



EPA Public Access

Author manuscript

Integr Environ Assess Manag. Author manuscript; available in PMC 2018 November 01.

About author manuscripts

Submit a manuscript

Published in final edited form as:

Integr Environ Assess Manag. 2017 November ; 13(6): 1038–1044. doi:10.1002/ieam.1954.

A WEIGHT OF EVIDENCE FRAMEWORK FOR ENVIRONMENTAL ASSESSMENTS: INFERRING QUALITIES

Glenn Suter^{1,†}, Susan Cormier[‡], and Mace Barron[§]

[†]U.S. Environmental Protection Agency, Office of Research and Development, National Center for Environmental Assessment, 26 W. Martin L. King Drive, Cincinnati, OH 45268, (513) 936-8763

[‡]U.S. Environmental Protection Agency, Office of Research and Development, National Center for Environmental Assessment, 26 W. Martin L. King Drive, Cincinnati, OH 45268

[§]U.S. Environmental Protection Agency, Office of Research and Development, National Health and Environmental Effects Research Laboratory, Gulf Ecology Division, 1 Sabine Island Drive, Gulf Breeze, FL 32561

Abstract

The weighing of heterogeneous evidence such as conventional laboratory toxicity tests, field tests, biomarkers, and community surveys is essential to environmental assessments. Evidence synthesis and weighing is needed to determine causes of observed effects, hazards posed by chemicals or other agents, the completeness of remediation, and other environmental qualities. As part of its guidelines for WoE in ecological assessments, the U.S. Environmental Protection Agency has developed a generally applicable framework. Its basic steps are: assemble evidence, weight the evidence, and weigh the body of evidence. Use of the framework can increase the consistency and rigor of WoE practices and provide greater transparency than ad hoc and narrative-based approaches.

Keywords

Weight of evidence; risk assessment; systematic review; ecological risk; causation

INTRODUCTION

Inferences in environmental assessment often involve multiple and heterogeneous pieces of evidence. For example, inferring the cause of an observed biological impairment could involve evidence derived from conventional laboratory tests, ambient media tests, biomarkers, biological surveys, chemical analyses, and models. Such inferences require weighing the evidence. Although the weighing is often done by unstructured narratives or narratives guided by a list of considerations, an explicit weight-of-evidence (WoE) process can increase the defensibility of results (Weed 2005).

¹Address correspondence to suter.glenn@epa.gov.

Disclaimer—Although this article has undergone Agency review, the views expressed herein are those of the authors and do not necessarily reflect the views or policies of the USEPA.

The USEPA (2016) has developed and adopted WoE guidelines for ecological assessments. The approach is potentially applicable to human health and welfare assessments, but ecological assessors have, in general, been more accepting of the formal weighing of multiple types of evidence. The framework presented here can be used to integrate multiple pieces of evidence and to infer qualities such as causation, teratogenicity, or impairment. However, WoE may also be employed to derive quantities such as benchmark concentrations or half-lives of chemicals (Suter et al. accepted).

WoE to infer qualities employs both qualitative and quantitative methods, and it inevitably requires subjective judgment. While quantitative analysis is essential to generating evidence, there is no way to quantitatively combine the heterogeneous evidence that appears in environmental assessments and serves different purposes. For example, some evidence may indicate that a cause and effect are co-located while other evidence indicates that affected organisms are altered in characteristic ways, so the evidence is not commensurable and cannot be quantitatively combined. In addition, some qualities, particularly causation, have no quantitative metric (Norton et al. 2014). Finally, the properties of evidence that are weighed such as relevance are not quantitative. Hence this framework transparently organizes and presents judgments and does not engage in pseudo-quantification (assigning numbers to qualities) or in ignoring the qualitative aspects of an issue so as to make it quantitative (e.g., limiting assessment of causation to correlations).

This framework is new, but it is derived from the diverse WoE approaches in ecological assessment (Fox 1991; Burton et al. 2002; Chapman 2007; McDonald et al. 2007; Suter and Cormier 2011) and health risk assessment (Rhomberg et al. 2013). Its major elements have been employed in externally peer-reviewed U.S. Environmental Protection Agency (USEPA) assessments. Like the framework and guidelines for ecological risk assessment (USEPA 1992; USEPA 1998), it is intended to provide a common process and terminology that can be adapted and applied to many different assessment problems and contexts. In practice, assessors should apply it flexibly to ensure that each application is fit for purpose.

