EPA Public Access

Author manuscript

Integr Environ Assess Manag. Author manuscript; available in PMC 2019 September 01.

About author manuscripts

Submit a manuscript

Published in final edited form as:

Integr Environ Assess Manag. 2018 September ; 14(5): 631–638. doi:10.1002/ieam.4059.

SPECIFYING THE DIMENSIONS OF AQUATIC LIFE BENCHMARK VALUES IN CLEAR, COMPLETE, JUSTIFIED PROBLEM FORMULATIONS

Glenn Suter^{*,†}

[†]National Center for Environmental Assessment, 26 W. Martin L King Drive, U.S. Environmental Protection Agency, Cincinnati, Ohio, USA, 45268, suter.glenn@epa.gov

Abstract

Nations that develop water quality benchmark values have relied primarily on standard data and methods. However, experience with chemicals such as Se, ammonia, and tributyltin has shown that standard methods do not adequately address some taxa, modes of exposure and effects. Development of benchmark values that are protective requires an explicit description of the issues, a problem formulation. In particular, the assessment endpoints and other dimensions should be specified for each chemical so that the necessary data will be obtained and appropriate analyses will be performed. Assessment endpoints specify the entity and attribute to be protected. In addition, the level of protection, including the magnitude of effect and the proportion affected is specified. Magnitude and proportion are included, because they are used to calculate the benchmark concentration. If uncertainty is considered in the benchmark, the proportion of the uncertainty distribution that is protected should be specified. Because effects are related to the duration of exposure and time for recovery, temporal dimensions should be specified. Clearly described exposure metrics are also needed, because the relevant exposure parameter is not always total aqueous concentration. Finally, the benchmark may be applicable to particular geographic or climatological areas, water chemistries, taxa, or habitat types. Considering and justifying all the dimensions is likely to result in protective and more easily communicated benchmarks.

Keywords

Criteria; Guidelines; Problem formulation; Endpoints; Standards

INTRODUCTION

Aquatic life benchmarks include standards, criteria, levels of concern, guidance values, and limit values used for compliance assessments of ambient waters, permitting of effluent discharges, and screening of site contaminants. They have been derived using standard methods published in guidance documents (CCME 2007; DWAF 1996; EC 2011a; Liu 2015;

^{*}Address correspondence to suter.glenn@epa.gov.

Publisher's Disclaimer: *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 US Environmental Protection Agency.

Data accessibility—there are no data to share.

Stephan et al. 1985; UKTAG 2008; Warne et al. 2015). That approach has served the environment well, but it has become clear that standard methods are not always appropriate (Buchwalter et al. 2017). Recent experience in the U.S. with criteria for ammonia, selenium, tributyltin, atrazine, and dissolved mineral ions have shown the importance of considering non-standard taxa, modes of exposure, sources of data, or data analyses (USEPA 2003a; 2003b; 2013; 2016a; 2016b). Some categories of chemicals such as endocrine disruptors, nanomaterials, and pharmaceuticals may require nonstandard data or methods. The authors of many guidelines acknowledged the need for flexibility, but without an explicit problem formulation the temptation to follow standard methods is strong. Problem formulation can serve as a pause to ask whether the default assumptions will provide adequate protection or whether nonstandard data and methods should be incorporated.

Clearly formulating the problem is an important part of ecological risk assessment and of risk-based benchmark development (USEPA 1998; Suter and Cormier 2008). Problem formulation assembles and weighs evidence to determine what potential receptors are sensitive to the pollutant, what sources and routes of exposure are important, what are the modes of action, etc. It includes tasks such as reviewing the literature and conceptually modeling the sources, routes of exposure and effects. Its fundamental task, however, is to define the dimensions of the benchmark to be derived in the analysis and characterization phases. Since 2012, the U.S. Environmental Protection Agency (EPA) has included problem formulation in criteria documents, but those problem formulations could benefit from greater formality and clarity. This paper proposes that the benefits of problem formulations could be clarified, if they focused on specifying and justifying the dimensions of the benchmark. However, currently the only dimensions that are routinely formally defined are the two dimensions of the assessment endpoint, the entity and attribute. The other dimensions either are not considered (e.g., uncertainty) or are defined by default and not explicitly considered or justified in the problem formulation except in unusual cases (e.g., duration of exposure). The paper describes how to derive dimensions that are organized in five categories.

1. The assessment endpoint consisting of an entity that is susceptible and worthy of protection and an attribute of the entity that is affected.
2. The threshold level of effect on the endpoint attribute is defined by a magnitude of effect and the proportion of entities affected.
3. Temporal dimensions may include the duration of the exposure and the frequency with which exceedances may occur.
4. Exposure metrics such as total concentration, dissolved concentration, free ionic concentration, temperature, or specific conductivity are specified.
5. Limits of applicability are defined on chemical, spatial, or ecological dimensions.

