An Australian Government Initiative

Australian & New Zealand GUIDELINES FOR FRESH & MARINE WATER QUALITY

Method for deriving Australian and New Zealand water quality guideline values for protecting aquatic ecosystems from toxicants – update of 2018 version

Prepared for the Australian and New Zealand Guidelines for Fresh and Marine Water Quality

August 2025

Water Quality Guidelines is a joint initiative of the Australian and New Zealand governments, in partnership with the Australian states and territories.

Cataloguing data
This publication (and any material sourced from it) should be attributed as: Warne MStJ, Batley GE, van Dam RA, Chapman JC, Fox DR, Hickey CW, Stauber JL and Fisher R 2025. *Method for deriving Australian and New Zealand water quality guideline values for protecting aquatic ecosystems from toxicants – update of 2018 version*. Prepared for the Australian and New Zealand Guidelines for Fresh and Marine Water Quality. CC BY 4.0. Australian and New Zealand Governments and Australian state and territory governments, Canberra, ACT, Australia.

Contact
Australian Government Department of Climate Change, Energy, the Environment and Water
GPO Box 858 Canberra ACT 2601
Switchboard +61 2 6272 3933 or 1800 900 090
Email waterquality@dcceew.gov.au

Acknowledgements
This report was peer reviewed by 3 independent reviewers and was also reviewed and approved by jurisdictional technical and policy oversight groups prior to being published. We also acknowledge the numerous individuals who provided advice throughout the revision of the method.

Contents

Figures

Tables

Appendix Figures

Appendix Tables

Preface

As part of the revision of the 2000 *Australian and New Zealand Guidelines for Fresh and Marine Water Quality* (ANZECC and ARMCANZ 2000a, b, c; referred to herein as the '2000 Guidelines'), a number of working groups were established to review particular sections of the Guidelines. A Toxicants and Sediments Working Group was established to investigate necessary revisions to the Toxicants section. This was initially done at a workshop at CSIRO Land and Water, Lucas Heights, NSW in April 2010. A contract to undertake these revisions was issued by the Council of Australian Government's Standing Council on Environment and Water in February 2013. The original version of the current report was prepared by the Queensland Department of Science, Information Technology and Innovation in consultation with selected members of the Toxicants and Sediments Working Group. The Toxicants and Sediments Working Group produced 2 reports. The first report, originally published in 2014 (Batley et al. 2018), described the technical rationale for the key changes made to the derivation method for toxicant guideline values set out in the 2000 Guidelines. The second report, which is this current report, provides the details of the method for deriving water-quality guideline values for metal, non-metallic inorganic and organic toxicants in Australia and New Zealand. The first version of this report was published in August 2015 (Warne et al. 2015) and was subsequently updated in 2016 and 2017, and the second version was published in 2018 (Warne et al. 2018). Further updates were made over the ensuing 6 years, and the third and current version was published in 2025. All versions of the report were peer reviewed by 3 reviewers and, subsequently, the jurisdictional oversight committee overseeing the technical aspects of the ANZG (2018) Guidelines. Key revisions between the 2015 and 2018 methods and, subsequently, the 2018 and current methods are documented in Appendix A.

It is important to note that the method for deriving Australian and New Zealand water quality guideline values for toxicants should be used in conjunction with the relevant information and guidance provided in the *Australian and New Zealand Guidelines for Fresh and Marine Water Quality* ANZG (2018).

Finally, this update to the toxicant guidelines derivation method is dedicated to the late Dr Michael St John Warne (17 July 1962 to 19 February 2024). Michael made an enormous contribution to the field of ecotoxicology in Australia and internationally. He introduced the species sensitivity distribution approach for deriving toxicant guideline values for Australia and New Zealand (Warne 1998, 2001) and played a pivotal role in its implementation and improvement over the ensuing 25+ years. He is greatly missed by the other co-authors of this report and, undoubtedly, the broader international ecotoxicology community.

Summary

This document provides the details (process and rules) of the updated method for deriving water-quality guideline values for metal, non-metallic inorganic and organic toxicants for the protection of aquatic ecosystems in Australia and New Zealand. The method is focused on the derivation of default (i.e. national or Australian and New Zealand) guideline values to protect against long-term (chronic) direct toxicity but provides additional guidance, where necessary, for the derivation of regional, site-specific and short-term (acute) guideline values. Definitions are provided for acute-toxicity and chronic-toxicity data types, while the method also specifies the type of toxicity test endpoints that are considered acceptable. All data need to meet minimum data-quality requirements as determined through a formal quality scoring process and, potentially, the use of professional judgement. The preferred method for deriving guideline values continues to be based on the use of a species sensitivity distribution of chronic toxicity data. While the less preferred assessment factor method for deriving guideline values is also covered, it is unchanged from the 2000 Guidelines and, hence, is not described in detail here. The minimum data requirements for using a species sensitivity distribution have increased from the previous version of the method – that is, toxicity data for ≥ 6 species that belong to ≥ 4 taxonomic groups. However, using toxicity data from ≥ 8 species is strongly encouraged, and from ≥ 15 species is considered optimal. Updated guidance is provided for determining if a dataset is bimodal and how to deal with this when deriving guideline values. The endorsed software packages for deriving guideline values, in particular for default guideline values, are ssdtools or shinyssdtools, both of which employ a technique termed 'model averaging'. Model-averaged guideline values can be derived for different levels of species protection, and 80%, 90%, 95% and 99% species-protection values should be derived for all default guideline values. The basis of the reliability classification for guideline values remains the same as the 2018 version of the method – where the number and types of toxicity data points and an estimate of the fit of the distribution to the data are considered. The method also provides guidance on how to account for bioaccumulation and chemical formulations, and links to other relevant guidance and information on the ANZG (2018) website.

1 Introduction

1.1 Background

The 2000 Australian and New Zealand Guidelines for Fresh and Marine Water Quality (ANZECC and ARMCANZ 2000a, b, c; referred to herein as the '2000 Guidelines') and the 2000 Australian Guidelines for Water Quality Monitoring and Reporting (ANZECC and ARMCANZ 2000d) represented a major step forward in the assessment and monitoring of water quality. Key advances at that time included the adoption of a risk-based approach to managing water quality, the notion of different levels of ecosystem condition and associated protection, new methods for deriving water-quality guideline values (GVs; termed 'trigger values' in the 2000 Guidelines) for toxicants based on species sensitivity distributions (SSDs), and the promotion of integrated assessment (i.e. assessments combining physicochemical, toxicological and biological indicators).

A review of the above 2 Guidelines commenced in 2009. Initial investigation of technical revision requirements and some high-priority revisions were conducted by a series of working groups, each consisting of appropriate experts. The Toxicants and Sediments Working Group (Working Group 4) was responsible for the method for deriving GVs for toxicants in surface waters. The method described in this report is the culmination of the deliberations of that working group and subsequent updates to the method in 2015 (Warne et al. 2015), 2018 (Warne et al. 2018) and the present. The method represents a component of, and should be used in conjunction with, the broader guidance provided in the *Australian and New Zealand Guidelines for Fresh and Marine Water Quality* (ANZG 2018).

1.2 Purpose of this report

This report presents the revised method for deriving water-quality GVs for toxicants in surface waters in Australia and New Zealand for the protection of aquatic ecosystems. It supersedes the guidance provided in the 2000 Guidelines, Warne et al. (2015) and Warne et al. (2018). It provides a step-by-step process for deriving GVs. The method is focused on the derivation of default (i.e. national or Australian and New Zealand) guideline values (DGVs) but also provides additional guidance and supporting references, where appropriate, for the derivation of regional, site-specific and short-term GVs. In this report, we use the term 'GV' to apply to the derivation of any type of guideline value and only use the term 'DGV' when referring to default guideline values.

Two distinctly different methods can be used to derive GVs: the SSD or assessment-factor (AF) methods. The current document mostly focuses on the SSD method, as it is the preferred method. For further background information on the less preferred AF method for calculating toxicant GVs, readers are referred to Warne (1998) and Warne (2001).

Although the method has been developed specifically for GVs for surface water quality, it can also serve as the starting point for methods for deriving toxicant GVs for sediment quality and groundwater quality. However, each of these will require different considerations that are not covered in the current methodology. Additionally, the method does not specifically preclude modifications to the derivation approach that may be needed to account for unusual characteristics of a toxicant; however, any aspects of a derivation that are not explicitly captured by the current method need to be clearly detailed and justified in the documentation such that they can be properly reviewed and approved.

2 Overview of the revised method

2.1 Key features

The revised method for toxicant GV derivation is very similar to that used in the 2000 Guidelines and that described in Warne et al. (2018) and has retained the following key features:

- the method is risk-based, in that (when using the SSD method) it produces GVs that are based on probability distributions of species' sensitivities and that specify a percent of species that is likely be protected (or, conversely, affected)
- the method uses a hierarchical, tiered framework that recommends the use of the SSD method instead of the AF method
- the method includes an assessment of the reliability of the GVs
- the method encourages the conduct of site-specific investigations and the derivation of site-specific GVs
- the method includes a policy of transparency so that it is clear how the GVs were derived.

The revised method includes the following key features compared to the method used in the 2000 Guidelines:

- revised definitions of acute and chronic toxicity and an altered classification of toxicity tests
- guidance on the derivation of GVs for short-term exposure and when their derivation is appropriate (also see Batley et al. 2018)
- an expanded suite and preferred hierarchy of statistical estimates of toxicity that are deemed acceptable to derive GVs
- reducing the reliance on NOEC (see Glossary for definitions) data for GV derivation where appropriate (e.g. NOEC data could be excluded when there are acceptable alternative data for ≥ 15 species that belong to ≥ 4 taxonomic groups)
- guidance to improve the design of toxicity tests for the purposes of concentration-response modelling and the associated estimation of effect concentrations (i.e. EC/IC/LC and NEC/NSEC data) (also see Batley et al. 2018)
- inclusion of non-traditional endpoints (e.g. behavioural, biochemical), provided their ecological relevance has been unequivocally demonstrated
- ability to combine chronic and acute (converted to chronic using conversion factors) toxicity data or fresh and marine data in one dataset for GV derivation under specific circumstances
- adoption of a new statistical approach, termed model averaging, and associated software, ssdtools (and its online web platform shinyssdtools), for deriving GVs (replacing Burrlioz 2.0)
- inclusion of a statistical mixture distribution within the model averaging approach as a means of detecting and, if appropriate, accommodating bimodality in the SSD

- an improved method for determining the reliability of GVs that considers (i) the hierarchy of acceptable data types, (ii) the sample size and (iii) a visual estimation of goodness of fit
- GVs calculated using an AF method are classified as having 'unknown reliability'.

Batley et al. (2018) details the rationale for the key revisions reflected in the Warne et al. (2018) method, while the rationale for the additional revisions reflected in the current method are summarised in Appendix A (Table A.2).

2.2 Default guideline values and analytical methods

Prior to deriving a DGV for a toxicant, confirm that there is an analytical method for measuring the toxicant (or surrogate chemical marker or metabolite) in water or that the toxicant is known to be in use and it is reasonable to expect that it will be released to aquatic environments (in which case an analytical method would also need to be developed). Where a newly developed DGV is lower than the current, routine analytical limits of reporting for the toxicant, the DGV can still be published and this limitation stated. It is anticipated that such occurrences will drive the development of more sensitive analytical methods for the toxicant. Information on the availability of appropriate analytical methods can be sought from standard methods published by organisations such as the United States Environmental Protection Agency, Organisation for Economic Co-operation and Development or American Society for Testing and Materials, or by consulting with appropriately accredited commercial analytical laboratories, preferably in Australia or New Zealand. Additional guidance on requirements for analytical methods can be found in ANZG (2018; see *Monitoring – Laboratory analysis*).

2.3 Preferred guideline value derivation method

Only 2 methods are recognised in Australia and New Zealand for the derivation of DGVs for toxicants: the SSD method and the AF method. Background information on these methods can be found in Warne (1998), Shao (2000), Campbell et al. (2000) and Warne (2001). The GV-derivation method does not combine the use of SSDs and AFs when deriving GVs, as occurs elsewhere (e.g. European Commission 2018) – either one or the other method is used, depending on data availability (see Section 3.4.5 and Section 4). Reasons for not applying AFs to GVs derived from SSDs are detailed in Fox and Batley (2022).

The SSD method is the preferred method for deriving GVs and should be used whenever the toxicity data for a toxicant meet the minimum data requirements for this method. This method should also be used when quantitative structure-activity relationships (QSARs) are used for non-polar narcotic chemicals (see Warne 2001 for additional details). As an SSD is the preferred method for calculating GVs, the current document focuses on this method. The Burrlioz 2.0 software used to calculate toxicant GVs since ANZECC and ARMCANZ (2000a) has been replaced by the open-source software package, ssdtools (Thorley and Schwarz 2018; Thorley et al. 2025) and its accompanying web-based interface, shinyssdtools (Dalgarno 2018) (https://www.waterquality.gov.au/anz-guidelines/guideline-values/default/water-quality-toxicants/shinyssdtool), which are collectively referred to herein as '(shiny)ssdtools'. The SSD modelling approach used by (shiny)ssdtools represents an improvement to that used by Burrlioz 2.0. While Burrlioz (1.0 and 2.0) provided a technically defensible approach to deriving toxicant GVs for over 20 years, the transition to (shiny)ssdtools significantly strengthens the

technical basis of toxicant GVs for Australia and New Zealand (Fox et al. 2023) and ensures that the derivation method reflects the current state of the science for SSD modelling. Section 0 provides further details on (shiny)ssdtools.

Additional guidance for deriving site-specific GVs is provided by ANZG (2018; see *Deriving guideline values*), Huynh and Hobbs (2019) and van Dam et al. (2019). Also, state and territory governments may have their own guidance for deriving and applying site-specific GVs and they should always be consulted on appropriate methods during the planning stage of the derivation process.

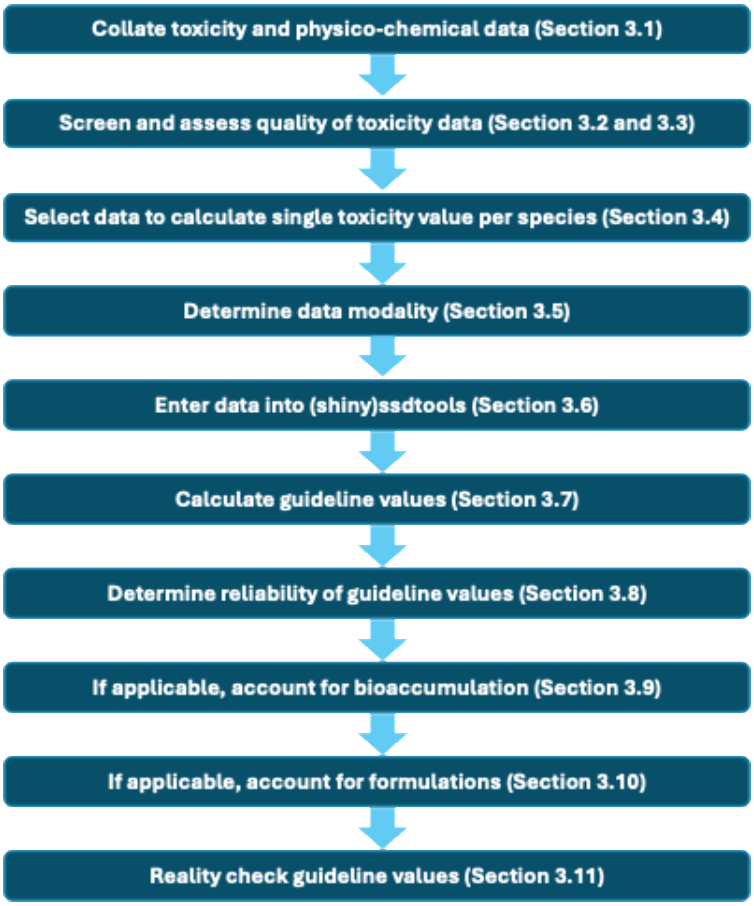
The method for deriving toxicant GVs should be used in conjunction and in accordance with the guidance for water-quality management and assessment that is provided in the current *Australian and New Zealand Guidelines for Fresh and Marine Water Quality* (ANZG 2018).

3 Method for calculating guideline values using the species sensitivity distribution approach

The development of GVs is a specialised process that requires sound knowledge of the method and professional judgement throughout. Therefore, it is strongly recommended that people deriving GVs should have detailed knowledge of ecotoxicology and environmental chemistry and, for DGVs, previous experience deriving GVs.

An overview of the revised method for calculating GVs using the SSD method is provided in Figure 1. Each step is subsequently described in detail below. While DGVs are derived to protect against harmful effects from long-term (i.e. chronic) exposure, the method set out in this report can also be used to derive GVs for short-term (i.e. acute) exposure, which may be useful at regional or site-specific scales or for other uses such as setting licence conditions or in prosecutions. Short-term GVs typically aim to protect most species against lethality or short-term sub-lethal effects during intermittent and transient exposures; thus, the toxicity data used to derive short-term GVs should adequately reflect such exposures. See Batley et al. (2018) for further guidance on the derivation of short-term GVs.

Figure 1 Schematic representation of the revised method for deriving guideline values using the species sensitivity distribution approach



3.1 Collating toxicity and physicochemical data

3.1.1 General

Acute, chronic (single-generation and multi-generation), laboratory, field, mesocosm or microcosm toxicity data should be obtained by conducting searches of the scientific literature, including water-quality documents from other jurisdictions (e.g. Canada, USA, Europe) and appropriate databases including, but not limited to, the ECOTOX database (US EPA 1994) and the Australasian Ecotoxicology Database (Warne et al. 1998; Warne and Westbury 1999; Markich et al. 2002; Langdon et al. 2009). For DGVs, toxicity data for both Australasian and overseas species should be collated (see Section 3.2).

For chemicals that do not bioaccumulate (i.e. chemicals with an octanol-water partition coefficient [log K_{ow}], bioconcentration [log BCF] or bioaccumulation [log BAF] factors of less than 4), only toxicity data related to the uptake of chemicals from water and from water and food combined should be collated (i.e. experiments where uptake is solely from food should not be included). For bioaccumulating chemicals, tests that assess uptake from water, water and food, and food only are appropriate (refer to Section 3.9).

All pesticides and many other chemicals are normally released into the environment as mixtures (e.g. pesticides will contain the active ingredient and additives or adjuvants designed to improve the effectiveness of the active ingredient). For such chemicals, only toxicity data generated by exposing test organisms to a relatively pure form of the active ingredient (refer to Section 3.10) should be used to calculate DGVs. However, toxicity data for formulations should also be collated in case it is decided to derive a formulation-corrected GV (refer to Section 3.10).

The 2000 Guidelines stipulated that only data from peer-reviewed scientific journals be used to derive GVs. However, for the current derivation method, any data (including from internal reports, consultancy reports and confidential registration data) can be used provided that:

- the document (and data) is publicly available

or

- the document (and data) is made publicly available as part of the derivation process (e.g. documents could be hosted on an organisation's website or, if practical, appended to the GV-derivation document).

and

- the data are assessed as being of acceptable quality (see Section 3.3.1).

Commercial-in-confidence data, such as those supplied by companies for assessments by the Australian Pesticides and Veterinary Medicine Authority or the Australian Industrial Chemicals Introduction Scheme, can be used, provided the owner of the data authorises their use and makes the data publicly available. Alternatively, where the owner approves the use of the data but not their release, the data can be used, provided an agreed independent assessor with expertise in GV derivation has assessed and confirmed the data are of acceptable quality (refer to Section 3.3.1). Evidence of such an assessment should be included with the final GVs (e.g. an appendix of the details and outcomes of the assessment).

As a general rule, toxicity data published prior to 1980 should not be included, as these data are considered more likely to be unreliable due to advances in experimental and analytical capabilities since that time (Warne 1998). Exceptions to this guidance can be made with appropriate professional judgement and justification.

3.1.2 Acute and chronic toxicity

The emphasis of the data search should be on chronic toxicity data, as these are most appropriate and preferred type of data for deriving long-term (chronic) GVs. However, acute toxicity data should also be collated if there are insufficient chronic data to meet the minimum data requirements for the SSD approach (Section 3.4) or if a short-term GV is being derived.

Data should be sorted into acute or chronic toxicity based on the following definitions:

Acute toxicity: A lethal or adverse sub-lethal effect that occurs due to exposure to a chemical for a short (relative to the organism's life span) period.

Chronic toxicity: A lethal or adverse sub-lethal effect that occurs due to exposure to a chemical for a period of time that is a substantial portion of the organism's life span, or an adverse effect on a critical early life stage.

Examples of endpoints and durations for different types of organisms that are considered acute and chronic are presented in Table 1. The recommended test durations in Table 1 apply to tests on temperate species, typically undertaken in water temperatures ranging from 15°C to 25°C. The information in Table 1 is comprehensive but not exhaustive, and there will be instances where toxicity tests will not fit into any of the listed categories (i.e. organism type, life stage, test endpoint, test duration). In such cases, professional judgement will need to be used to classify a toxicity test as acute or chronic. Related to this, the duration of acute and chronic tests for polar and tropical species will differ from those presented in Table 1 (typically longer for polar and shorter for tropical) (Batley et al. 2018). For example, chronic test durations for tropical *Lemna* and *Hydra* species are 4 days (Trenfield et al. 2020), while the requirement for similar temperate species would be 7 days. In contrast, chronic-test durations for polar (Antarctic) species can range from several weeks to more than a year, and acute-test durations can be 2 or more weeks (King et al. 2024). Consequently, professional judgement may also be needed to determine acute and chronic durations for polar and tropical species.

The diverse life-history strategies of invertebrate taxa means that it is not possible to make a general rule when defining chronic and acute test durations; hence, invertebrates have been divided into 2 groups: microinvertebrates and macroinvertebrates (see Table 1). Given the above, it is likely that best professional judgement may be needed to determine whether a particular test should be regarded as acute or chronic. The basis for all professional-judgement decisions must be transparent, understandable and documented.

