	2	4	3
Hope and Clarkson	3	3	3
Hypothesis based	2	3	3
IARC	3	3	4
INCa	3	3	4
Multi-criteria analysis	2	4	3
Modified Bradford Hill	3	3	3
Navigation Guide	1	3	3
OHAT	3	3	4
SCENIHR	2	3	4
WCRF/AICR	3	3	4
Weighted Bradford Hill	3	4	4

Note: IARC, International Agency for Research on Cancer; INCa, Institut National du Cancer/French National Cancer Institute; OHAT, Office of Health Assessment and Translation; SCENIHR, Scientific Committee on Emerging and Newly Identified Health Risks; WCRF/AICR, World Cancer Research Fund and American Institute for Cancer Research.

[a]The rankings were assigned to the methods by the authors collectively and reflect relative consideration of each of the three aspects defined and outlined in the Methods and Table 1: the extent of prescriptive nature contributing to transparency and reproducibility, relevance to be broadly applied within ANSES, and ease of implementation in terms of time and material/human resources (feasibility). Each aspect is ranked from 1 (i.e., the least) to 4 (i.e., the most).

Of the qualitative methods, OHAT is the most prescriptive, drawing upon a number of previously documented approaches in clinical medicine. The quality of individual studies (Step 2 of Stage 2) is evaluated based on responses to up to 15 questions (depending on study type) to assess the risk of bias. In Stage 3, preliminary confidence scores developed on this basis are either downgraded through the assessment of 5 properties of the body of evidence (risk of bias, unexplained inconsistency, indirectness, imprecision, and publication bias) or upgraded based on the consideration of another 4 properties (large magnitude of effect, dose response, residual confounding, and cross-species/population/study consistency). A comparison of OHAT and IARC, through a feasibility study of their application in an ANSES assessment of airborne particulates, indicated that the more prescriptive nature of OHAT led to greater ease of application, consistency, and reproducibility (Table 6).

A third category includes the decision tree method and multicriteria decision analysis. Multicriteria decision analysis is relevant when combining any type of data (qualitative or quantitative). However, considerable expert knowledge is required for its implementation to identify criteria and their associated weights (i.e., limited feasibility), and the results are highly expert-dependent. For the decision tree method, classification rules are expert-derived and based on acquired experience, taking into consideration diverse types of information, such as experimental studies, observations, and model outputs. Although feasible and relevant, decision trees are less prescriptive because there are no associated evaluation rules (Table 6).

Stage 4. Expressing WOE conclusions. The 13 methods/frameworks reviewed here address the expression of WOE conclusions. Most methods/frameworks use four classes, with an additional class to indicate that the available data preclude evaluation. Examples of classification in methods/frameworks are presented in Table 7.

Conclusions are illustrated or presented in different formats. OHAT presents the intermediate results in the form of graphs. NRC conducts an uncertainty analysis on WOE. SCENIHR expresses uncertainty in the WOE analysis in five classes (i.e., certain, probable, confident, possible, and uncertain). For multicriteria decision analysis, Hristozov et al. (2014a, 2014b) conducted a quantitative uncertainty analysis on data and expert judgments, whereas Linkov et al. (2011) conducted a sensitivity analysis on weightings and some input data. Although not necessarily increasing consistency, due to their being mostly dependent on varying expert input with the often-limited prescription of decision rules, these methods promote transparency in communicating the basis for the conclusion.

With regard to communication of the outcome, SR-EFSA and SR-Cochrane specify the topics to be addressed in the discussion and conclusions. SR-Cochrane relies on EPICOT (i.e., the PICO structure completed outlining the current state of the evidence and the date of recommendation) to identify the need and priorities for research, whereas GRADE structures the conclusions according to PICO, both of which are addressed initially in problem formulation. Both PICO and EPICOT offer cohesive consideration of communication at the outset and throughout the assessment.

All the examples of classifications reviewed here (Table 7) are considered relevant and feasible for expressing conclusions of WOE analysis, with varying degrees of prescription (Table 8).

Based on the ratings presented in Tables 4, 5, 6, and 8, several of the methods performed well for the three criteria considered, i.e., OHAT, modified Bradford Hill, AMSTAR, and WCRF/AICR. Systematic reviews (i.e., SR-Cochrane, SR-EFSA) and meta-analysis methods are considered prescriptive and relevant but less feasible.

Table 7. Example of weight of evidence classifications.

Method/framework	Reference	Number of Classes	Class title
Bayesian Inference	Schleier et al. (2015)	NA	NA
Epid-Tox	Adami et al. (2011)	4	Likely, Uncertain, Uncertain but plausible, Unlikely (Used to qualify the causal relationship between the environmental factor and the disease condition)
GRADE	Andrews et al. (2013b)	4	Strong Against, Weak Against, Weak For, Strong For
Hope and Clarkson	Hope and Clarkson (2014)	5	Weak, Not indicated, Not indicated, Not indicated, Strong
IARC	IARC (2006)	5	Group 1: The agent is carcinogenic to humans Group 2A: The agent is probably carcinogenic to humans Group 2B: The agent is possibly carcinogenic to humans Group 3: The agent is not classifiable as to its carcinogenicity to humans Group 4: The agent is probably not carcinogenic to humans
Modified Bradford Hill	Meek et al. (2014a); OECD (2014)	3	Weak, Moderate, Strong
Multi-criteria analysis	Linkov et al. (2011)	6	Do nothing, Institutional control, Clay capping, Mechanical dredging, Hydraulic dredging, Hot spot dredging
NRC	NRC (2014)	5	Carcinogenic to humans, Likely to be carcinogenic to humans, Suggestive evidence of carcinogenic potential, Inadequate information to assess carcinogenic potential, Not likely to be carcinogenic to humans
OHAT	OHAT (2015)	5	(1) Known to be a hazard to humans, (2) Presumed to be a hazard to humans, (3) Suspected to be a hazard to humans, (4) Not classifiable as a hazard to humans, or (5) Not identified as a hazard to humans
SR–Cochrane	Higgins and Green (2011)	4	Very low, Low, Moderate, Strong
SR-EFSA	EFSA (2010)	NA	NA
SCENIHR	SCENIHR (2012)	5	Weighting not possible, Uncertain, Weak, Moderate, Strong
WCRF/AICR	WCRF/AICR (2014)	5	Convincing/Probable/Limited - suggestive/Limited - no conclusion/Substantial effect on risk unlikely