Explicit consideration of these dimensions has three advantages. First, they make the intent of the benchmark clear to the audiences of the problem formulation. One type of audience includes decision makers who will be using the benchmark in regulatory or management actions. Another is stakeholders and the public who will be required to comply with the benchmark or will be relying on the benchmark to protect their valued resources. They may

consider whether to buy in to the benchmark derivation before it is performed or they may accept or challenge the benchmark after it is derived. A third type of audience is the toxicologists, chemists, and modelers who will be responsible for generating data, assembling data from the literature, and analyzing the data to derive the benchmarks. The dimensions and their justification provide the blueprint for the process of benchmark derivation.

The second advantage of an explicit list of dimensions is to guide the problem formulation process. At minimum the dimensions are a checklist that ensures that essential issues are addressed. More fundamentally, considering the dimensions can ensure that the assessors determine whether the defaults are sufficient or whether nonstandard dimensions are needed for a chemical or other agent. That is, is the available evidence supportive of the assumptions that underlie the default, or does it suggest that nonstandard data, models, or analyses should be employed to derive a different endpoint, magnitude of effect, duration of exposure, etc.? For example, field observations showed that, to derive protective ammonia criteria, it was necessary to perform tests of unionid mussels (USEPA 2013).

Third, the dimensions may promote international harmonization. Aquatic life benchmarks from different agencies can differ by orders of magnitude, even when the goals, basic approach (species sensitivity distributions), and available data are effectively the same (Schlekat et al. 2017). Clearly and specifically defining and justifying the benchmark dimensions for each chemical should facilitate identifying the critical differences.

The dimensions are explained primarily in terms of benchmarks derived from laboratory test results in species sensitivity distributions, which is the standard approach in Australia/New Zealand, Canada, Europe, South Africa, and the United States. Alternative expressions of the dimensions are also suggested that may be more appropriate for some chemicals or other agents. Benchmarks for effects on humans and wildlife that feed on aquatic biota and issues of ensuring compliance are not included.

ASSESSMENT ENDPOINT

Definitions of assessment endpoints for benchmark assessments, like those for conventional risk assessments, include ecological entities and attributes of those entities that are to be protected (Suter 1989; USEPA 1998). They should be sufficiently clear to be operational and they must be suited to the case. Most national benchmarks have very general endpoints so as to be broadly applicable, but more local benchmarks and benchmarks for chemicals with specific effects may be quite specific. For example, where migrating salmonids are present in Canada, a fluoride benchmark based on behavioral disruption is recommended (CCME 2002).

It is important to recognize that more than one level of biological organization is often involved in an assessment endpoint definition. Typically, there is a lower level for which test or observational data are available and a higher level which is of greater interest and may be treated as an aggregate of the lower level attributes. For example, community effects are typically estimated as the proportion of species or affected genera (e.g., the proportion of

species affected by a kill, estimated from acute toxicity tests of individual species). Hence, quantifying an assessment endpoint typically involves combining the responses of organisms to derive a population effect and combining those population effects into a community effect (based on the direct organismal effects, without population or community dynamics). Effects at higher levels of organization can also be tested or observed directly (e.g., loss of species in a mesocosm or a stream community) (CCME 2007; USEPA 2011; Warne et al. 2015). These different ways of generating endpoints from data result in different definitions of the endpoint dimensions. It is helpful to suggest tests and test endpoints that are appropriate to the assessment endpoint.

Entity—

The most common entity for aquatic life benchmarks is the biotic community. It typically includes either all aquatic species or only aquatic animals. For herbicides and other chemicals that are particularly toxic to plants, it may include only the community of aquatic plants. Populations of threatened or endangered species or commercially or recreationally important species may be entities for location-specific benchmarks.

Attribute—

The most common attributes for assessment endpoints are survival, growth, and reproduction, the standard aquatic toxicity test endpoint attributes. However, the test endpoints are attributes of test organisms, and the assessment endpoints are attributes of the species comprising biotic communities extrapolated from test data. That is, the typical benchmark assessment attribute is survival, growth or reproduction of the members of the populations in a biotic community. Some agencies allow other attributes (e.g., behavior, deformities, physiology) if they are shown to be relevant (CCME 2007; EC 2011a; Warne et al. 2015) or biologically important and from an important species (Stephan et al. 1985). For plants, growth or primary production are standard attributes. When field survey data are available, extirpation (effective absence of a population) is an attribute that can be derived from occurrences of species or genera in an area (Cormier et al. 2013; USEPA 2011). Although the default attributes dominate benchmark development, it is important to remember that the specific effects of a pollutant may make other attributes more appropriate. Openness to nonstandard attributes is mandated by the U.S. Clean Water Act's section on Criteria Development and Publication (U.S. Code 1314 (a)(1)) which requires "the latest scientific knowledge on the kind and extent of all identifiable effects on health and welfare including, but not limited to, plankton, fish, shellfish, wildlife, plant life, ... and on the effects of pollutants on biological community diversity, productivity, and stability" Other nations have similar broad mandates for protection.