Table 1 Indicative classification of acute and chronic toxicity tests for temperate species

Organism type	Life stage[a]	Relevant endpoint[b]	Test duration
Acute toxicity			
Fish and amphibians	Adults, juveniles	All[c]	< 21 d
	Embryos, larvae	All	< 7 d
Macroinvertebrates[d]	Adults, juveniles	All	< 14 d
	Embryos, larvae	All (except fertilisation, larval development, metamorphosis)	< 7 d
	Embryos, larvae	Larval development/metamorphosis	< 48 h
Microinvertebrates[e]	Adults, juveniles, larvae	All (except fertilisation, larval development – see microinvertebrate chronic)	< 7 d
Macrophytes	Mature	All	< 7 d
Macroalgae	Mature	Lethality and growth	< 7 d
Microalgae	Not applicable	All	≤ 24 h
Microorganisms	Not applicable	All	≤ 24 h
Chronic toxicity			
Fish and amphibians	Adults, juveniles	All[f]	≥ 21 d
	Embryos, larvae	All	≥ 7 d
Macroinvertebrates	Adults, juveniles	All (except reproduction, larval development, metamorphosis, fertilisation – see below)	≥ 14 d
	Embryos, larvae	All (except reproduction, larval development, metamorphosis, fertilisation – see below)	≥ 7 d
	Adults, juveniles, larvae	Reproduction	≥ 14 d (or ≥ 3 broods for large cladocerans)
	Embryos	Larval development, metamorphosis	≥ 48 h
	Gametes	Embryo fertilisation	≥ 1 h
Microinvertebrates	Adults, juveniles, larvae	Reproduction	≥ 7 d (or ≥ 3 broods for small cladocerans)
	Adults, juveniles, larvae	Lethality, immobilisation	≥7 d
	Embryos	Larval development	≥ 48 h
	Gametes	Embryo fertilisation	≥ 1 h
Macrophytes	Mature	All	≥ 7 d
Macroalgae	Mature	All	≥ 7 d
	Early life stages	Lethality	≥ 7 d
	Early life stages	Development	≥ 48 h
	Early life stages	Fertilisation	≥ 1 h
Microalgae	Not applicable	All	> 24 h
Microorganisms	Not applicable	All	> 24 h

[a] The life stage at the start of the toxicity test.

[b] Endpoints need to be ecologically relevant – see Section 3.2.
[c] For acute tests, 'all' refers to all ecologically relevant endpoints for a particular life stage of a particular species.
[d] Macroinvertebrates include invertebrates where full-grown adults are ≥ 2 mm long (e.g. decapods, echinoderms, molluscs, annelids, corals, amphipods, larger cladocerans [e.g. *Daphnia magna*, *Daphnia carinata* and *Daphnia pulex*] and insect species where larvae are ≥ 2 mm long).
[e] Microinvertebrates are defined here as invertebrate species where full-grown adults are typically < 2 mm long (e.g. some cladocerans [e.g. *Ceriodaphnia dubia* and *Moina australiensis*], copepods, conchostracans, rotifers, acari, bryozoa, hydra).
[f] For chronic tests, 'all' encompasses all ecologically relevant endpoints measured in both single- generation and multi-generation tests.

3.1.3 Supporting information

Wherever possible, relevant physicochemical information should be obtained from reliable sources for every chemical for which a GV is being derived. Relevant supporting information includes Chemical Abstract Services number, International Union of Pure and Applied Chemistry name, common name, aqueous solubility (at standard temperature 25°C), boiling and melting point, chemical formula, key degradation processes (including any metabolites of concern), half-life in water and sediment, molecular weight, octanol–water partition coefficient, organic carbon–water partition coefficient, partition coefficient, bioconcentration factor, specific gravity and vapour pressure. In relation to the specifics of the chemical used in a toxicity test from which data are used for a GV derivation, information on the form (e.g. the salt that was tested and its hydration) and purity of the chemical that was used should also be documented. The oxidation state of the chemical may also be important information (e.g. chromium (III) or (VI)). Other types of information on the chemical that should be obtained include its mode of action, whether it is an essential element, and if there are any factors known to modify its toxicity. The supporting information will provide relevant context for the derivation as well as for users of the GV, including (but not limited to):

- need to derive a short-term (acute) or long-term (chronic) GV or if bioaccumulation needs to be taken into account
- likely behaviour of the toxicant in the environment and whether a short-term (acute) or long-term (chronic) GV is needed or the type of matrix for which a GV will be more relevant (e.g. water, sediment)
- information to support decisions on data quality, such as the aqueous solubility limit, and purity and form of the toxicant
- important toxicity-modifying factors (TMFs) that might need to be accounted for in the derivation or application of the GV
- clarity on the exact toxicant for which the GV applies.

The following physicochemical parameters of the control/diluent water used for toxicity testing should also be documented where available: pH, salinity (or conductivity), dissolved oxygen, hardness, alkalinity, dissolved organic carbon, temperature, light level and source (especially for algal and plant testing), diurnal period and any additives to the water (e.g. culture medium, food). Depending on the chemical, it may also be worthwhile documenting major ion concentrations if they are reported (e.g. this will be important for metals and other inorganics for which bioavailability can be influenced by specific major ions).

3.2 Screening toxicity data

Once the toxicity data have been collated, they should be screened to determine their suitability for use in GV derivation. Key considerations for data suitability are described below and are also relevant to Section 3.3.

3.2.1 Ecological relevance

Endpoints that are considered to be ecologically relevant (e.g. lethality, immobilisation, growth, development, population growth, reproduction) can be used to derive GVs. Non-traditional endpoints, such as photosynthesis inhibition, *in-vivo* biochemical and physiological endpoints, behavioural endpoints, and genotoxicity and mutagenicity, may also be used provided that their ecological relevance for the species, or closely related species, has been unequivocally demonstrated. An endpoint is considered to have ecological relevance when it negatively affects a species' ecological competitiveness (i.e. its ability to increase the frequency of its genes in subsequent generations). What is considered ecologically relevant will be both species- and toxicant-specific. An effect on a species' competitiveness can be direct or indirect in the case of a symbiotic organism such as zooxanthellae in corals. Organism-level endpoints (also referred to as 'apical endpoints'), such as survival, immobilisation, growth, development and reproduction, are typically considered to be ecologically relevant endpoints in that they are likely to manifest in effects at the population level. Non-traditional endpoints that have had their ecological relevance demonstrated, through clear evidence that effects on these endpoints will manifest in effects at the organism level, can be used to derive GVs. Otherwise, they should only be used as an additional line of evidence in risk assessments based on weight of evidence (WoE). When deriving site-specific GVs, the onus of proving ecological relevance of an endpoint lies with the organisation or person deriving the GV. Special consideration can be given to the use of non-traditional endpoints for which ecological relevance has not been demonstrated if they are the only data available for unique environments (e.g. polar environments) for which regional or site-specific GVs are to be derived. This extends to the use of genomic and multi-omic data in environmental assessments and, potentially, GV derivation. Again, appropriate justification for all decisions should be provided.

When searching for and compiling data for GV derivation, it is advisable that the literature search be restricted to data based on traditional endpoints and that data from non-traditional endpoints are evaluated only in exceptional circumstances; for example, where there are insufficient traditional data or to address particular site-specific concerns.

In some cases, toxicity data may exist for species that are considered nuisance species, such as regional invasive pest species and toxin-producing, bloom-forming blue-green algae. Unless a professional judgement to exclude potential nuisance species can be clearly made and documented (e.g. where a highly sensitive nuisance species will skew the overall SSD and associated DGVs), then such species should be included in the dataset for GV derivation.

There are few geographical restrictions to the use of toxicity data for deriving DGVs. While toxicity data for Australian and New Zealand species are most relevant and should be used (provided they pass the quality-assessment criteria), toxicity data for non-Australian/New Zealand species should also be used (provided they pass the quality-assessment criteria), to maximise the number of species and taxonomic groups represented in the SSD. Data for temperate and tropical species can be used to derive a single set of DGVs for a toxicant, unless there is sufficient evidence to indicate that there

are clear differences in the sensitivity of species between these climatic zones, in which case DGVs for different climatic zones might need to be derived. Also, given the very different physiological characteristics of polar species and polar environments compared with temperate and tropical regions (Gissi et al. 2020), it is recommended that data for polar species under polar conditions are not used to derive DGVs for other regions. Geographical restrictions will most likely be needed for site-specific GVs (see van Dam et al. 2019 for relevant guidance). Both western and Indigenous knowledge on local species and relevant environmental conditions are likely to be important for determining the need and requirements for site-specific GVs.

3.2.2 Checking for significant limitations of toxicity data

Data with any of the characteristics presented in

Table 2 should not be used, unless a strong justification is provided. A number of these characteristics are discussed throughout Section 3.2 and are also covered in the process for data-quality assessment described in Section 3.3.1. An example of possible justifications to include data that have the types of limitations listed in

Table 2 include that the data represent the only available data for a species or that the exclusion rule results in insufficient data being available to derive the GV. Guideline values derived using data with limited reliability should be accompanied by a caveat describing the nature of the data limitations (also see Section 3.8 regarding caveats on DGVs).

Table 2 Circumstances and types of toxicity data that should not be used for derivation of guideline values[a]

Test characteristic	Conditions that warrant exclusion of the data from default guideline value derivation
Experimental design	Where the test concentrations differ by a large amount (e.g. ≥ 10-fold differences such as 1, 10, 100, 1,000 and 10,000 µg/L) (refer to Section 3.3.1) Where < 3 test concentrations (plus a control) were tested[b] (refer to Section 3.3.1)
Test conditions	Where test water-quality conditions are outside acceptable environmental ranges (e.g. pH < 6 or > 9, salinity > 5‰ for freshwater, < 25‰ or > 36‰ for marine water) (refer to Section 0) Where toxicity-modifying factors (e.g. pH, hardness, dissolved organic carbon) relevant to the toxicant being tested were not measured in the test waters (refer to Section 3.3.1)
Duration of exposure	If not stated or does not conform with Table 1 (refer to Section 3.3.1)
Toxicological endpoint	If endpoint not stated (refer to Section 3.3.1) Endpoints did not include lethality, immobilisation, reproduction, growth, population growth or the equivalent, unless the endpoint has been unequivocally demonstrated to be ecologically relevant (refer to Section 0)
Measure of toxicity (e.g. ECx, NOEC)	If not stated or able to be inferred (refer to Section 3.3.1)
Type of toxicity data	Acute toxicity data that are not EC/IC/LC50s (refer to sections 0 and 3.4.2.2). Note: this condition does not apply if short-term guideline values are being derived (refer to sections 0 and 3.4.2.3) Left-censored data (i.e. '<' toxicity values) where (i) there are uncensored data for that species, (ii) the data point does not sit at the lower end of the distribution of species sensitivities, or (iii) the exclusion of the data would result in a less conservative guideline value (refer to Section 3.4.2) Right-censored data (i.e. '>' toxicity values) that (i) are too far outside the existing data range or (ii) have an overly large influence on the final guideline value (refer to Section 3.4.2)

Test characteristic	Conditions that warrant exclusion of the data from default guideline value derivation
Chemical purity	Where tests are performed using chemical reagents with a purity of < 80% for the toxicant or using chemical formulations (refer to Section 3.10)
Measurement of test chemical	If concentrations of the test chemical were not measured in the test treatments at least once[c] (refer to Section 3.3.1)
Test water contamination	Where significant contamination of the test water was observed in one or more treatments (refer to Section 3.3.1)
Test acceptability criteria	Where a test failed test acceptability criteria (including reference toxicity-testing criteria) and there were no mitigating reasons for the failure (refer to Section 3.3.1)
Concentration-response relationship	Where there was no concentration-response relationship (but also see information above regarding right-censored values) (refer to Section 3.3.1)
Aqueous solubility	If toxicity values are greater than twice the aqueous solubility at the exposure temperature (on the basis that precipitated forms of the chemical do not contribute to toxicity)

[a] Except where the data are of particular significance and a strong justification for their inclusion can be provided and is deemed to be of net benefit to the derivation of the guideline value.

[b] Ideally, ≥ 6 concentrations should have been tested, but some allowance for fewer than this (i.e. 3–5) may be made, depending on the overall reliability of the data and amount of data.

[c] This requirement could be relaxed for persistent and some highly soluble toxicants, and measurement of the stock solution possibly being acceptable, but not for toxicants such as metals. Also, tests where some of the treatments are measured and the remainder are estimated from a regression of the measured concentrations may be acceptable as long as sufficient test concentrations across the full range were measured to enable an adequate representation of the nominal versus measured concentration relationship (e.g. > 4).

3.2.3 Censored data

Selection of censored values (i.e. values expressed as greater than [>] or as less than [<]) will typically require some professional judgement. Note that < and > values should not be used if other non-censored toxicity data are available for the same species. While (shiny)ssdtools can accommodate the incorporation of left-censored (<) and interval-censored (> <) data, it does not currently accommodate right-censored (>) data. Moreover, the censoring capability of (shiny)ssdtools requires validation in the context of model averaging and, at the time of publication of this method, has a range of technical limitations. Thus, when using the current method and until further updates are made to (shiny)ssdtools, the following guidance should be followed.

Toxicity values expressed as greater-than (>) can be used, subject to professional judgement being applied to determine whether they (i) are too far outside the existing data range or (ii) have an overly large influence on the final GV. If the data do not meet either of these criteria, they can be used. When using > values for deriving a GV, the actual value should be used (e.g. a value of > 20 µg/L would, for the purposes of deriving a GV, be used as 20 µg/L).

Less-than (< or ≤) values should be excluded, unless (i) there are no other data for a species, (ii) the data point sits at the lower end of the distribution of species sensitivities, or (iii) the exclusion of the data would result in a less conservative GV. Toxicity values reported as < or ≤ values are typically concentrations at which a measurable (statistically significant) effect may still be observed and, thus, in theory, a conversion factor should still be applied to convert them to a negligible-effect equivalent. For example, where there is a statistically significant effect (relative to the control) at all concentrations including the lowest concentration, then a NOEC cannot be calculated (thus, the NOEC can be referred to as < the lowest concentration), and the LOEC might be the lowest

concentration or a lower concentration that was not tested (thus, the LOEC can be referred to as ≤ the lowest concentration). In such cases, the lowest concentration should be treated as a LOEC, and the LOEC-to-negligible-effect factor of 2.5 (see Section 3.4.2.1) should be applied to this value (e.g. where a LOEC is reported as ≤ 10 µg/L, the value used in the final derivation dataset is 10 µg/L ÷ 2.5, being 4 µg/L). The lower value of a range of values for toxicity (e.g. EC50 = 25–50 µg/L) can also be used, subject to professional judgement. For all the above decisions, apply best professional judgement, and document the reasoning behind all decisions (e.g. where the effect at the lowest concentration is significant but the effect size is ≤ 10%, a conversion factor might not be applied to this concentration).

3.2.4 Hierarchy of statistical estimates of toxicity

The most preferred types of statistical estimates are negligible-effect concentrations that have been directly estimated from the concentration-response relationship and that include, but are not necessarily limited to:

- NEC (van der Hoeven et al. 1997; Fox 2009; Fox and Billoir 2011)
- NSEC (Fisher and Fox 2023, Fisher et al. 2024).

Other appropriate estimates include:

- x% effect/inhibition/lethal concentration where x ≤10 (wherever possible, ECx or ICx data should be used in preference to LCx data, assuming the values for the former estimates are lower than those for the latter estimate)
- BEC10 (Hoekstra and Van Ewijk 1993).

Less preferred estimates include (in order of preference):

- x% effect/inhibition/lethal concentration (EC/IC/LCx), where x > 10 and ≤ 20, and NOEC – these estimates do not require any form of conversion to negligible-effect values
- LOEC, MATC (geometric mean of the NOEC and LOEC) and EC50/IC50/LC50 – these estimates require application of a conversion factor to derive a negligible-effect value (see Section 3.4.2.1)
- acute EC/IC/LC50 converted using an acute-to-chronic ratio (ACR; see Section 3.4.2.2)

The same hierarchy of statistical estimates of toxicity applies to both default (chronic) GVs and short-term (acute) GVs. Notably, the LOEC, MATC and EC/LC50 would all need to be converted to a negligible-effect concentration using an appropriate conversion factor before they can be used for a GV derivation (see Section 3.4.2.1). The limitations of the NOEC, LOEC and MATC, which are derived from hypothesis-testing approaches, have been well documented elsewhere (e.g. Warne and van Dam 2008; van Dam et al. 2012a, b; Fox and Landis 2016).

When toxicity values for multiple statistical estimates of toxicity are reported for a specific endpoint for a species, the rigour and associated defensibility of each of the values should be considered in addition to whether the values are based on a preferred or less preferred estimate when making a decision about which value to use. For example, while as a general rule it is preferable to not use NOEC data to derive GVs (refer to Batley et al. 2018), an EC/IC/LC10 value can be less reliable than a NOEC value if it has been derived from concentration-response data based on too few concentrations to properly characterise the concentration-response relationship. All decisions that include an element of professional judgement should be clearly explained and justified. Further

guidance on the selection of the most appropriate statistical estimate of toxicity is provided in Section 0.

3.2.5 Physicochemical water-quality ranges

The acceptable pH range for freshwaters for which toxicity data have been generated is 6.0–9.0. Typically, tests conducted outside of this pH range should not be included in the generic dataset or in species-specific geometric means. However, exceptions may be made where such data will clearly improve the reliability of the GV or add numerous Australian or New Zealand species to the dataset (with all decisions needing to be transparent and appropriately justified). Moreover, as pH can be a significant modifier of toxicity, it may be useful to derive DGVs for different pH ranges but only if pH is known to significantly affect toxicant bioavailability (e.g. many metals, and ionisable substances such as many pharmaceuticals) and if sufficient data exist (also see Section 0 on the derivation of bioavailability-based GVs). Site-specific GV derivations may also be undertaken for conditions within specific pH ranges for specific sites or regions.

The salinity at which toxicity tests are conducted is important as it defines the ecosystem type (i.e. fresh or marine) and can also be a modifier of toxicity. To date, GVs for Australia and New Zealand have only been derived for freshwater and marine ecosystems. Freshwater GVs should only be derived using ecotoxicity data from tests where the salinity is < 0.5 parts per thousand (‰), while marine GVs should only use toxicity data from tests where the salinity is between 25‰ and 36‰. It is worth noting that pH may vary across this salinity range, which may need to be considered in the event that pH is a TMF of the chemical in question (see Section 0 for more details on TMFs). If there are sufficient ecotoxicity data tested under estuarine conditions, estuarine DGVs (or site-specific GVs) could be derived using the methods set out in this report. Where there are insufficient data to derive estuarine GVs, ANZG (2018) provides guidance on the selection of appropriate DGVs for estuarine waters. The salinity of marine waters can exceed 36‰ under certain conditions or times of year. For such situations, if salinity is known to be a toxicity modifier for the toxicant(s) in question, site-specific GVs are recommended. More than one GV may be required if significant seasonal variability exists.

Guideline values derived for toxicants in one medium (fresh, estuarine or marine) typically should not include data from other media. However, toxicity data from multiple media can be used when there are insufficient data for the medium in question and either (i) statistical analysis reveals there is no difference in the toxicity in the different media or (ii) based on the chemistry of the chemical or its mode of action, there is no reason to expect differences in toxicity. For toxicants where there is no difference in sensitivity between freshwater species and marine/estuarine species, it may be preferable to derive a single combined freshwater and marine SSD and associated GVs (refer to Section 3.4.5 for further details).

Some chemicals, such as pesticides, are released into the environment in the form of commercial formulations. For such chemicals, refer to Section 3.10 to determine which data to use when deriving DGVs and formulation-corrected GVs.

3.3 Assessing the quality of toxicity data

3.3.1 Laboratory-based toxicity data

The quality of all laboratory-based toxicity data being considered in the derivation of GVs should be assessed, apart from those that have already been assessed. Data that have previously been

assessed include those used to derive the ANZECC and ARMCANZ (2000a, b) GVs, water-quality guidelines of other jurisdictions that state that the data have been assessed (e.g. Canada, USA, Europe), and those in the Australasian Ecotoxicology database (Warne et al. 1998; Warne and Westbury 1999; Markich et al. 2002; Langdon et al. 2009). Note that data 'quality' as assessed here is analogous to data 'reliability' as described in toxicity-data assessment approaches elsewhere (e.g. the Moermond et al. 2016 criteria for evaluating ecotoxicity data).

The data quality assessment should be conducted using the Excel™ data-quality spreadsheet that has been developed for this purpose and that can be made available by the Australian Department of Climate Change, Energy, the Environment and Water. The data-quality assessment is based on the method of Hobbs et al. (2005), although it has been updated for the current version of the derivation method. Every toxicity value must have its quality assessed as, even within a single study, it is possible for toxicity data to have been generated using different methods and hence be of different quality.

The data-quality assessment scheme examines how each toxicity value was generated and awards a quality score and quality grade on the basis of answers to a series of questions (

Table 3 and Appendix B). One of 6 different questionnaires is completed for each toxicity value, depending on the environmental medium (freshwater, marine or estuarine), type of toxicant (metal/non-metallic inorganic or organic) and type of test organism (plant or non-plant) used.

Table 3 provides the table for freshwater, metals/non-metallic inorganics and non-plants, and the corresponding tables for the other questionnaires are presented in **Error! Reference source not found.**. The quality score for each toxicity value is represented as a percentage, calculated as follows:

Quality score = (total score from questionnaire ÷ the maximum possible score) × 100

Toxicity data with a quality score ≥ 80% are classed as 'high' quality, data with a quality score of ≥ 50 to < 80% are classed as 'acceptable' quality, and data with a quality score of < 50% are classed as 'unacceptable' quality. High-quality and acceptable-quality data can be used to derive GVs, while unacceptable-quality data should not be used to derive GVs. Normally, toxicity data calculated using nominal-concentration data would not be used to derive GVs; however, professional judgement can be used to include such data, provided a justification for their use is given (

Table 3 and Appendix B).

Professional judgement often needs to be used when assessing the quality of data, particularly where one or more of the aspects of the experimental design is less than optimal, as it could be a fundamental flaw. For example, researchers may have measured and stated the pH of the test media, thereby scoring full marks, but if the pH drifted by 3 units during the test, this would be considered a fundamental flaw. In such cases, it would be appropriate to score the quality as unacceptable. Or, a toxicity test may have used only a small number of test treatments (e.g. 2–5), and a decision will be required about the overall quality of the associated data. When professional judgement is used in assessing the quality of toxicity data, justifications for the decisions should be provided in the data-quality assessment spreadsheet.