Basic Concepts and Terminology

Much of the confusion and controversy surrounding WoE results from vague and inconsistent terminology used to describe basic concepts. We begin by clarifying terms and the concepts to which they apply.

An environmental *WoE* is an inferential process that assembles, evaluates, and integrates evidence to perform a technical inference in an assessment. Contrary to some usages, WoE is not a type of assessment. It is an inferential process that may be used throughout an entire assessment, such as causal assessments using the CADDIS framework (<https://www3.epa.gov/caddis>). However, it is used more often to make an inference about a component of an assessment, particularly in the problem formulation phase. For example, should teratogenicity in frogs be an assessment endpoint? Should the chemical of concern be considered bioaccumulative?

Evidence is information that informs inferences regarding a condition, cause, prediction, or outcome. In general, generating evidence requires identifying a relationship that is relevant

to the inference such as increase in chemical concentration relative to the appearance of an effect. Data such as a chemical concentration are not evidence. Concentration might contribute to evidence of the cause of a fish kill, a water quality criteria exceedence, regional background, etc., depending on the inferential context and associated information. This point is emphasized by the term *evidence group* which was defined as the combination of an exposure-response relationship, information concerning environmental conditions that influence that relationship, and either an exposure estimate or a response level (Hope and Clarkson 2014). The basic unit is a *piece of evidence*. Pieces are generally assigned to a *type* such as biotic community surveys, including both exposure and condition data. All of the applicable evidence used to make an inference concerning a proposition constitutes the *body of evidence*.

Not all information provides useful evidence. Evidence must have *explanatory implications* for the quality being inferred, in the context of the issue being assessed. Because most WoE analyses deal with some question of causation, the most generally useful implication is that the evidence indicates one or more of the characteristics of causation (Cormier et al. 2010). For example, a toxicity test may indicate what level of exposure is sufficient and a field survey may demonstrate co-occurrence. Other qualities such as protection, contaminant of concern, impairment, and remediation also have identifiable characteristics that can be used to determine what implications evidence may have for an inference (USEPA 2016).

Evidence is weighted and weighed with respect to its *properties*. The most commonly used properties are *relevance* and *reliability*. These are the properties against which scientific evidence is judged in U.S. courts of law under the Daubert rule (Supreme Court of the United States 1993; Annas 1999) and they are commonly used internationally (Rudén et al. 2017; Moermond et al. 2017). In environmental assessments, relevance expresses the degree of correspondence between the evidence and the assessment endpoint and context to which it is applied. Reliability expresses the degree to which a study is well designed or has other attributes that inspire confidence. Another important property of evidence in most environmental assessments is its *strength*, the degree of differentiation from randomness or from control, background, or reference conditions. In addition to these three properties of pieces or types of evidence, the body of evidence is weighed with respect to *collective properties* potentially including number, coherence, diversity, and absence of signs of bias (Norton et al. 2014; USEPA 2016).

FRAMEWORK

WoE has three basic steps (Suter and Cormier 2011; Figure 1). First, evidence is assembled. Second, the evidence is weighted. Third, the body of evidence for each alternative result is weighed and the weighed alternatives are compared. The framework does not include a planning step or a communication or decision-making step, because WoE is an analytical process embedded in a larger assessment process. In many cases, it is a component of the problem formulation. In particular, WoE is used to determine the hazard posed by the agent of concern. In some cases, such as causal assessments, it carries through the entire assessment to derive a qualitative result (e.g., identification of the cause of impairment as in CADDIS).

Assemble evidence

Successful WoE requires that the relevant evidence be identified and assembled. All scientists have performed literature reviews, but, since the rise of evidence-based-medicine, it has become apparent that systematic reviews can give more complete and unbiased results than informal methods (Bilotta et al. 2014). The method for searching the literature should be specified including the topic and the strategy of the search. Ideally, an information specialist will assist in this effort and the results will be systematically compiled and documented. Nevertheless, the assessors have the primary responsibility to ensure that the full breadth of the topic is identified and that obscure sources such as unpublished data from government agencies or data in prior assessments are found and included.