LEVEL OF PROTECTION

Benchmark assessments must specify the level of protection which conventionally includes both the magnitude of effect and the proportion of entities experiencing effects of that magnitude. Magnitude, however, can define the acceptable level of effects with or without proportion.

Magnitude of effect—

Magnitude is the degree of effects on designated attributes of the components of the assessment endpoint entity. Because of the hierarchical nature of ecological endpoints, the magnitudes of effect are typically hierarchical. For benchmarks based on laboratory toxicity testing, the foundation is the magnitude of effects on attributes of the test organisms. It is the x in LC_x (lethal concentration) or EC_x (effective concentration). A quantal magnitude is the absolute proportion of the dichotomous effect itself (e.g., proportion dead, deformed, extinct, eutrophic, etc.). The magnitudes of count or continuous effects may be the measured parameter itself (e.g., mean mass, length, young/female), or it may be expressed as the proportion of control. Conventionally, the magnitudes of effects on the test endpoint are directly extrapolated from the test population to field populations (i.e., the magnitude of effect on survival, growth or reproduction in the laboratory are applied to the field without population modeling). However, a simple counter-example is provided by the U.S. acute criteria, in which dividing the 50% lethality concentrations by 2 brings lethality below 10%, on average. If mesocosm or field data are used, the magnitudes of effects need not be extrapolated up to community attributes. Various magnitudes could be used for effects on communities derived from test or field data such as 10% reduction in productivity or species number. Note that test endpoints based on hypothesis testing statistics (NOECs and LOECs) have been used in place of magnitudes of effects; that practice is undesirable but sometimes unavoidable (CCME 2007; USEPA 2013; Warne et al. 2015).

Proportion—

This is the proportion of organisms in an endpoint population or of populations in an endpoint community that would experience the designated magnitude of effects. Endpoints for quantal effects are defined at the population level as the proportion of organisms responding (referred to as incidence in human health risk assessments). For community-level endpoints, the conventional proportion is 5% of species or genera experiencing the population-level effect. That is, populations of the 5th centile species or genus would experience an effect of the prescribed magnitude on the specified attribute. For example, 5% of species in an aquatic animal community (the proportion) have at least 10% mortality (the magnitude). Since Stephan et al. (1985), the 5th centile has been the default proportion for communities in the U.S. and elsewhere. However, from the beginning, the need to lower the proportion when important species are likely to be affected has been recognized. Also, it may be appropriate to vary the proportion for communities requiring different levels of protection based on community properties (Warne et al. 2015).

Although conventional water quality benchmarks based on species sensitivity distributions (SSDs) require definition of both the magnitude of effects on species and proportions of species affected, in some cases a proportion is not required. If the test endpoint entity corresponds to the assessment endpoint entity, defining magnitude is sufficient. Similarly, if the assessment endpoint entity is an aquatic community (e.g., freshwater lotic communities) and a mesocosm test is used with an appropriate whole-system attribute (e.g., number of species or primary production), the magnitude of effect on that whole-system attribute is sufficient to quantify the assessment endpoint. The endpoint effect is not a composite of levels of organization from which a proportion can be calculated.

If endangered or otherwise highly valued species were protected as individual members of an exposed population, the dimensions defining levels of protection could be defined as in human health risk assessment (IPCS 2014). That is, the magnitude could be death, low weight, or infertility, and the proportion could be proportion of individuals in the population.

UNCERTAINTY

If uncertainty is estimated for the data and model used to derive a benchmark, the benchmark could be set to protect the endpoint with a certain probability (IPCS 2014). In the current laboratory-based methods, the uncertainty distributions of the magnitude or proportion are not explicitly considered so the benchmarks are the “best estimate.” We may assume it is the central estimate, so half the time the true protective benchmark is higher and half the time lower than the level that is identified by the endpoint effect. This is in contrast to the lower confidence limit on the benchmark dose (BMDL) in the USEPA’s human health risk assessments which provide 95% coverage (only 5% probability that the dose causing a 10% magnitude of effect is lower than the benchmark dose) (USEPA 2012). If uncertainty is also estimated for extrapolations (e.g., from chronic test results to real population responses and from sensitivity distributions of test species to real communities), that raises further complications and the potential for unrealistically wide distributions. Further, the best estimate is not necessarily the central estimate, so the actual coverage may be considerably different from 50% when uncertainty is not estimated.