Note: EFSA, European Food Safety Authority; GRADE, Grading of Recommendations Assessment, Development and Evaluation; NA, Not applicable because the corresponding step was not addressed by the approach; NRC, U.S. National Research Council; OHAT, Office of Health Assessment and Translation; SCENIHR, Scientific Committee on Emerging and Newly Identified Health Risks; SR, Systematic Review; WCRF/AICR, World Cancer Research Fund and American Institute for Cancer Research.

Discussion

The results of the review described here have illustrated that a wide range of methods is applied when assessing WOE in hazard identification, most notably and broadly in the environmental field (i.e., to assess effects on human and ecological health in the general environment).

Developing and documenting the three assessment planning steps (i.e., scoping, question formulation and protocol for assessment) has contributed to the efficient and cohesive consideration of priority areas and their transparent communication in assessing WOE. Assessment protocols are designed considering appropriate and available associated resourcing, taking into account urgency, potential public, and environmental health impacts, available data, societal issues, and the level of acceptable uncertainty. Consideration of the relative resourcing of various stages in WOE assessment (i.e., how complex the approach is at each stage) should also be addressed, commensurate with their likely impact on the outcome (e.g., the more direct impact of a systematic assessment of integration based on prescriptive approaches at stage 3 versus a systematic review of the literature at stage 1). Depending on the issues addressed, the existing reporting templates (PICO, PECO, etc.) reviewed here have contributed to but have not fully delineated the nature of the required documentation. In addition, although conceptual models normally address risk resulting from identified pathways of exposure, similar figurative representation of the envisaged steps in assessing WOE in hazard identification may facilitate assessment planning, as part of formal planning in the iterative definition of areas of focus and critical questions and subquestions (U.S. EPA 1998, 2014). The development of reporting templates for delineating protocols in assessment planning would facilitate greater transparency, and potentially consistency, on the basis of the selection of specific approaches to weight of evidence assessment, depending on available resources.

Aspects to be documented in assessment protocols include the type of literature review (namely, a formally systematic review), or an in-depth review taking into account the considerations proposed, for example, by EFSA (2010). Protocols for assessing the quantity and quality of available evidence, including sources and potential confidentiality, and for integrating conflicting results should be specified, as well as resources needed to carry out the review. The assessment protocol should also specify criteria for the inclusion and exclusion of relevant data based on consideration of the quality and weighting, for integrating studies of a similar type. The protocol for establishing and integrating LOEs

Table 8. Ranking of the methods for expressing weight of evidence conclusions.

Approach	Prescriptive nature[a]	Relevance[a]	Feasibility[a]
Bayesian inference	3	4	2
Epid-Tox	3	4	4
GRADE	3	4	4
Hope and Clarkson	3	3	3
IARC	3	4	4
Modified Bradford Hill	3	2	3
Multi-criteria analysis	3	3	3
NRC	3	4	3
OHAT	3	4	4
SR-Cochrane	4	4	3
SR-EFSA	4	4	3
SCENIHR	3	4	3
WCRF/AICR	4	3	4

Note: EFSA, European Food Safety Authority; GRADE, Grading of Recommendations Assessment, Development and Evaluation; IARC, International Agency for Research on Cancer; NRC, U.S. National Research Council; OHAT, Office of Health Assessment and Translation; SCENIHR, Scientific Committee on Emerging and Newly Identified Health Risks; SR, Systematic Review.

[a]The rankings were assigned to the methods by the authors collectively and reflect relative consideration of each of the three aspects defined and outlined in the Methods and Table 1: the extent of prescriptive nature contributing to transparency and reproducibility, relevance to be broadly applied within ANSES, and ease of implementation in terms of time and material/human resources (feasibility). Each aspect is ranked from 1 (i.e., the least) to 4 (i.e., the most).

should also be specified, along with estimated resources for conducting the assessment. Developing and completing prerequisite reporting templates for the assessment protocol would improve transparency for the rationale for selecting methods in WOE assessment, such as partial or full/directed systematic reviews, meta-analyses, and Bayesian inference.

We analyzed the frameworks and methods (namely, to identify the extent of the prescriptive nature, relevance, and feasibility) and found that preferred methods are often the least feasible (i.e., the most complex requiring, for example, specialized expertise), due to limited resources (e.g., lack of expertise or time). This finding underscores the need for transparent, easily adaptable, and broadly applicable communicative methods that draw on collective expertise. Due to limited resources, it is expected that the application of the more complex approaches for which feasibility has been judged as low in the current study (e.g., Bayesian analysis and meta-analysis) will understandably be limited, based on careful consideration of the abovementioned factors, including the importance of the question at hand, urgency, and available resources. However, implementation of these complex approaches can lead to greater efficiency in public health protection through a more systematic allocation of resources than is currently made. The availability of a documented assessment protocol addressing delineated considerations in reporting templates should also enhance common understanding of (sometimes limited) objectives and facilitate the provision of early input to modify the selection of appropriate methods and allocation of associated resources.