Table 3 Scoring system for assessing the quality of toxicity data for metals/non-metallic inorganics to freshwater non-plants, to be used in the derivation of guideline values for toxicants

The corresponding sets of questions for other combinations of medium/toxicant type/organism type are provided in Appendix B.

No.	Question	Score[a]
A. General test design information		
1	Was the duration of the exposure stated (e.g. 48 h, 96 h)?	Yes (3), No (0/FAIL)
2	Was the biological endpoint (e.g. immobilisation, population growth) stated? Note: ensure that the endpoint is considered to be ecologically relevant. Do not use endpoints with no demonstrated ecological relevance to derive guideline values, and they do not need their quality assessed.	Yes (3), No (0/FAIL)
3	Was the measure of toxicity reported (e.g. NEC, ECx, NOEC) **and** its associated biological-effect size reported or quantifiable from the data (e.g. EC10, LC50, 15% effect at NOEC)?	Yes (3), No (0/FAIL)
4	Were appropriate controls (e.g. no-toxicant control, solvent control) used?	Yes (3), No (0/FAIL)[b]
5	How may treatment concentrations were used (in addition to the control)?	≥9 (3), 6–8 (2), 3–5 (1), < 3 (0/FAIL)[b]
6	What was the test concentration spacing? Note: a spacing of < 3.2 is highly preferred, while a spacing of ≥ 10 is too great.	≤ 3.2 (3), 3.3 9 (1), ≥ 10 (0/FAIL)[b]
7	Was each control and chemical concentration at least duplicated? Note: for concentration-response modelling, it is also acceptable to have many concentrations (e.g. > 15) without replication (assuming the controls are replicated).	Yes (3), No (0/FAIL)
8	Were the characteristics of the test organism (e.g. length, mass, age) stated?	Yes (3), No (0)
9	Was the type of test medium used stated (e.g. synthetic or natural water? If synthetic, to what recipe? If natural, what source? Filtered or unfiltered?)?	Yes (3), No (0)
10	Was the type of exposure (e.g. static, flow-through) stated?	Yes (3), No (0)
	Maximum sub-total for Part A = 30	
B. Test performance/results		
11	Were analytical reagent-grade chemicals or the highest possible purity chemicals used for the experiment?	Yes (3), No/not stated (0)
12 (A)	Were test solutions, blanks and/or controls tested for common contamination (e.g. elevated naturally occurring substances, such as nutrients, metals, metalloids) or other suspect contaminants?	Yes (2), No/not stated (0)
12 (B)	If so, were any significant contamination issues identified?	No (2), Yes (0/FAIL)
13	Were the contaminant concentrations measured at the beginning and end of the exposure? Note: normally, toxicity data calculated using nominal-concentration data would not be used to derive guideline values; however, professional judgement can be used to include such data, provided a justification for their use is provided.	Yes (8), Measured once (4), Not measured or stated (0/FAIL)[b]
14 (A)	Were test-acceptability criteria stated (e.g. mortality in controls must not exceed a certain percentage) or inferred (e.g. test methods used were United States Environmental Protection Agency or Organisation for Economic Co-operation and Development)?	Stated (2), Inferred (1), Neither (0)
14 (B)	If so, were test-acceptability criteria met? Note: data that fail the acceptability criteria are automatically deemed to be of unacceptable quality and must not be used.	Yes (6), Not reported (2), No (0/FAIL)
15 (A)	Were parallel reference-toxicant toxicity tests conducted?	Yes (2), No (0)
15 (B)	If so, did the reference toxicity results fall within the acceptable limits?	Yes (6), Not reported (2), No (0/FAIL)

No.	Question	Score[a]
16	Was there a concentration-response (C-R) relationship, either observable or stated (i.e. was there a larger response in the endpoint of interest at higher treatment concentrations than at lower treatment concentrations?)? Note: a full C-R relationship is one that includes zero, partial and full effects, while a partial C-R relationship is one that is missing either zero, partial or full effects. Professional judgement might be required to make decisions associated with C-R relationships that exhibit characteristics such as non-monotonicity and low-dose effects.	Yes – full (8), Yes – partial (4), No – (0/FAIL)
17 (A)	Was the statistical method or model used to determine the toxicity estimate stated **and** appropriate (e.g. refer to Green et al. 2018, noting that new methods will/have emerged that may not be captured in past publications)?	Yes (6), No (0)
17 (B)	For point estimates from concentration-response modelling (e.g. LC/EC/NEC/BEC/NSEC), was an estimate of variability provided? For estimates based on hypothesis testing (e.g. NOEC/LOEC/MATC), was the significance level 0.05 or less?	Yes (2), No (0)
	Maximum sub-total for Part B = 47	
C. Test water chemistry/test conditions		
18	Were the following parameters measured and reported?	
18.1	pH – was pH measured at least at the beginning and end of the toxicity test?	Yes (3), No (0/FAIL)
18.2	Hardness	Yes (3), No (0/FAIL)
18.3	Alkalinity	Yes (3), No (0/FAIL)
18.4	Dissolved organic carbon concentration	Yes (3), No (0/FAIL)
18.5	Dissolved oxygen (DO) – was DO measured at least at the beginning and end of the toxicity test?	Yes (2), No (0/FAIL)
18.6	Conductivity	Yes (2), No (0/FAIL)
18.7	Temperature	Yes (2), Temperature of room or chamber was reported (1), No (0/FAIL)
18.8	Any other parameters of relevance (e.g. other toxicity-modifying factors)	Yes (2), No (0/FAIL)
	Maximum sub-total for Part C = 20	
	Maximum total for parts A + B + C = 97	

Source: Updated from Warne et al. (2018).

[a] Where a question has a possible 'No (0/FAIL)' outcome, professional judgement may be required to determine if a 0 is assigned or if the test fails the quality-assessment process.

[b] See

Table 2 for additional guidance.

3.3.2 Field-based, microcosm and mesocosm data

Field-based, microcosm and mesocosm data are generated using different methods to those used to generate laboratory-based data, as they are trying to be more environmentally realistic. Therefore, a different (non-quantitative) quality-assessment scheme is used, although many of the key elements are the same as the laboratory-based data assessment (

Table 3). The quality of field-based, microcosm and mesocosm data should be assessed using a combination of factors considered crucial by the OECD (1992) and the European Commission (2018), as summarised below. Note that some of these factors are also relevant to laboratory, single-species toxicity testing, and assessment of the overall reliability of such tests should also take these factors into consideration (most of which are captured in the formal quality-assessment process – see Section 3.3.1).

For field-based, microcosm and mesocosm data to be considered of acceptable quality and, therefore, suitable to derive GVs (either by themselves or in combination with laboratory-based data) or to ground-truth laboratory-based GVs, they should meet all or at least most of the following criteria:

- have an adequate and unambiguous experimental set up, including a toxicant exposure regime, that reflects:
  - exposure in the field, including the relevant exposure pathways and fate of the toxicant
  - measurement of chemicals
- have ≥ 3 concentration treatments, a suitable control and appropriate replication. Control performance should be reported and deemed to be acceptable. The required degree of replication depends on the statistical method used to calculate the toxicity. If a hypothesis-based method is used to calculate NOECs or LOECs, treatments should be replicated at least in triplicate. However, if a regression method is used to calculate the toxicity, no replication is required, but the number of treatments should be increased to ≥ 6. Ultimately, the design should permit a sound statistical evaluation
- have a realistic biological community that
  - is representative of the taxa distribution and trophic structure in the ecosystem being assessed and should contain at least invertebrates, phototrophs and organisms associated with nutrient cycling. While fish could be included, this may not be possible for either practical reasons (the fish may eat the other test organisms) or ethical reasons (the use of fish may be precluded by animal ethics)
  - contains taxa sensitive to the mode of action of the toxicant, as best as can reasonably be predicted
- measure sensitive endpoints consistent with the mode of action of the toxicant, as best as can reasonably be predicted
- measure chemical and physical properties that are known to, or are likely to, affect exposure to the toxicant or the bioavailability, at appropriate intervals throughout the duration of the study
- permit concentration-response curves for individual contaminants to be derived
- measure individual-level, population-level or community-level endpoints
- be of sufficient duration to account for a significant proportion of the test organisms' life spans
- be of sufficient duration to reflect the persistence of the toxicant, as best as can be reasonably achieved.

Although it is possible to generate sufficient toxicity data from a single field-based, microcosm or mesocosm study to derive a DGV, this is not acceptable. Data from multiple such studies are required to derive a DGV that is based on microcosm or mesocosm data only. Alternatively, data from

microcosm or mesocosm studies can be combined in an SSD with data from single-species laboratory toxicity tests, provided that the microcosm or mesocosm data are reported for individual species, and it can be reasonably assumed that the individual responses are a direct result of exposure to the toxicant and not a result of overall community or ecosystem changes in the mesocosm across concentrations. Effects data reported for higher taxonomic levels (e.g. genus, family) should not be used in an SSD that includes individual-species data. If data from microcosm or mesocosm studies are not suitable for inclusion in an SSD, they can be used as an additional line of evidence for helping to validate the final GVs. Where DGVs are to be derived using data solely from field-based, microcosm or mesocosm studies, the method detailed in this report should be used unless an appropriate alternative method can be justified.

When deriving site-specific GVs, constraints around the type and amounts of data/studies may require alternative, but still scientifically defensible, approaches to be used, as described in ANZG (2018; see *Deriving guideline values for water quality*) and by Huynh and Hobbs (2019) and van Dam et al. (2019). In such cases, all decisions and associated justifications need to be documented, and peer review is necessary. Also, the relevant government agency should always be consulted on appropriate methods during the planning stage of the derivation process.

3.4 Selection of data to derive guideline values

3.4.1 General

The collated data that have successfully passed the quality-assurance procedures should be entered into the Excel™ data-entry spreadsheet that has been developed for this purpose and that can be made available by the Australian Department of Climate Change, Energy, the Environment and Water. The associated information that should be included in the spreadsheet includes (where applicable): source of the data; species (scientific and common) name; phyla of the species; type of organism (refer to Table 1); life stage of test organism (refer to Table 1); media type (fresh, marine or estuarine); key water quality parameters such as pH and temperature (also see sections 3.2 and 3.1, respectively); exposure duration; exposure type (acute or chronic); statistical estimate of toxicity (e.g. EC10, IC10 and NOEC); endpoint (e.g. immobilisation, population growth); concentration at the estimate of toxicity; the conversion factor used to convert chronic median (i.e. EC/IC/LC50) and low-effect (i.e. LOEC) toxicity values to negligible-effect values; the chronic estimated negligible-effect values; the conversion factor to convert acute toxicity values to chronic toxicity values; and the converted-acute toxicity values.

For particular groups of toxicants, additional information may be required. For example, for metals affected by TMFs (see Section **Error! Reference source not found.**), the spreadsheet should include the values of the TMFs at which the test was conducted. Also, where TMFs have been built into a GV derivation (i.e. a bioavailability-based GV – see Section 0), the toxicity values adjusted for the relevant TMFs should also be included.

The data should first be sorted in this order: media type, species, endpoint, statistical estimate of toxicity, exposure type.

In general, the more closely a laboratory or field experiment mimics the exposure of organisms in the field and the complexity of the ecosystems in which they live, the more relevant the resulting data should be for GV derivation (see Section 3.3.2 for more details).

When deriving site-specific GVs, a WoE approach, such as used by Cormier et al. (2008), van Dam et al. (2014) and Moore et al. (2017), might be appropriate, provided a rigorous and justified approach is followed. Guidance on when and how to use WoE to help derive a GV has been provided by US EPA (2016), ANZG (2018; see *Deriving guideline values using multiple lines of evidence*) and van Dam et al. (2019).

3.4.2 Conversion of toxicity data

3.4.2.1. Conversion of low-effects and median-effects estimates to negligible-effects estimates

Section 3.2 outlined a general hierarchy of preferred, appropriate and less preferred statistical estimates of chronic toxicity. Although the preferred toxicity estimates (i.e. NEC, NSEC), appropriate toxicity estimates (i.e. EC/IC/LCx where x ≤ 10; BEC10) and some of the less preferred toxicity estimates (i.e. EC/IC/LCx where x > 10 and ≤ 20, NOEC) are not numerically the same, they are all treated as representing negligible-effect estimates for the purposes of deriving GVs. This means that no conversion factors are applied to these types of data, and they are used as is. The rigour and associated defensibility of all statistical estimates of toxicity should be assessed as part of the data-selection process. This would include (but not be limited to) the standard error of a point estimate of toxicity (e.g. EC/IC/LCx) or even a coefficient of variation (i.e. the standard error of a point estimate of toxicity relative to the magnitude of the estimate), particularly EC/IC/LCx data, where x ≤ 10, as well as the measured-effect size of a NOEC or NSEC, to consider whether such data are appropriate to use. For example, any NOECs and NSECs with an effect size > 20% should probably not be used to derive GVs. The exclusion of any data on this basis needs to be appropriately justified and documented.

In cases where there are insufficient chronic NEC, NSEC, EC/IC/LCx (where x ≤ 10), BEC10, EC/IC/LC15–20 and NOEC data to derive a GV using the SSD method, chronic LC/IC/EC50, LOEC and MATC values can be used but should be divided by a conversion factor of 5, 2.5 or 2, respectively, to provide estimates of chronic negligible-effect values (ANZECC and ARMCANZ 2000b; Warne 2001). Alternatively, if the concentration-response data are available in the source reference, the data can be re-modelled in order to estimate a preferred or other appropriate estimate of toxicity. Such analyses should always be presented as supporting information for the GV derivation. If estimated chronic values are used (i.e. those for which a conversion factor has been applied), this information should be recorded in the spreadsheets used for data calculation and in the SSD plots or the accompanying text and tables of toxicity values.

3.4.2.2. Conversion of acute data to chronic data

The use of chronic toxicity data is always preferred. However, in cases where there are insufficient chronic data to derive a GV, there are often considerable acute toxicity data that can be converted to provide an estimate of chronic toxicity. In such cases, chronic and converted-acute toxicity data should be combined to derive a GV. An ACR is the ratio of the acute toxicity (reported as the LC50 or EC50) to the chronic toxicity data (reported as the NOEC, EC10 or other negligible-effect value) for a particular chemical. Note that acute EC/IC/LCx (where x ≠ 50), NOEC, LOEC and MATC values must not be used to derive chronic GVs.

Acute-to-chronic ratios can be expected to be both chemical (or chemical group) specific and species (or taxon) specific and, therefore, highly variable. Limitations to the use of ACRs are discussed by Warne (1998). If acute data are to be used for the derivation of chronic SSDs, it is preferable to calculate chemical and taxon-specific ACRs. The data used to calculate an ACR do not have to be for

the same statistical estimates of toxicity or endpoints, but they must be for the same species and have been presented in the same paper or at least determined in the same laboratory. Acute-to-chronic ratios should be calculated directly from experimental toxicity data. While it may be possible to use the chemical-specific ACRs reported in the detailed chemical descriptions in Section 8.3.7 of ANZECC and ARMCANZ (2000b), they should be validated or updated based on new data and knowledge from the past 25 years. The following rules should be applied when applying ACRs to acute toxicity data for a chemical:

- If there is only one ACR, that ACR should be used for all species.
- If there is more than one ACR, the geometric mean of ACR values for each taxonomic group should be determined, and the appropriate taxonomic group ACR values should then be applied to acute data for that taxonomic group.
- If there is more than one ACR but none for the taxonomic group with acute toxicity data, the geometric mean of all the ACR values for the chemical should be used.

In the absence of an ACR for a particular toxicant, a default ACR of 10 should be used to convert acute LC/EC/IC/50 values to chronic negligible-effect values (originally recommended by US EPA 1986 and OECD 1992). Use of default ACRs should be carefully considered, taking into account whether the chemical is known to have similar acute and chronic toxicity, whether acute toxicity is more likely to occur than chronic toxicity in natural situations (e.g. chlorine), or whether the chemical is an essential element (e.g. boron, copper, iron, manganese, molybdenum, nickel, selenium, zinc) that can be physiologically regulated when environmental concentrations are low. Justification for any professional judgement decisions is required. If converted-acute values are used, this information should be recorded in the spreadsheet used for data calculation and in the (shiny)ssdtools SSD plots or the accompanying text and tables of toxicity values.

3.4.2.3. Conversion of toxicity data for short-term guideline values

As noted in Section 0, the preferred hierarchy of statistical estimates of toxicity also applies to acute toxicity data for deriving short-term GVs. Table 1 provides details of what would constitute acute toxicity data. Acute toxicity data are most often reported as LC/IC/EC50 values rather than negligible-effect values. Acute LC/IC/EC50 values can be converted to acute negligible-effect values by dividing by a default conversion factor of 5, as applied to chronic data conversion (Section 3.4.2.1). Alternatively, species-specific conversion factors can be derived if there are matched acute negligible-effect (e.g. EC10, NOEC) and median-effect (e.g. EC50) data available for species from the same acute toxicity study. When applying conversion factors to the median-effect data, the same rules should be adopted as described for ACRs in Section 3.4.2.2. As for long-term (chronic) GVs, if acute ECx or ICx data are available for a species, they should be used in preference to acute LCx data, unless a reliable LCx value is lower than the ECx or ICx values.

This approach was used for the chlorine marine water DGVs (ANZG 2025a) and for short-term GVs for copper and zinc for New Zealand (Gadd et al. 2024). When converting acute low-effects and median-effects data to acute negligible-effects data, care needs to be taken that the conversion does not reduce the converted value below values associated with chronic toxicity.

Notably, if short-term GVs are to be used for regulatory purposes, such as setting licence conditions or in prosecutions, the data preferences may change to reflect the purpose of the GV.

3.4.3 The use of bioavailability models for metal default guideline values

It is well established that there are a number of abiotic TMFs that can modify the bioavailability and toxicity of chemicals (particularly metals and metalloids) to aquatic organisms. These include, but are not limited to organic carbon, pH, temperature, alkalinity, chloride, hardness (i.e. the aqueous concentration of calcium and magnesium ions) and inorganic ligands (e.g. Wang 1987). Bioavailability models aim to predict the effect that a TMF or set of TMFs will have on the toxicity of a chemical and can also be used to adjust GVs based on the values or concentrations of the TMF at the site of interest. When bioavailability models are incorporated into GV derivations, the GVs are initially derived for a specific 'reference' condition that is representative of a high-bioavailability condition, and will include look-up tables or a spreadsheet-based or online calculator that can be used to provide the GV that is relevant to the combination of TMF conditions at the site of interest. To date, bioavailability-based DGVs have been predominantly developed for metals but also for ammonia (ANZG 2025b), nitrate (ANZG 2025c) and hydrogen sulfide (ANZECC and ARMCANZ 2000e).

In the 2000 Guidelines (ANZECC and ARMCANZ 2000a, b), DGVs for cadmium, chromium (III), copper, lead, nickel and zinc were normalised to a water hardness of 30 mg/L $CaCO_3$ and could subsequently be adjusted using hardness-based algorithms (adopted from US EPA 1996), which were largely based on acute toxicity data for fish (see Table 3.4.3 and Table 3.4.4 in ANZECC and ARMCANZ 2000a). Since then, considerable research has shown the limitations of these hardness algorithms. For example, Markich et al. (2005) showed that hardness has either no effect or a limited effect on copper toxicity for a variety of freshwater organisms (i.e. an alga, a bacterium and a cladoceran). Similarly, currently in-draft DGVs for chromium (III), nickel and zinc have reported that the hardness algorithms for these metals have numerous limitations and should not be used to adjust DGVs (ANZG, unpublished data). Common criticisms of the hardness algorithms include that they are based on narrow ranges of water hardness for a limited range of species (with many Australian and New Zealand waters having values outside the tested ranges) and then are extrapolated to all species. A Dutch review of the use of hardness-based environmental quality standards concluded that they were not predictors of ecological risk posed by metals (RIVM 2004). Updated guidance on the use of the hardness algorithms is provided in ANZG (2018).

The above limitations contributed to the development of biotic ligand models (BLMs) that consider the effect of TMFs, including water hardness, that compete for the biotic ligand and affect metal toxicity. Some BLMs have been adopted by national and international regulatory bodies, including the nickel BLM adopted by the European Commission to derive environmental quality standards for freshwaters (European Commission 2019). At present, no BLMs have been endorsed for the derivation of DGVs for Australia and New Zealand. Considerably more work is required to develop and validate scientifically robust method(s) for modifying or deriving GVs for metals.

An alternative approach, known as a multiple linear regression (MLR) model, predicts the toxicity of a metal based on empirical data for TMFs such as hardness, pH and DOC (Brix et al. 2020). Multiple linear regression models have been developed and used for aluminium in the United States and Canada (US EPA 2018) and for iron, lead and zinc in Canada (ECCC 2020, 2024; CCME 2018). They have also been proposed for use for DGVs for nickel (Stauber et al. 2021) and zinc (Stauber et al. 2022) in freshwaters for Australia and New Zealand. Multiple linear regression models are also used in the Australian ecological investigation levels for contaminated sites (NEPC 2013). Similar to the BLM approach, toxicity data from a single species can be used to develop relationships, or data for

multiple species can be pooled to select the best MLR, provided that the effects of water chemistry on toxicity are similar across species. Such approaches may assist in the transition from simple algorithm approaches to more complex BLM approaches for deriving GVs for metals. The use of non-linear models for accounting for TMFs is less common (e.g. Erickson et al. 1987; Cui et al. 2023) but does represent an option. Such models are likely to be more complex than linear models and would equally need appropriate validation and, ultimately, to be simple enough to encourage implementation by stakeholders.

Models (e.g. BLMs, MLRs) for predicting the effects of TMFs on metal toxicity and adjusting GVs accordingly can be included in the development of DGVs. However, if this is to be done, it needs to take into account the following considerations:

- There needs to be sufficient evidence that the model incorporates TMFs that are relevant to the metal of interest – i.e. that the TMFs exhibit relationships with the chronic toxicity of the metal that can be accurately modelled.
- It may be beneficial if the role of the TMFs in affecting the toxicity of the metal, and any associated model that has been developed to predict this, has been recognised by other national or international regulatory authorities (e.g. United States Environmental Protection Agency, Environment and Climate Change Canada) or has been published in the peer-reviewed literature.
- The validated ranges of the TMFs in the model should be relevant to Australian and New Zealand water-quality conditions.
- A model is more likely to be accepted for use in the development of a DGV if it has been validated using water-quality conditions and species relevant for Australia and New Zealand, as was done for nickel (Peters et al. 2018; Stauber et al. 2021) and zinc (Stauber et al. 2022). Validation may include assessing model outputs for both predictiveness and protectiveness.
- Additional lines of evidence supporting the approach and appropriate caveats (e.g. non-validation to local conditions or species) should also be documented.