So far, none of the aquatic benchmark derivation methods has used quantitative uncertainty as a dimension of a benchmark. However, assessment factors to add precaution when the data set is not optimal are employed by the EC (2011a). In Australia/New Zealand, uncertainty due to suboptimal data results in downgrading the reliability of trigger values and increased flexibility in site-specific implementation (ANZECC 2000). Canada and the U.S. field-based method for specific conductivity report confidence intervals as additional information.

BENCHMARK TEMPORAL DIMENSIONS

Because of the temporal dynamics of aqueous emissions, receiving water flows, and biological responses, U.S. aquatic life criteria consider temporal dimensions. The same is true of air quality benchmarks, but temporal variability is not generally considered for soils and sediments. Other nations treat temporal dynamics differently from the U.S. For example, rather than define duration as averaging time, Australia and New Zealand recommend that action be triggered when at least 5% of measured concentrations exceed the guideline value (ANZECC 2000). The UK uses the same centile for some substances to protect against higher concentrations (UKTAG 2008).

Duration—

Duration is the time period over which concentrations are averaged when determining whether a benchmark value is exceeded. In general, longer durations of exposure result in effects of greater magnitude in a larger proportion of entities. Therefore, assuring a particular degree of protection requires specifying the duration of exposure. In the U.S.,

Canada, and South Africa, water quality benchmarks include acute and chronic categories. In the U.S., the default durations are 1 hour (criterion maximum concentration—CMC) and 96 hours (criterion continuous concentration(CCC)). In Europe the equivalent durations are defined as maximum and annual average, but they are both applied to chronic toxicity and they may be modified depending on expected exposure patterns (EC 2011a). These are durations of exposures in the field, expressed as averaging times for concentrations, not the durations of toxicity tests. Durations may be conceptualized in at least two ways. First, they may be durations of exposure to a constant concentration that begin to induce the endpoint effect. Second, if concentrations are assumed to fluctuate, durations may be the averaging time for which a fluctuating concentration may be assumed to induce the endpoint effect. Both conceptualizations are discussed as contributing to the default durations in Stephan et al. (1985). Bioaccumulative chemicals require longer durations, such as 30 days for aqueous selenium in the U.S. (USEPA 2016a). Chemical-specific durations can be derived by toxicokinetic modeling or by time-to-event modeling of test data. However, when whole organism or tissue concentrations are used, as with selenium, the duration is instantaneous, because a point sample integrates prior exposures (USEPA 2016a). Detailed guidance on biota monitoring is provided for European Environmental Quality Standards based on concentrations in biota (EQS_{BIOTA}) (EC 2014). Compliance is based on the geometric mean concentration with the timing and number of samples based on the conditions in the case.

Frequency—

Frequency is the temporal frequency with which exceedance of benchmarks may occur. The U.S. EPA bases frequency on the assumption that exceedances are due to random variance in emissions and dilution flows rather than accidents, so the exceedance is assumed to be small. In the U.S., frequency is necessary because emission limits are designed based on probability distributions of emissions and flows, so the probability of a particular concentration is non-zero. However, most other jurisdictions do not set frequencies with which a benchmark concentration may be exceeded. The frequency in the U.S. (no more than once in three years) is based on time to recovery. The results of a literature review, indicated that, “Most aquatic ecosystems can probably recover from most exceedances in about three years” (Stephan et al. 1985). Although there is considerable variance in time to recovery (Gergs et al. 2016), it is reasonable to choose a default frequency value that protects most cases most of the time. A problem formulation, however, could consider whether persistence of the contaminant, the life histories and dispersal abilities of the sensitive taxa, or the nature of the effects should be taken into consideration when deriving an acceptable frequency.

BENCHMARK EXPOSURE METRICS

Exposure metrics depend on the mode of exposure. Most water quality benchmarks are expressed as aqueous concentrations, either total or dissolved, and the mode of exposure is assumed to be direct contact. When exposure occurs through diet, water quality benchmarks may be expressed as aqueous, dietary, tissue, or whole-body concentrations (EC 2011a; EC 2014). Other exposure metrics include pH, temperature (degrees Celsius), and aggregate

measures such as conductivity. For nanoparticles and plastic microparticles, it is still not clear how to appropriately express aqueous exposures (Kahn et al. 2017).