Assessors may have the opportunity to obtain evidence by designing studies to produce information specific to the case. Case-specific information can fill data gaps in the literature review, can be highly relevant, and can be designed to meet data quality requirements.

Once papers, reports, data sets, and other information sources have been identified and obtained, they are screened. Screening eliminates sources that do not meet minimal criteria for relevance and reliability. Care is taken to eliminate information that is in fact uninformative or even misleading without eliminating marginal information that may contribute to the inference but with low weight.

Sorting studies into distinct categories permits more systematic weighting and weighing. Conventional types of evidence provide the most common categories. A typology for use at contaminated sites includes single-chemical toxicity tests, body burdens, ambient media toxicity tests, biomarkers and pathologies, and biological surveys (Suter 1996; Suter et al. 2000). Other categories may be useful, particularly the characteristics of causation (Cormier et al. 2010; Suter and Cormier 2011).

Having obtained, screened and categorized the available information, assessors derive the evidence by extracting the data or other information from its source, organizing it, and analyzing it to derive useful evidence. The information includes observational and experimental data, derived values such as growth rates of populations, narrative data such as interpretation of behavior as avoidance, and accepted knowledge such as physical laws. The analysis of data to generate evidence includes processes such as partitioning data sets, performing basic statistical analyses such as calculating means and confidence intervals, relating variables with contingency tables or regression analysis, or determining spatial and temporal patterns.

Weight the evidence

Because the various pieces and types of evidence are generally not equally worthy of influence on the inference, it is beneficial to assign weights to the evidence. Weighting involves evaluating the evidence with respect to the properties (relevance, reliability, and strength) and assigning a score that reflects the evaluation. The weights may then be combined to produce an overall weight for each piece or type of evidence.

Relevance has three potential components. (1) Biological relevance—correspondence between the taxa, life stages, and processes measured or observed and the assessment endpoint. (2) Physical/chemical relevance—correspondence between the chemical or physical agent tested or measured and the chemical or physical agent constituting the stressor of concern. (3) Environmental relevance—correspondence between test conditions and conditions at the assessed site or between the environmental conditions in a studied location and in the region or other area of concern.

A strong signal is better differentiated from noise than is a weak signal, so a strong signal is given more weight. Strong evidence shows (1) a large magnitude of difference between a treatment and control in an experiment or between exposed and reference conditions in an observational study, (2) a high degree of association between a putative cause and effect, or (3) a large number of elements in a category of evidence (e.g., multiple biomarkers give evidence of interaction). Strength is a property of the evidence obtained from the results of a study, not the reliability of the design of the study or its methods. Most statistical analyses relate to strength, and it is easier to define standard scoring criteria for strength than for the other properties.

Reliability consists of inherent properties that can make evidence trustworthy. There are many potential subproperties of reliability including: design and execution, abundance of data, minimal confounding, specificity, signs of bias, peer review, transparency, corroboration, consistency, and consilience. Assessors must consider which subproperties are most important to distinguishing the reliability of evidence in their case.

Scoring of the evidence can (1) reduce the ambiguity of verbal evaluations of evidence, (2) clarify the pattern of results, and (3) facilitate the combining of weights across properties and, in the next step, across bodies of evidence for alternative hypotheses. We recommend a system of +, – and 0, symbols to represent evidence that, respectively, supports, weakens, or has no effect on the credibility of a hypothesis. More symbols represent greater weight. For example, +++ could represent extremely reliable supporting evidence, and 0 could represent ambiguous relevance. When possible and appropriate, criteria for the scores should be specified. This scoring system has been used in epidemiology, ecological causal assessment, and risk assessment and is recommended in the WoE guidelines for ecological assessments (USEPA 2016).

