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): 1045–1051. doi:10.1002/ieam.1953.

A WEIGHT OF EVIDENCE FRAMEWORK FOR ENVIRONMENTAL ASSESSMENTS: INFERRING QUANTITIES

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

[§]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 U.S. Environmental Protection Agency (USEPA) has developed a generally applicable framework for a weight-of-evidence (WoE) process for deriving quantitative values from multiple estimates. These guidelines are intended for environmental assessments that require the generation of quantitative parameters such as degradation rates or that develop quantitative products such as criterion values or magnitudes of effects. The basic steps are to weigh evidence for the environmental quality to be quantified, generate the value by merging estimates or by identifying the best estimate, and weight the results to determine confidence in the numerical value. When multiple data sets or outputs of multiple models are available, it may be appropriate to weigh the evidence. Use of the framework to weigh multiple estimates may increase the accuracy of quantitative results compared to a single estimate from a default method. Its use can provide greater transparency compared to ad hoc weighing of evidence.

Keywords

Weight of evidence; risk assessment; water quality criteria; parameter estimation; meta-analysis

INTRODUCTION

Environmental assessments typically involve the derivation of quantities such as levels of effects or numerical criteria. Often those quantities could be estimated by multiple data analysis methods applied to multiple data sets from multiple experimental or observational approaches. None of the multiple estimates is the unknown true value, but each provides evidence concerning the true value. Hence, it is appropriate to weigh each piece of evidence (estimate) to determine the best-supported value of the effect level, threshold concentration, or whatever value is being derived (USEPA 2016).

¹Corresponding Author: Glenn Suter, U.S. Environmental Protection Agency: 26 W. Martin L. King Drive, Cincinnati, OH 45268, (513)936-8763, 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 U.S. Environmental Protection Agency.

We present a formal framework that uses the richness of the body of information to derive quantities by weighing qualitative and quantitative evidence. Note this is not a quantification of the WoE process to derive qualities (e.g., Linkov et al. 2009), rather it is the derivation by WoE of a numerical output that is an actual quantity. This approach is illustrated by examples from ecological assessments, but it is equally applicable to human health assessments.

Table 1 presents examples to illustrate how qualities and quantities are associated in environmental assessments. When we estimate quantities such as cleanup levels, those numerical values apply to some quality that is being quantified. Before deriving quantities, we must determine whether the system actually has a quality to be quantified (e.g., impairment or causation) or alternative qualities that could be quantified (e.g., assessment endpoints or hazards) using the method described in the companion paper (Suter et al. this issue). Note that even if a quality is assessed probabilistically (e.g., the probability of impairment), it is still a quality. The quality may be determined a priori (e.g., a standard assessment endpoint). If it is not, it may be derived by WoE, because there are likely to be multiple pieces of evidence concerning the quality that must be considered. Having defined the quality, the appropriate quantity is derived.

This framework for inferring quantities by WoE, like other USEPA frameworks, is intended to be sufficiently general and flexible to apply to the range of assessment problems addressed by the various offices of the USEPA and other organizations that may find it useful.

THE FRAMEWORK

The framework for WoE to derive a quantity has three basic steps (Figure 1). First, the quality to be quantified is determined. This step uses the WoE framework for inferring a quality (Suter et al. this issue). The other steps are distinct to a WoE that derives a quantity. In the second step, analyses are performed to derive the quantity using either of two approaches: merging the numerical estimates or choosing the best estimate. Third, confidence in the quantity is weighed based on the relative merits and inadequacies of the body of evidence. This WoE framework for inferring a quantity may serve different roles. It may represent the entire process for deriving a criterion or standard or it may represent the choice of a parameter value for a numerical model or an intermediate result for which diverse evidence is available.

Weighing evidence for the quality to be quantified

Before assembling or weighing evidence of a quantity, it is necessary to identify the assessment endpoint(s) or other qualities to be quantified and to credibly establish their relevance and reliability. If the qualities are assumed or are defined by policy, then the initial WoE step is not needed. For example, unless the assessment endpoint is standardized or otherwise pre-determined, the entity, attribute and other aspects that determine what will be quantified in a risk assessment should be defined by weighing evidence about the appropriate concerns. The framework for this WoE for qualities is presented in Suter et al. (this issue).