Reflecting these considerations, the following 4-step approach is recommended for assessing and validating bioavailability models for Australian and New Zealand conditions:

1. **Confirm model rigour:** Has the model been developed and adopted by a key overseas jurisdiction, such as the United States Environmental Protection Agency, Environment and Climate Change Canada or the European Commission, or has the model been published in the peer-reviewed literature? This represents a starting point for identifying potentially suitable models. The validation of the model's performance needs to be examined and should include, as a minimum, consideration of the appropriateness of the model structure, the quality and representativeness of the data used for calibration or parametrisation, and graphical and quantitative model-residual-based assessments of bias following the methods in Garman et al. (2020). A description of the technical elements of the model development and validation should be provided.
2. **Assess water quality representation:** What are the validated ranges of the TMFs in the model, and are they relevant to Australian and New Zealand water types? If no, are the gaps in water quality coverage significant enough to render the model not relevant for Australian

and New Zealand water quality conditions? Is more research required to extend the models to a sufficient coverage of Australian and New Zealand water-quality conditions (e.g. as done for nickel by Peters et al. 2018)? Note that it is more important for a model to be representative of Australian and New Zealand water-quality conditions than it is to have incorporated data for Australian and New Zealand species. This is because water-quality conditions (i.e. TMFs) are more likely to affect the toxicity of a toxicant than geographical differences in biological communities.

3. **Assess species representation:** Did model development or validation include any data from Australian or New Zealand species? Note that this should not be a criterion for acceptance of a model but will help to understand the overall relevance of the model to Australian and New Zealand species.

4. **Undertake Australian and New Zealand water-quality validation:** Even if steps 1 to 3, above, indicate that the model is relevant to Australian and New Zealand conditions, some level of validation of the model for a range of Australian and New Zealand water-quality conditions is likely to be required. The objective is to determine whether the model accurately predicts toxicity in Australian and New Zealand water types with sufficient confidence and provides sufficient protection. If possible, this should rely on existing or new data for toxicity of the toxicant in question to Australian and New Zealand species in different Australian and New Zealand water-quality types that cover the range from low to high TMF levels. As an example, Peters et al. (2018) assessed the applicability of a nickel BLM for Australian conditions by comparing the BLM toxicity predictions with the measured toxicity of nickel to various Australian freshwater species in a range of different Australian water-quality conditions.

If bioavailability-based GVs are being derived from data that require conversion (e.g. acute-to-chronic conversions, median-effect to negligible-effect conversions; see Section 3.4.2), it is essential that any necessary normalisation of the data to a standard water-quality condition based on the TMFs (e.g. hardness, DOC, pH, temperature) is performed prior to the data conversions. This is because the original values are generally the basis for TMF-modifier analyses.

3.4.4 Obtaining a single toxicity value for each species

- A single toxicity value is used to represent the sensitivity of each species in an SSD. However, there are often multiple toxicity values for each species, including data for several endpoints and exposure durations. Consequently, some selection, manipulation and reduction in the number of the toxicity data are required. An example of the application of these procedures to a dataset is presented for the cladoceran *Ceriodaphnia* cf. *dubia* in

- Determine the lowest value for each species (column 8, **Error! Not a valid bookmark self-reference.**). The lowest endpoint value is adopted as the toxicity value to represent the sensitivity of the species in the final dataset. For example, there are 4 *C. dubia* endpoint combinations (development, growth, immobilisation, reproduction) with toxicity values of 5.1 µg/L, 7 µg/L, 3.9 µg/L and 0.19 µg/L, respectively. Thus, the value of 0.19 µg/L for reproduction would be adopted as the toxicity value for *C. dubia*.

Table 4. The following rules for data manipulation should be applied to all toxicity data for each species:

- Determine the toxicity value for each combination of endpoint, statistical estimate of toxicity, life stage and duration (column 6,
- Determine the lowest value for each species (column 8, **Error! Not a valid bookmark self-reference.**). The lowest endpoint value is adopted as the toxicity value to represent the sensitivity of the species in the final dataset. For example, there are 4 *C. dubia* endpoint combinations (development, growth, immobilisation, reproduction) with toxicity values of 5.1 µg/L, 7 µg/L, 3.9 µg/L and 0.19 µg/L, respectively. Thus, the value of 0.19 µg/L for reproduction would be adopted as the toxicity value for *C. dubia*.

- Table 4). If there is a single value for a combination, it is adopted (e.g. row 1,
- Determine the lowest value for each species (column 8, **Error! Not a valid bookmark self-reference.**). The lowest endpoint value is adopted as the toxicity value to represent the sensitivity of the species in the final dataset. For example, there are 4 *C. dubia* endpoint combinations (development, growth, immobilisation, reproduction) with toxicity values of 5.1 µg/L, 7 µg/L, 3.9 µg/L and 0.19 µg/L, respectively. Thus, the value of 0.19 µg/L for reproduction would be adopted as the toxicity value for *C. dubia*.

- Table 4). If there are multiple values for a combination, the geometric mean of the values is calculated and adopted for that combination (i.e. rows 4–5, 7–8, 10–11 and 12–14,
- Determine the lowest value for each species (column 8, **Error! Not a valid bookmark self-reference.**). The lowest endpoint value is adopted as the toxicity value to represent the sensitivity of the species in the final dataset. For example, there are 4 *C. dubia* endpoint combinations (development, growth, immobilisation, reproduction) with toxicity values of 5.1 µg/L, 7 µg/L, 3.9 µg/L and 0.19 µg/L, respectively. Thus, the value of 0.19 µg/L for reproduction would be adopted as the toxicity value for *C. dubia*.

- Table 4). Geometric means should not be calculated across different statistical estimates of toxicity, life stages and test durations.
- Determine the lowest toxicity value for each endpoint (column 7,
- Determine the lowest value for each species (column 8, **Error! Not a valid bookmark self-reference.**). The lowest endpoint value is adopted as the toxicity value to represent the sensitivity of the species in the final dataset. For example, there are 4 *C. dubia* endpoint combinations (development, growth, immobilisation, reproduction) with toxicity values of 5.1 µg/L, 7 µg/L, 3.9 µg/L and 0.19 µg/L, respectively. Thus, the value of 0.19 µg/L for reproduction would be adopted as the toxicity value for *C. dubia*.

- Table 4). This will be the lowest of the values for each combination of endpoint, statistical estimate of toxicity and duration (column 6,
- Determine the lowest value for each species (column 8, **Error! Not a valid bookmark self-reference.**). The lowest endpoint value is adopted as the toxicity value to represent the sensitivity of the species in the final dataset. For example, there are 4 *C. dubia* endpoint combinations (development, growth, immobilisation, reproduction) with toxicity values of 5.1 µg/L, 7 µg/L, 3.9 µg/L and 0.19 µg/L, respectively. Thus, the value of 0.19 µg/L for reproduction would be adopted as the toxicity value for *C. dubia*.

- Table 4). Generally, the longest duration will have the lowest toxicity values, but this is not always the case. If there is a single value for each combination of species, endpoint and duration, it is adopted (e.g. row 1,
- Determine the lowest value for each species (column 8, **Error! Not a valid bookmark self-reference.**). The lowest endpoint value is adopted as the toxicity value to represent the sensitivity of the species in the final dataset. For example, there are 4 *C. dubia* endpoint combinations (development, growth, immobilisation, reproduction) with toxicity values of 5.1 µg/L, 7 µg/L, 3.9 µg/L and 0.19 µg/L, respectively. Thus, the value of 0.19 µg/L for reproduction would be adopted as the toxicity value for *C. dubia*.

- Table 4). If there are multiple combinations of species, endpoint, statistical estimate of toxicity and duration, the lowest value is adopted. For example, for the combination of immobilisation (rows 4–8,
- Determine the lowest value for each species (column 8, **Error! Not a valid bookmark self-reference.**). The lowest endpoint value is adopted as the toxicity value to represent the sensitivity of the species in the final dataset. For example, there are 4 *C. dubia* endpoint combinations (development, growth, immobilisation, reproduction) with toxicity values of 5.1 µg/L, 7 µg/L, 3.9 µg/L and 0.19 µg/L, respectively. Thus, the value of 0.19 µg/L for reproduction would be adopted as the toxicity value for *C. dubia*.

- Table 4), there is a single 192-h NOEC value of 10 µg/L and 2 geometric mean EC10 values for 2 durations (168 h and 192 h) of 27.4 µg/L and 3.9 µg/L; thus, the lowest value of 3.9 µg/L would be adopted as the lowest geometric-mean value for the immobilisation endpoint.
- Determine the lowest value for each species (column 8, **Error! Not a valid bookmark self-reference.**). The lowest endpoint value is adopted as the toxicity value to represent the sensitivity of the species in the final dataset. For example, there are 4 *C. dubia* endpoint combinations (development, growth, immobilisation, reproduction) with toxicity values of 5.1 µg/L, 7 µg/L, 3.9 µg/L and 0.19 µg/L, respectively. Thus, the value of 0.19 µg/L for reproduction would be adopted as the toxicity value for *C. dubia*.

Table 4 Example of the application of data-manipulation rules to obtain a single toxicity value for a species – in this case the microcrustacean *Ceriodaphnia* cf. *dubia*

	1 Endpoint	2 Statistical estimate of toxicity	3 Life stage	4 Duration (h)	5 Toxicity value (µg/L)	6 Value for each combination of endpoint, estimate of toxicity, life stage and duration (µg/L)	7 Lowest value for each endpoint (µg/L)	8 Lowest value for species (µg/L)
1	Development	NEC	Neonate	168	5.1	5.1	5.1	0.19
2	Growth	EC10	Neonate	168	7	7	7	
3	Growth	EC10	Adult	240	23	23		
4	Immobilisation	EC10	Neonate	168	25	27.4[a]	3.9	
5	Immobilisation	EC10	Neonate	168	30			
6	Immobilisation	NOEC	Neonate	192	10	10		
7	Immobilisation	EC10	Neonate	192	5	3.9[a]		
8	Immobilisation	EC10	Neonate	192	3			
9	Reproduction	EC20	Neonate	240	5.0	5.0	0.19	
10	Reproduction	EC10	Neonate	240	2.0	1.3[a]		
11	Reproduction	EC10	Neonate	240	0.9			
12	Reproduction	EC10	Neonate	480	0.2	0.19[a]		
13	Reproduction	EC10	Neonate	480	0.15			
14	Reproduction	EC10	Neonate	480	0.24			

Source: modified from Batley et al. (2018)

[a] Values represent geometric means.

It is inappropriate to calculate a geometric mean for 2 or more values for the same species, endpoint, statistical estimate of toxicity, life stage and test duration if they differ by a large amount (e.g. by more than one order of magnitude). In such cases, professional judgement will be required to select the value. Potential options for selecting a value include (i) selecting the lowest value and (ii) comparing the reliability of the studies and using the value from the study deemed the most reliable. The justification for any such decisions should be documented. Chapman (2015) provided some useful general guidance and 7 rules regarding the calculation of geometric means; however, where there is a difference between the above rules and those of Chapman (2015), the above take precedence.

Where water quality may have significantly varied across the tests for some reason (e.g. in studies specifically designed to assess the effects of physicochemical variables, such as pH, hardness, alkalinity or dissolved organic carbon, on toxicity), best professional judgement will need to be applied as to whether the geometric mean or the lowest toxicity value from across the tests should be used for the GV derivation. Where tests for individual species have demonstrated a significant dependence of toxicity on a physicochemical variable, the toxicity data that correspond to the most toxic set of conditions should be used for GV derivation. Justification for all decisions relating to

these issues needs to be provided. Where the measured value of an important physicochemical variable (i.e. one that affects the toxicity of the contaminant in question) in the toxicity-test dilution water is well outside of the typical range of that variable in Australia and New Zealand (see Table 3 in Batley et al. 2018), best professional judgement should be applied to determine whether or not the toxicity value associated with that test should be included in the dataset.

In many cases, available data will still be dominated by NOECs and, although less preferred, these values may be used. Although the phasing out of NOEC data to derive GVs has been widely recommended (e.g. Warne and van Dam 2008, van Dam et al. 2012a, b), in many cases datasets for GV derivation will be very small if NOEC data are excluded. Ultimately, it is highly desirable to include as many data in an SSD as possible. Moreover, under appropriate experimental designs and test performance, NOEC values can still represent robust estimates of negligible effects. As a guide, if there are preferred toxicity estimates (i.e. NEC, NSEC) or other appropriate toxicity estimates (i.e. EC/IC/LCx where x ≤ 10, BEC10) for ≥ 15 species that belong to ≥ 4 taxonomic groups, available NOEC data could be excluded from the final dataset. In addition, it is preferable to include a NOEC for a species if there are no more-preferred values for that species, irrespective of the sample size of the overall dataset. The effect of the omission of NOEC data from a toxicity dataset on the resultant SSD and GVs needs to be examined on a case-by-case basis, primarily in the context of any changes to the reliability of the resulting GV (see Section 3.8). All related decisions on the inclusion or exclusion of NOEC data need to be appropriately justified and documented.

3.4.5 Do the data meet the minimum data requirements of the species sensitivity distribution method?

The minimum data requirements to use the SSD method have been updated as a result of the transition from the use of Burrlioz to (shiny)ssdtools for deriving GVs. Toxicity data are required for ≥ 6 species that belong to ≥ 4 different taxonomic groups (Fox et al. 2024b).

Taxonomic groups are generally considered to be defined by phylum (or sometimes class), although this is not always possible or meaningful.

Table 5 provides guidance on the assignment of taxonomic groups. Although the recommended taxonomic groupings may not be fully based on a formal scientific designation, they nevertheless represent taxonomically distinct groupings that are useful for non-experts to understand the types of organisms represented in a GV derivation. Moreover, as it is not possible to include taxonomic groups for all organisms that have been used for toxicity testing, some professional judgement may be required to assign additional taxonomic groups for some species.

Table 5 Examples of distinct taxonomic groups

Dash (–) indicates no additional comments for that organism.

Major types of organisms	Phylum or other classification level	Recommended taxonomic grouping[a]	Comments, including example test organisms
Vertebrates	Chordata	Fish	–
		Amphibian	–
Invertebrates	Arthropoda	Crustacean	Cladoceran, shrimp, amphipod, copepod, barnacle, crab, crayfish, lobster
		Insect	Mayfly, chironomid

Major types of organisms	Phylum or other classification level	Recommended taxonomic grouping[a]	Comments, including example test organisms
	Mollusca	Bivalve	Mussel, oyster, scallop, clam
		Gastropod	Snail, abalone
		Cephalopod	Cuttlefish, octopus
	Annelida	Annelid	Polychaete
	Echinodermata	Echinoderm	Sea urchin, sand dollar
	Cnidaria	Cnidarian	Coral, anemone, hydra
	Rotifera	Rotifer	–
Unicellular plants[b]	Heterokontophyta	Diatom	–
	Chlorophyta	Green microalga	–
	Haptophyta	Golden-brown microalga	–
	Cryptista	Cryptomonad	–
	Dinoflagellata	Dinoflagellate	–
Multicellular plants[b]	Chlorophyta	Green macroalga	–
	Ochrophyta	Brown macroalga	–
	Rhodophyta	Red macroalga	–
	Tracheophyta	Macrophyte	Duckweed, watermilfoil, *Hydrilla*, seagrass
Others	Cyanobacteria	Cyanobacterium	–
	Bacteria	Bacterium	–
	Protozoa	Protozoan	–
	Fungi	Fungus	–

[a] List is not exhaustive for all test species, and professional judgement may be required to assign other taxonomic groups that are not listed.

[b] Phylum classifications may vary between different information sources.

The minimum acceptable data requirement is, however, not optimal, and the use of toxicity data for more than 6 species and more than 4 taxonomic groups is highly desirable (see Section 3.8). For the purposes of the current derivation method, and assuming that ≥ 4 taxonomic groups are represented, datasets that contain data for 6–7 species are termed 'adequate', datasets that contain data for 8–14 species are termed 'good', and datasets that contain data for ≥ 15 species are termed 'preferred'. These data categories are used to inform the reliability classification for GVs (see Section 3.8).

For toxicity datasets that meet the minimum data requirements, GVs should be derived by model-averaging a set of candidate distributions using the 'ssd_fit_bcanz()' function in ssdtools or via shinyssdtools. The use of this method will ensure that the currently approved ssdtools settings and associated methods are used for the derivation (see Section 0 for further details). As previously specified, DGVs should be derived using chronic ecotoxicity data. If, however, there are insufficient chronic ecotoxicity data to generate DGVs using the SSD method or there are sufficient data but the fit of the SSD to the data is poor, there are 2 potential methods to overcome these situations:

1) The first method involves supplementing chronic data with acute data that have been converted to chronic equivalent data, as discussed in Section 3.4.2.2. This method can be applied to all types of toxicants (e.g. metals, non-metallic inorganics, organics). It was not prescribed by the 2000 Guidelines for the derivation of GVs but it has been used since in deriving site-specific GVs in order to overcome a lack of data. The value of this approach for toxicants with small chronic datasets is now recognised and, as such, it is included in the current derivation method and can be used to derive GVs at any spatial scale (e.g. national, regional, site-specific). Using this method will result in a lower reliability being assigned to the resulting GVs, as described in Section 3.8.
2) The second method is to combine chronic ecotoxicity data from more than one medium (i.e. freshwater, estuarine water, marine water). This method can only be applied to organic toxicants when (i) statistical analysis indicates that there is no difference in the bioavailability or toxicity of the chemical in the media being combined or (ii) based on knowledge of the chemistry of the chemical or its mode of action, there is no reason to expect differences in the toxicity of the chemical in the media being combined. Although separate DGVs for freshwater and marine water are typically derived, it may be possible, and in some cases preferable, to derive a single DGV for both freshwater and marine water where the above criteria have been met. Such a decision should be based on the circumstances of each specific case. Notably, if it is deemed valid to pool data from both media (based on the above criteria), the GV reliability is not downgraded as it is for a dataset containing a mix of chronic and (converted) acute data.

The use of either method for supplementing the chronic dataset should be justified. In the case of organic chemicals, the justification should include consideration of the number of species and taxa represented in each combined dataset and the fit of the SSD to the combined data. The method selected for use should be the one that results in data for more species and taxa or a better fit of the SSD to the data.

If there are still insufficient data to derive GVs using the SSD method, the less-favoured AF method should be used (see Section 4 and Warne 2001).

3.5 Checking the toxicity data for multimodal distributions

Toxicity datasets for chemicals with specific modes of action that target certain processes that occur in some taxa but not others can often be bimodal or multimodal. For example, toxicity datasets for many herbicides are bimodal – the sensitivity of plants is significantly greater than that of animals due to the chemicals targeting biochemical processes that only occur in plants (e.g. photosynthesis). The statistical distributions typically used in SSD methods (e.g. log-normal, log-logistic, Burr Type III) are unimodal. The fitting of such distributions to bimodal or multimodal datasets can lead to unrepresentative results. (shiny)ssdtools includes 2 5-parameter distributions that incorporate 2 distributions connected by a mixing proportion, which are appropriate for modelling bimodal datasets. One of the 2 mixture distributions, the log-normal–log-normal, is included in the set of default distributions (see Section 0). The current section describes how to check for, and deal with, bimodal datasets.

3.5.1 Weight-of-evidence approach

A weight-of-evidence (WoE) approach should be adopted to determine if a dataset exhibits bimodality or multimodality as a result of differential sensitivity of a specific taxonomic group or groups. The key factors to consider are the chemical's mode of action, indications of bimodality or multimodality, and the presence of taxa-specific sensitivity. Typically, consideration of the 4 following questions should provide sufficient information with which to make a decision. Professional judgement will often be needed to guide the analysis and interpretation and, as such, the details of the decision processes should be fully documented.

Question 1: Is there a specific mode of action that could result in taxa-specific sensitivity?

The mode of action of a chemical is a key indicator of whether a dataset might exhibit bimodality or multimodality and of the likely relative sensitivity of different taxa. Chemicals with non-specific modes of action (e.g. non-polar and polar narcosis) are more likely to have unimodal toxicity datasets, while those with highly specific modes of action (e.g. acetylcholinesterase and photosystem-II inhibitors) are more likely to have bimodal or multimodal datasets. While the existence of a specific mode of action increases the likelihood that there will be differences in the sensitivity of different types of organisms, this will not always be the case, particularly where the site of action is present in many types of organisms. Evidence of a specific mode of action that could result in taxa-specific sensitivity should be sought from the literature. Where no such information exists, or where the information is conflicting, the answer to this question would be 'uncertain'.

Question 2: Does the dataset suggest bimodality?

Modality of the data can be assessed both visually and statistically.

Inspection of the dataset using an SSD or a frequency histogram will provide an initial indication of modality. In addition to the presence of more than one distinct sub-grouping of data, datasets spanning very large ranges (e.g. > 4 orders of magnitude) can suggest bimodality or multimodality.

Statistical support for bimodality will be evident as a high weight assigned to the 5-parameter bimodal log-normal–log-normal distribution in the set of default distributions in (shiny)ssdtools when compared to the unimodal distributions. As a general guide, a corrected Akaike Information Criterion weight for the bimodal distribution that is equal to, or better than, the most highly weighted unimodal distribution will indicate the dataset could be bimodal. However, the log-normal–log-normal distribution can also be assigned a high weight for datasets that are not bimodal but for which different groups of the dataset have markedly different variances. Such characteristics of a dataset may be due to various reasons, including simple artefacts in the dataset that have no toxicological basis (e.g. a series of identical NOECs for a number of species, sourced from the same study). Thus, a high weight assigned to the log-normal–log-normal distribution should not automatically be assumed to indicate that the dataset is bimodal. Future versions of (shiny)ssdtools may include a wider range of distributions that capture these different causes of apparent bimodality. Moreover, given that the model-averaging process applies a penalty to distributions with more parameters, it will be unlikely for the log-normal–log-normal distribution to be assigned a high weight for smaller datasets (e.g. $n < 15$). Therefore, for smaller datasets (e.g. $n < 15$), a low weight assigned to the log-normal–log-normal distribution should not automatically be assumed to indicate that the dataset is not bimodal.