The standard tool for weighting evidence is the scoring table (Table 1). It displays individual pieces of evidence or categories of evidence for a hypothesis, such as the types shown here, scored with respect to properties or subproperties and the overall weight. The first line presents hypothetical scores that illustrate an important point. The score for the overall weight is not the average of the property scores. In this case, the overall score is low because, even though the laboratory tests show a strong relationship and are extremely reliable, because the species, conditions, or responses have low relevance to the case, the overall weight of laboratory tests would be low. These judgements concerning heterogeneous bodies of evidence are subjective, but, when many similar assessments are performed, guidance on scoring can be developed and used in future assessments.

The weighting step is often skipped, but that practice should be justified. Omission of the weighting step is justified if all evidence is highly relevant, reliable and strong. Highly relevant and reliable evidence may be obtained when studies are conducted for the assessment and are well designed and conducted. Uniformly strong evidence is obtained when there are strong relationships between causal agents and effects.

Weigh the body of evidence

Having assigned weights to the evidence, the bodies of evidence for each hypothesis are weighed. The first step is to integrate the weighted evidence and generate a weighed body of evidence. Once again, the primary tool is a table, this time called a weight-of-evidence table (Table 2). The organization of the table depends on the application. Pieces or types of evidence may carry over directly from the scoring table (Table 1). However, it is often useful to combine them into types (e.g., all laboratory studies and all field studies) or into characteristics of causation or other logical implications of the evidence as in Table 2. Logical implications explain how the evidence supports or weakens a hypothesis. The bodies of evidence also have collective properties: (1) the number of pieces or categories of evidence, (2) the coherence or logical consistency of the evidence, (3) the absence of bias as indicated by consistent results across funders and types of investigators, and (4) diversity of the biota, responses, and conditions. For example, both laboratory toxicity tests of specific agents and effluent toxicity tests (Table 1) could contribute to the weight of evidence for sufficiency of a chemical pollutant as a hypothesized cause (a causal characteristic) and could contribute to the number and diversity of evidence (collective properties; Table 2).

Interpretation is the step in which the hypothesis that is best supported by the evidence is identified and confidence in its truth is evaluated. As in civil legal proceedings and typically in public policy in the U.S., the standard of proof is the preponderance of the evidence and scoring of WoE tables can reveal which hypothesis meets that criterion (e.g., which has most of the evidence with one or more + scores but no – scores). However, there is another important criterion, adequacy. Ideally, data quality objectives will be developed during problem formulation and adequate evidence will be generated. However, many real assessments depend on a meager body of existing evidence. Further, some hypotheses are extraordinary (e.g., it has not been previously demonstrated or is implausible) and therefore require extraordinary evidence. In either of those cases, the adequacy of the evidence is judged alongside the judgement of which hypothesis is best supported.

The interpretation of evidence often involves applying a three-value logic to each hypothesis (e.g., yes, no, maybe; true, false, uncertain; high, low, intermediate). In the easiest situation, evidence with overwhelming weight supports one hypothesis and the job is done. In the absence of a positive conclusion, any hypotheses that are clearly false can be eliminated from further consideration. Uncertain hypotheses get carried forward to further evidence generation or assessment. Whatever the result, the inference must be explained. That is, assessors describe what the bodies of evidence say about the hypotheses, what is unknown, and how much confidence the evidence provides.

In some cases, the WoE is ambiguous or there are significant discrepancies within the bodies of evidence. In such cases, one can stop the assessment process and request more data or

perform further data analyses to try to resolve the problem. However, it is commonly helpful to reformulate the problem. This could involve redefining the endpoint or the sources and agents considered, reinterpreting the evidence, reformulating hypotheses, or adding hypotheses. These changes may be justified by the deeper understanding of the case that comes with having performed the WoE analysis. However, reformulating the problem post hoc can lead to bias or self-deception in the interest of reaching a conclusion. If possible, independent evidence should be found or generated to confirm that the changes and resulting conclusions are justified.

DISCUSSION

Most commonly, WoE refers to a summary interpretation or synthesis of the evidence without any actual weighting and weighing (Weed 2005; Linkov et al. 2009). In contrast, this formal WoE approach involves a framework with three basic steps: assembly, weighting and weighing of the evidence. Beyond that level or elaboration, inherent conflicts arise between the desire for clear, consistent and formal methods that minimize subjective judgment and the need for sufficient flexibility to suite the WoE process to the problem. The elaborated framework presented here is based on experience in the USEPA and elsewhere in applying WoE to various ecological assessment problems and is believed to be generally useful. Examples of experience in formal WoE include the many causal assessments at <https://www.epa.gov/caddis>, ecological risk assessments such as the Bristol Bay watershed assessment (USEPA 2014), Superfund remedial investigations (Johnston et al. 2002), and assessments of contaminated sediments in Canada (COA Sediment Task Group 2008).