The use of WoE as part of a problem formulation is relatively common but is usually not as formal as described here. A good example occurs at the beginning of reviews of the United States secondary (welfare-based) National Ambient Air Quality Standards (NAAQS), where appropriate endpoints such as terrestrial vertebrate reproduction are determined by weighing evidence in an Integrated Science Assessment (ISA) (USEPA 2013a; Lassiter et al. 2015). The quality may also be a process in a model. For example, evidence from various studies was weighed to determine whether a bioavailability adjustment factor should be used and therefore must be estimated for a Superfund site, based on WoE (Integral Consulting Inc. 2013).

The framework for WoE for qualities is presented elsewhere (Suter et al. this issue; USEPA 2016). It includes the systematic accumulation of information and generation of evidence, the weighting of the evidence, and the weighing of the body of evidence. The WoE process may be adapted for the nature of the case and the amount and diversity of evidence. For example, ISAs weigh evidence from hundreds, sometimes thousands, of papers and reports and place more emphasis on systematic literature review than most assessments (USEPA 2015).

Even when WoE is not needed to identify the quality to be quantified (e.g., it is a parameter in a model), care in evidence generation can provide reliable inputs for the derivation of the quantity. The systematic literature review, screening and weighting for relevance and reliability, and reanalysis of the data to generate relevant evidence can also provide the alternative estimates of the quantity of interest and associated weights. In addition, it can clarify which WoE approach (i.e., merge estimates or choose the best estimate) is appropriate to the available quantitative estimates.

Merging quantitative evidence

If the quantitative estimates obtained from multiple data sets are of sufficient relevance and reliability, they can be merged into a single joint estimate. The statistical techniques for this purpose are called meta-analysis (Borenstein et al. 2009). The simplest method is averaging, as for example in the calculation of a species mean acute value when multiple acute lethality tests have been reported for a species/chemical combination (Stephan et al. 1985). Weighted means are commonly used in meta-analysis, particularly inverse variance weighting when a fixed effects model is appropriate. Meta-analysis also provides techniques for combining studies into a common regression model, such as an exposure-response relationship, which may then be used to derive the quantity.

In addition to weighting by variance or other statistical properties, weighting may be based on relevance or reliability of the methods used to generate each estimate. This approach is appropriate when, as is often the case, the differences in methods are more important than the error variance. For example, if studies are distinguished by relevance (e.g., laboratory tests versus field studies), a relevance weighting scale could be devised.

Meta-analyses have been criticized for biases in the input data, loss of information, statistical errors, and other potential problems (Thompson and Pocock 1991). If those problems cannot be corrected, they can be avoided by choosing the best estimate.

Choosing the best quantitative evidence

If the range of relevance or reliability of estimates is large, the best estimate may be chosen based on the weights rather than merging estimates from poor studies with those from excellent studies. In addition, broadly defined qualities, such as a protective concentration, may be estimated by values that are so heterogeneous that their mean would be hard to interpret (e.g., primary production of algae, diversity of invertebrates, and fecundity of brook trout). In either case, the weighting and weighing of evidence for estimates would be based on relevance, reliability, and collective properties as when weighing qualities. Collective properties are properties of a set of evidence such as diversity (multiple types or sources of studies), coherence (logical consistency in the evidence), and corroboration (validation by independent studies). Strength, however, would not be used as a weighting property, because the strongest relationship between exposure and response or other relationships does not necessarily provide the best quantitative estimate of the quality of interest. For example, relationships derived from field data are valuable because they represent the real world even though they often have lower strength (e.g., low correlation coefficients).

Table 2 presents a hypothetical example of a WoE table that might be used to weigh alternative estimates of a protective benchmark concentration. All estimates might be included, but in this case, the evidence is weighed after having been aggregated by type. Types of evidence with multiple values (e.g., multiple laboratory chronic tests) may have been merged by meta-analysis or the best estimate within a type may have been determined by weighing. In any case, the scores must be explained and justified. Preferably, criteria for scoring the estimates would be determined in advance to minimize the subjectivity of the judgments. Subjective judgements are inevitable when weighing heterogeneous bodies of evidence, but they can be objective in the sense of being transparent, disinterested, and impartial (USEPA 2002).