BENCHMARK APPLICABILITY

The applicability of benchmarks may be constrained by chemical or physical characteristics, location, or ecosystem type. Depending on the chemical, a benchmark may not be applicable to waters with extreme temperature, pH, or other conditions. For example, the 1988 U.S. aluminum criteria apply only to pH 6.5-9.0 waters (USEPA 1988). Some benchmarks, such as the European nickel environmental quality standard, are functions of toxicity modifying factors such as pH, dissolved organic carbon, and calcium (EC 2011b). Those models may extend the range of conditions to which a benchmark is applicable relative to a single concentration, but they have their limits which must be recognized. Some benchmarks may be applicable only if sensitive taxa occur. For example, U.S. ammonia criteria may be adjusted locally contingent on occurrence of salmonids or freshwater mussels. For naturally occurring chemicals, the background levels should be considered to avoid setting benchmarks that will be exceeded naturally. Benchmarks derived from field data may be applied to the region from which the data were obtained or sufficiently similar regions. For example, the field-based conductivity benchmark was applied only to streams in Central Appalachia with pH >6 (USEPA 2011). Field or mesocosms data may reveal that benchmarks are specific to certain ecosystem types. For example, chronic water column criterion concentrations for selenium may be set separately for lotic and lentic ecosystems (USEPA 2016a). U.S. criteria are nationally applicable, although they may be modified for specific sites. Guideline trigger values in Australia/New Zealand are location-specific and the level of protection varies depending on whether the site is a high conservation area, moderately modified ecosystem, or highly modified ecosystem (ANZECC 2000). Criteria for China are watershed-specific, based on test species that are representative of local species (Liu 2015).

OTHER DIMENSIONS

As the science and practice of benchmark derivation evolve, other dimensions might become relevant. In particular, if benchmarks were derived by weight of evidence, the overall weight of the evidence contributing to the derived value might be reported, along with uncertainty, as part of the expression of confidence in the value (USEPA 2016c).

EXAMPLES FROM GUIDANCE

Neither in the U.S. nor in other nations do guidelines for benchmark derivation explicitly or fully define the dimensions listed here. However, most of the dimensions can be inferred from the methods (Tables 1, 2). Because they are generic, the dimensions in guidance may be somewhat vague or ambiguous so the tables involve some interpretation. The guidance may offer options including deferment to local decisions. Some enable more or less protection, allow non-laboratory evidence, or include consideration of natural history or fluctuating concentrations.

A CASE EXAMPLE

An unconventional specific example is provided by selenium, based on the USEPA freshwater criterion which includes a 29-page problem formulation (USEPA 2016a). Selenium is bioaccumulative and the best measure of exposure is concentrations in eggs and ovaries followed by whole fish or muscle concentrations. Aqueous concentrations are a less desirable alternative. The assessment endpoint is the most sensitive effect on aquatic life, which is deformities and associated early life-stage mortality in fish. Relative to dimensions from guidelines, those from cases should be less general or ambiguous.

Entity—fish community

Attribute—gross deformity or resulting mortality

Magnitude—10% deformed or dead (EC₁₀)

Proportion—95% of genera protected (HC₀₅)

Uncertainty—unspecified but presumably 50% of the uncertainty distribution is covered by the best estimate

Exposure Metric—fish tissue dry concentration (mg/kg)

Duration—instantaneous (30 days when average aqueous concentrations are used)

Frequency—not to be exceeded (recovery requires 10 years)

Applicability—national fresh waters (lotic and lentic are distinguished for aqueous concentrations)

DISCUSSION AND CONCLUSIONS

Currently, only the USEPA explicitly includes problem formulation or assessment endpoints in their derivation of water quality criteria. The USEPA has used various assessment endpoints for criteria, ranging from the vague goal “ecosystem health” to the much clearer, “the survival, growth, and reproduction of a high percentage of species of a diverse assemblage of freshwater aquatic animals (fish, amphibians, and invertebrates) and plants” (USEPA 2016a). However, for benchmark assessment endpoints to operationally defined, they should be paired with the level of protection (i.e., the magnitude, proportion, or both). The probability of protection given uncertainty or variability might also be considered as a dimension of the benchmark, although that is not current practice in any nation. Problem formulations should also explicitly define the temporal dimensions, the measure of exposure, and the range of applicability. By moving beyond entity and attributes, fully specifying the dimensions of benchmarks clarifies how the endpoint is protected.

To determine the appropriate values of the dimensions of a benchmark, problem formulations should assemble, weight and weigh the available evidence (USEPA 2016c). A formal weight of evidence process could provide a consistent rationale for selecting standard

or alternative data and methods for data analysis, modeling, and interpretation. Narrative descriptions of the evidence do not provide confidence or transparency. Also, completely and consistently defined endpoints and levels of protection can more reliably reveal differences in sensitivity among taxa, attributes, and test systems.

Finally, by clarifying what dimensions should be determined for each benchmark, problem formulation can reveal the need for new data or innovative assessment methods. The temporal dimensions, for example, seem to be ripe for reconsideration, given the lack of continuous monitoring data and the variation among chemicals and taxa in the kinetics of uptake and the time to response and recovery.