This WoE approach minimizes errors in judgment by providing:

- A standard inferential structure in the form of a framework so that steps are not skipped without reason
- Clearly defined alternative hypotheses concerning the qualities being analyzed
- A standard set of properties for weighting the evidence
- A scoring system that makes the weighting explicit without pseudo-quantification (i.e., we do not recommend assigning numbers to qualities and combining them as if they were quantities with common units)
- Verbal explanations of the relationships of the scored evidence with the hypotheses

A formal WoE framework and methodology has the particular advantage of dealing with the problem of narrative fallacy, the tendency to form a narrative from any set of information that makes sense of it, even when there is no causal relationship (Kahnman 2011; Norton et al. 2015). This WoE methodology provides a non-narrative method of weighing a body of evidence and reaching a conclusion. Having reached a conclusion by non-narrative means, one is free to use a narrative to explain any relationships that have been identified by WoE (Norton et al. 2015).

In addition to reducing errors in judgment, the framework increases the credibility of WoE by making the process more transparent. Unlike traditional narrative reviews, this approach shows the reader how the evidence was assembled, weighted and weighed and affords reviewers and stakeholders the opportunity to explore alternative judgments. Hence, it is responsive to the U.S. National Research Council's call for the USEPA to develop a more transparent and defensible methodology for WoE (NAS 2011).

This approach is qualitative in the sense that it weighs evidence to infer qualities and also in the sense that it requires qualitative inference. There is no numerical solution to causation, to impairment, to membership in an adverse outcome pathway, to remediation, etc. Similarly, there is no quantitative method to combine incommensurable evidence such as the symptoms displayed by dying fish and the spatial distribution of a contaminant relative to a fish kill, but both pieces of evidence are clearly relevant to inferring the cause of the kill. Statistical analyses and modeling are, however, essential to generating evidence and to characterizing some properties of evidence. In addition, this assessment of qualities can help identify the assessment focus (e.g., assessment endpoint, hazard) prior to weighing evidence to derive a quantity (e.g., criterion, remedial goal) as an output of the assessment (Suter et al., accepted).

Potential benefits of WoE

1. WoE provides greater confidence in results obtained by considering all relevant and reliable evidence. For example, it is not uncommon for causal assessments to consider only statistical evidence of co-occurrence of an effect and its potential causes. That approach provides much less confidence than one that also considers evidence of temporal sequence, specific alterations, and other characteristics of causal relationships. In many cases, no single type of evidence is sufficient to reach a conclusion.
2. A formal WoE method increases defensibility by demonstrating that all relevant evidence has been considered and no credible evidence has been arbitrarily dismissed. Without an explicit process planned in advance, reviewers might criticize or even dismiss an assessment for excluding data or evidence that they believe deserved greater consideration.
3. Transparency of the processes also increases defensibility of assessments. A formal WoE method enables reviewers and readers to understand and critique the processes of assembling, weighting, and weighing the evidence.

Potential challenges and solutions

1. A formal WoE process can require considerable time and effort, which could lead to performance of fewer assessments or delayed decisions. The solution to this challenge is to tailor the WoE method to the assessment. For example, in some circumstances the full three-step WoE process is not necessary or practical, but a two-step process is useful. The systematic assembly of evidence is always appropriate, but weighting might be unnecessary if all evidence is equivalent. Weighing a body of evidence is unnecessary if only one relevant piece of

evidence is available, although it may be appropriate to weight its relevance, reliability and strength.

2. A formal WoE process may be unclear to readers or may discourage engagement. To avoid this, conclusions of each section may be presented as a short narrative up front, and details of evidence and analysis may follow, for those who are engaged by the conclusions. If scoring and WoE tables are large or numerous, they may be presented in an appendix.