Question 3: Do the data show taxa-specific sensitivity (i.e. through distinct grouping of different taxa types)?

In addition to conducting the above check for multimodality, the data should be examined for signs of taxa-specific sensitivity. Information from Question 1, above, can inform this step (i.e. a known specific mode of action will point to potentially sensitive taxa groups). Examine the data or an SSD of the data with the species or taxonomic group names tagged to the values and look for distinct sub-groups of different taxon types. If the mode of action of the chemical is known, consider whether the composition and distribution of the sub-groups is consistent with the mode of action. If distinct or even overlapping sub-groups can be identified, use a graphical method such as a box-and-whisker plot and confidence intervals to further assess the differences between the subsets of data.

Question 4: Is it likely that indications of bimodality or multimodality or distinct clustering of taxa groups are *not* due to artefacts of data selection, small sample size, test procedures or other reasons unrelated to a specific mode of action?

As a final check, it is important to consider the strength (including quality) of evidence associated with the answers to each of the above questions. Specifically, for Question 2 and Question 3, the likelihood that an indication of bimodality or taxa-specific sensitivity is due to an artefact associated with data selection (e.g. inclusion and exclusion of certain data during the data-screening process, or a distinct division between chronic and converted-acute data), small sample size (i.e. insufficient data to indicate biologically meaningful differences in sensitivity between taxa groups), or test procedures (e.g. differences in sensitivity reported from chronic short-term versus long-term tests) should be considered. Other factors to consider include:

- whether data for the same chemical in the alternative medium (e.g. marine where freshwater is the medium of interest and vice versa) are unimodal, bimodal or multimodal. Note that this will only be relevant where the relative sensitivity of different taxa groups is considered to be similar in fresh and marine water
- whether data for other chemicals with the same mode of action are unimodal, bimodal or multimodal.

The response to this question requires a judgement about whether indications of bimodality or multimodality or distinct groupings of specific taxon types are biologically meaningful and potentially associated with a specific mode of action or are artefacts of the above types of issues.

3.5.2 Decision making

If the WoE associated with the above 4 questions indicates that the dataset is bimodal or multimodal or that there is distinct taxa-specific sensitivity, and this is known or thought likely to be due to a specific mode of action, the model-averaged estimates that include the bimodal distribution should be used to derive the GV.

However, the WoE process may result in equivocal outcomes for which professional judgement will be required. Guidance on this is provided below, although it is unlikely to cover all possible outcomes.

If the WoE indicates that a dataset exhibits no toxicologically-related bimodality, but the bimodal distribution is assigned a high weight relative to the other distributions, a 2-step process needs to be undertaken, as follows:

1) If question 4 concluded that indications of bimodality or taxa-specific sensitivity were due to artefacts associated with the dataset, determine if the rigour or representativeness of the dataset can be improved. If the dataset can be improved, re-fit it with the full set of default distributions and re-assess the modelling outputs.
2) If there are no defensible options for improving the dataset, consider whether removal of the log-normal–log-normal distribution from the set of default distributions is warranted (e.g. if its inclusion is having a marked influence on the GVs or their reliability rating). If so, the dataset should be re-fitted using only the default unimodal distributions and their model-averaged estimates used to derive the GVs.

To help guide users with this process, several examples are detailed in Appendix C.

If a dataset is relatively small (e.g. n < 15), and the bimodal distribution is accordingly assigned a low weight relative to the unimodal distributions, but there is a clear split in the dataset that is consistent with a well-understood mode of action, only the data for the most sensitive taxonomic group should be used to derive the GVs. This assumes that (i) there are sufficient data for the most sensitive taxonomic group to do this (i.e. n ≥ 6), (ii) the original, full dataset met the minimum requirement for the number of taxonomic groups (i.e. ≥ 4 taxonomic groups), and (iii) the output is more defensible than that based on retaining all the data (this would include consideration of GV reliability and possibly professional judgement associated with the respective SSDs and resultant GVs). Caution should be exercised with such cases, given that apparent splits in small datasets can be due to random artefacts.

3.6 Use of (shiny)ssdtools

The open-source ssdtools software (Thorley and Schwarz 2018, Thorley et al. 2025) or its online interface, shinyssdtools (Dalgarno 2018) (collectively referred to as '(shiny)ssdtools'), should be used to calculate the GVs for all chemicals that meet the minimum data requirements of the SSD method. It can also be used to derive low-reliability GVs for non-polar narcotics (see use of QSARs in Warne 2001). Entry of toxicity data and calculation of GVs should follow the (shiny)ssdtools user instructions (https://www.waterquality.gov.au/anz-guidelines/guideline-values/default/water-quality-toxicants/shinyssdtool).

The key feature of (shiny)ssdtools is that it employs a technique known as model averaging. In the context of SSD modelling, the approach fits multiple distributions to a toxicity dataset and uses the weights of the fits of each distribution to construct a model-averaged SSD, from which GVs can be estimated. Thus, there is no longer a reliance on just a single distribution that may not provide an adequate representation of the dataset, as was the case with Burrlioz 2.0. Full statistical details of (shiny)ssdtools and model averaging in SSDs can be found in Thorley and Schwarz (2018), Schwarz and Tillmanns (2019), Fox et al. (2021, 2024a, b) and Thorley et al. (2025).

(shiny)ssdtools contains a set of 'default' distributions (currently the gamma, log-Gumbel, log-logistic, log-normal, log-normal–log-normal and Weibull) that are automatically used to construct a model-averaged SSD. Apart from the log-normal–log-normal distribution, they are all 2-parameter distributions. The log-normal–log-normal distribution is a 5-parameter mixture distribution (i.e. 2 log-normal distributions with a mixing proportion) that is suitable for modelling bimodal datasets (as described in Section 0). (shiny)ssdtools also includes a number of additional distributions outside of

the default distributions, namely, the 2-parameter Gompertz distribution, the 3-parameter Burr(III) distribution and the 5-parameter log-logistic–log-logistic distribution. The currently listed non-default distributions were not included in the set of default distributions for specific reasons detailed by Fox et al. (2024a, b). It should be noted that the set of default distributions may change over time, while additional non-default distributions may also be added. In addition, there may be further technical improvements to ssdtools. Provided that either the default (shiny)ssdtools settings or the 'ssd_fit_bcanz()' function (within R) are used for derivation using the most recent version of ssdtools on the Comprehensive R Archive Network, resulting estimates will be generated using the most recently approved methodology. The specific version of (shiny)ssdtools used to generate the output should always be reported as part of the derivation, to ensure complete reproducibility.

When entering data into the GV derivation software, each toxicity value should be entered to no more than 3 significant figures.

When deriving GVs (including DGVs), the full set of default distributions in (shiny)ssdtools must be used unless there is an over-riding reason to remove or add distributions. For example, Section 0 provides guidance on the potential removal of the log-normal–log-normal distribution where it is modelling a dataset artefact and affecting the GVs or reliability rating. Alternatively, one of the non-default distributions could be included in the distribution set if it yields a demonstrably better relative fit to the dataset than all of the default distributions. The DGVs should be based on the model-averaged SSD – there is no need to exclude poor-fitting default distributions, as the model averaging will account for the relative fits of the distributions (i.e. a poor-fitting distribution will be assigned a very low weight and will have minimal influence on the final DGVs). Selecting only one or a small subset of the default distributions based on the goodness-of-fit statistics is not recommended (and should not be done for DGVs).

Alternative software may be used for the derivation of site-specific GVs; however, justification would need to be provided as to the technical validity of the alternative method and why it is superior to using (shiny)ssdtools. As part of this, and for SSD-based GVs, it would need to be shown that the alternative software is a recognised tool for generating SSDs. At the time of publication of this method document, (shiny)ssdtools is the only software package endorsed by the *Australian and New Zealand Guidelines for Fresh and Marine Water Quality* for the derivation of DGVs.

3.7 Calculate guideline values

3.7.1 Levels of species protection

The ANZG (2018) Guidelines specify 3 different levels of ecosystem condition, for which different levels of protection are recommended. Four different protective concentrations (PCx values; where x = the percentage of species to be protected) are derived from the SSD and used as GVs to cover the different levels of protection for these ecosystem conditions: PC99, PC95, PC90 and PC80. To protect ecosystems of high conservation value and ecosystems that are slightly to moderately disturbed, PC99 and PC95 values (i.e. concentrations protective of 99% and 95% of species) are typically recommended, respectively. The values for slightly to moderately disturbed ecosystems are also generally recommended for highly disturbed ecosystems; however, depending on the extent of the disturbance and prospects for improvement in ecosystem condition, and agreement among stakeholders about the desired ecosystem condition and associated level of protection, the PC90 or PC80 values may be applied instead. Guidance related to the application of GVs for toxicants in the

context of the ecosystem condition and level of protection is provided in ANZG (2018; see *Level of protection*).

Although (shiny)ssdtools can estimate the 95% confidence limits for each PCx value, the current position for the reporting of DGVs is that the confidence limits are not reported. This is largely due to the fact that, currently, there is no agreed approach on how confidence limits for DGVs should be interpreted and used in a regulatory context. For other types of GVs (e.g. regional, site-specific), the 95% confidence limits can be reported; however, clear guidance or agreement on how these limits are to be interpreted and used is strongly advisable and would most likely require consultation with relevant local jurisdictions.

It should be noted that when the toxicity data for a toxicant are multimodal, and the data for the most sensitive group of taxa need to be used to derive GVs, the levels of protection indicated by the resulting PC values only apply to the most sensitive group. For example, a PC95 value for a multimodal toxicant that is more toxic to insects and, thus, where only the insect data have been used in the SSD, will protect 95% of insects and a higher percentage of other organisms (as they are less sensitive).

3.7.2 Significant figures and rounding

The precision to which a GV is reported should be commensurate with the magnitude of the uncertainty in the value, which can be approximated by the standard error (SE) associated with the GV (as reported in the (shiny)ssdtools output). Thus, the SE of the GV should be used as the basis for determining the number of significant figures to which the GV is reported. Normal scientific practice is that the significant figures to which a value is reported should be no more than the first significant figure in the SE of the value. Therefore, the following rule for determining the number of significant figures for DGVs should be applied:

- The significant figure to which a DGV is reported must correspond to the place of the first significant figure in the SE of the DGV. Where the SE is greater than the DGV, the DGV must be reported to one significant figure.

Rounding of a GV to the appropriate number of significant figures should be done in accordance with the 'rounding to evens' rule set out in Australian Standard SAA 2706-2003 (Standards Australia 2003). The 'rounding to evens' rule adopts the most common convention for rounding – that is, rounding up when the value following the last value to be retained is > 5 and rounding down when the value following the last value to be retained is < 5. However, when the value following the last value to be retained is exactly 5 (i.e. there are no non-zero numbers to the right of this 5), the value should be rounded to the closest even number.

Full details of the significant figures and rounding rules for toxicant GVs, including worked examples, are provided in ANZG (2018).

3.8 Determine the reliability of the guideline values

The classification scheme for assigning the reliability of GVs derived using the SSD method is based on 3 factors:

1) the type of toxicity data (chronic, a mixture of chronic and converted-acute or only converted-acute values)

2) the number of species for which toxicity data are available (i.e. 6–7, 8–14 or ≥ 15)

3) a visual assessment of the fit of the SSD to the toxicity data (i.e. good or poor).

There are 6 classes of reliability: very high, high, moderate, low and very low, and the sixth class of unknown reliability is assigned to GVs derived using the AF method (see Section 1.1). The reliability of GVs associated with various combinations of these 3 factors is presented in Table 6. It is worth noting that the log-normal–log-normal distribution in (shiny)ssdtools, having 5 parameters, can provide markedly better fits to datasets than the other, 2-parameter default distributions in (shiny)ssdtools. However, where the log-normal–log-normal distribution is specifically modelling dataset artefacts that are associated with data-selection decisions and have no toxicological basis (e.g. clusters of identical or almost identical values), and these artefacts cannot be defensibly removed, the log-normal–log-normal distribution may need to be removed, as described in Section 0 and Appendix C.

It is recognised that other factors not considered in the assessment of GV reliability may affect the accuracy of the GV. Two such factors, and appropriate caveats that should accompany the reliability classification if applicable to the chemical being assessed, are:

- Toxicity datasets spanning 4 or more orders of magnitude – an appropriate caveat would be that GVs based on data with such a large range of values tend to be highly conservative and uncertain, especially at the 99% species-protection level.
- A heavy reliance on standard, single-generation toxicity studies for persistent, bioaccumulative and toxic (PBT) chemicals – an appropriate caveat would be that GVs for PBT substances that are not based on multi-generation tests are likely to not provide sufficient protection to aquatic ecosystems.

In the event that other factors are deemed to significantly affect the accuracy of the GVs, similarly appropriate caveats should be provided.

To assist in determining whether the fit of the distribution to the toxicity data is good or poor, examples are presented in Figure 2. Figure 2a, Figure 2b and Figure 2c represent poor fits, characterised by the SSD curve not following the data very well. In contrast, the SSD curves in Figure 2d, Figure 2e and Figure 2f generally follow the data very well and, thus, represent good fits. Although the model fit and the associated GVs are independent of the plotting positions of the toxicity values in the SSD (see Batley et al. 2018 for details), a visual check of the adequacy of the fitted distribution is a valuable exercise for determining the reliability classification (notwithstanding the limitations associated with fitting such models to, typically, very few data). As the current GV derivation method does not weight the toxicity values in the SSD, the entire range of the distribution contributes equally to the model. Thus, any visual assessment of goodness of fit should take into account the whole of the distribution, not only the lower left portion. Nevertheless, the GVs are estimated from the lower end of the distribution and, thus, a poor model fit in this region is cause for concern and should not be ignored.

Table 6 Classification of the reliability of guideline values using the species sensitivity distribution method

Data type	Sample size (adequacy)[a]	Adequacy of fit of species sensitivity distribution model	Reliability
Chronic[b, c]	≥ 15 (preferred)	Good	Very high
	≥ 15 (preferred)	Poor	Moderate
	8–14 (good)	Good	High
	8–14 (good)	Poor	Moderate
	6–7 (adequate)	Good	Moderate
	6–7 (adequate)	Poor	Low
Combined chronic and converted-acute, or combined chronic fresh and chronic marine where the 2 criteria in Section 3.2.5 are not met	≥ 15 (preferred)	Good	Moderate
	≥ 15 (preferred)	Poor	Low
	8–14 (good)	Good	Moderate
	8–14 (good)	Poor	Low
	6–7 (adequate)	Good	Moderate
	6–7 (adequate)	Poor	Low
Converted acute	≥ 15 (preferred)	Good	Moderate
	≥ 15 (preferred)	Poor	Low
	8–14 (good)	Good	Moderate
	8–14 (good)	Poor	Low
	6–7 (adequate)	Good	Low
	6–7 (adequate)	Poor	Very low

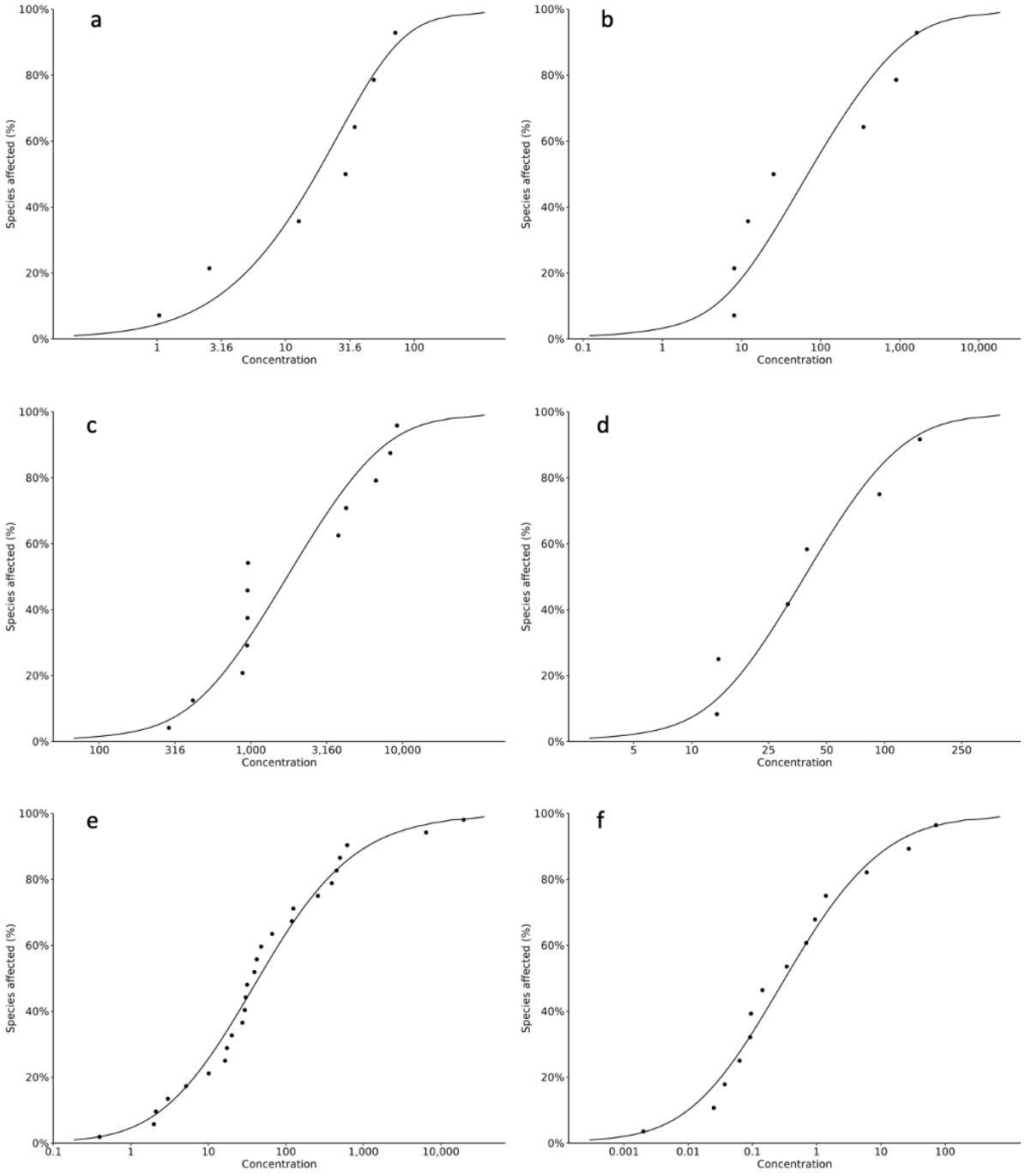
[a] The sample size is assumed to comprise data from ≥ 4 taxonomic groups.

[b] This includes all types of chronic data, irrespective of whether they are chronic NEC, BEC10, EC10 and NOEC values or estimates of chronic EC10 and NOEC values that were converted from chronic LOEC, MATC or EC50 data. This also includes combined freshwater and marine data where either of the 2 criteria specified in Section 3.2.5 (i.e. [i] no statistically significant difference between the freshwater and marine toxicity datasets and [ii] based on chemistry and mode of action, no reason to expect differences in toxicity between freshwater and marine water) have been met.

[c] The sample size is assumed to comprise data from ≥ 4 taxonomic groups.

Given the level of subjectivity in determining the functional form and fit of the SSD model, it is recommended that a panel of ≥ 3 relevant experts agree on the fit, especially where the decision is unclear. Irrespective of how the goodness of fit is decided, a statement explaining the selected category should be provided. Moreover, for DGVs, the independent review process will provide a further assessment of the decision on the model fit. Ideally, site-specific GVs should also be independently reviewed, while further review would also be made by the relevant regulatory body in the event that such GVs are submitted as part of an environmentally regulated activity (e.g. environmental-discharge licence application). These review processes should ensure that the final decision on SSD model fit is appropriate and defensible.

Figure 2 Examples of poor (a, b, c) and good (d, e, f) fits to datasets obtained, using (shiny)ssdtools v2.0 software



The reliability-classification scheme provides a quick and transparent means of indicating the general level of confidence in a GV. It also provides an indication of GVs that would benefit from the addition of more toxicity data. The reliability of DGVs can be improved by third parties, by obtaining new ecotoxicity data (from more recent literature, recent water-quality GV documents or their equivalent from other jurisdictions, or by generating new data), combining them with the existing data, and using the method described in the current report to derive a new GV. Improved GVs can be submitted, via the third-party GV-derivation process, for national consideration and endorsement as DGVs (ANZG 2018; see *Pathway for toxicant default guideline value publication*). Information on the recommended ways that DGVs should be applied is provided in ANZG (2018; see *Toxicant default guideline values for water quality in aquatic ecosystems*).

3.9 Accounting for the potential for chemicals to bioaccumulate

Some chemicals are known to bioaccumulate in aquatic organisms via water or dietary exposure. Long-term bioaccumulation of a chemical can lead to toxicity to these organisms that toxicity tests based on relatively short-term or single-generation water-only exposures will not fully account for, and can also result in toxicity (via secondary poisoning) to their predators. Thus, water-quality GVs may not appropriately protect aquatic organisms from the effects of bioaccumulative chemicals, and there is a need to account for this when deriving and applying GVs. This section mostly deals with the consideration of bioaccumulation when deriving GVs, while ANZG (2018; see *Accounting for local conditions – bioaccumulation*) provides background information on bioaccumulation and additional guidance on how to deal with bioaccumulation when applying GVs.

Typically, chemicals with $\log_{10}$ values for octanol–water partition coefficient (log-K_{OW}), bioconcentration factor (log-BCF) or bioaccumulation factor (log-BAF) ≥ 4 are considered to have the potential to bioaccumulate. However, these measures are not always reliable indicators of bioaccumulative potential (e.g. for chemicals that preferentially accumulate in tissues other than lipid, such as poly-fluoroalkyl substances [PFAS]) and, thus, field-based measurements of actual bioaccumulation are always preferable.

For potentially bioaccumulative chemicals (e.g. PFAS, dioxins), it is particularly important that data from long-term mesocosm or field-effects data or multi-generational tests where both bioaccumulation and ecologically relevant endpoints were assessed are used to derive the GVs. ANZG (2018) suggests, as an indication, that where (i) ≥ 30% of the dataset for a potentially bioaccumulative chemical is represented by such data and (ii) these data represent ≥ 3 taxonomic groups, the GVs could be considered to sufficiently account for the potential effects of bioaccumulation.

