Society of Environmental Toxicology and Chemistry
Technical Issue Paper

Weight-of-Evidence in Environmental Risk Assessment of Chemicals

The phrase weight-of-evidence can refer to a concept or a procedure. As a concept, it refers to situations where several pieces of evidence are used to reach a conclusion. As a procedure, it refers to the process of assembling, weighing, and evaluating evidence to come to a scientifically defensible conclusion. The weight-of-evidence approach is used when scientific questions can only be answered by using several different pieces of evidence.

The weight-of-evidence framework represents the process of assembling, weighing, and evaluating information to come to a scientifically defensible conclusion.

Environmental risk assessment of chemicals [1] is the practice of determining the nature, likelihood, and magnitude of harmful effects occurring to humans and ecosystems from exposure to current or future chemicals. Here, we focus on the procedural aspects of applying the weight-of-evidence approach to environmental assessments of chemicals.

Why and When Is It Done?

Some scientific questions can be easily answered, such as information about length or temperature. Other scientific questions require several different pieces of evidence to answer. Environmental scientists often deal with the second type of question. They need to present a conclusion that is best supported by the available body of evidence to provide a science-based recommendation to environmental decision makers. They use weight-of-evidence to collectively assess available research to arrive at that conclusion.

When conducting environmental risk assessments, scientists use weight-of-evidence to inform decisions such as: Should contaminated sediment be dredged; should a certain product be allowed on the market; or, what limit should be set for a chemical concentration in air or water? Assessments support such decisions by answering scientific questions such as: Are the chemicals at those concentrations likely to harm human or ecosystem health?

When the decision is made to use a weight-of-evidence approach in an environmental risk assessment, it is important to introduce basic concepts of weight-of-evidence early in the process during discussions with stakeholders, including the local population, subject matter experts, and regulators. These discussions would typically include the reason behind the decision to use a weight-of-evidence approach and how it will be implemented, including identifying the lines of evidence that will be used and the criteria for weighing them.

[1] View the SETAC Technical Issue Paper: Environmental Risk Assessment of Chemicals at https://www.setac.org/page/TIPS.

Why Are There Multiple Pieces of Evidence?

In an environmental assessment of chemicals, there are multiple pieces of evidence that can be used because there are so many ways to study the impact of chemicals in the environment. We may study chemicals in the laboratory or in the field (the environment). We may study chemicals one at a time or chemical mixtures. We may estimate exposure to chemicals by measuring concentrations in the field or by using mathematical models to predict concentrations. We may estimate exposure to chemicals in terms of concentrations in various environmental matrices (such as water, air, or soil), or in plants and animals. Each of these different types of studies produces evidence that has its own strengths and weaknesses, which need to be considered.

In addition, we may have results from multiple studies of the same type. Examples of replicate studies range from multiple studies of the effects of a chemical on salmon eggs to multiple studies of the effects of oil spills on shore-line ecosystems. When multiple studies of the same type give the same result, they make us more confident. When the results differ, we must understand why. That is also part of the weight-of-evidence process.

So, how do we provide a scientific answer to a question about the possibility of a chemical harming human health or the environment? We might imagine that a laboratory experiment or particular measurements of an ecosystem (such as a chemical concentration or density and diversity of a species) would provide a clear, reliable answer, but, most often, they do not do so by themselves.

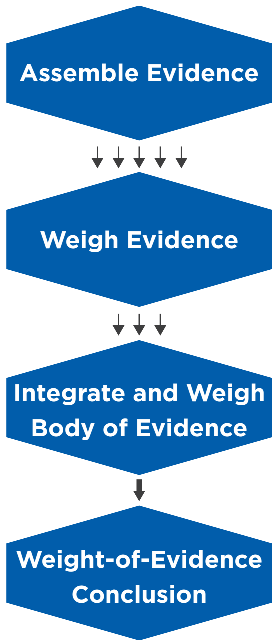
There are typically multiple ways to assess chemical effects in the environment but not one of them can give a complete answer. Consider the following scenarios to evaluate contaminated sediment:

1. We might collect the contaminated sediment and analyze it to determine the concentrations of some chemicals in the sediment, then compare our results to concentrations considered to represent a risk of adverse effects based on controlled laboratory toxicity tests. This process is clear and well understood; however, it lacks realism; it does not account for the combined effects of multiple chemicals; the forms of the chemicals; the chemicals that could be there but were not analyzed; the sediment's physical properties (such as grain size); the differences between the few test species in laboratories and the many species in the ecosystem; and their diverse life cycles, responses to chemicals, and interactions with one another.
2. We might collect the contaminated sediment and test its effects on laboratory organisms. This allows for the evaluation of the specific chemical mixture present in the sediment and accounts for the properties of the sediment, but other disadvantages of using simple laboratory tests remain, such as use of laboratory species that do not reflect native species found in the natural environment.
3. We could take measurements of the biota in a contaminated area (such as measurement of abundance and diversity) and a similar but uncontaminated area and compare them. For example, we could count the number of shellfish species that are growing in a contaminated area versus an uncontaminated area. This method provides site-specific information on environmental health of the sediment, but the complexity of the real world means that we cannot be sure why there are apparent differences or whether we have measured the right differences.