CONCLUSIONS

Since ecological risk assessment was formalized in the 1990s, more types of evidence are available to environmental assessors and assessors have more experience with weighing various types of evidence. As a result, WoE appears in environmental assessments, either as an organizing principle for the entire assessment or to inform individual steps in the process (Figure 2). The time has come for WoE in environmental assessments to be performed in a more consistent and rigorous manner. The basic WoE framework can help to achieve that goal without being so prescriptive or onerous that it inhibits the production of useful environmental assessments.

Acknowledgments

This paper draws on the contributions of the members of the USEPA Risk Assessment Forum's Technical Panel on Weight of Evidence: Mace Barron, David Charters, Susan Cormier, Karen Eisenreich, Kris Garber, Wade Lehman, Scott Lynn, Chris Sarsony and Glenn Suter and the panel's coordinator, Lawrence Martin. Beth Owens, Glenn Rice, and anonymous reviewers provided valuable comments on earlier versions of the manuscript. This paper is adapted from (USEPA 2016) which was reviewed by external peers (Bryan Brooks, Peter Chapman, Valery Forbes, Igor Linkov) and by many USEPA scientists and managers. The authors received no financial support other than their salaries.

References

- Annas GJ. Burden of proof: Judging science and protecting public health in (and out of) the courtroom. *Pub Health Pol Forum*. 1999; 89:490–493.
- Bilotta G, Milner A, Boyd I. On the use of systematic reviews to inform environmental policies. *Environ Sci Pol*. 2014; 42:67–77.
- Burton GA, Chapman PM, Smith EP. Weight-of-evidence approaches for assessing ecosystem impairment. *Human & Ecol Risk Assess*. 2002; 8:1657–1673.
- Chapman PM. Determining when contamination is pollution - weight of evidence determinations for sediments and effluents. *Environ Int*. 2007; 33:492–501. [PubMed: 17027966]
- COA Sediment Task Group. Canada-Ontario decision-making framework for assessment of Great Lakes contaminated sediments. Environment Canada and Ontario Ministry of the Environment; 2008.
- Cormier SM, Suter GW II. A framework for fully integrating environmental assessment. *Environ Manage*. 2008; 42:543–556. [PubMed: 18506517]
- Cormier SM, Suter GW II, Norton SB. Causal characteristics for ecotoxicology. *Hum Ecol Risk Assess*. 2010; 16:53–73.
- Fox GA. Practical causal inference for ecotoxicologists. *J Toxicol Environ Health*. 1991; 33:359–374. [PubMed: 1875428]
- Hope BK, Clarkson JR. A strategy for using weight-of-evidence methods in ecological risk assessments. *Hum Ecol Risk Assess*. 2014; 20:290–315.