In contrast, a dataset comprising only short-term chronic, single-generation toxicity tests and any form of acute toxicity test are unlikely to fully characterise the long-term toxic effects of such substances in aquatic systems. For such GVs, it should be clearly stated in the GV documentation that the GV only considers relatively short-term effects and may not provide adequate protection from the potential effects of bioaccumulation. Additionally, the GV documentation should recommend a more conservative GV to account for the potential inadequate protection. For example, this could involve recommending the 99% species-protection GV for an ecosystem that is slightly to moderately disturbed, where the 95% species-protection GV would normally apply. It is acknowledged that this is an arbitrary approach, and ANZG (2018; see *Accounting for local conditions – bioaccumulation*)

provides more detailed guidance on how to apply an appropriate GV for potentially bioaccumulative chemicals.

It is important to note that toxicant DGVs derived using the current method are not intended to specify species-protection concentrations for air-breathing animals that inhabit aquatic ecosystems and prey on aquatic organisms. Consequently, the DGVs may not account for effects that result from the biomagnification of toxicants such as PFAS in air-breathing animals. For addressing this, alternate approaches are necessary. HEPA (2025) provides a nationally accepted example of developing guidance based on acceptable wildlife-diet concentrations for toxicants when the risk for air-breathing animals cannot reliably be determined from a water or sediment concentration. Users of the guidelines are encouraged to develop site-specific GVs for potentially bioaccumulative chemicals using other methods that can be scientifically justified. As noted in ANZG (2018), more comprehensive guidance on methods for deriving GVs that incorporate the impacts of bioaccumulative chemicals may need to be developed in the future.

3.10 Accounting for formulations

Many chemicals, in particular pesticides, are released into the environment as a constituent of a commercial formulation. These formulations contain an active ingredient (AI) as well as various additives such as emulsifiers, wetting agents and adjuvants to improve the performance of the AI. The toxicity of a formulation may be different to that of the AI due to the additives contributing to, or modifying the toxicity of, the AI. The mobility and decomposition rates of the various additives can differ, such that an environmental water sample may have a different composition to the composition of the original formulation, unless the formulation was applied directly to the water (e.g. direct-spray drift, accidental spill) and a water sample was taken within a relatively short timeframe. Moreover, multiple formulations often exist for toxicants such as pesticides, and it would be very difficult to know with confidence which formulations were being used within a catchment and in what proportions. For these reasons, DGVs are derived based on the active ingredient only and, as such, DGVs for such chemicals should be based on toxicity tests where the test organism is exposed to a reasonably pure form of the active ingredient (≥ 80% AI). Consequently, ecotoxicity data for toxicants that are used as formulations (e.g. pesticides) can be used to derive DGVs, provided that it is clear from the derivation that the test substance (i) was the AI and was not a formulation (e.g. that it was a technical material, technical grade, technical reagent, analytical grade or analytical reagent) and (ii) had a purity of ≥ 80%.

If the type of formulation of a toxicant being used at a site or in a catchment is known, it may be possible to correct the DGV based on the difference between the toxicity of the AI and that of the formulation, or to derive a formulation-specific GV if the formulation is known to modify the toxicity of the AI. However, for the reasons described in the previous paragraph, the situations where this will be a meaningful exercise are likely to be uncommon. To derive a 'formulation-corrected' GV, the toxicity of the commercial formulation and the AI should be compared. All toxicity data used in such a comparison must pass the screening and quality-assurance process and be based on the same species under the same test conditions (i.e. paired data). The ratio of formulation toxicity to AI toxicity should be determined for each paired set of data and then the geometric mean of these differences determined. If the geometric mean of the ratio is ≥ 3 or ≤ 0.33, it would be appropriate to correct the DGV by the difference – thus, the resulting formulation-specific GV might be larger or smaller than the DGV based on the AI. Note that the 2 ratio criteria are somewhat arbitrary and were

defined by expert judgement. Alternatively, it may be possible to derive a formulation-specific GV if sufficient toxicity data for the relevant formulation are available or generated. However, as with formulation-corrected GVs, the situations where the derivation of formulation-specific GVs will be a meaningful exercise are likely to be uncommon.

To the extent that the data allow, publication of DGVs for toxicants that are used as formulations should be accompanied by a summary of the toxicity of the formulations and if or how this differs from the toxicity of the AI. The DGVs for pesticides or other chemicals that are used as formulations should be expressed as a concentration of the AI, for example, 'x' µg/L AI.

3.11 Reality-checking the guideline values

Once the GVs have been derived, their suitability should be evaluated by comparing them to the available chronic or converted-acute toxicity data used to derive them or to data from field surveys/monitoring or field-based or laboratory-based, microcosm or mesocosm toxicity experiments. The aim of this is to determine whether any species, for which toxicity data are available, might be affected if exposed at the GV concentration. If either of the following conditions are met, the GV should be considered to provide inadequate protection:

- if a GV is greater than the geometric mean of experimental chronic-negligible-effect data for any key species (e.g. economically, culturally or ecologically important)
- if there is a discrepancy between the theoretical level of protection that should be provided and that indicated as being offered, based on the available experimental chronic (or converted-acute) toxicity data (e.g. if more than 5% of the experimental data are below the PC95 value). This will be difficult to assess for small datasets.

Determining whether a species is of particular ecological, economic or cultural importance may require consultation with relevant stakeholders, such as academia, government agencies, the aquaculture industry and local Indigenous communities.

In cases where the protection provided by an SSD-derived GV is deemed inadequate, a GV with a higher level of species protection could be recommended – for example, the PC99 recommended to be used for ecosystems that are slightly to moderately disturbed and the PC95 used for ecosystems that are highly disturbed. If this does not provide sufficient protection, additional toxicity data are required.

The GVs for naturally occurring elements (e.g. metals) and compounds (e.g. some hydrocarbons) should be checked against background concentrations to ensure that unrealistically low GVs (lower than typical background concentrations) are not derived. A default set of background data for metals and metalloids is presented in the 2000 Guidelines (Table 8.3.2, ANZECC and ARMCANZ 2000b). Alternatively, site-specific or regional GVs based on background concentrations could be derived (see ANZG 2018: *Deriving guideline values using reference-site data*).

Finally, it is important to note that the method for deriving toxicant GVs is part of the current *Australian and New Zealand Guidelines for Fresh and Marine Water Quality* (ANZG 2018) that, in turn, is part of a much broader framework for managing and assessing water and sediment quality (Australian Government 2018). As such, the method should be used in conjunction and accordance with the guiding principles and overall guidance provided in these guidelines.

4 Assessment-factor method

This method should only be used for deriving GVs when there are insufficient data to meet the minimum data requirements of the SSD method. The guidance provided in Warne (1998, 2001) should be used to calculate GVs by this method. The method has little scientific rigour, and the resulting values are generally, but not necessarily, conservative (i.e. low concentrations and hence very protective of ecosystems) due to the magnitude of the AFs used. All GVs derived using the AF method are classed as having 'unknown reliability'. The 2000 Guidelines contained GVs with a variety of different terms (e.g. low-reliability environmental-concern levels), which created confusion among users. Thus, GVs, default or site-specific, are to be referred to only by their reliability category (i.e. very high, high, moderate, low, very low and unknown reliability).

Unknown reliability GVs derived using the AF method should not be used as DGVs. Where possible, more toxicity data should be acquired (from the literature or generated) to enable these GVs to be updated and, consequently, their reliability improved. New GVs, generated because of the existence of GVs of unknown reliability, could be submitted via the third-party GV derivation process, for national consideration and endorsement as DGVs (see Section 3.1).

Guideline values derived using the AF method do not need to undergo the reality-check procedure except when checking them against natural background concentrations. This is because the GV is the most sensitive toxicity value divided by an assessment factor and, therefore, the GV provides protection to all species for which there is toxicity data.

5 Ensuring transparency in the derivation of guideline values

The electronic toxicity data-quality assessment sheets that are generated as part of deriving GVs should be supplied along with other documents when the proposed GV is submitted for consideration and approval.

All the data used to derive GVs (e.g. toxicity data, ACRs) and the corresponding physicochemical data must be included as part of the documentation for proposed GVs. All decisions based on professional judgement must be fully explained and justified, including the presentation of data that support the decision. This entails the documentation of a defensible scientific case as to why a particular action is recommended. Some examples of the documentation of decisions based on professional judgement for DGVs are provided below:

- Rationale for selecting the final dataset for deriving the DGVs for metolachlor in freshwater – ANZG (2020)
- Rationale for selecting the final dataset for deriving the DGVs for zinc in marine water – ANZG (2021)
- Decision to base the dioxins in freshwater DGVs on fish data only – ANZG (2023)
- Rationale for selecting the final dataset for deriving the DGVs for simazine in marine water – ANZG (2025d)
- Re-analysis of data for an echinoderm species and exclusion of data for 10 bivalves for the DGVs for aluminium in marine water – ANZG (2025e)

Moreover, most of the DGVs published since 2020 on the ANZG (2018) website include numerous professional judgement decisions on individual data selections, which can be also used as a guide for the degree of documentation that is required.

Appendix A. Details of updates to the derivation method from 2015 to 2018 and from 2018 to 2025

Table A.1 Details of updates between the 2015 and 2018 versions of the method for deriving Australian and New Zealand water quality guideline values for toxicants

Details of updates between 2015 and 2018
The citation of the report was changed
Section 1.1: the final sentence was modified
Section 1.2: a sentence was added to clarify the difference between guideline values and default guideline values
Section 2: a new final paragraph was inserted
Section 3.1, first paragraph: • Text was added to clarify that chronic toxicity data includes data generated from both single-generation and multi-generation toxicity tests • A sentence was added to clarify when toxicity data related to uptake from water and from food are appropriate to use • Text was added to clarify what type of toxicity data can be used to derive guideline values for chemicals that are normally released in mixtures, such as pesticides • Text was added to clarify the types of exposure routes that are appropriate for bioaccumulating and non-bioaccumulating chemicals
Section 3.1: • Text on the mode or mechanism of action of the test chemical was added to the information to be collated • Text on deriving guideline values for estuarine, freshwater and marine waters was added
Table 1: • The content of the rows for acute microinvertebrates, chronic macroinvertebrates and chronic microinvertebrates were changed • Changes were made to the footnotes, particularly the definitions of macroinvertebrates and microinvertebrates • There are now rows for 3 early life-stage endpoints (lethality, development and fertilisation), each with their own definition of the minimum exposure duration to be considered chronic • The chronic section was modified to accommodate both single-generation and multi-generation toxicity data, and definitions of these terms were provided in the Glossary
Section 3.2: • First paragraph: text was added to indicate that data with characteristics presented in Table 2 can be used if a strong justification is provided • Last paragraph: this paragraph was added to explain how ecotoxicity data for chemicals that occur in formulation should be screened
Table 2: • The caption to the table was changed and a footnote added to the caption of the table • An example was provided in row 2 • The definition in row 3 was expanded
Section 3.3.2: the text for the second dot point on the degree of replication required was modified
Section 3.4: 2 new paragraphs were added (the last 2 in this section). These provide guidance on the use of field, mesocosm and microcosm data, how much replication is required, appropriate experimental design, and they introduce the concept of using a weight-of-evidence approach to derive site-specific guideline values
Section 3.4.2: text was added that relates to the use of acute-to-chronic ratios (ACRs) to convert acute NOEC, LOEC and MATC values to chronic NOEC/EC10 values. The ACR for essential elements was deleted

Details of updates between 2015 and 2018
Section 3.4.3: the title of this section was changed and the text extensively rewritten to highlight the limitations of the available methods that attempt to account for modifying factors of metal toxicity – the hardness-modified method, multiple linear regressions and biotic ligand models – and to permit the use of all methods to derive or modify guideline values
Section 3.4.5: • The number of parameters in log-logistic and Burr type III distributions (2 and 3, respectively) was added and a reference to Batley et al. (2018) was added • The third, fourth and fifth paragraphs (in the August 2015 version) have been replaced by 4 paragraphs. These paragraphs explain the ability to combine chronic and converted chronic data and to combine fresh and marine chronic ecotoxicity data in cases where there are insufficient chronic ecotoxicity data
Section 3.5: this section was replaced. The new section contains a weight-of-evidence approach for assessing if data are unimodal or multimodal
Section 3.7: this section was modified to explain the links between grades of ecosystem protection and the levels of protection provided by guideline values. Text about rounding off guideline values was added
Section 3.8: • The first sentence was modified to clarify that it refers to guideline values derived using a species sensitivity distribution method • The first sentence of the second paragraph (after the dot points) was modified to include the reliability allocated to guideline values derived using the assessment-factor method • After the text describing how the reliability of guideline values is determined, new material was added on how other factors can affect the accuracy of guideline values and how to address this • The paragraph following Table 7 was expanded considerably to provide additional context regarding the visual assessment of how well the distribution fits the toxicity data • New text was added after Figure 2. This explains how the reliability classification of default guideline values can be improved and provides a link to a site that states how to use default guideline values
Section 3.9: a paragraph was added on the appropriateness of single-generation and multi-generation toxicity data for bioaccumulating and 'persistent, bioaccumulative and toxic' substances. Additional text was added to indicate when toxicity data based on food-exposure might be appropriate
Section 3.10: this section was moved to Section 3.11 and a new Section 3.10 'Accounting for formulations' was added. The current Section 3.10 explains how to deal with chemicals that occur in commercial formulations; for example, pesticides
Section 3.11: a second paragraph was added to clarify how to use guideline values derived using the assessment-factor method
Section 3.12: • The second-last paragraph has had some clarifying text added • A new final paragraph was added to explain the links between this document and the Australian and New Zealand Guidelines for Fresh and Marine Water Quality and the Australian National Water Quality Management Strategy

Table A.2 Details of updates between the 2018 and current (2025) versions of the method for deriving Australian and New Zealand water quality guideline values for toxicants

Section	Details of update in 2025 version	Rationale for update
Throughout	Numerous minor editorial updates	*To clarify and improve specific aspects of the derivation method* Numerous minor editorial updates are proposed throughout this report to improve clarity and consistency
Title	Minor revisions to the title	*For improved clarity*
Preface	Addition of a Preface that details the history of the derivation method	*For improved clarity and layout* The historical context was previously included in the Summary; however, this was not the appropriate location so it was moved to a new Preface section and updated to reflect the new update to the method
Summary	Updated to reflect the latest update to the method	*For improved clarity and layout*
Section on *Modifications made since 2015*	Moved this section to a new appendix (Appendix A) and added another table that lists the key changes between the 2018 and 2024 versions	*For improved clarity and layout* Considered more appropriate to move the information on updates to an appendix, which is then referred to early in the document (i.e. in the Introduction). Readers can now go to the appendix if they are interested, without the details cluttering the front end of the document
Section 1.2 *'Purpose of this report'*	Added text to recognise that the method could be used as a starting point for deriving sediment or groundwater quality guideline values, and that modifications to the method may be needed to account for unusual characteristics of a toxicant	*To provide some flexibility for the method to be applied for different purposes* It is acknowledged that the method may not be fit for purpose and may therefore require modification for different types of media and for certain types of toxicants. The added text provides this flexibility but clearly states that full technical justification for any modifications must be documented
Section 2 'Overview of the revised method'	Minor changes made to reflect updates to the rest of the method	*For improved clarity and layout* Added 3 sub-section headings to better separate the text Most of the edits in this section reflect the recent changes made to the method
Section 2 'Overview of the revised method'	Inclusion of text (new Section 2.2) that emphasises the need to check that an appropriate analytical method is available for a toxicant before embarking on a guideline value derivation	*To clarify and improve specific aspects of the derivation method* As a result of committee review of the mancozeb and alpha-cypermethrin default guideline values, it was recommended that formal processes be put in place to ensure that default guideline values cannot be derived for toxicants for which there is no appropriate analytical method available. Such guidance was added to the website (*Third party process for proposing default guideline values*), and it was necessary that a note to this effect also be added to Warne et al. (2018), early in the document

Section	Details of update in 2025 version	Rationale for update
Section 0 'Preferred guideline value derivation method'	Added text regarding replacement of Burrlioz 2.0 by (shiny)ssdtools	*To clarify and improve specific aspects of the derivation method; to reflect current state of the science* Needed to alert users that Burrlioz 2.0 has been replaced by (shiny)ssdtools
Section 2, Section 3.3.2, Section 3.4.1	Referred to more recent and more detailed sources of guidance on deriving site-specific guideline values	*To reflect updated guidance on the ANZG (2018) website and associated content (e.g. technical briefs)* In late 2019, the *Australian and New Zealand Guidelines* website was updated to include reference to 2 recently published sources of guidance on deriving site-specific guideline values (Huynh and Hobbs 2019 and van Dam et al. 2019). Although Warne et al. (2018) provided some information on deriving site-specific guideline values, it did not go into much detail and, as such, now benefits from referring readers to other sources of recent and more detailed guidance on this, including the ANZG (2018) website
Section 3.1 'Collating toxicity and physicochemical data'	Added sub-headings to improve readability	*For improved clarity and layout*
Section 3.1 'Collating toxicity and physicochemical data'	Added a sentence to highlight that, for default guideline values, toxicity data for both Australasian and overseas species should be collated	*To clarify and improve specific aspects of the derivation method* It was brought to attention that Warne et al. (2018) does not explicitly state that data from both Australian and New Zealand species and overseas species can be used (although it is implied in the first paragraph). The added sentence made it explicit. Also, additional text on this issue was added in Section 3.2 (see further below)
Section 3.1 'Collating toxicity and physicochemical data'	Updated guidance now in Section 0 relating to what to do for estuarine/brackish waters to reflect updated guidance on the ANZG (2018) website: • *Toxicant default guideline values for water quality in aquatic ecosystems* • *Salinity effects*	*To reflect updated guidance on ANZG (2018) website and associated content (e.g. technical briefs)* Website guidance on what to do for toxicant guideline values for estuarine/brackish waters was updated following a committee request (from Queensland). The updated guidance is more detailed and appropriate than the original limited guidance in Warne et al. (2018), which simply stated that 'Marine guideline values should be used for estuarine ecosystems (≥ 0.5‰ to < 25‰); however, if there are sufficient ecotoxicity data tested under estuarine conditions then estuarine guideline values should be derived using the methods set out in this report'. Therefore, the guidance on this issue in Warne et al. (2018) was updated
Section 3.1 'Collating toxicity and physicochemical data'	Removed sentence stating the preference to use ecotoxicity data published in peer-reviewed papers	*To clarify and improve specific aspects of the derivation method* The sentence partially contradicted earlier criteria for acceptability of publication types. If data are assessed to be reliable, they should be treated equally, regardless of where they were reported

Section	Details of update in 2025 version	Rationale for update
Section 3.1 'Collating toxicity and physicochemical data'	Added sentence stating that evidence of any independent assessment of unpublished data needs to be provided with the final guideline values	*To clarify and improve specific aspects of the derivation method* To ensure sufficient supporting evidence is provided with guideline value documentation
Section 0 'Acute and chronic toxicity'	Example test durations for tropical and polar species added	*To clarify and improve specific aspects of the derivation method* To provide more information and guidance for users
Section 0 'Acute and chronic toxicity', Table 1	Minor revision to chronic test durations for specific embryo/larval macroinvertebrate endpoints, from ≥ 14 d to ≥ 7 d	*To clarify and improve specific aspects of the derivation method*
Section 3.1.3 'Supporting information'	Details added on the various roles that supporting information plays	*To clarify and improve specific aspects of the derivation method* To provide more information and guidance for users
Section 3.2 'Screening toxicity data'	Added sub-headings and re-arranged the sections to improve readability	*For improved clarity and layout*
Section 3.2 'Screening toxicity data'	Added a new section (3.2.2) titled 'Checking for significant limitations of toxicity data', which captures Table 2 Added additional items to Table 2 relating to characteristics that warrant the rejection of toxicity data, and added reference to the relevant sections of the method that relate to each issue. Also added 2 new characteristics (not previously in the method): • A study tested only 1–2 concentration (plus a control) of a toxicant – with an accompanying table note describing the preferred number of test treatments • Test concentrations were not measured In the text, examples of where data with limitations listed in Table 2 might be included in default guideline value derivations was added	*To clarify and improve specific aspects of the derivation method* To make the table more comprehensive and to better link this section with Section 3.3.1 (assessing quality of laboratory toxicity data) For the 2 new characteristics that were added: • A test based on 1–2 concentration (plus a control) is effectively meaningless for estimating a negligible-effect estimate • Studies without measured test concentrations are far less reliable. However, there may be instances in which such data are acceptable

Section	Details of update in 2025 version	Rationale for update
Section 3.2 'Screening toxicity data' (new Section **Error! Reference source not found.**)	Updates to guidance on how to use censored data (i.e. '<' and '>' values)	*To clarify and improve specific aspects of the derivation method* When using a '<' value (i.e. where a significant effect is still observed at the lowest concentration tested), a conversion factor of 2.5 (i.e. conversion from a LOEC to a NOEC) should be recommended but professional judgement as to whether or not a conversion factor is applied (e.g. where the effect at the lowest concentration is significant but < 10%) is allowed Added information on capability of (shiny)ssdtools to accommodate censored data
Section 3.2 'Screening toxicity data' (new Section 0)	Created new Section 0 titled 'Ecological relevance' Added text to clarify how to treat data for potential nuisance species	*To clarify and improve specific aspects of the derivation method*
Section 3.2 'Screening toxicity data' (new Section 0)	Added paragraph to more clearly address geographical considerations for species selection for default guideline values	*To clarify and improve specific aspects of the derivation method* Warne et al. (2018) provide little explicit guidance about species relevance and associated selection for default guideline values. This was added. It includes mention of combining temperate and tropical, as well as not using polar species for non-polar default guideline values
Section 3.2 'Screening toxicity data' (new Section 0)	Updated guidance on the preferred hierarchy of statistical estimates of toxicity that includes a clearer hierarchy and more consideration on the use or otherwise of NOECs	*To clarify and improve specific aspects of the derivation method* The hierarchy was improved to reflect that fact that there is not necessarily a single preferred statistical estimate of toxicity but rather, there is a group of preferred ones, some appropriate ones and some less preferred ones. Also, the guidance now emphasises the need to assess the rigour and associated defensibility of each of the values in addition to whether the values are based on a preferred or less preferred estimate when making a decision about which value to use
Section 3.2 'Screening toxicity data' (new Section 0)	The term 'negligible-effect' estimate or concentration has replaced 'NOEC/EC10'. This term is also now used in all default guideline value technical briefs. It was added to the Glossary, defined as follows: 'Negligible effect: corresponding to an estimated no-effect or acceptable-effect level; estimates of a negligible-effect are typically restricted to the NEC, BEC10, EC10, IC10 and NOEC'	*To clarify and improve specific aspects of the derivation method* Throughout Warne et al. (2018) and also in most draft default guideline value technical briefs, the term 'NOEC/EC10 equivalent' (or similar) was used to denote a value that has been converted from a chronic LOEC or EC50 or an acute EC/LC50. The term is not accurate, because in fact we accept both no-effect estimates (i.e. NEC, NOEC) and what one might term 'acceptable-effect' (e.g. EC10, EC20, BEC10) estimates. Thus, a new, more accurate term to encompass these types of estimates was adopted