Problem formulation may appear to be an impediment to efficient benchmarks development. However, practices such as the development of well-justified default dimensions and clear criteria for accepting or departing from the defaults, particularly in the context of tiered benchmark values, can minimize the added effort. At minimum, even if defaults are used, the dimensions serve as a check list to demonstrate that assessors have considered each dimension and for this chemical the defaults are not inappropriate. More importantly, considering all dimensions during problem formulation can increase the assurance of appropriately protective benchmarks. Specifying each dimension assures that the evidence relevant to each is considered, choices are justified, and when appropriate non-standard data and methods are applied.

Acknowledgements—

The author has no conflict of interest and received no financial support other than his salary. Constructive comments from Susan Cormier, Russell Erickson, Susan Norton, Michael Elias, and Kathryn Gallagher as well as anonymous reviewers greatly improved the manuscript.

REFERENCES

- [ANZECC] Australia and New Zealand Environment and Conservation Council. 2000 Australian and New Zealand guidelines for fresh and marine water quality, Vol. 1 The guidelines. Canberra, (Australia): Agriculture and Resource Management Council of Australia and New Zealand Paper No. 4.
- [CCME] Canadian Council of Ministers of the Environment. 2002 Canadian water quality guidelines for the protection of aquatic life, inorganic fluorides. <http://ceqg-rcqe.ccme.ca/download/en/180> (accessed March 29, 2018).
- [CCME] Canadian Council of Ministers of the Environment. 2007 A protocol for the derivation of water quality guidelines for the protection of aquatic life 2007. Winnipeg (Manitoba, Canada): CCME.
- Buchwalter DB, Clements W, Luoma S. 2017 Modernizing water quality criteria in the united states: a need to expand the definition of acceptable data. *Environ Toxicol Chem* 36:285–291. [PubMed: 28117949]
- Cormier S, Suter G, Zheng L. 2013 Derivation of a benchmark for freshwater ionic strength. *Environ Toxicol Chem* 32:263–271. [PubMed: 23161648]
- [DWAF] Department of Water Affairs and Forestry. 1996 South African Water Quality Guidelines, Volume 7: Aquatic Ecosystems. Pretoria (South Africa).
- [EC] European Community. 2011a Common implementation strategy for the Water Framework Directive (2000/60/EC), guidance document no. 27, technical guidance for deriving environmental quality standards. Brussels (Belgium): EC Technical Report-2011-055.