The results of this review have also indicated that the principles of the limited range of methods identified as being relevant to potentially the most influential stages of WOE assessment—namely, the later steps of stage 2 (integration within an LOE) and 3 (integration of LOEs)—are similar and relate essentially to expert-informed weighting of components. These methods range from qualitative to semiquantitative to fully quantitative. Expert-informed experience is derived from a formal analysis of previous examples in defining relevant considerations and their relative weighting. This analysis is distinct from expert judgment of an individual or group, for which relevant criteria and weightings are often not well specified.

Bradford Hill considerations have figured prominently in this integration but have varied depending on the extent of their prescriptive nature and the field of application (e.g., epidemiological studies, mode of action or integration of epidemiological and toxicological data, based on the consideration of mode of action). The extent to which these methods have been prescribed, taking into account previous expert-informed experience, contributes most to their consistency and reproducibility.

The approaches identified have been based on qualitative, semiquantitative, or quantitative techniques to establish LOEs and WOE, consistent with the WOE classification system proposed by Linkov et al. (2009). Although quantitative methods are more rigorous (i.e., prescriptive), their implementation (stages 2 and 3) requires specific knowledge of elicitation and statistical methodology. In contrast, purely qualitative methods for establishing LOEs, such as Bradford Hill considerations in assessing causality in epidemiological studies, require fewer resources to implement. However, their transparency is limited, often leading to different conclusions by different groups, the basis for which is unclear. Semiquantitative, more prescriptive methods, such as OHAT and modified Bradford Hill, offer, then, a valuable intermediate option that conserves resources but also increases the transparency and consistency of assessments.

The delineation of conclusions in various defined classifications also contributes to transparency. The nature of these descriptions requires careful consideration, to avoid, as far as possible, the misinterpretation that higher classifications infer greater hazard; rather, they indicate greater preponderance of evidence. Brief, plain-language descriptions of the nature and extent of evidence and graphical illustrations may be preferred over less clear descriptors such as "probably," "possibly," and "potentially."

The results of this review have also indicated that methods have been broadly applied in some application fields, such as environmental health or human food and nutrition (cf. Table 3). However, the seeming lack of application in some fields for certain methods may be a function of specific assessment needs or, for example, the restricted date range of the literature review. For example, the sole method identified here as enabling an assessment of the quality of study syntheses is R-AMSTAR; all the other identified methods are based exclusively on the quality of individual studies to establish LOEs. Although R-AMSTAR has mostly been adopted in the medical sector, it could be applied in a range of disciplines, given the broad relevance of its rather generic contents. In other applications and disciplines (e.g., Plant Health), WOE analysis is not referenced. This finding relates in part to variations in terminology and requirements in different application fields (e.g., although not explicitly mentioned, the decision-support scheme for pest risk analysis developed by the European and Mediterranean Plant Protection Organization addresses WOE).

Consequently, the current work contributes to relate primarily to considering the principles of existing methods and assessing their potential utility in a broad range of application areas relevant to the purview of ANSES. We considered a number of characteristics of each method as a basis for method selection in assessment planning, including the extent of prescription, relevance, and ease of implementation (feasibility). We additionally used a series of case studies for selected ongoing or completed assessments in a range of different applications at ANSES (ANSES 2017) to further evaluate the value of these characteristics to screening for planning and conducting assessments. Two limitations of the current study were the fact that we restricted our consideration of WOE analysis to hazard identification alone and the fact that the interrelationships between WOE and uncertainty analysis were not explicitly considered. We plan to develop and integrate these aspects in future research, in further consideration of the working group's recommendations by ANSES.

In addition, it is important to note that the scores developed for the prescriptive nature, relevance, and feasibility of various methods are meaningful in a relative context only and are limited to generalized considerations for assessment. They mostly reflect the extent and documentation of expert judgment and ease of application across a broad range of applications. Applying each of the methods to specific assessments is necessarily dependent on case-specific objectives and conditions, as indicated in problem formulation.

The results of the current analysis indicate that ultimately, over the short term, transparency is critical in increasing confidence and, over the long term, is critical to potentially increasing consistency in WOE analysis, within defined constraints of assessment planning. Identified outstanding areas that are relevant for considering the quality of studies when establishing LOEs and their weighted integration include delineating criteria for the consideration of additional factors, such as selection criteria for experts.

Conclusions

The documentation of planning, taking into account factors outlined for each of the approaches reviewed here (namely, extent of prescription, relevance to the question at hand and ease of

implementation), considerably increases transparency in the rationale for the justified adoption of different approaches based on factors such as urgency and the extent of available resources. This finding should increase common understanding of the constraints that provide a legitimate basis for variations in the approaches taken when considering WOE. Development and application of reporting templates for assessment planning outlining specific aspects to be addressed will likely increase common understanding of the appropriate nature of required transparency in the selection of assessment methods.