Weighting for most estimated values is adequately addressed by weighting relevance, reliability, and collective properties, but in some cases, the magnitude of the estimate is also a consideration. For example, screening values to determine whether a chemical requires additional study typically use lowest values to ensure that chemicals are not screened out that do require more attention. For regulatory benchmarks, precautionary policies may require that the lowest credible estimate be used to ensure protection. Finally, the magnitude of the estimate may be a tie breaker. That is, if multiple estimates have the same relevance and reliability, the lowest value may be chosen.

Choosing the best value by WoE has not been an established practice, but it is recommended by the European Chemical Agency for choosing among alternative test data for the same species, endpoint, duration, life stage, and testing conditions (ECHA 2008). They do not provide a weighting method but they do indicate that other studies may add weight to a particular test.

Weighting and confidence in quantitative results

Decision makers want to know how much confidence should be placed in a result, and not just the uncertainty (Science Policy Council 2000). Confidence in a derived quantity

depends on both the statistical scatter associated with the number and the weight of the evidence that contributed to deriving the number (Figure 2). The concepts of variability and uncertainty are familiar, as are their expressions in terms of statistical measures of the scatter of data (e.g., confidence interval on a mean) and scatter of derived values (e.g., range of chronic values). Assigning weight to quantitative results is much less familiar but is also important. A quantity derived by meta-analysis may have a small standard deviation, but may be from studies with little relevance to the case being assessed, may be based on studies that are unreliable due to bad designs and poor quality assurance, or may have meager collective properties due to being based on few tests with low replication conducted in one laboratory. Those properties are at least as important as statistical scatter to the confidence that may be placed in a result.

The weight of a quantitative result is a function of the relevance and reliability of the individual estimates and of the gain in reliability due to weighing multiple estimates. If a best estimate is chosen, the fact that it was compared to other available estimates can provide greater confidence. If estimates are statistically merged, the result should have greater reliability than any individual estimate. The overall weight of the quantitative result should be based on collective properties such as the number of estimates, their quantitative similarity, and diversity of study types.

Evidence other than quantitative estimates may still contribute information that influences confidence. For example, a study might not provide an estimate of the mean degradation rate but might indicate how high it could be. Other studies may indicate the scope of habitats and taxa to which a quantity is applicable, better indicating the relevance. Also, studies that are not adequate to derive an estimate may increase the reliability of a quantity if they give similar results.

The weight of a quantitative result may be used to determine its applicability. For example, a screening value for soil, sediment or water might also be used as a remedial goal if it is very relevant and reliable and is based on weighing evidence from several studies. More generally, reporting the weight of evidence along with statistical scatter more fully communicates confidence in assessment results.

Because of the need to recognize that derived quantities have weight as well as scatter, presentations of quantitative results may have four parts. As a hypothetical example:

1. The quality expressed—threshold concentration for avoidance.
2. The numerical result and units— x $\mu\text{g/L}$.
3. Scatter—95% CI of $\pm 0.3x$ mg/L .
4. Weight—the body of evidence is moderately relevant, highly reliable, diverse, and coherent.

EXAMPLE 1, WOE FOR AN OZONE AIR QUALITY STANDARD

The USEPA has developed Secondary National Ambient Air Quality Standards for major air pollutants using a process similar to the framework presented here. It includes two

assessments: one corresponds to the determination of the quality to be quantified and the other corresponds to the analysis and the determination of confidence (Figure 1). This example focuses on the secondary standard as defined by the Clean Air Act which is based on adverse effects on public welfare including ecosystem components such as vegetation, soils, wildlife, and water.

This current example of WoE for a quantity does not formalize the process to the extent recommended by USEPA (2016). It does, however, clearly illustrate the steps in the framework and does properly combine qualitative and quantitative weighing to derive defensible quantitative results.

Identifying the quality to be quantified

The process begins with an ISA that determines what welfare effects occur in a relevant range of pollutant levels (i.e., the qualities to be quantified; USEPA 2015). The ISA infers those qualities using a consideration-guided WoE process (USEPA 2016). It assembles evidence by performing a systematic literature review with sequential screening of studies for general scientific quality and relevance. The studies are categorized by effect and type of study. The evidence is evaluated and synthesized to identify exposure/response relationships and to estimate exposures. Evidence is weighed in the ISA primarily to determine whether the apparent relationships between exposure to the pollutant and effects are causal. The evidence is weighted relative to nine “aspects” which are adapted from Hill’s (1965) considerations. The bodies of evidence are then weighed into one of five bins ranging from “causal relationship” to “not likely to be a causal relationship” (USEPA 2015).