Each of these three scenarios is a line of evidence. Each line of evidence has its own strengths and weaknesses, and by obtaining multiple lines of evidence and weighing and evaluating their results, we can find the conclusion that is best supported by all evidence. This is a weight-of-evidence process.

How Is It Done?

Weight-of-evidence is a process that may be performed in various ways depending on the complexity of the issue, the amount of evidence, and the potential for harm. In any case, weight-of-evidence has three basic steps:



Identify, filter, and summarize available lines of evidence

Assign each piece of evidence a weight based on the evidence's strength, relevance, and reliability.

Evaluate the lines of evidence together as a whole and assess consistency to identify the conclusion.

We begin assembling evidence by searching the literature and consulting stakeholders to determine what is known that is potentially pertinent. When possible and necessary, we perform new studies to fill gaps in the available information. Once obtained, we then assess the evidence and give it a weight. Finally, we integrate and weigh all the evidence (the body of evidence) collectively.

A line of evidence is assigned a weight based on its relevance, strength, and reliability.

In the second step, we apply weight to each line of evidence to determine how much influence it should have on the outcome or conclusion based on its relevance, strength, and reliability to the specific question being considered. Some evidence is more relevant than others. Relevant evidence could exemplify the appropriate species (such as a saltwater organism if we are interested in marine environments), the relevant types of effects (such as reduced growth or death), or represent the applicable conditions (such as typical concentrations in a stream). Some evidence is stronger than others. Strong evidence shows a clear differentiation between exposed systems and controls, reference (un-impacted), or randomness, such as a strong exposure–response relationship in a laboratory test or large differences between locations (contaminated vs. uncontaminated). Similarly, some evidence is more reliable than others. Reliable evidence comes from abundant data obtained using good scientific practives, which are clearly reported. In summary, a weight is applied to each line of evidence based on these considerations.

Weights may be represented in a number of different ways from more qualitative based on best professional judgements to the use of complex quantitative methods. The most important factor is transparency about how these weights are assigned and used. Weights can be numeric values on a scale, or words such as "highly" or "weakly," "positive" or "negative," or symbols such as +, -, and 0. At this point in a weight-of-evidence assessment, it is sometimes helpful to construct a weight-of-evidence matrix. The format of weight-of-evidence matrices should be suited to the assessment and should clearly communicate the transparency, consistency, and reasonableness of interpretation. This process of weighing evidence works best when it can be done *before* data are collected and agreed upon by all interested and informed stakeholders.

Example of a weight-of-evidence matrix

Line of Evidence	Relevance	Strength	Reliability
1.	+	+	-
2.	++	0	+++
3.	-	++	+

Legend:

+ Relevant/strong/reliable line of evidence

- NOT relevant/strong/reliable line of evidence

0 Neutral

Finally, we evaluate the entire body of evidence, considering each weighted line of evidence, to generate a balanced, weight-of-evidence interpretation for the question at hand. This analysis step can be challenging, but the weight-of-evidence framework provides a systematic approach to get it done. It begins by integrating all of the evidence that has been assembled and weighted for a question. Then, the integrated evidence is interpreted. We use a logical process to compare the evidence against the original question and determine if there is a single clear answer.

At times, a body of evidence will appear to be inconsistent. In such cases, the interpretation may explain the inconsistencies. For example, if a chemical concentration in sediment at a wetland is at a level shown to cause toxicity in a laboratory test where that chemical has been added to an artificial laboratory sediment, but a test of the actual contaminated sediment in the laboratory shows no toxicity, it is likely that the chemical in the actual contaminated sediment is in a form that is less toxic. However, if the actual sediment causes harm but none of the individual chemicals do so, then combined toxic effects are likely. These interpretations depend on knowledge and experience, preferably applied by a team of scientists with diverse expertise.

When there are inconsistencies among the lines of evidence, they should be interpreted based on knowledge and experience, preferably applied by a team of scientists with diverse expertise.