- Johnston RK, Munns WR, Tyler PL, Marajh-Whittemore PR, Finkelstein K, Munney K, Short FT, Melville A, Hahn SP. Weighing the evidence of ecological risk from chemical contamination in the estuarine environment adjacent to the Portsmouth Naval Shipyard, Kittery, Maine, USA. *Environ Toxicol Chem.* 2002; 21:182–194. [PubMed: 11804053]
- Kahnman, D. *Thinking fast and slow*. Penguin Books; New York: 2011.
- Linkov I, Long EB, Cormier SM, Satterstrom FK, Bridges T. Weight-of-evidence evaluation in environmental assessment: Review of qualitative and quantitative approaches. *Sci Total Env.* 2009; 497:5199–5205.
- McDonald BG, DeBruyn AMH, Wernick BG, Patterson L, Pellerin N, Chapman PM. Design and application of a transparent and scalable weight-of-evidence framework: an example from Wabamun Lake, Alberta, Canada. *Integr Environ Assess Manag.* 2007; 3:476–483. [PubMed: 18046796]
- Moermond C, Beasley A, Breton R, Junghans M, Laskowski R, Solomon K, Zahner H. Assessing the reliability of ecotoxicological studies: An overview of current needs and approaches. *Integr Environ Assess Manag.* 2017; 13 in press.
- [NAS] National Academy of Sciences. Review of the Environmental Protection Agency’s draft IRIS assessment of formaldehyde. Washington, D.C: National Academy of Sciences; 2011.
- Norton, SB, Cormier, SM., Suter, GW., editors. *Ecological causal assessment*. CRC Press; Boca Raton, FL: p. 2015
- Rhomberg LR, Goodman JE, Bailey LA, Prueitt RL, Beck NB, Bevan C, et al. A survey of frameworks for best practices in weight-of-evidence analyses. *Crit Rev Toxicol.* 2013; 43:753–784. [PubMed: 24040995]
- Rudén C, Adams J, Ågerstrand M, Brock TCM, Poulsen V, Schlekot CE, Wheeler JR, Henry TR. Assessing the relevance of ecological studies for regulatory decision making. *Integr Environ Assess Manag.* 2017; 13 in press.
- Supreme Court of the United States. *Daubert v Merrell Dow Pharmaceuticals*. 509 U.S. 579, 113 S.Ct. 2786, 125 L.Ed. 2d 469, (U.S. 2828, 1993) (NO. 92-102). 1993.
- Suter, GW, II. Risk characterization for ecological risk assessment of contaminated sites. Oak Ridge, TN: Oak Ridge National Laboratory; 1996. Report no. ES/ER/TM-200
- Suter, GW., II, Efrogmson, RA., Sample, BE., Jones, DS. *Ecological risk assessment for contaminated sites*. Lewis Publishers; Boca Raton, FL: 2000.
- Suter GW II, Cormier S. Why and how to combine evidence in environmental assessments: Weighing evidence and building cases. *Sci Total Environ.* 2011; 409:1406–1417. [PubMed: 21277006]
- Suter GW II, Cormier SM, Barron MG. accepted. A weight of evidence framework for environmental assessments: Inferring quantities. *Integr Environ Assess Manag.*
- [USEPA] United States Environmental Protection Agency. Framework for ecological risk assessment. Washington, D.C: Risk Assessment Forum; 1992. EPA/630/R-92/001
- [USEPA] United States Environmental Protection Agency. Guidelines for ecological risk assessment. Washington, D.C: Risk Assessment Forum; 1998. EPA/630/R-95/002F
- [USEPA] United States Environmental Protection Agency. Integrating ecological assessment and decision-making at EPA: A path forward Results of a colloquium in response to Science Advisory Board and National Research Council recommendations. Washington, D.C: Risk Assessment Forum; 2010. Report no. EPAS/100/R-10/004
- [USEPA] United States Environmental Protection Agency. An assessment of potential mining impacts on salmon ecosystems of Bristol Bay, Alaska. Washington, D.C: U.S. Environmental Protection Agency; 2014. EPA 910-R-14-001A
- [USEPA] United States Environmental Protection Agency. Weight of evidence in ecological assessment. Washington, D.C: Risk Assessment Forum; 2016. Report no. EPA/100/R-16/001
- Weed DL. Weight of evidence: A review of concept and methods. *Risk Anal.* 2005; 25:1545–1557. [PubMed: 16506981]