The ozone ISA provides an example (USEPA 2013b). The welfare section focuses on ozone’s direct effects on vegetation and secondary effects on ecosystem processes, because prior ozone assessments had shown that they are the most sensitive welfare effects. Thousands of studies were evaluated including exposures in laboratories, green houses, field chambers, chamberless field systems, ambient gradients, and site comparisons. Current exposures were estimated from ambient measurements and models. The effects that were found to be relevant and reliably causal were: visible foliar injury, reduced vegetation growth, reduced ecosystem productivity, reduced carbon sequestration, reduced crop yield and quality, altered terrestrial water cycling, altered below-ground biogeochemical cycles, and altered community composition.

Choosing the best quantitative estimate

The analytical step is performed in the Welfare Risk and Exposure Assessment (WREA), which quantitatively estimates risks to human welfare based on ecological effects of exposure to an air pollutant. Specifically, it estimates magnitude of effects of current pollutant levels and of different potential standards in various environments on the various endpoints. The choice of exposure/response relationship for each endpoint is based on the WoE for each, including consideration of the concordance across the body of evidence. For tree growth in response to ozone, seedling studies in open-top chambers were compared to free air studies and found to be consistent (USEPA 2014a). Then the empirical studies were compared to multiple modeling studies for adult trees.

The results of the seedling studies were chosen as providing the best exposure/response models and the other studies were used to evaluate uncertainty and confidence. Those models were used to estimate current growth losses and losses with alternative ozone levels (Table 3). This is an example of choosing the best estimate, but merging estimates was also used in the WREA. For example, a published meta-analysis of multiple willingness-to-pay studies was used to estimate the value of potentially affected threatened and endangered species. Similarly, estimates from multiple models of carbon sequestration by trees were merged when a species-specific model was not available.

Assigning weights and confidence to quantities

As in the last step of the framework (define confidence in quantitative results—Figure 1), WREA “summarizes overall confidence in the results, as well as relative confidence between the different analyses” (USEPA 2014b). This includes an analysis of “qualitative uncertainty” in terms of the direction of potential bias and scores for magnitude of uncertainty and for knowledge base. It also includes quantification of variability and uncertainty (expressed as frequency distributions and standard errors).

The WREA completes the scientific assessment of the evidence. However, it is followed by a Policy Assessment, which summarizes ISA and WREA, and lays out the decision alternatives for the USEPA Administrator’s own weighing process (USEPA 2014b).

EXAMPLE 2, WOE FOR AN ATRAZINE LEVEL OF CONCERN

During preparation of this paper, an example of WoE for quantities appeared that derives a level of concern (LoC) for atrazine risks to aquatic plants in a way that is remarkably consistent with this framework (Moore et al. 2017). The endpoint was pre-defined, so no WoE was needed to identify the quality to be quantified. The authors derived potential LoCs by four methods, using either laboratory data or microcosm and mesocosm data and different analytical approaches. They used both of the WoE methods recommended in this paper for inferring a quantity. Taking the merge quantitative evidence approach, they weighted the four LoC estimates (range = 19.6–26 µg/L) for seven reliability attributes and obtained a weighted mean LoC of 23.6 µg/L. In addition, they used the evidence weights to choose the Comprehensive Aquatic Systems Model (CASM) result (19.6 µg/L) as the individual best estimate. Because all four methods have strengths and weaknesses, they proposed using the merged estimate rather than the best individual estimate. That subjective judgment implies an informal analysis of confidence which concluded that none of the individual estimates had high or low confidence, so averaging multiple moderate-confidence estimates would increase overall confidence in the final LoC.