The current review also highlighted the value of acquired experience in contributing to expert-informed prescription of the relevant factors to be considered in reporting templates, as a basis for increasing the transparency and defensibility of WOE analysis. This aspect is particularly important for establishing and integrating LOEs. All WOE assessments include elements of expert-informed judgments. However, transparency regarding the nature and basis of those judgments in individual assessments (often attributed to "expert judgment") is often lacking. Developing prescriptive reporting templates based on collective expert experience increases common understanding of important elements for consideration and their relative weighting. This, in turn, contributes to more consistent evaluations, but necessarily requires that contributing experts be much more explicit about the factors being taken into consideration. For example, OHAT provides a relatively prescriptive and transparent approach to assessment planning, review and evaluation, which facilitates adoption and is likely to increase common understanding of relevant elements for consideration. The generic utility of this approach has been illustrated in assessments of the National Toxicology Program (U.S. NTP 2015) and in a range of case studies conducted by various organizations (e.g., EFSA and ANSES). However, it does not yet robustly address mechanistic data. Integrating experience from more mechanistically driven approaches, such as the application of modified Bradford Hill considerations as a basis for considering patterns of epidemiological, toxicological, and mechanistic data in mode of action analyses, may well inform its additional development. IARC classifications, on the other hand, result from the consideration of a much less prescriptive approach by a convened group of experts and, as such, reflect less documented and variable expert judgment, as does multicriteria decision analysis. Explicit criteria for selecting experts and process considerations concerning the weighting of their input seem essential to ensure greater transparency in these more judgment-dependent methodologies. However, reporting templates that draw much more broadly on previous collective experience, defining specific aspects taken into consideration and the nature of their relative weighting, are preferred.

Prescriptive generic approaches providing an encompassing framework, such as OHAT, that draw broadly on an analysis of experience acquired in application and less on consensus expert opinion, are likely to offer the greatest transparency and consistency in WOE analysis. Specific issues identified, when planning an assessment, require a combination of these more generic frameworks with specialized approaches (e.g., those used to consider the extent of mechanistic support for competing hypotheses in mechanistically motivated integration of LOEs). Selecting expert-informed prescriptive approaches (versus consensus based on expert judgment) is likely to provide the greatest transparency and, potentially, the greatest consistency of evaluations within identified constraints.

Acknowledgments

The authors thank I. Albert, Institut National de la Recherche Agronomique/French National Institute of Agricultural Research (INRA); N. Bonvallot, Institut national de la santé et de la recherche médicale/French National Institute of Health and Medical Research (INSERM); C. Brochot, Institut national de l'environnement industriel et des risques/French National Institute For Industrial Environment And Risks (INERIS); S. Fraize-Frontier, Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail/French Agency for Food, Environmental and Occupational Health and Safety (ANSES); P. Glorennec, INSERM; C. Saegerman, Liège University; and M. Sanaa, ANSES, for suggestions and comments.

This publication was made possible by ANSES Request No 2015-SA-0089 (2015–2017). Its contents are solely the responsibility of the authors and do not necessarily represent the official view of the ANSES.

References

Adami HO, Sir Berry CL, Breckenridge CB, Smith LL, Swenberg JA, Trichopoulos D. 2011. Toxicology and epidemiology: improving the science with a framework for combining toxicological and epidemiological evidence to establish causal inference. Toxicol Sci 122(2):223–234, PMID: 21561883, https://doi.org/10.1093/toxsci/kfr113.

Akl EA, Maroun N, Guyatt G, Oxman AD, Alonso-Coello P, Vist GE, et al. 2007. Symbols were superior to numbers for presenting strength of recommendations to health care consumers: a randomized trial. J Clin Epidemiol 60(12):1298–1305, PMID: 17998085, https://doi.org/10.1016/j.jclinepi.2007.03.011.

Andrews JC, Schünemann HJ, Oxman AD, Pottie K, Meerpohl JJ, Coello PA, et al. 2013a. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. J Clin Epidemiol 66(7):726–735, PMID: 23570745, https://doi.org/10.1016/j.jclinepi.2013.02.003.

Andrews J, Guyatt GH, Oxman AD, Alderson P, Dahm P, Falck-Ytter Y, et al. 2013b. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. J Clin Epidemiol 66(7):719–725, PMID: 23312392, https://doi.org/10.1016/j.jclinepi.2012.03.013.

ANSES (Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail). 2012. IAvis et rapport de l'ANSES relatif à l'Étude des liens entre facteurs de croissance, consommation de lait et de produits laitiers et cancers [in French]. Maisons-Alfort, France: ANSES.

ANSES. 2013a. Avis et rapport de l'ANSES relatif à l'Évaluation des risques du bisphénol A (BPA) pour la santé humaine. – Tome 1: Évaluation des risques du bisphénol A (BPA) pour la santé humaine et aux données toxicologiques et d'usage des bisphénols S, F, M, B, AP, AF, et BADGE [in French]. Maisons-Alfort, France: ANSES.

ANSES. 2013b. Avis et rapport de l'ANSES relatif à la mise à jour de l'expertise « Radiofréquences et santé » [in French]. Maisons-Alfort, France: ANSES.

ANSES. 2015. Bisphenol A: EFSA recommends lowering the Total Daily Intake (TDI) while considering that current exposure levels are without risk for human health. Maisons-Alfort, France: ANSES. http://www.anses.fr/en/content/bisphenol-efsa-recommends-lowering-total-daily-intake-tdi-while-conside-ring-current-exposure [accessed 11 July 2017].

ANSES. 2017. Illustrations et actualisation des recommandations pour l'évaluation du poids des preuves et l'analyse d'incertitude à l'ANSES [in French]. Maisons-Alfort, France: ANSES.

Bailey LA, Nascarella MA, Kerper LE, Rhomberg LR. 2016. Hypothesis-based weight-of-evidence evaluation and risk assessment for naphthalene carcinogenesis. Crit Rev Toxicol 46(1):1–42, PMID: 26202831, https://doi.org/10.3109/10408444.2015.1061477.

Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. 2011. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 64(4):401–406, PMID: 21208779, https://doi.org/10.1016/j.jclinepi.2010.07.015.

Bergman A, Becher G, Blumberg B, Bjerregaard P, Bornman R, Brandt I, et al. 2015. Manufacturing doubt about endocrine disrupter science – a rebuttal of industry-sponsored critical comments on the UNEP/WHO report – State of the Science of Endocrine Disrupting Chemicals 2012. Regul Toxicol Pharmacol 73(3):1007–1017, PMID: 26239693, https://doi.org/10.1016/j.yrtph.2015.07.026.

Berkman ND, Lohr KN, Ansari M, McDonagh M, Balk E, Whitlock E, et al. 2012. Grading the Strength of a Body of Evidence When Assessing Health Care Interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: An Update. In: *Methods Guide for Comparative Effectiveness Reviews (Prepared by the RTI-UNC Evidence-based Practice Center under Contract No. 290 - 2007 - 10056 - I)*. Rockville, MD: Agency for Healthcare Research and Quality.

Bilotta GS, Milner AM, Boyd I. 2014. On the use of systematic reviews to inform environmental policies. Environ Sci Policy 42:67–77, https://doi.org/10.1016/j.envsci.2014.05.010.

BioBayes Group. 2015. *Initiation à la statistique bayésienne* [in French]. Paris: Ellipses.

Boobis AR, Cohen SM, Dellarco V, McGregor D, Meek ME, Vickers C, et al. 2006. IPCS framework for analyzing the relevance of a cancer mode of action for humans. Crit Rev Toxicol 36(10):781–792, PMID: 17118728, https://doi.org/10.1080/10408440600977677.

Boobis AR, Doe JE, Heinrich-Hirsch B, Meek ME, Munn S, Ruchirawat M, et al. 2008. IPCS framework for analyzing the relevance of a noncancer mode of action for humans. Crit Rev Toxicol 38(2):87–96, PMID: 18259981, https://doi.org/10.1080/10408440701749421.

Chalmers I, Hedges LV, Cooper H. 2002. A brief history of research synthesis. Eval Health Prof 25(1):12–37, PMID: 11868442, https://doi.org/10.1177/016327870202501003.

ECETOC (European Centre for Ecotoxicology and Toxicology of Chemicals). 2009. Framework for the Integration of Human and Animal Data in Chemical Risk Assessment. TR. Brussels, Belgium: European Centre for Ecotoxicology and Toxicology of Chemicals.

ECHA (European Chemicals Agency). 2011. Guidance on information requirements and chemical safety assessment. Chapter R.4 Evaluation of available information. Helsinki, Finland: European Chemicals Agency.

EFSA (European Food Safety Agency). 2010. Application of systematic review methodology to food and feed safety assessments to support decision making. EFSA J 8(6):1637.

EFSA. 2014. Scientific opinion on the risk of Phyllosticta citricarpa (Guignardia citricarpa) for the EU territory with identification and evaluation of risk reduction options. EFSA J 12(2):3557, https://doi.org/10.2903/j.efsa.2014.3557.

EFSA. 2017. Glyphosate: EFSA responds to critics. http://www.efsa.europa.eu/en/press/-news/160113. [accessed 4 July 2017].

FAO/WHO (Food and Agriculture Organization of the United Nations/World Health Organization). 2001. Evaluation of allergenicity of genetically modified foods - Report of a Joint FAO/WHO expert consultation on allergenicity of foods derived from biotechnology. Rome, Italy: Food and Agriculture Organization of the United Nations.

FDA (U.S. Food and Drug Administration). 2009. Guidance for Industry: Evidence-Based Review System for the Scientific Evaluation of Health Claims - Final. Washington, DC: U.S. Food and Drug Administration.

Goodman M, Squibb K, Youngstrom E, Anthony JG, Kenworthy L, Lipkin PH, et al. 2010. Using systematic reviews and meta-analyses to support regulatory decision making for neurotoxicants: lessons learned from a case study of PCBs. Environ Health Perspect 118(6):727–734, PMID: 20176542, https://doi.org/10.1289/ehp.0901835.

Gosling JP, Hart A, Owen H, Davies M, Li J, MacKay C. 2013. A Bayes linear approach to weight-of-evidence risk assessment for skin allergy. Bayesian Anal 8(1):169–186, https://doi.org/10.1214/13-BA807.

Guha N, Roy A, Kopylev L, Fox J, Spassova M, White P. 2013. Nonparametric Bayesian methods for benchmark dose estimation. Risk Anal 33(9):1608–1619, PMID: 23339666, https://doi.org/10.1111/risa.12004.

Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. 2011a. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 64(4):383–394, PMID: 21195583, https://doi.org/10.1016/j.jclinepi.2010.04.026.

Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, et al. 2011b. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol 64(4):395–400, PMID: 21194891, https://doi.org/10.1016/j.jclinepi.2010.09.012.

Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. 2011c. GRADE guidelines 6. Rating the quality of evidence – imprecision. J Clin Epidemiol 64(12):1283–1293, PMID: 21839614, https://doi.org/10.1016/j.jclinepi.2011.01.012.

Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. 2011d. GRADE guidelines: 7. Rating the quality of evidence – inconsistency. J Clin Epidemiol 64(12):1294–1302, PMID: 21803546, https://doi.org/10.1016/j.jclinepi.2011.03.017.

Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. 2011e. GRADE guidelines: 5. Rating the quality of evidence – publication bias. J Clin Epidemiol 64(12):1277–1282, PMID: 21802904, https://doi.org/10.1016/j.jclinepi.2011.01.011.

Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, et al. 2011f. GRADE guidelines: 9. Rating up the quality of evidence. J Clin Epidemiol 64(12):1311–1316, PMID: 21802902, https://doi.org/10.1016/j.jclinepi.2011.06.004.

Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. 2011g. GRADE guidelines: 4. Rating the quality of evidence - study limitations (risk of bias). J Clin Epidemiol 64(4):407–415, PMID: 21247734, https://doi.org/10.1016/j.jclinepi.2010.07.017.

Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. 2011h. GRADE guidelines: 8. Rating the quality of evidence – indirectness. J Clin Epidemiol 64(12):1303–1310, PMID: 21802903, https://doi.org/10.1016/j.jclinepi.2011.04.014.

Guyatt GH, Oxman AD, Sultan S, Brozek J, Glasziou P, Alonso-Coello P, et al. 2011i. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. J Clin Epidemiol 66(2):151–157, PMID: 22542023, https://doi.org/10.1016/j.jclinepi.2012.01.006.

Guzelian PS, Victoroff MS, Halmes NC, James RC, Guzelian CP. 2005. Evidence-based toxicology: a comprehensive framework for causation. Hum Exp Toxicol 24(4):161–201, PMID: 15957536, https://doi.org/10.1191/0960327105ht517oa.

Hardy A, Dorne JL, Aiassa E, Alexander J, Bottex B, Chaudhry Q, et al. 2015. Editorial: increasing robustness, transparency and openness of scientific assessments. EFSA J 13(3):3. , https://doi.org/10.2903/j.efsa.2015.e13031.

HAS (Haute Autorité de Santé). 2013. Niveau de preuve et gradation des recommandations de bonne pratique Haute Autorité de santé [in French]. Saint-Denis, France: Haute Autorité de Santé.

Higgins JPT, Green S (eds). 2011. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. The Cochrane Collaboration. http://handbook.cochrane.org [accessed 11 July 2017].

Hill AB. 1965. The *environment and disease: association or causation?* Proc R Soc Med 58 (5):295–300, PMID: 14283879.

Hope BK, Clarkson JR. 2014. A strategy for using weight-of-evidence methods in ecological risk assessments. Hum Ecol Risk Assess 20(2):290–315, https://doi.org/10.1080/10807039.2013.781849.

Howard BE, Shah R, Walker K, Pelch K, Holmgren S, Thayer K. 2014. Use of text-mining and machine learning to prioritize the results of a complex literature search. In: *53rd annual meeting of the Society of Toxicology, March 23–27, 2014*. Phoenix, AZ.

Hristozov DR, Gottardo S, Cinelli M, Isigonis P, Zabeo A, Critto A, et al. 2014a. Application of a quantitative weight of evidence approach for ranking and prioritising occupational exposure scenarios for titanium dioxide and carbon nanomaterials. Nanotoxicology 8(2):117–131, PMID: 23244341, https://doi.org/10.3109/17435390.2012.760013.

Hristozov DR, Zabeo A, Foran C, Isigonis P, Critto A, Marcomini A, et al. 2014b. A weight of evidence approach for hazard screening of engineered nanomaterials. Nanotoxicology 8(1):72–87, PMID: 23153309, https://doi.org/10.3109/17435390.2012.750695.

IARC (International Agency for Research on Cancer). 2006. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans – Preamble. Lyon, France: World Health Organization International Agency for Research on Cancer.

INCa (Institut National du Cancer). 2015. Nutrition et prévention primaire des cancers: actualisation des données. Boulogne-Billancourt, France: Institut national du cancer. http://www.e-cancer.fr/Expertises-et-publications/Catalogue-des-publications/Nutrition-et-prevention-primaire-des-cancers-actualisation-des-donnees [accessed 11 July 2017].

Kho ME, Brouwers MC. 2012. The systematic review and bibliometric network analysis (SeBriNA) is a new method to contextualize evidence. Part 1: description. J Clin Epidemiol 65(9):1010–1015, PMID: 22742919, https://doi.org/10.1016/j.jclinepi.2012.03.009.

Khosrovyan A, Rodríguez-Romero A, Antequera Ramos M, DelValls TA, Riba I. 2015. Comparative analysis of two weight-of-evidence methodologies for integrated sediment quality assessment. Chemosphere 120:138–144, PMID: 25016337, https://doi.org/10.1016/j.chemosphere.2014.06.043.

Klimisch HJ, Andreae M, Tillmann U. 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regul Toxicol Pharmacol 25(1):1–5, PMID: 9056496, https://doi.org/10.1006/rtph.1996.1076.

Krimsky S. 2005. The weight of scientific evidence in policy and law. Am J Public Health 95(suppl1):S129–S136, PMID: 16030328, https://doi.org/10.2105/AJPH.2004.044727.

Kung J, Chiappelli F, Cajulis OO, Avezova R, Kossan G, Chew L, et al. 2010. From systematic reviews to clinical recommendations for evidence-based health care: validation of revised assessment of multiple systematic reviews (R-AMSTAR) for grading of clinical relevance. Open Dent J 4:84–91, PMID: 21088686, https://doi.org/10.2174/1874210601004020084.

Linkov I, Loney D, Cormier S, Satterstrom FK, Bridges T. 2009. Weight-of-evidence evaluation in environmental assessment: review of qualitative and quantitative approaches. Sci Total Environ 407(19):5199–5205, PMID: 19619890, https://doi.org/10.1016/j.scitotenv.2009.05.004.