Section	Details of update in 2025 version	Rationale for update
Section 3.2 'Screening toxicity data' (new Section 0) and Section 3.4.2 'Conversion of toxicity data'	Added text to recommend that, where there is a NOEC and a more preferred estimate (e.g. NEC, EC10) available for a species, that a check of the defensibility of the values is undertaken before the NOEC is rejected	*To clarify and improve specific aspects of the derivation method* NOECs are less favoured statistical estimates than NECs, NSECs and ECx where x < 20%. However, depending on the nature of a dataset, the NOEC may, in some cases, still be a more defensible estimate than one of the more preferred estimates. Before rejecting a NOEC for a NEC, NSEC or ECx (where x < 20%), the rigour and associated defensibility of the values should be considered, and professional judgement applied accordingly
Section 3.2 'Screening toxicity data' (new Section 3.2.5)	Created new section titled 'Physicochemical water quality ranges' Added text from earlier in the document relating acceptable salinity ranges for freshwater and marine water default guideline derivations	*To clarify and improve specific aspects of the derivation method* New section captures existing and new information on acceptable ranges for key physicochemical variables in toxicity testing waters Moved text from an earlier section to this section, where it is more relevant
Section 3.2 'Screening toxicity data' (new Section 3.2.5)	Added new guidance that, for some toxicants (where the specific criteria are met) it might be preferable to derive a single combined freshwater/marine species sensitivity distribution and associated guideline values (with reference to more details in Section 3.4.5)	*To clarify and improve specific aspects of the derivation method* New section captures existing and new information on acceptable ranges for key physicochemical variables in toxicity testing waters Moved text from an earlier section to this section, where it is more relevant
Section 3.3.1 'Laboratory-based toxicity data' (Table 3 and **Error! Reference source not found.**)	Updated the data-quality assessment to be more weighted towards test quality/performance and less weighted towards the provision of basic test information. Also slight revision to the 2 toxicant types – from 'metals' and 'non-metals' in Warne et al. (2018) to 'metals/non-metallic inorganics' and 'organics'	*To clarify and improve specific aspects of the derivation method* Limitations of the data-quality assessment process were highlighted in the public submissions for the draft perfluorooctane sulfonate freshwater default guideline values. Acceptable test quality (i.e. a score of 50%) could be achieved just through the provision of basic test information, before any aspects associated with test quality or performance were considered. This was addressed by giving less weighting to basic test information and more weighting to test performance. Additionally, some additional questions were added to better assess test performance. Revision to the toxicant types was to reflect the fact that a number of non-metallic inorganics can have similar toxicity modifying factors to metals (e.g. fluoride, boron, nitrate, ammonia)
Section 3.3.2 'Field-based, microcosm and mesocosm data'	Section updated to include more guidance for using microcosm or mesocosm data in conjunction with single-species toxicity-test data, in species sensitivity distributions	*To clarify and improve specific aspects of the derivation method* The guidance in Warne et al. (2018) lacked detail regarding what type of microcosm and mesocosm data are suitable for combining in a species sensitivity distribution with single-species toxicity-test data. This guidance has now been added. Other minor improvements were also made to the guidance

Section	Details of update in 2025 version	Rationale for update
Section 3.4 'Selection of data to derive guideline values'	Added and revised sub-section headings to improve readability	*For improved clarity and layout*
Section 3.4 'Selection of data to derive guideline values' (new Section 3.4.1)	New Section 3.4.1 titled 'General' Text related to field and 'cosm' studies was deleted because it duplicated text in Section 3.3.2	*To clarify and improve specific aspects of the derivation method*
Section 3.4 'Selection of data to derive guideline values' (new Section 3.4.2.3)	Created new section for addressing how to convert toxicity data for the derivation of short-term guideline values	*To clarify and improve specific aspects of the derivation method* The guidance in Warne et al. (2018) was ambiguous and needed more clarity. It also did not recognise some additional issues that may need to be considered, which have now been added
Section 0 'The use of bioavailability models for metal default guideline values'	Significant revisions to this section to better reflect current knowledge/practices and be more focused on guidance for deriving guideline values rather than guidance for applying bioavailability adjustments to existing default guideline values. Revisions and additions include: • updated text on multiple linear regression models to reflect current status • acknowledgement of non-linear approaches to modelling bioavailability • process for validating a bioavailability model for Australia and New Zealand • transfer of information and table from the superseded 2000 Guidelines hardness corrections to the ANZG (2018) website	*To reflect updated guidance on ANZG (2018) website and associated content (e.g. technical briefs); to reflect current state of the science; to clarify and improve specific aspects of the derivation method* This section needed to address what is required if one wants to incorporate bioavailability into the derivation of a guideline value. Guidance on how to adjust existing default guideline values for bioavailability was moved to the relevant location on the ANZG (2018) website, where such guidance is more relevant (i.e. Water hardness section in Accounting for local conditions page):

Section	Details of update in 2025 version	Rationale for update
Section 0 'Obtaining a single toxicity value for each species'	Updated guidance on data requirements for calculating a geometric mean, including the addition of life stage to this process	*To clarify and improve specific aspects of the derivation method* The guidance in Warne et al (2018) was too vague and led to at least 2 different interpretations of data requirements for calculating a geometric mean. The guidance was clarified, including an updated Table 4. Geometric means of data can only be calculated where the data are for the same species, life stage, endpoint, statistical estimate of toxicity and duration. Guidance has also been added regarding the approach to take when 2 values that otherwise meet the criteria for a geometric mean are very different (e.g. > 1 order of magnitude)
Section 0 'Obtaining a single toxicity value for each species'	Inserted paragraph regarding use of NOEC (and similar) data from data-conversion section into this section	*For improved clarity and layout* The paragraph was about the use of NOEC data rather than conversion of NOEC data and, thus, was more appropriately located in Section 0
Section 3.4.5 'Do the data meet the minimum data requirements of the species sensitivity distribution method?'	Added text to note that the minimum data requirements have been updated as a result of the transition to (shiny)ssdtools and replaced references to Burrlioz with appropriate references to (shiny)ssdtools	*To clarify and improve specific aspects of the derivation method* Need to alert users to the fact that the minimum data requirements have increased, and that (shiny)ssdtools should be used instead of Burrlioz 2.0
Section 3.4.5 'Do the data meet the minimum data requirements of the species sensitivity distribution method?' (Table 5)	Table 5 'Examples of distinct taxonomic groups' was updated	*To clarify and improve specific aspects of the derivation method* The corresponding table (Table 6) in Warne et al. (2018) did not promote consistent assignment of taxonomic groups between default guideline value derivations and was also inconsistent in the resolution at which it distinguished some taxonomic groupings. This resulted in inconsistency in how service providers assigned taxonomic groups to species for default guideline value derivations. Ideally, assignment of taxonomic groups between default guideline values should be as consistent as possible. Thus, Table 6 was updated

Section	Details of update in 2025 version	Rationale for update
Section 3.4.5 'Do the data meet the minimum data requirements of the species sensitivity distribution method?'	Guidance regarding when data from 2 different media (i.e. freshwater and marine water) can be combined was expanded to explicitly allow the derivation of a single set of default guideline values for both media types, assuming the relevant criteria have been met	*To clarify and improve specific aspects of the derivation method* Warne et al. (2018) made no mention of situations where it might be appropriate to derive a single set of default guideline values for both media types rather than different default guideline values for each medium. Allowance for this approach was added to the revised method. Notably, where the criteria to combine fresh and marine data have been met, the reliability of the default guideline values will not be down-graded
Section 0 'Checking the toxicity data for multimodal distributions'	Updated process for assessing datasets for bimodality to accommodate adoption of bimodal distributions in (shiny)ssdtools Includes re-structuring of the section into 2 sub-sections: (i) the weight-of-evidence approach and (ii) decision making	*To clarify and improve specific aspects of the derivation method; to reflect current state of the science.* Inclusion of the log-normal–log-normal mixture distribution in the (shiny)ssdtools default model set required changes to be made to the bimodality-assessment process to capture the use of this model and information on bimodality that it can provide. Assuming certain requirements are met, a bimodal dataset can now be modelled as the full dataset (i.e. not cut down to only the most sensitive taxonomic group) using the log-normal–log-normal mixture distribution. Note: this distribution is very new to species sensitivity distribution modelling, and there remain some uncertainties over its performance and use for this purpose. To aid with the use of this distribution, a new appendix (Appendix C) was added that provides several examples of where the log-normal–log-normal distribution has an influence on the species sensitivity distribution and how this can be interpreted and dealt with. Note that there may be a need for further guidance to be developed in the future
Section 0 'Enter toxicity data into Burrlioz'	Whole section updated to reflect replacement of Burrlioz 2.0 by (shiny)ssdtools (new title 'Use of (shiny)ssdtools')	*To clarify and improve specific aspects of the derivation method; to reflect current state of the science* Appropriate new guidance on how to use (shiny)ssdtools for the derivation of guideline values – in particular, default guideline values – was required
Section 3.7 'Calculate guideline values'	Restructured to create 2 sub-sections – 3.7 'Levels of species protection' and 0 'Significant figures and rounding'	*For improved clarity and layout; to clarify and improve specific aspects of the derivation method; to reflect current state of the science.* Needed to add new sub-section that explained the new significant figures and rounding rules for guideline values and default guideline values, as reviewed and approved by the Australian and New Zealand Guidelines jurisdictional committees. These rules reflect proper statistical practice and provide a defensible basis for determining the number of significant figures and associated rounding of guideline values and default guideline values

Section	Details of update in 2025 version	Rationale for update
Section 3.7 'Calculate guideline values' (new Section 3.7)	Corrected guidance around the application of guideline values to highly modified systems to reflect existing ANZG (2018) website guidance that the values for slightly to moderately disturbed systems are also generally recommended for highly disturbed systems, but that the 90% or 80% species-protection guideline values can be applied under certain circumstances	*To clarify and improve specific aspects of the derivation method* Warne et al. (2018) stated that the 90% or 80% species-protection guideline values are applied to highly disturbed systems. This is only partially correct. The ANZG (2018) guidance (and original 2000 Guidelines) on this aspect specifies that the values for slightly to moderately disturbed systems are generally recommended for highly disturbed systems, but that the 90% or 80% species-protection guideline values can be applied under certain circumstances. Text was updated to reflect this
Added Section 3.8 'Determine the reliability of the guideline values'	Improved the clarity of Table 6 and added a sentence at the end of the section to refer the reader to guidance on the ANZG (2018) website on how to apply default guideline values depending on their reliability rating. Note: no changes were made to the reliability classification itself (other than the case of combined fresh and marine data – see rationale for update) Also updated the number of species for 'adequate' datasets (from 5–7 to 6–7) Also updated Figure 2 to reflect (shiny)ssdtools SSD outputs and to add another species sensitivity distribution depicting a 'good' fit	*To clarify and improve specific aspects of the derivation method* In Table 6, 2 columns – column 1 (Sample size) and column 3 (Adequacy of sample size) – related to the sample size. The table benefited from bringing these 2 columns together, and the table re-ordered from left to right based on (i) type of toxicity data, (ii) sample size and (iii) adequacy of species sensitivity distribution model fit Also added reference to the case where the reliability of guideline values based on combined freshwater and marine data are not down-graded if either of the 2 criteria specified in Section 0 have been met The number of species for 'adequate' datasets (from 5–7 to 6–7) also needed to be updated to reflect the new minimum dataset size
Section 3.9 'Accounting for the potential for chemicals to bioaccumulate	Updated this section to reflect updated guidance on the *Bioaccumulation* webpage	*To reflect updated guidance on ANZG (2018) website and associated content (e.g. technical briefs)* The guidance on the ANZG (2018) website for how to deal with bioaccumulation was significantly updated in mid/late 2019 following concerns of misapplication, particularly for assessments of perfluorooctane sulfonate. The updated guidance underwent a rigorous jurisdictional review process. The guidance in Warne et al. (2018) was relatively detailed and mostly consistent with the original website guidance and, thus, needed to be revised to align with the new website guidance. The updated guidance now focuses on how to deal with bioaccumulation when deriving guideline values rather than also when applying guideline values, the latter being the role of the ANZG (2018) website rather than the guideline value derivation method
Section 3.10 'Accounting for formulations'	Section completely updated	*To clarify and improve specific aspects of the derivation method* This section was completely revised following jurisdictional committee comments on a draft default guideline values technical brief, in relation to how formulation data were treated. The existing text was deemed to provide impractical or even non-implementable guidance. From an implementation or practicality perspective, the revised text is considered to be fit for purpose

Section	Details of update in 2025 version	Rationale for update
Section 3.11 'Reality-checking the guideline values'	Improved the clarity of the guidance on how to 'reality-check' the draft default guideline values	*To clarify and improve specific aspects of the derivation method* The text in Warne et al. (2018) was ambiguous in parts, and some users had expressed some confusion over the method to reality check the draft default guideline values. The text was updated
Section 5 'Ensuring transparency in the derivation of guideline values'	Added several examples of documentation of professional judgement decisions from recently published default guideline values	*To clarify and improve specific aspects of the derivation method* Examples of documented professional judgement processes from recent default guideline value derivations were added to help users understand what level of documentation is required. This addition was in response to a peer-review comment.
References	Supporting references for revisions added	*To provide sufficient supporting information*

Appendix B. Data-quality scoring system

Table B.1 Scoring system for assessing the quality of toxicity data for organics to freshwater non-plants, to be used in the derivation of guideline values for toxicants

No.	Question	Score[a]
A. General test design information		
1	Was the duration of the exposure stated (e.g. 48 h, 96 h)?	Yes (3), No (0/FAIL)
2	Was the biological endpoint (e.g. immobilisation, population growth) stated? Note: ensure that the endpoint is considered to be ecologically relevant. Do not use endpoints with no demonstrated ecological relevance to derive guideline values, and they do not need their quality assessed.	Yes (3), No (0/FAIL)
3	Was the measure of toxicity reported (e.g. NEC, ECx, NOEC) **and** its associated biological-effect size reported or quantifiable from the data (e.g. EC10, LC50, 15% effect at NOEC)?	Yes (3), No (0/FAIL)
4	Were appropriate controls (e.g. no-toxicant control, solvent control) used?	Yes (3), No (0/FAIL)[b]
5	How may treatment concentrations were used (in addition to the control)?	≥ 9 (3), 6–8 (2), 3–5 (1), < 3 (0/FAIL)[b]
6	What was the test concentration spacing? Note: a spacing of < 3.2 is highly preferred, while a spacing of ≥ 10 is too great.	≤ 3.2 (3), 3.3–9 (1), ≥ 10 (0/FAIL)[b]
7	Was each control and chemical concentration at least duplicated? Note: for concentration-response modelling, it is also acceptable to have many concentrations (e.g. > 15) without replication (assuming the controls are replicated).	Yes (3), No (0/FAIL)
8	Were the characteristics of the test organism (e.g. length, mass, age) stated?	Yes (3), No (0)
9	Was the type of test medium used stated (e.g. synthetic or natural water? If synthetic, to what recipe? If natural, what source? Filtered or unfiltered?)?	Yes (3), No (0)
10	Was the type of exposure (e.g. static, flow-through) stated?	Yes (3), No (0)
	Maximum sub-total for Part A = 30	
B. Test performance/results		
11	Were analytical reagent-grade chemicals or the highest possible purity chemicals used for the experiment?	Yes (3), No/not stated (0)
12 (A)	Were test solutions, blanks and/or controls tested for common contamination (e.g. elevated naturally occurring substances, such as nutrients, metals, metalloids) or other suspect contaminants?	Yes (2), No/not stated (0)
12 (B)	If so, were any significant contamination issues identified?	No (2), Yes (0/FAIL)
13	Were the contaminant concentrations measured at the beginning and end of the exposure? Note: normally, toxicity data calculated using nominal-concentration data would not be used to derive guideline values; however, professional judgement can be used to include such data, provided a justification for their use is provided.	Yes (8), Measured once (4), Not measured or stated (0/FAIL)[b]
14 (A)	Were test-acceptability criteria stated (e.g. mortality in controls must not exceed a certain percentage) or inferred (e.g. test methods used were United States Environmental Protection Agency or Organisation for Economic Co-operation and Development)?	Stated (2), Inferred (1), Neither (0)
14 (B)	If so, were test-acceptability criteria met? Note: data that fail the acceptability criteria are automatically deemed to be of unacceptable quality and must not be used.	Yes (6), Not reported (2), No (0/FAIL)

No.	Question	Score[a]
15 (A)	Were parallel reference-toxicant toxicity tests conducted?	Yes (2), No (0)
15 (B)	If so, did the reference toxicity results fall within the acceptable limits?	Yes (6), Not reported (2), No (0/FAIL)
16	Was there a concentration-response (C-R) relationship, either observable or stated (i.e. was there a larger response in the endpoint of interest at higher treatment concentrations than at lower treatment concentrations?)? Note: a full C-R relationship is one that includes zero, partial and full effects, while a partial C-R relationship is one that is missing either zero, partial or full effects. Professional judgement might be required to make decisions associated with C-R relationships that exhibit characteristics such as non-monotonicity and low-dose effects.	Yes – full (8), Yes – partial (4), No – (0/FAIL)
17 (A)	Was the statistical method or model used to determine the toxicity estimate stated **and** appropriate (e.g. refer to Green et al. 2018, noting that new methods will/have emerged that may not be captured in past publications)?	Yes (6), No (0)
17 (B)	For point estimates from concentration-response modelling (e.g. LC/EC/NEC/BEC/NSEC), was an estimate of variability provided? For estimates based on hypothesis testing (e.g. NOEC/LOEC/MATC), was the significance level 0.05 or less?	Yes (2), No (0)
	Maximum sub-total for Part B = 47	
C. Test water chemistry/test conditions		
18	Were the following parameters measured and reported?	
18.1	pH – was pH measured at least at the beginning and end of the toxicity test?	Yes (3), No (0/FAIL)
18.2	Dissolved oxygen (DO) – was DO measured at least at the beginning and end of the toxicity test?	Yes (2), No (0/FAIL)
18.3	Conductivity	Yes (2), No (0/FAIL)
18.4	Temperature	Yes (2), Temperature of room or chamber was reported (1), No (0/FAIL)
18.5	Any other parameters of relevance (e.g. other toxicity-modifying factors)?	Yes (2), No (0/FAIL)
	Maximum sub-total for Part C = 11	
	Maximum total for parts A + B + C = 88	

Source: Updated from Warne et al. (2018).

[a] Where a question has a possible 'No (0/FAIL)' outcome, professional judgement may be required to determine if a 0 is assigned or if the test fails the quality-assessment process.

[b] See
Table 2 for additional guidance.