- [EC] European Community. 2011b Nickel EQS dossier. <https://circabc.europa.eu/sd/d/1e2ae66f-25dd-4fd7-828d-9fd5cf91f466/Nickel%20EQS%20dossier%202011.pdf> (accessed March 27, 2018).
- [EC] European Community. 2014 Guidance document no. 32 on biota monitoring (the implementation of EQS_{BIOTA}) under the Water Framework Directive. Brussels (Belgium).
- Gergs A, Classen S, Strauss T, Ottermanns R, Brock TC, Ratte HT, Hommen U and Preuss TG, 2016 Ecological recovery potential of freshwater organisms: Consequences for environmental risk assessment of chemicals In: de Voogt P, editor. Reviews of environmental contamination and toxicology Volume 236 Basel, Switzerland: Springer International Publishing p 259–294. [PubMed: 26423077]
- [IPCS] International Programme on Chemical Safety. 2014 Guidance document on evaluating and expressing uncertainty in hazard characterization. Geneva (Switzerland): World Health Organization Harmonization Project Document 11.
- Kahn F, Syberg K, Palmqvist A. 2017 Are standardized test guidelines adequate for assessing waterborne particulate contaminants? *Environ Sci Technol* 51:1948. [PubMed: 28177624]
- Liu Z 2015 Water quality criteria green book of China. Dordrecht (Netherlands): Springer 161p.
- Schlekat CE, Merrington G, Keverett D, Peters A 2017 Chemical standard derivation for the protection of aquatic life: A guided world tour. *Integr Environ Assess Manag* 13:794–796. [PubMed: 28613029]
- Stephan CE, Mount DI, Hanson DJ, Gentile JH, Chapman GA, Brungs WA. 1985 Guidelines for deriving numeric national water quality criteria for the protection of aquatic organisms and their uses. Washington (DC): U.S. Environmental Protection Agency.
- Suter GW, II. 1989 Ecological endpoints In: Warren-Hicks W, Parkhurst BR, Baker JS, editors. Ecological assessment of hazardous waste sites: a field and laboratory reference document. Corvallis (OR): U.S. Environmental Protection Agency EPA 600/3–89/013. p 2–1–2–8.
- Suter GW, II, Cormier SM. 2008 What is meant by risk-based environmental quality criteria? *Integr Environ Assess Manag* 4:486–489. [PubMed: 18593215]
- [UKTAG] United Kingdom Technical Advisory Group on the Water Framework Directive. 2008 Proposals for environmental quality standards for Annex VIII substance. London (UK): SR1–2007.
- [USEPA] U.S. Environmental Protection Agency. 1988 Ambient water quality criteria for aluminum - 1988. Washington (DC): Office of Water, USEPA EPA 440/5-86-008.
- [USEPA] U.S. Environmental Protection Agency. 1998 Guidelines for ecological risk assessment. Washington (DC): Risk Assessment Forum, USEPA EPA/630/R-95/002F.
- [USEPA] U.S. Environmental Protection Agency. 2003a Ambient aquatic life water quality criteria for atrazine - revised draft. Washington (DC): Office of Water, USEPA EPA-822-R-03-023.
- [USEPA] U.S. Environmental Protection Agency. 2003b Ambient aquatic life water quality criteria for tributyltin (TBT) - final. Washington (DC): Office of Water, USEPA EPA 822-R-03-031.
- [USEPA] U.S. Environmental Protection Agency. 2011 A field-based aquatic life benchmark for conductivity in central appalachian streams. Cincinnati (OH): Office of Research and Development, National Center for Environmental Assessment, USEPA EPA/600/R-10/023F.
- [USEPA] U.S. Environmental Protection Agency. 2012 Benchmark dose technical guidance. Washington (DC): Risk Assessment Forum, USEPA EPA/100/R-12/001.
- [USEPA] U.S. Environmental Protection Agency. 2013 Aquatic life ambient water quality criteria for ammonia – freshwater. Washington (DC): Office of Water, USEPA EPA 822-R-13-001.
- [USEPA] U.S. Environmental Protection Agency. 2016a Aquatic life ambient water quality criterion for selenium – freshwater. Washington (DC): Office of Water, USEPA EPA 822-R-16-006.
- [USEPA] U.S. Environmental Protection Agency. 2016b Public review draft: field-based methods for developing aquatic life criteria for specific conductivity. Washington (DC): Office of Water, USEPA EPA-822-R-07-010.
- [USEPA] U.S. Environmental Protection Agency. 2016c Weight of evidence in ecological assessment. Washington (DC): Risk Assessment Forum, USEPA EPA/100/R-16/001.

Warne M, Batley G, van Dam R, Chapman J, Fox D, Hickey C, Stauber J. 2015 Revised method for deriving Australian and New Zealand water quality guideline values for toxicants. Brisbane, Queensland (Australia): Department of Science, Information Technology and Innovation.

Table 1.

Generic dimensions inferred from international water quality guidelines (CCME 2007; DWAF 1996; EC 2011; Warne et al. 2015). Unspecified dimensions highlighted in grey.