Linkov I, Welle P, Loney D, Tkachuk A, Canis L, Kim JB, et al. 2011. Use of multicriteria decision analysis to support weight of evidence evaluation. Risk Anal 31(8):1211–1225, PMID: 21371061, https://doi.org/10.1111/j.1539-6924.2011.01585.x.

Mandrioli D, Silbergeld EK. 2016. Evidence from toxicology: the most essential science for prevention. Environ Health Perspect 124(1):6–11, PMID: 26091173, https://doi.org/10.1289/ehp.1509880.

Marvier M, McCreedy C, Regetz J, Kareiva P. 2007. A meta-analysis of effects of Bt cotton and maize on non-target invertebrates. Science 316(5830):1475–1477, PMID: 17556584, https://doi.org/10.1126/science.1139208.

Marvier M. 2011. Using meta-analysis to inform risk assessment and risk management. J Verbr Lebensm 6(S1):113–118, https://doi.org/10.1007/s00003-011-0675-6.

Meek ME, Bucher JR, Cohen SM, Dellarco V, Hill RN, Lehman-McKeeman LD, et al. 2003. A framework for human relevance analysis of information on carcinogenic

modes of action. Crit Rev Toxicol 33 (6):591–653, PMID: 14727733, https://doi.org/10.1080/713608373.
Meek ME. 2008. Recent developments in frameworks to consider human relevance of hypothesized modes of action for tumours in animals. Environ Mol Mutagen 49(2):110–116, PMID: 18213650, https://doi.org/10.1002/em.20369.
Meek ME, Boobis A, Cote I, Dellarco V, Fotakis G, Munn S, et al. 2014a. New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. J Appl Toxicol 34(1):1–18, PMID: 24166207, https://doi.org/10.1002/jat.2949.
Meek ME, Palermo CM, Bachman AN, North CM, Lewis RJ. 2014b. Mode of action human relevance (species concordance) framework: evolution of the Bradford Hill considerations and comparative analysis of weight of evidence. J Appl Toxicol 34(6):595–606, PMID: 24777878, https://doi.org/10.1002/jat.2984.
Metcalfe DD. 2005. Genetically modified crops and allergenicity. Nat Immunol 6(9):857–860, PMID: 16116460, https://doi.org/10.1038/ni0905-857.
Moher D, Liberati A, Tetzlaff J, Altman DG. The PRISMA Group, 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. PLoS Med 6(7):e1000097, https://doi.org/10.1371/journal.pmed.1000097.
Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. 2015. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. Syst Rev 4(1):1, PMID: 25554246, https://doi.org/10.1186/2046-4053-4-1.
Money CD, Tomenson JA, Penman MG, Boogaard PJ, Lewis RJ. 2013. A systematic approach for evaluating and scoring human data. Regul Toxicol Pharmacol 66(2):241–247, PMID: 23579077, https://doi.org/10.1016/j.yrtph.2013.03.011.
Murad MH, Montori VM, Ioannidis JP, Jaeschke R, Devereaux PJ, Prasad K, et al. 2014. How to read a systematic review and meta-analysis and apply the results to patient care: users' guides to the medical literature. JAMA 312 (2):171–179, PMID: 25005654, https://doi.org/10.1001/jama.2014.5559.
NRC (U.S. National Research Council). 2014. *Review of EPA's Integrated Risk Information System (IRIS) Process*. Washington, DC: The National Academies Press; https://doi.org/https://doi.org/10.17226/18764.
O'Connor D, Green S, Higgins JPT. 2011. Chapter 5: Defining the review question and developing criteria for including studies. In: *Cochrane Handbook of Systematic Reviews of Intervention*. http://handbook-5-1.cochrane.org/front_page.htm [accessed 14 March 2017].
OECD (Organisation for Economic Cooperation and Development). 2014. *Users' Handbook Supplement to The Guidance Document for Developing and Assessing Adverse Outcome Pathways*. Paris: Organisation for Economic Cooperation and Development.
OHAT (Office of Health Assessment and Translation). 2015. *Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration*. Research Triangle Park, NC: OHAT.
Pieper D, Buechter RB, Li L, Prediger B, Eikermann M. 2015. Systematic review found AMSTAR, but not R(evised)-AMSTAR, to have good measurement properties. J Clin Epidemiol 68(5):574–583, PMID: 25638457, https://doi.org/10.1016/j.jclinepi.2014.12.009.
Rhomberg LR, Goodman JE, Bailey LA, Prueitt RL, Beck NB, Bevan C, et al. 2013. A survey of frameworks for best practices in weight-of-evidence analyses. Crit Rev Toxicol 43(9):753–784, PMID: 24040995, https://doi.org/10.3109/10408444.2013.832727.
Rhomberg L. 2015. Hypothesis-based weight of evidence: an approach to assessing causation and its application to regulatory toxicology. Risk Anal 35(6):1114–1124, PMID: 24724710, https://doi.org/10.1111/risa.12206.
Rooney AA, Boyles AL, Wolfe MS, Bucher JR, Thayer KA. 2014. Systematic review and evidence integration for literature-based environmental health science assessments. Environ Health Perspect (122(7):711–718, PMID: 24755067, https://doi.org/10.1289/ehp.1307972.
Rothman KJ, Greenland S. 2005. Causation and causal inference in epidemiology. Am J Public Health 95 Suppl 1(suppl 1):S144–S150, PMID: 16030331, https://doi.org/10.2105/AJPH.2004.059204.
Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. 1996. Evidence based medicine: what it is and what it isn't. BMJ 312(7023):71–72, PMID: 8555924.