DISCUSSION AND CONCLUSIONS

The WoE approach treats estimates derived from different combinations of data and analytical methods as separate pieces of evidence concerning the desired true value. The estimates may be weighted and statistically merged to generate a consensus estimate or may be weighted and weighed to identify the best (i.e., weightiest) estimate. Although there have been assessments that weigh evidence to derive a quantity (e.g., examples 1 and 2) and

frameworks for deriving specific types of quantities that employed meta-analysis (e.g., USEPA 2006, Cormier et al. 2008), this is the first formal framework for weighing evidence to derive a quantity for a range of environmental assessments. We believe that it will help to standardize methods and will often be preferable to using a single quantitative estimation method or using ad hoc judgement. It potentially provides greater accuracy, transparency, defensibility, flexibility, and confidence. In addition, the WoE framework to infer a quantity provides a context for the analytical methods that ensures that the appropriate quality is quantified and that the confidence in the quantity is fully described.

This WoE framework for quantities is an extension of conventional WoE for identifying causes, hazards, and other qualities (Suter et al. this issue; USEPA 2016). Having a connection between WoE for qualities and quantities provides advantages that a purely quantitative WoE method (e.g., meta-analysis alone) lacks. First, by weighing the evidence concerning the quality to be quantified, it minimizes type 3 errors (i.e., answering the wrong question). Second, it ensures that relevant information is obtained and considered, not just the studies that generate a numerical estimate. As a result, aspects of the problem are revealed that might otherwise be overlooked such as limits on the range of applicability. Third, the evidence for the quality being quantified provides a context for the quantitative results that conveys what is known, how well it is known, and why it is important.

In addition to the link to conventional WoE, weighing evidence to derive quantities has a major inherent advantage. It provides more flexibility than application of a standard method to standard data. That flexibility makes more data and methods available to assessors and potentially results in more accurate quantitative values. However, flexibility implies subjective judgments and inconsistent methods. This critique is minimized by using a standard WoE framework to engender greater consistency and defensibility than an ad hoc approach to weighing evidence, while potentially enabling more relevance and reliability than a method that does not weigh multiple estimates. In addition, a formal WoE method provides more transparency than an informal narrative weighing of evidence. Regulatory agencies have relied on standard methods for deriving benchmark values such as Stephan et al. (1985), because their application is objective and consistent for all pollutants. A similar standard of fairness may be achieved by following a consistent WoE framework and methodology to maximize objectivity and consistency.

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 of the panel's coordinator, Lawrence Martin. Glenn Rice, Beth Owens, Jeffrey Herrick, Meredith Lassiter, 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 many USEPA scientists and managers. The authors received no financial support other than their salaries.

References

Borenstein, M., Hedges, L., Higgins, J., Rothstein, H. Introduction to meta-analysis. Chichester, U.K: Wiley; 2009.