If the question cannot be answered by the available evidence, the assessors must reconsider other ways to apply information in the environmental risk assessment. This involves reconsidering the original question. Are we looking at the wrong effects, sources, chemicals or other stressors, or relationships? For example, if the effect of concern is the loss of half the species in a sediment community, it may be that some were eliminated by one cause and others were eliminated by a second cause. These reconsiderations can lead to repeating the weight-of-evidence process after obtaining critical missing evidence, or with a better definition of the question.

Uses of Weight-of-Evidence in Risk Assessment of Chemicals

Weight-of-evidence can be used to assess the likelihood that existing conditions are causing adverse effects or that a hypothetical future action or condition will cause adverse effects. Examples of existing conditions include contaminated sites, existing effluents, and current uses of a chemical. Future conditions result from permitting the use of a new chemical product or permitting an effluent that contains certain chemicals. In both types of risk assessments, multiple pieces of evidence may be weighed, but existing conditions can provide more diverse evidence. Weight-of-evidence may be used to derive quantitative benchmarks, such as an air quality standard or a cleanup goal based in part on a review of the toxicity information that is available for each specific chemical. However, it is most often used to derive a qualitative conclusion, such as whether a chemical accidently released into a stream is the cause of adverse effects to fish in the stream or whether a chemical may cause birth defects.

It is often clear from the preliminary stage of an environmental risk assessment—the problem formulation stage—that a weight-of-evidence approach is needed. During the problem formulation stage, the risk assessor identifies assessment endpoints (the environmental entities and attributes being assessed, for example, successful reproduction of song birds)

CC BY

and measurement endpoints related to it (measured attributes related to the assessment endpoint, for example, successful reproduction and fledging of a specific song bird from monitored nest boxes). Often, there are multiple measurement endpoints for one assessment endpoint and that is where a weight-of-evidence approach may be applied to integrate the results to come to a conclusion. Essentially, all measurement endpoints serve as lines of evidence. Therefore, the problem formulation stage of an environmental risk assessment provides an excellent opportunity for the risk assessor to transparently define the weight-of-evidence approach that will be used to draw the conclusion, before conducting the risk assessment. This includes defining the lines of evidence, the weighing schemes, and how the data will be integrated.

A robust problem formulation of an environmental risk assessment clearly defines the weight-of-evidence approach to be applied.

The final step of an environmental risk assessment—the risk characterization step—integrates both an exposure assessment and a hazard (effect) assessment. Therefore, in a risk assessment, lines of evidence for exposure must always be paired with lines of evidence for effects, and the weight-of-evidence approach must be applied to both exposure and effects lines of evidence concurrently.

In environmental risk assessment of chemicals, the weight-of-evidence approach should be applied to both exposure and hazard assessment and effects!

Weight-of-Evidence in Chemical Exposure Assessment

Exposure assessment is the component of an environmental risk assessment of a chemical that determines how much, for how long, and how often a chemical comes into contact with a receptor, such as a plant, an animal, or a human. A chemical in water may reach a bird through the bird drinking the water, swimming in the water, or its food—the plants, insects or fish that have been in contact with the chemical in the water. The bird's exposure can be estimated in many ways, for example through a combination of measured and estimated concentrations in water, food items, bird feathers and other bird tissue. It can be estimated for an individual, modeled for a population of individuals, or sampled from a population of individuals. In situations where the assessment is made for a chemical prior to its use and release in the environment, measured concentrations are unavailable, so exposures are typically estimated using predictive tools and models. The quantity and types of measurements available for the exposure assessment will be different and will need to be weighed to arrive at robust conclusions regarding exposure.

Weight-of-Evidence in Chemical Hazard Assessment

Chemical hazard assessment describes both the environmental fate properties of the chemical and its ability to cause harm (toxicity) and is a component of the problem formulation for an environmental risk assessment of a chemical. That is, before quantifying the risks, we must determine a chemical's fate in the environment, how organisms are exposed to the chemical, and what types of effects may result at specific exposure levels. A chemical may dissolve in water and enter a fish's gills, or it may become part of the food web should a fish consume it depending upon the chemicals' physical and chemical characteristics, among other factors. The fish may die due to toxic effects or may accumulate the chemical and transfer it to its eggs, causing deformities in the young, depending on the possible expressions of a chemical's toxicity to fish.

Weight-of-Evidence in Chemical Environmental Fate Properties

Some chemicals are likely to degrade when exposed to sunlight while other are more toxic with exposure to light. Some chemicals are likely to partition into air while others sink into sediment. Some chemicals have a high affinity to soils and bind with them, making transport to groundwater unlikely, while others easily percolate to the groundwater from soil. The environmental fate properties of a chemical dictate how it behaves in the environment and can help us predict its environmental breakdown and transport. Examples of these properties include rate of degradation or potential for bioaccumulation in the environment. These properties can be experimentally measured, can be predicted based on the structure of the chemical, or

CC BY

can be estimated based on other properties for which there are field or laboratory data. The various ways these properties can be identified constitute lines of evidence that should be considered together to come to a conclusion about the chemical property.