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. Brussels: European Commission.
Schleier JJ, Marshall LA, Davis RS, Peterson RK. 2015. A quantitative approach for integrating multiple lines of evidence for the evaluation of environmental health risks. PeerJ 3:e730, PMID: 25648367, https://doi.org/10.7717/peerj.730.
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 189(2):138–144, PMID: 19477248, https://doi.org/10.1016/j.toxlet.2009.05.013.
Schünemann HJ, Oxman AD, Vist GE, Higgins JPT, Deeks JJ, Glasziou P, et al. 2011. Chapter 12: Interpreting results and drawing conclusions. In: *Cochrane Handbook of Systematic Reviews of Intervention*. http://handbook-5-1.cochrane.org/front_page.htm [accessed 14 March 2017].
Shea BJ, Bouter LM, Peterson J, Boers M, Andersson N, Ortiz Z, et al. 2007a. External validation of a measurement tool to assess systematic reviews (AMSTAR). PLoS One 2(12):e1350, PMID: 18159233, https://doi.org/10.1371/journal.pone.0001350.
Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. 2007b. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 7:10, PMID: 17302989, https://doi.org/10.1186/1471-2288-7-10.
Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. 2009. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol 62(10):1013–1020, PMID: 19230606, https://doi.org/10.1016/j.jclinepi.2008.10.009.
Spiegelhalter DJ, Abrams KR, Myles JP. 2004. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Chichester, England: John Wiley & Sons.
Suter GW, Cormier SM. 2011. Why and how to combine evidence in environmental assessments: weighing evidence and building cases. Sci Total Environ 409(8):1406–1417, PMID: 21277006, https://doi.org/10.1016/j.scitotenv.2010.12.029.
Swaen G, van Amelsvoort L. 2009. A weight of evidence approach to causal inference. J Clin Epidemiol 62(3):270–277, PMID: 18834711, https://doi.org/10.1016/j.jclinepi.2008.06.013.
U.S. EPA (U.S. Environmental Protection Agency). 1997. Rules of thumb for superfund remedy selection. Washington, DC: U.S. Environmental Protection Agency.
U.S. EPA. 1998. Guidelines for Ecological Risk Assessment. Washington, DC: U.S. Environmental Protection Agency.
U.S. EPA. 2001. HED Standard Operating Procedure: Executive Summaries for Toxicology Data Evaluation Record (DERs). Washington, DC: U.S. Environmental Protection Agency.
U.S. EPA. 2003. A summary of general assessment factors for evaluating the quality of scientific and technical information. Washington, DC: U.S. Environmental Protection Agency
U.S. EPA. 2014. Framework for Human Health Risk Assessment to Inform Decision Making. edited by EPA Risk Assessment Forum. Washington, DC: U. S. Environmental Protection Agency.
U.S. EPA. 2017. EPA Releases Draft Risk Assessments for Glyphosate. Washington, DC: U.S. Environmental Protection Agency.
U.S. NTP (U.S. National Toxicology Program). 2008. NTP-CERHR Monograph on the Potential Human Reproductive and Developmental Effects of Bisphenol A. TR 08–5994. Research Triangle Park, NC: NIH Publication.
U.S. NTP. 2015. Handbook for Preparing Report on Carcinogens Monographs. Morrisville, NC: U.S. National Toxicology Program.
van Bilsen JH, Ronsmans S, Crevel RW, Rona RJ, Przyrembel H, Penninks AH, et al. 2011. Evaluation of scientific criteria for identifying allergenic foods of public health importance. Regul Toxicol Pharmacol 60(3):281–289, PMID: 20837076, https://doi.org/10.1016/j.yrtph.2010.08.024.
Vinken M. 2013. The adverse outcome pathway concept: a pragmatic tool in toxicology. Toxicology 312:158–165, PMID: 23978457, https://doi.org/10.1016/j.tox.2013.08.011.
Viswanathan M, Ansari MT, Berkman ND, Chang S, Hartling L, McPheeters LM, et al. 2012. *Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions. TR 12-EHC047-EF*. Rockville, MD: Agency for Healthcare Research and Quality.
WHO (World Health Organization). 2012. *Daily Iron and Folic Acid Supplementation in Pregnant Women – Guideline*. Geneva: World Health Organization.
WCRF/AICR (World Cancer Research Fund and American Institute for Cancer Research). 2014. Continuous Update Project Report. Food, Nutrition, Physical Activity, and the Prevention of Ovarian Cancer 2014. http://www.dietandcancerreport.org/cup/cup_resources.php [accessed 11 July 2017].
Weed DL. 2005. Weight of evidence: a review of concept and methods. Risk Anal 25(6):1545–1557, PMID: 16506981, https://doi.org/10.1111/j.1539-6924.2005.00699.x.
Williams MS, Ebel ED, Vose D. 2011. Framework for microbial food-safety risk assessments amenable to Bayesian modelling. Risk Anal 31(4):548–565, PMID: 21105883, https://doi.org/10.1111/j.1539-6924.2010.01532.x.
Woodruff TJ, Sutton P. 2011. An evidence-based medicine methodology to bridge the gap between clinical and environmental health sciences. Health Aff (Millwood) 30(5):931–937, PMID: 21555477, https://doi.org/10.1377/hlthaff.2010.1219.
Woodruff TJ, Sutton P. 2014. The Navigation Guide systematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes. Environ Health Perspect 122(10):1007–1014, PMID: 24968373, https://doi.org/10.1289/ehp.1307175.