- Cormier SM, Paul JF, Spehar RL, Shaw-Allen P, Berry WJ, Suter GW. Using field data and weight of evidence to develop water quality criteria. *Integr Environ Assess Manag*. 2008; 4:490–504. [PubMed: 18630992]
- [ECHA] European Chemicals Agency. Guidance on information requirements and chemical safety assessment, Chapter R. 10: Characterization of dose [concentration]-response for environment. Helsinki, Finland: European Chemical Agency; 2008.
- Hill AB. The environment and disease: Association or causation. *Proceed Royal Soc Medicine*. 1965; 58:295–300.
- Integral Consulting Inc. Baseline Ecological Risk Assessment: San Jacinto River Waste Pits Superfund Site. Seattle, WA: U.S Environmental Protection Agency; 2013. No 9559469
- Lassiter MG, Owens EO, Patel MM, Kirrane E, Madden M, Richmond-Bryant J, Hines EP, Davis JA, Vinikoor-Imler L, Dubois J-J. Cross-species coherence in effects and modes of action in support of causality determination in the U.S. Environmental Protection Agency's integrated science assessment for lead. *Toxicology*. 2015; 330:19–40. [PubMed: 25637851]
- 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.
- Moore D, Greer C, Manning G, Wooding K, Beckett K, Brain R, Marshall G. A weight-of-evidence approach for deriving a level of concern for atrazine that is protective of aquatic plant communities. *Integr Environ Assess Manag*. 2017; doi: 10.1002/ieam.1865
- Science Policy Council. Risk characterization handbook. Washington, D.C: U.S. Environmental Protection Agency; 2000. No. EPA 100-B-00-002
- Stephan, CE., Mount, DI., Hanson, DJ., Gentile, JH., Chapman, GA., Brungs, WA. Guidelines for deriving numeric National Water Quality Criteria for the protection of aquatic organisms and their uses. Washington, D.C: U.S. Environmental Protection Agency; 1985. No. PB85-227049
- Suter GW, Cormier SM, Barron MG. A weight of evidence framework for environmental assessments: Inferring qualities. *Integr Environ Assess Manag*. this issue.
- Thompson SG, Pocock SJ. Can meta-analysis be trusted? *The Lancet*. 1991; 338:1127–1130.
- [USEPA] United States Environmental Protection Agency. Guidelines for ensuring and maximizing the quality, objectivity, utility, and integrity, of information disseminated by the Environmental Protection Agency. Washington, D.C: Office of Environmental Information; 2002. No. EPA/260R-02-008
- [USEPA] United States Environmental Protection Agency. Framework for developing suspended and bedded sediments (SABS) water quality criteria. Washington, DC: Office of Water, US EPA; 2006. EPA-822-R-06-001
- [USEPA] United States Environmental Protection Agency. Integrated science assessment for lead. Research Triangle Park NC: National Center for Environmental Assessment; 2013a. No.EPA/600/R-10/075F
- [USEPA] United States Environmental Protection Agency. Final report: Integrated science assessment for ozone and related photochemical oxidants. Washington, D.C: National Center for Environmental Assessment; 2013b. No. EPA/600/R-10/076F
- [USEPA] United States Environmental Protection Agency. Welfare risk and exposure assessment for ozone, final. Washington, D.C: Office of Air and Radiation; 2014a. No. EPA-452/R-14-005a
- [USEPA] United States Environmental Protection Agency. policy assessment for the review of the ozone national ambient air quality standards. Washington, D.C: Office of Air and Radiation; 2014b. No. EPA-452R-14-006
- [USEPA] United States Environmental Protection Agency. Preamble to the integrated science assessments. Research Triangle Park NC: National Center for Environmental Assessment; 2015. No. EPA/600/R-15/067
- [USEPA] United States Environmental Protection Agency. Weight of evidence in ecological assessment. Washington, D.C: Risk Assessment Forum; 2016. No. EPA/100/R-16/001

Figure 1.
A framework for deriving quantitative values by WoE in environmental assessments (USEPA 2016).

Figure 2.

A diagram of the combination of statistical scatter and qualitative weight to define the confidence that should be afforded an assessment result (USEPA 2016).

Table 1

Examples of qualities to be quantified and associated quantities from environmental assessments. Both may be derived by Weight of Evidence, but the methods differ.

Qualities	Quantities
A site is impaired by a waste material	Cleanup level
An assessment endpoint for a risk assessment is appropriate and relevant	Magnitude of effects
An assessment endpoint for a criterion assessment is appropriate and relevant	Criterion value
A hazard is credible	Probability and magnitude of the risk of the hazard
A chemical is bioaccumulative	Bioaccumulation factor for algae
An agent is the cause of observed impairment	The level of exposure below which it does not cause impairment

Table 2

WoE table to summarize quantitative benchmark estimates from four types of evidence with hypothetical scores and estimated benchmark values

Evidence	Relevance	Reliability	Collective properties	Overall WoE	Estimate (mg/L)
Acute SSD and A/C ratio ^a	^b ++	++	+++	++	15
Laboratory chronic tests	++	+	++	+	2
Mesocosms	++	+	+	++	0.5
Field surveys	+++	+++	++	+++	0.1

^a SSD= species sensitivity distribution, A/C = Acute/Chronic

^b +, ++, +++ indicate some, strong or very strong support, respectively, for the estimate

Table 3

Relative biomass losses (median/75th centile/maximum percentages) for tree species exposed to ozone at recent levels and two alternative levels (measured as weighted parts per million-hours [ppm-hrs]). The original table presents more species and levels (USEPA 2014b).

Species	Recent conditions	15 ppm-hrs	7 ppm-hrs
Red maple	0.95/1.25/3.49 (%)	0.08/0.17/0.77	0.05/0.08/0.39
Ponderosa pine	3.71/5.93/24.3	0.65/0.94/3.25	0.50/0.09/0.93
Loblolly pine	0.30/0.36/0.71	0.05/0.07/0.17	0.04/0.05/0.09
Black cherry	23.9/28.5/51.5	4.89/7.94/23.9	3.41/4.41/13.7