Dimension	Definition (black circle) and Derivation (open circle)			
	Australia/New Zealand	Canada	European Union	South Africa
Entity	<ul style="list-style-type: none"> • Unspecified ◦ None 	<ul style="list-style-type: none"> • All aquatic animals and plants ◦ Default 	<ul style="list-style-type: none"> • Pelagic communities ◦ Default 	<ul style="list-style-type: none"> • Aquatic ecosystems ◦ Default
Attribute	<ul style="list-style-type: none"> • Survival, growth, reproduction and others that are ecologically relevant ◦ Defaults and others if demonstrated relevance 	<ul style="list-style-type: none"> • No adverse toxic effects, including embryonic development, hatching, or germination success; survival of juvenile stages; growth, reproduction, and survival of adult stages ◦ Default effects and others if ecologically relevant and scientifically sound 	<ul style="list-style-type: none"> • Maximum concentration—survival, development, reproduction ◦ Defaults and others if relevant to population sustainability 	<ul style="list-style-type: none"> • Apparently, the conventional survival, growth and reproduction ◦ Unspecified
Magnitude	<ul style="list-style-type: none"> • No effect concentration, 10% reduction, 10–20% reduction, or 50% reduction ◦ Toxicity tests—LC/EC/ICx 	<ul style="list-style-type: none"> • Short term—50% reduction ◦ Acute lethality or immobilization tests—LC/EC50 • Long term—no negative effect preferred, then 10% reduction; then 11–25% reduction, if necessary ◦ Chronic toxicity test—ECx 	<ul style="list-style-type: none"> • Maximum concentration—50% reduction ◦ Acute lethality/immobilization tests—L(E)C50 • Continuous concentration—10% reduction ◦ Chronic toxicity test—EC10 	<ul style="list-style-type: none"> • Maximum concentration—50% reduction ◦ Acute lethality, LC50 • Continuous concentration—unspecified ◦ Chronic toxicity test—LOEC or NOEC
Proportion	<ul style="list-style-type: none"> • 99%, 95%, 90%, or 80% of genera protected, depending on conservation value and level of disturbance ◦ SSD 	<ul style="list-style-type: none"> • 95% of genera protected ◦ SSD 	<ul style="list-style-type: none"> • 95% of genera protected ◦ SSD 	<ul style="list-style-type: none"> • 95% of genera protected ◦ SSD
Uncertainty	<ul style="list-style-type: none"> • Unspecified ◦ None 	<ul style="list-style-type: none"> • Unspecified but with uncertainty on SSD reported ◦ 95% confidence interval on 5th centile of SSD 	<ul style="list-style-type: none"> • Unspecified, but factors may be applied ◦ None 	<ul style="list-style-type: none"> • Unspecified ◦ None
Duration	<ul style="list-style-type: none"> • Unspecified ◦ Defined & determined by states & territories 	<ul style="list-style-type: none"> • Unspecified ◦ None 	<ul style="list-style-type: none"> • Maximum—maximum or centile ◦ Continuous—annual average ◦ Default principles; member nations vary in implementation 	<ul style="list-style-type: none"> • Unspecified ◦ None
Frequency	<ul style="list-style-type: none"> • 95% of values at a site fall below the guideline value ◦ Determined by states & territories 	<ul style="list-style-type: none"> • Unspecified ◦ None 	<ul style="list-style-type: none"> • Unspecified ◦ None 	<ul style="list-style-type: none"> • Unspecified ◦ None
Applicability	<ul style="list-style-type: none"> • National, by categories based on ecosystem value and level of disturbance ◦ Categorization method unspecified; left to states & territories 	<ul style="list-style-type: none"> • Above natural background ◦ Background definition unspecified 	<ul style="list-style-type: none"> • European freshwaters and salt waters ◦ 5 ppt threshold salinity • Above natural background ◦ Concentrations in headwaters or pristine areas or the regional 10th centile 	<ul style="list-style-type: none"> • National fresh waters ◦ Unspecified

Table 2.

Generic dimensions extracted from USEPA Water Quality Guidelines (Stephan et al. 1985, USEPA 2016b) and recent criteria. Unspecified dimensions highlighted in grey.

Dimension	Definition (black circle) and Derivation (open circle)		
	U.S. conventional	U.S. phytotoxics	U.S. field-based for specific conductivity
Entity	<ul style="list-style-type: none"> Community of aquatic animals Default 	<ul style="list-style-type: none"> Aquatic plants Default for herbicides and other primarily phytotoxic chemicals 	<ul style="list-style-type: none"> Communities of aquatic invertebrates Invertebrates are most sensitive and provide the best data
Attribute	<ul style="list-style-type: none"> Maximum concentration—survival Continuous concentration—survival, growth, reproduction Defaults 	<ul style="list-style-type: none"> Biomass, production, and other biologically important attributes Default for plants 	<ul style="list-style-type: none"> Local extirpation Practical given field data for occurrences
Magnitude	<ul style="list-style-type: none"> Maximum concentration—90% survival Acute lethality or immobilization tests—L₁(EC50/2) Continuous concentration—90% or 80% survival, growth or reproduction Chronic toxicity tests—EC10 or EC20 	<ul style="list-style-type: none"> 20% or 50% reduction 96-hour toxicity tests 	<ul style="list-style-type: none"> 95% of occurrence Distribution of occurrences
Proportion	<ul style="list-style-type: none"> 95% of genera protected SSD 	<ul style="list-style-type: none"> Unspecified Lowest value 	<ul style="list-style-type: none"> 95% of genera protected SSD
Uncertainty	<ul style="list-style-type: none"> Unspecified None 	<ul style="list-style-type: none"> Unspecified None 	<ul style="list-style-type: none"> Unspecified, but confidence intervals reported Bootstrap of SSD
Duration	<ul style="list-style-type: none"> Maximum concentration—1 hour Continuous concentration—96 hours Based on time to effects 	<ul style="list-style-type: none"> Maximum concentration—1 hour Continuous concentration—96 hours None 	<ul style="list-style-type: none"> Maximum—1 day Continuous—1 year Based on consideration of time to effects and sampling frequency
Frequency	<ul style="list-style-type: none"> Once per 3 years Based on time to recovery 	<ul style="list-style-type: none"> Once per 3 years Based on time to recovery 	<ul style="list-style-type: none"> Once per 3 years Based on time to recovery considering affected biota
Applicability	<ul style="list-style-type: none"> National freshwater and saltwater, site-specific possible None 	<ul style="list-style-type: none"> National freshwater and salt water None 	<ul style="list-style-type: none"> Regional streams, pH>6 Region of derivation of sufficiently similar area