Weight-of-Evidence for Toxicity Characterization

An understanding of chemical toxicity is essential to environmental assessment and protection. Chemical toxicity is best described as the relationship between the amount of a chemical and the amount needed to cause harm (often described as a dose–response or concentration–response relationship). Chemical toxicity can also be described by various types of toxicity values which could be estimates of exposure doses or concentrations that are not likely to cause appreciable risk of harm. In a risk assessment, toxicity estimates must be identified for specific chemicals and taxa, and there are many ways to derive these estimates from various laboratory, field and modeling studies. Therefore, the various data need to be weighed and integrated to identify an appropriate toxicity estimate.

Rarely is there perfect information available to characterize the toxicity of a specific chemical to a specific species. However, data from a variety of related studies may be available. Differences in the studies may be the result of variation in the chemical form, the animal species, the life stages of the animal, the length of study, the observed effects, the ways of exposing the animal, sampling and analytical measurement techniques, and other aspects related to the way the data were collected (study design). Results may even be contradictory, and there is also the possibility for false positive and negative responses. To help search for meaningful patterns in these studies, weight-of-evidence criteria can be applied to ascertain which results are more informative than others in the context of the risk assessment being conducted.

Example: Some factors to be considered when assigning a weight to toxicity tests used as lines of evidence to identify a toxicity value.

- **Relevance**
  - Chemical state and range in concentration used in the laboratory toxicity test compared with the chemical in the field
  - Type and life stage of test species compared with the species of interest
  - Conditions of laboratory test (example, length of exposure) compared with the field
  - Sensitivity of the test species compared with the species of interest
  - Route of exposure (example, oral, inhalation) in the test compared with the field
  - Test endpoint (example: weight loss, growth) compared with the effect of concern in the field
- **Strength**
  - Strength of association between the exposure and effect
  - Study power (example, appropriate sample size)
- **Reliability**
  - Size of dataset
  - Diversity of information
  - Adherence to Good Laboratory Practices (well-documented)
  - Use of established peer-reviewed methods
  - Use of analytical chemistry techniques to measure concentrations
  - Transparency of the study (provides enough details and discloses potential conflicts of interest)
  - Corroboration of data from one study to other results
  - Reproducibility of the study (can be repeated with similar results)

CC BY

Weight-of-Evidence in Environmental Risk Assessment of New Chemicals or New Sources

To protect environmental quality, regulatory agencies often set chemical concentrations as criteria for standards such as air quality standards or soil screening levels. These concentrations are based on results of prospective environmental assessments to evaluate whether the chemicals will cause certain effects at certain concentrations. When conducting such assessments, we encounter multiple lines of evidence for both the amount of exposure expected and the toxicity of the chemical. For example, for exposure, lines of evidence could include estimated concentrations based on specific uses as well as measured concentrations in controlled application studies. Similarly, toxicity lines of evidence could be numerous; they may include laboratory studies on various animals for various durations through various exposure routes or could be based on large-scale studies in the field (macrocosms). The lines of evidence for both exposure and effect must all be weighed simultaneously to come to a robust conclusion.

Weight-of-Evidence in Environmental Risk Assessment of Existing Conditions

After chemicals are released, we weigh evidence to determine whether or not they are causing unacceptable damage as well as the type and severity of the damage at various locations. If the answers to those questions suggest that the area of release must be remediated, we must determine how much and what kind of efforts are needed. Usually, we set goals in terms of acceptable concentrations (based upon estimated levels of risk) in sediment, soil, or water.

Example: Lines of evidence used in a risk assessment for a stream.

To determine whether selenium in a stream is affecting tree swallows in the vicinity of a specific site, an environmental assessor evaluates the following lines of evidence:

1. Compare the concentration of selenium in sediment and water with screening toxicity levels
2. Compare the tree swallow's estimated exposure dose using concentration of selenium in its food (insects) with known avian toxicity reference value
3. Compare concentration of selenium in tree swallow eggs collected in the area with a known egg benchmark value for another species of bird obtained from published studies
4. Monitor and compare a reproduction metric (example: number of eggs hatched) of tree swallows in the area with those at a reference site (a similar site that is not contaminated)

When we are dealing with a contaminated site and potentially affected organisms, there is ample opportunity for generating evidence, but even then, real systems are complex. Remedial efforts are expensive; it is important that all available information is reviewed, weighed, and that its relevance is agreed upon before additional data are collected. Evidence from the field may seem to be superior to evidence from the laboratory, but for many reasons field data may be difficult to interpret. The chemicals may be associated with effects, but the actual cause of those effects may be an extraneous factor such as soil compaction or organic matter from sewage rather than industrial waste. On the other hand, the biological community may appear to be unaffected because the sensitive species are gone and only the resistant species remain. We must carefully weigh all the lines of evidence and get agreement beforehand from all stakeholders during the problem formulation stage to avoid conflicts in interpretation. Careful and thorough study designs will go a long way to address such issues so the data obtained is clear.