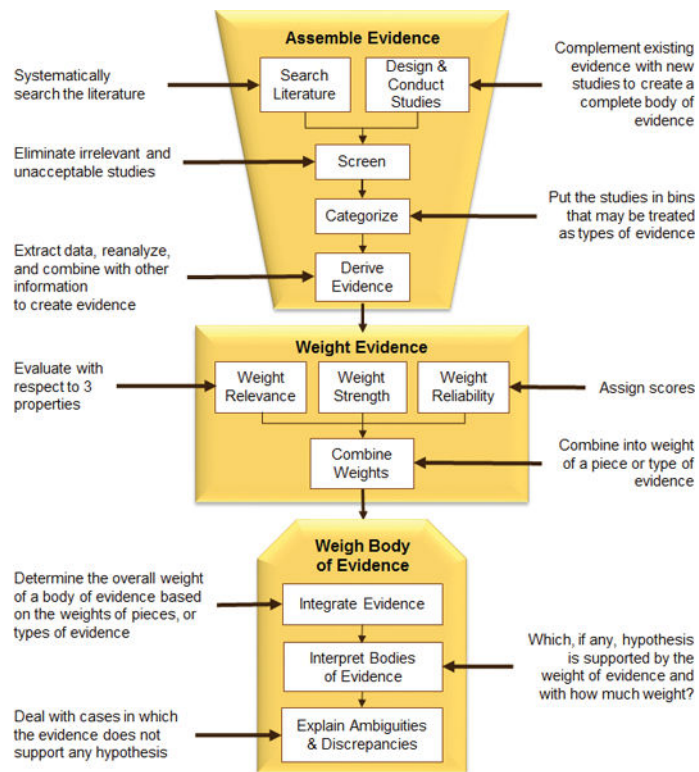


Figure 1. The detailed framework for the qualitative WoE process with explanatory annotations (USEPA 2016).

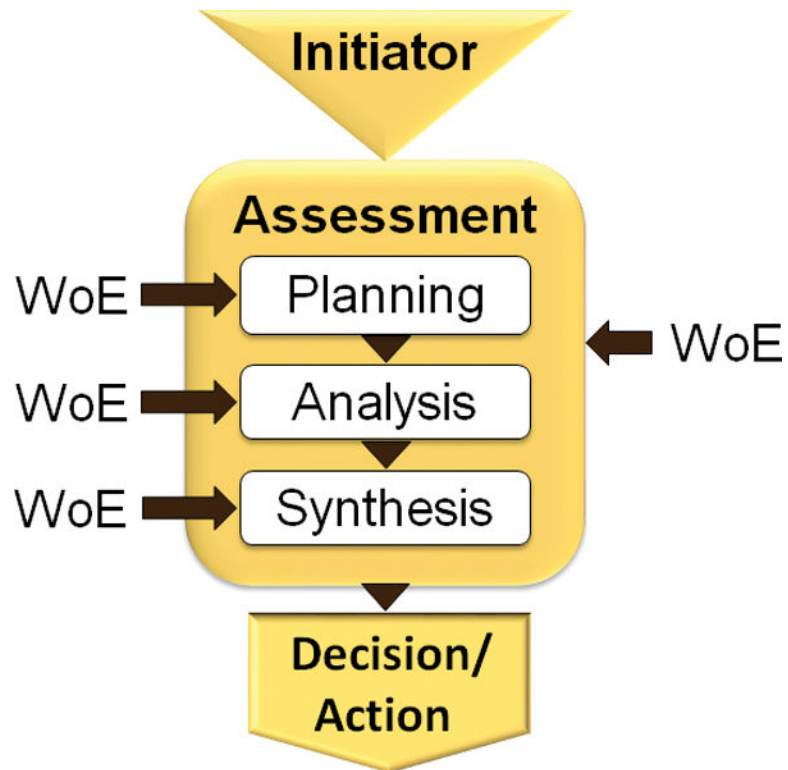


Figure 2.

A basic framework for all types of environmental assessments (Cormier and Suter 2008; USEPA 2010). Weight of evidence can contribute to one or more individual steps (WoE on the left) or can be the basis for the entire assessment (WoE on the right) (USEPA 2016).

Table 1

Generic scoring table for a hypothesis with evidence aggregated into conventional types, with first line completed with hypothetical scores. The overall weight is positive but low because the test relevance is low. From USEPA (2016).

Types of Evidence	Relevance	Strength	Reliability	Overall Weight
Laboratory toxicity tests of a specific agent	+	++	+++	+
Effluent toxicity tests				
Ambient media toxicity tests				
Field biological surveys				
Field biomarkers and organ or whole-body concentrations				
Field symptoms				

Table 2

A generic weight of evidence table for n alternative causal hypotheses ($H_1, H_2, \dots H_n$), based on causal characteristics and collective properties of the bodies of evidence USEPA (2016). The cells would be scored with the +, -, 0 system. Terms are defined in USEPA (2016).

	Combined Weight		
	H ₁	H ₂	H _n
Characteristics of Causation			
Co-occurrence			
Sufficiency			
Time order			
Interaction			
Specific alteration			
Antecedents			
Body of Evidence, Collective Properties			
Number			
Coherence			
Absence of bias			
Diversity			
Integrated WoE			
WoE for the hypothesis			

EPA Author Manuscript

EPA Author Manuscript

EPA Author Manuscript