Summary

When we provide a science-based recommendation to environmental decision makers, we want the conclusion that is best supported by the available body of evidence. We find that some evidence illuminates one aspect of the problem and other evidence illuminates other aspects. In addition, some evidence is more relevant, more reliable, or stronger than another. We deal with this diversity of evidence by a process of (1) carefully assembling all of the available and potentially

CC BY

useful research, (2) assigning a weight to each piece of evidence based on our judgement of the influence that it should have on the conclusion, and (3) integrating and weighing the body of evidence to determine the conclusion that is best supported. To minimize bias and ensure transparency, we present all of the evidence collected and the methods used to complete the weight-of-evidence process.

Resources

European Food Safety Authority (EFSA). 2017. Guidance on the use of the weight of evidence approach in scientific assessments. doi: 10.2903/j.efsa.2017.4971.

European Commission Joint Research Center (JRC). ToxRTool - Toxicological data Reliability Assessment Tool.

Hall AT, Martin O, Belanger SE, Galay-Burgos M, Guiney PD, Maack G, Stubblefield B. 2017. New Approach for Weight-of-Evidence Assessment for Ecotoxicological Effects in Regulatory Decision-Making. Integ Environ Assess and Manage. 13(4): 573–579. DOI: 10.1002/ieam.1936.

McDonald BG, deBruyn AM, Wernick BG, et al. 2007. Design and application of a transparent and scalable weight-of-evidence framework: An example from Wabamun Lake, Alberta, Canada. Integ Environ Assess Manage 3(4):476–83.

Moermond CTA, Beasley A, Breton R, Junghans M, Laskowski R, Solomon K, Zahner H. 2017. Assessing the reliability of ecotoxicological studies: An overview of current needs and approaches. Integ Environ Assess and Manage. 13(3): 640–651.DOI: 10.1002/ieam.1870.

Ruden C, Adams J, Ågerstrand M, Brock T, Poulsen V, Schlekat C, Wheeler J, Henry T. 2017. Assessing the relevance of ecotoxicological studies for regulatory decision-making. Integ Environ Assess and Manage. 13(4): 652–663. DOI: 10.1002/ieam.1846.

Golder Associates Ltd. (North Vancouver, British Columbia, Canada). Guidance for a Weight of Evidence Approach in Conducting Detailed Ecological Risk Assessments (DERA) in British Columbia. Final report 31 March 2006. Vancouver (BC): Ministry of Environment. Report No.: 05-1421-040. Submitted to Science Advisory Board for Contaminated Sites in British Columbia.

Suter G, S Cormier, M Barron. 2017a. A weight of evidence framework for environmental assessments: Inferring qualities. Integ Environ Assess Manage 13(6): 1038–1044.

Suter G, S Cormier, M Barron. 2017b. A weight of evidence framework for environmental assessments: Inferring quantities. Integ Environ Assess Manage 13(6): 1045–1051.

United States Environmental Protection Agency (USEPA). 2016. Weight of Evidence in Ecological Risk Assessment. EPA/100/R-16/001.

Government of Canada. 2012. Federal Contaminated Sites Action Plan (FCSAP). Ecological Risk Assessment Guidance.

Please contact setac@setac.org for guidance on relevant SETAC publications and experts in the subject.

Acknowledgments

SETAC is grateful for the efforts and contributions of those involved in the publication of this document: Pat Guiney, Mark Johnson, Tamar Schlekat, Jane Staveley, and Glenn Suter.

Copyright

When referring to this text, please use the following citation:

[SETAC] Society of Environmental Toxicology and Chemistry. 2018. Technical Issue Paper: Weight-of-Evidence in Environmental Risk Assessment of Chemicals. Pensacola (FL): SETAC. 8 pp.

SETAC encourages you to reproduce and distribute the document in the interest of enhancing communication about the environmental sciences. This document is copyright protected under the Creative Commons 4.0 international license, and the only requirements are attribution to SETAC and to not modify the information.