Table B.2 Scoring system for assessing the quality of toxicity data for metals/non-metallic inorganics to freshwater plants, to be used in the derivation of guideline values for toxicants

No.	Question	Score[a]
A. General test design information		
1	Was the duration of the exposure stated (e.g. 48 h, 96 h)?	Yes (3), No (0/FAIL)
2	Was the biological endpoint (e.g. immobilisation, population growth) stated? Note: ensure that the endpoint is considered to be ecologically relevant. Do not use endpoints with no demonstrated ecological relevance to derive guideline values, and they do not need their quality assessed.	Yes (3), No (0/FAIL)
3	Was the measure of toxicity reported (e.g. NEC, ECx, NOEC) **and** its associated biological-effect size reported or quantifiable from the data (e.g. EC10, LC50, 15% effect at NOEC)?	Yes (3), No (0/FAIL)
4	Were appropriate controls (e.g. no-toxicant control, solvent control) used?	Yes (3), No (0/FAIL)[b]
5	How may treatment concentrations were used (in addition to the control)?	≥ 9 (3), 6–8 (2), 3–5 (1), < 3 (0/FAIL)[b]
6	What was the test concentration spacing? Note: a spacing of < 3.2 is highly preferred, while a spacing of ≥ 10 is too great.	≤ 3.2 (3), 3.3–9 (1), ≥ 10 (0/FAIL)[b]
7	Was each control and chemical concentration at least duplicated? Note: for concentration-response modelling, it is also acceptable to have many concentrations (e.g. > 15) without replication (assuming the controls are replicated).	Yes (3), No (0/FAIL)
8	Were the characteristics of the test organism (e.g. length, mass, age) stated?	Yes (3), No (0)
9	Was the type of test medium used stated (e.g. synthetic or natural water? If synthetic, to what recipe? If natural, what source? Filtered or unfiltered?)?	Yes (3), No (0)
10	Was the type of exposure (e.g. static, flow-through) stated?	Yes (3), No (0)
	Maximum sub-total for Part A = 30	
B. Test performance/results		
11	Were analytical reagent-grade chemicals or the highest possible purity chemicals used for the experiment?	Yes (3), No/not stated (0)
12 (A)	Were test solutions, blanks and/or controls tested for common contamination (e.g. elevated naturally occurring substances, such as nutrients, metals, metalloids) or other suspect contaminants?	Yes (2), No/not stated (0)
12 (B)	If so, were any significant contamination issues identified?	No (2), Yes (0/FAIL)
13	Were the contaminant concentrations measured at the beginning and end of the exposure? Note: normally, toxicity data calculated using nominal-concentration data would not be used to derive guideline values; however, professional judgement can be used to include such data, provided a justification for their use is provided.	Yes (8), Measured once (4), Not measured or stated (0/FAIL)[b]
14 (A)	Were test-acceptability criteria stated (e.g. mortality in controls must not exceed a certain percentage) or inferred (e.g. test methods used were United States Environmental Protection Agency or Organisation for Economic Co-operation and Development)?	Stated (2), Inferred (1), Neither (0)
14 (B)	If so, were test-acceptability criteria met? Note: data that fail the acceptability criteria are automatically deemed to be of unacceptable quality and must not be used.	Yes (6), Not reported (2), No (0/FAIL)
15 (A)	Were parallel reference-toxicant toxicity tests conducted?	Yes (2), No (0)
15 (B)	If so, did the reference toxicity results fall within the acceptable limits?	Yes (6), Not reported (2), No (0/FAIL)

No.	Question	Score[a]
16	Was there a concentration-response (C-R) relationship, either observable or stated (i.e. was there a larger response in the endpoint of interest at higher treatment concentrations than at lower treatment concentrations?)? Note: a full C-R relationship is one that includes zero, partial and full effects, while a partial C-R relationship is one that is missing either zero, partial or full effects. Professional judgement might be required to make decisions associated with C-R relationships that exhibit characteristics such as non-monotonicity and low-dose effects.	Yes – full (8), Yes – partial (4), No – (0/FAIL)
17 (A)	Was the statistical method or model used to determine the toxicity estimate stated **and** appropriate (e.g. refer to Green et al. 2018, noting that new methods will/have emerged that may not be captured in past publications)?	Yes (6), No (0)
17 (B)	For point estimates from concentration-response modelling (e.g. LC/EC/NEC/BEC/NSEC), was an estimate of variability provided? For estimates based on hypothesis testing (e.g. NOEC/LOEC/MATC), was the significance level 0.05 or less?	Yes (2), No (0)
	Maximum sub-total for Part B = 47	
C. Test water chemistry/test conditions		
18	Were the following parameters measured and reported?	
18.1	pH – was pH measured at least at the beginning and end of the toxicity test?	Yes (3), No (0/FAIL)
18.2	Hardness	Yes (3), No (0/FAIL)
18.3	Alkalinity	Yes (3), No (0/FAIL)
18.4	Dissolved organic carbon concentration	Yes (3), No (0/FAIL)
18.5	Conductivity	Yes (2), No (0/FAIL)
18.6	Temperature	Yes (2), Temperature of room or chamber was reported (1), No (0/FAIL)
18.7	Light intensity/illuminance	Yes (2), No (0/FAIL)
18.8	Any other parameters of relevance (e.g. other toxicity-modifying factors)?	Yes (2), No (0/FAIL)
	Maximum sub-total for Part C = 20	
	Maximum total for parts A + B + C = 97	

Source: Updated from Warne et al. (2018).

[a] Where a question has a possible 'No (0/FAIL)' outcome, professional judgement may be required to determine if a 0 is assigned or if the test fails the quality-assessment process.

[b] See

Table 2 for additional guidance.

Table B.3 Scoring system for assessing the quality of toxicity data for organics to freshwater plants to be used in the derivation of guideline values for toxicants

No.	Question	Score[a]
A. General test design information		
1	Was the duration of the exposure stated (e.g. 48 h, 96 h)?	Yes (3), No (0/FAIL)
2	Was the biological endpoint (e.g. immobilisation, population growth) stated? Note: ensure that the endpoint is considered to be ecologically relevant. Do not use endpoints with no demonstrated ecological relevance to derive guideline values, and they do not need their quality assessed.	Yes (3), No (0/FAIL)
3	Was the measure of toxicity reported (e.g. NEC, ECx, NOEC) **and** its associated biological-effect size reported or quantifiable from the data (e.g. EC10, LC50, 15% effect at NOEC)?	Yes (3), No (0/FAIL)
4	Were appropriate controls (e.g. no-toxicant control, solvent control) used?	Yes (3), No (0/FAIL)[b]
5	How may treatment concentrations were used (in addition to the control)?	≥ 9 (3), 6–8 (2), 3–5 (1), < 3 (0/FAIL)[b]
6	What was the test concentration spacing? Note: a spacing of < 3.2 is highly preferred, while a spacing of ≥ 10 is too great.	≤ 3.2 (3), 3.3–9 (1), ≥ 10 (0/FAIL)[b]
7	Was each control and chemical concentration at least duplicated? Note: for concentration-response modelling, it is also acceptable to have many concentrations (e.g. > 15) without replication (assuming the controls are replicated).	Yes (3), No (0/FAIL)
8	Were the characteristics of the test organism (e.g. length, mass, age) stated?	Yes (3), No (0)
9	Was the type of test medium used stated (e.g. synthetic or natural water? If synthetic, to what recipe? If natural, what source? Filtered or unfiltered?)?	Yes (3), No (0)
10	Was the type of exposure (e.g. static, flow-through) stated?	Yes (3), No (0)
	Maximum sub-total for Part A = 30	
B. Test performance/results		
11	Were analytical reagent-grade chemicals or the highest possible purity chemicals used for the experiment?	Yes (3), No/not stated (0)
12 (A)	Were test solutions, blanks and/or controls tested for common contamination (e.g. elevated naturally occurring substances, such as nutrients, metals, metalloids) or other suspect contaminants?	Yes (2), No/not stated (0)
12 (B)	If so, were any significant contamination issues identified?	No (2), Yes (0/FAIL)
13	Were the contaminant concentrations measured at the beginning and end of the exposure? Note: normally, toxicity data calculated using nominal-concentration data would not be used to derive guideline values; however, professional judgement can be used to include such data, provided a justification for their use is provided.	Yes (8), Measured once (4), Not measured or stated (0/FAIL)[b]
14 (A)	Were test-acceptability criteria stated (e.g. mortality in controls must not exceed a certain percentage) or inferred (e.g. test methods used were United States Environmental Protection Agency or Organisation for Economic Co-operation and Development)?	Stated (2), Inferred (1), Neither (0)
14 (B)	If so, were test-acceptability criteria met? Note: data that fail the acceptability criteria are automatically deemed to be of unacceptable quality and must not be used.	Yes (6), Not reported (2), No (0/FAIL)
15 (A)	Were parallel reference-toxicant toxicity tests conducted?	Yes (2), No (0)
15 (B)	If so, did the reference toxicity results fall within the acceptable limits?	Yes (6), Not reported (2), No (0/FAIL)

No.	Question	Score[a]
16	Was there a concentration-response (C-R) relationship, either observable or stated (i.e. was there a larger response in the endpoint of interest at higher treatment concentrations than at lower treatment concentrations?)? Note: a full C-R relationship is one that includes zero, partial and full effects, while a partial C-R relationship is one that is missing either zero, partial or full effects. Professional judgement might be required to make decisions associated with C-R relationships that exhibit characteristics such as non-monotonicity and low-dose effects.	Yes – full (8), Yes – partial (4), No – (0/FAIL)
17 (A)	Was the statistical method or model used to determine the toxicity estimate stated **and** appropriate (e.g. refer to Green et al. 2018, noting that new methods will/have emerged that may not be captured in past publications)?	Yes (6), No (0)
17 (B)	For point estimates from concentration-response modelling (e.g. LC/EC/NEC/BEC/NSEC), was an estimate of variability provided? For estimates based on hypothesis testing (e.g. NOEC/LOEC/MATC), was the significance level 0.05 or less?	Yes (2), No (0)
	Maximum sub-total for Part B = 47	
C. Test water chemistry/test conditions		
18	Were the following parameters measured and reported?	
18.1	pH – was pH measured at least at the beginning and end of the toxicity test?	Yes (3), No (0/FAIL)
18.2	Conductivity	Yes (2), No (0/FAIL)
18.3	Temperature	Yes (2), Temperature of room or chamber was reported (1), No (0/FAIL)
18.4	Light intensity/illuminance	Yes (2), No (0/FAIL)
18.5	Any other parameters of relevance (e.g. other toxicity-modifying factors)?	Yes (2), No (0/FAIL)
	Maximum sub-total for Part C = 11	
	Maximum total for parts A + B + C = 88	

Source: Updated from Warne et al. (2018).

[a] Where a question has a possible 'No (0/FAIL)' outcome, professional judgement may be required to determine if a 0 is assigned or if the test fails the quality-assessment process.

[b] See
Table 2 for additional guidance.

Table B.4 Scoring system for assessing the quality of toxicity data for contaminants (metal or non-metal) to marine and estuarine non-plant species, to be used in the derivation of guideline values for toxicants

No.	Question	Score[a]
A. General test design information		
1	Was the duration of the exposure stated (e.g. 48 h, 96 h)?	Yes (3), No (0/FAIL)
2	Was the biological endpoint (e.g. immobilisation, population growth) stated? Note: ensure that the endpoint is considered to be ecologically relevant. Do not use endpoints with no demonstrated ecological relevance to derive guideline values, and they do not need their quality assessed.	Yes (3), No (0/FAIL)
3	Was the measure of toxicity reported (e.g. NEC, ECx, NOEC) **and** its associated biological-effect size reported or quantifiable from the data (e.g. EC10, LC50, 15% effect at NOEC)?	Yes (3), No (0/FAIL)
4	Were appropriate controls (e.g. no-toxicant control, solvent control) used?	Yes (3), No (0/FAIL)[b]
5	How may treatment concentrations were used (in addition to the control)?	≥ 9 (3), 6–8 (2), 3–5 (1), < 3 (0/FAIL)[b]
6	What was the test concentration spacing? Note: a spacing of < 3.2 is highly preferred, while a spacing of ≥ 10 is too great.	≤ 3.2 (3), 3.3–9 (1), ≥ 10 (0/FAIL)[b]
7	Was each control and chemical concentration at least duplicated? Note: for concentration-response modelling, it is also acceptable to have many concentrations (e.g. > 15) without replication (assuming the controls are replicated).	Yes (3), No (0/FAIL)
8	Were the characteristics of the test organism (e.g. length, mass, age) stated?	Yes (3), No (0)
9	Was the type of test medium used stated (e.g. synthetic or natural water? If synthetic, to what recipe? If natural, what source? Filtered or unfiltered?)?	Yes (3), No (0)
10	Was the type of exposure (e.g. static, flow-through) stated?	Yes (3), No (0)
	Maximum sub-total for Part A = 30	
B. Test performance/results		
11	Were analytical reagent-grade chemicals or the highest possible purity chemicals used for the experiment?	Yes (3), No/not stated (0)
12 (A)	Were test solutions, blanks and/or controls tested for common contamination (e.g. elevated naturally occurring substances, such as nutrients, metals, metalloids) or other suspect contaminants?	Yes (2), No/not stated (0)
12 (B)	If so, were any significant contamination issues identified?	No (2), Yes (0/FAIL)
13	Were the contaminant concentrations measured at the beginning and end of the exposure? Note: normally, toxicity data calculated using nominal-concentration data would not be used to derive guideline values; however, professional judgement can be used to include such data, provided a justification for their use is provided.	Yes (8), Measured once (4), Not measured or stated (0/FAIL)[b]
14 (A)	Were test-acceptability criteria stated (e.g. mortality in controls must not exceed a certain percentage) or inferred (e.g. test methods used were United States Environmental Protection Agency or Organisation for Economic Co-operation and Development)?	Stated (2), Inferred (1), Neither (0)
14 (B)	If so, were test-acceptability criteria met? Note: data that fail the acceptability criteria are automatically deemed to be of unacceptable quality and must not be used.	Yes (6), Not reported (2), No (0/FAIL)
15 (A)	Were parallel reference-toxicant toxicity tests conducted?	Yes (2), No (0)
15 (B)	If so, did the reference toxicity results fall within the acceptable limits?	Yes (6), Not reported (2), No (0/FAIL)

No.	Question	Score[a]
16	Was there a concentration-response (C-R) relationship, either observable or stated (i.e. was there a larger response in the endpoint of interest at higher treatment concentrations than at lower treatment concentrations?)? Note: a full C-R relationship is one that includes zero, partial and full effects, while a partial C-R relationship is one that is missing either zero, partial or full effects. Professional judgement might be required to make decisions associated with C-R relationships that exhibit characteristics such as non-monotonicity and low-dose effects.	Yes – full (8), Yes – partial (4), No – (0/FAIL)
17 (A)	Was the statistical method or model used to determine the toxicity estimate stated **and** appropriate (e.g. refer to Green et al. 2018, noting that new methods will/have emerged that may not be captured in past publications)?	Yes (6), No (0)
17 (B)	For point estimates from concentration-response modelling (e.g. LC/EC/NEC/BEC/NSEC), was an estimate of variability provided? For estimates based on hypothesis testing (e.g. NOEC/LOEC/MATC), was the significance level 0.05 or less?	Yes (2), No (0)
	Maximum sub-total for Part B = 47	
C. Test water chemistry/test conditions		
18	Were the following parameters measured and reported?	
18.1	pH – was pH measured at least at the beginning and end of the toxicity test?	Yes (3), No (0/FAIL)
18.2	Dissolved organic carbon (for metals/non-metallic inorganics only)	Yes (3), No (0/FAIL)
18.3	Salinity/conductivity	Yes (3), No (0/FAIL)
18.4	Dissolved oxygen (DO) – was DO measured at least at the beginning and end of the toxicity test?	Yes (2), No (0/FAIL)
18.5	Temperature	Yes (2), Temperature of room or chamber was reported (1), No (0/FAIL)
18.6	Any other parameters of relevance (e.g. other toxicity-modifying factors)?	Yes (2), No (0/FAIL)
	Maximum sub-total for Part C = 15	
	Maximum total for parts A + B + C = 92	

Source: Updated from Warne et al. (2018).

[a] Where a question has a possible 'No (0/FAIL)' outcome, professional judgement may be required to determine if a 0 is assigned or if the test fails the quality assessment process.

[b] See
Table 2 for additional guidance.

Table B.5 Scoring system for assessing the quality of toxicity data for contaminants (metal or non-metal) to marine and estuarine plant species, to be used in the derivation of guideline values for toxicants

No.	Question	Score[a]
A. General test design information		
1	Was the duration of the exposure stated (e.g. 48 h, 96 h)?	Yes (3), No (0/FAIL)
2	Was the biological endpoint (e.g. immobilisation, population growth) stated? Note: ensure that the endpoint is considered to be ecologically relevant. Do not use endpoints with no demonstrated ecological relevance to derive guideline values, and they do not need their quality assessed.	Yes (3), No (0/FAIL)
3	Was the measure of toxicity reported (e.g. NEC, ECx, NOEC) **and** its associated biological-effect size reported or quantifiable from the data (e.g. EC10, LC50, 15% effect at NOEC)?	Yes (3), No (0/FAIL)
4	Were appropriate controls (e.g. no-toxicant control, solvent control) used?	Yes (3), No (0/FAIL)[b]
5	How may treatment concentrations were used (in addition to the control)?	≥ 9 (3), 6–8 (2), 3–5 (1), < 3 (0/FAIL)[b]
6	What was the test concentration spacing? Note: a spacing of < 3.2 is highly preferred, while a spacing of ≥ 10 is too great.	≤ 3.2 (3), 3.3–9 (1), ≥ 10 (0/FAIL)[b]
7	Was each control and chemical concentration at least duplicated? Note: for concentration-response modelling, it is also acceptable to have many concentrations (e.g. > 15) without replication (assuming the controls are replicated).	Yes (3), No (0/FAIL)
8	Were the characteristics of the test organism (e.g. length, mass, age) stated?	Yes (3), No (0)
9	Was the type of test medium used stated (e.g. synthetic or natural water? If synthetic, to what recipe? If natural, what source? Filtered or unfiltered?)?	Yes (3), No (0)
10	Was the type of exposure (e.g. static, flow-through) stated?	Yes (3), No (0)
	Maximum sub-total for Part A = 30	
B. Test performance/results		
11	Were analytical reagent-grade chemicals or the highest possible purity chemicals used for the experiment?	Yes (3), No/not stated (0)
12 (A)	Were test solutions, blanks and/or controls tested for common contamination (e.g. elevated naturally occurring substances, such as nutrients, metals, metalloids) or other suspect contaminants?	Yes (2), No/not stated (0)
12 (B)	If so, were any significant contamination issues identified?	No (2), Yes (0/FAIL)
13	Were the contaminant concentrations measured at the beginning and end of the exposure? Note: normally, toxicity data calculated using nominal-concentration data would not be used to derive guideline values; however, professional judgement can be used to include such data, provided a justification for their use is provided.	Yes (8), Measured once (4), Not measured or stated (0/FAIL)[b]
14 (A)	Were test-acceptability criteria stated (e.g. mortality in controls must not exceed a certain percentage) or inferred (e.g. test methods used were United States Environmental Protection Agency or Organisation for Economic Co-operation and Development)?	Stated (2), Inferred (1), Neither (0)
14 (B)	If so, were test-acceptability criteria met? Note: data that fail the acceptability criteria are automatically deemed to be of unacceptable quality and must not be used.	Yes (6), Not reported (2), No (0/FAIL)
15 (A)	Were parallel reference-toxicant toxicity tests conducted?	Yes (2), No (0)
15 (B)	If so, did the reference toxicity results fall within the acceptable limits?	Yes (6), Not reported (2), No (0/FAIL)

No.	Question	Score[a]
16	Was there a concentration-response (C-R) relationship, either observable or stated (i.e. was there a larger response in the endpoint of interest at higher treatment concentrations than at lower treatment concentrations?)? Note: a full C-R relationship is one that includes zero, partial and full effects, while a partial C-R relationship is one that is missing either zero, partial or full effects. Professional judgement might be required to make decisions associated with C-R relationships that exhibit characteristics such as non-monotonicity and low-dose effects.	Yes – full (8), Yes – partial (4), No – (0/FAIL)
17 (A)	Was the statistical method or model used to determine the toxicity estimate stated **and** appropriate (e.g. refer to Green et al. 2018, noting that new methods will/have emerged that may not be captured in past publications)?	Yes (6), No (0)
17 (B)	For point estimates from concentration-response modelling (e.g. LC/EC/NEC/BEC/NSEC), was an estimate of variability provided? For estimates based on hypothesis testing (e.g. NOEC/LOEC/MATC), was the significance level 0.05 or less?	Yes (2), No (0)
	Maximum sub-total for Part B = 47	
C. Test water chemistry/test conditions		
18	Were the following parameters measured and reported?	
18.1	pH – was pH measured at least at the beginning and end of the toxicity test?	Yes (3), No (0/FAIL)
18.2	Salinity/conductivity	Yes (3), No (0/FAIL)
18.3	Dissolved organic carbon (for metals/non-metallic inorganics only)	Yes (3), No (0/FAIL)
18.4	Temperature	Yes (2), Temperature of room or chamber was reported (1), No (0/FAIL)
18.5	Light intensity/illuminance	Yes (2), No (0/FAIL)
18.6	Any other parameters of relevance (e.g. other toxicity-modifying factors)?	Yes (2), No (0/FAIL)
	Maximum sub-total for Part C = 15	
	Maximum total for parts A + B + C = 92	

Source: Updated from Warne et al. (2018).

[a] Where a question has a possible 'No (0/FAIL)' outcome, professional judgement may be required to determine if a 0 is assigned or if the test fails the quality-assessment process.

[b] See
Table 2 for additional guidance.

Appendix C. Examples of data assessment decisions relating to bimodality and the log-normal–log-normal distribution

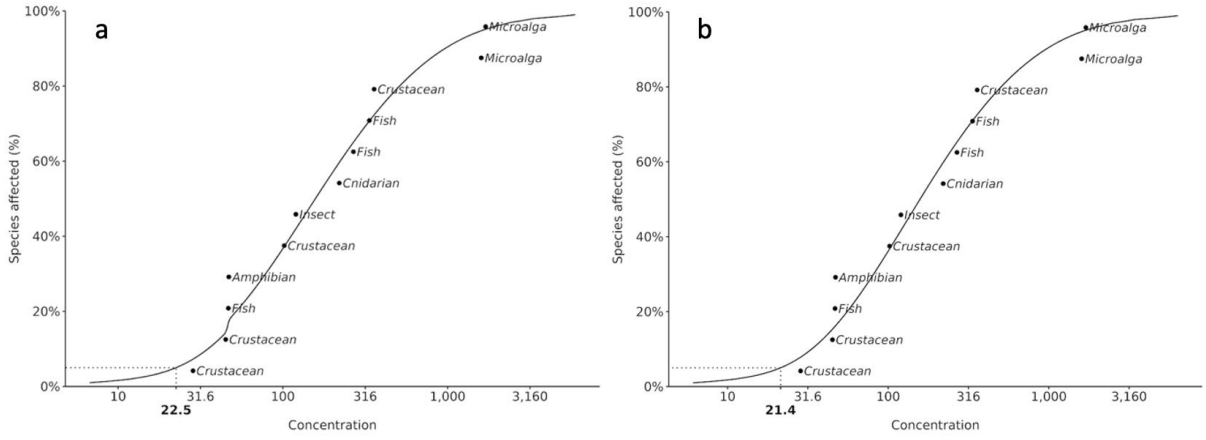
This appendix provides a number of examples of decision-making processes for datasets where the log-normal–log-normal (l-norm–l-norm) distribution in (shiny)ssdtools is having an influence, from small to dominant, on the model-averaged SSD and associated protective concentration (PC) values. The examples represent hypothetical datasets based on situations that have been encountered for real datasets. All datasets were analysed using (shiny)ssdtools v 0.3.1 (which uses ssdtools v 2.1.0) using the full set of default models. For the purpose of simplicity, only the PC95 values are considered in the examples. All units are assumed to be in μg/L. It is likely that issues that are not captured in the examples provided below will arise for other datasets; in many cases, some degree of professional judgement will be required, and the associated decision-making processes will need to be defensible and documented.

Dataset A – non-bimodal dataset, low weight for l-norm–l-norm

Dataset A contained 12 chronic values from 6 taxonomic groups (Figure C.1a). The l-norm–l-norm was assigned a relatively low weight (0.11), although its influence on the SSD is evident in the small deviation around the 20th percentile. The PC95 was 22.5, compared to 21.4 if the l-norm–l-norm was removed from the default model set (Figure C.1b). The toxicant's mode of action is not specific to any subset of taxonomic groups, and the dataset was determined to not be bimodal.

Given the low weight of the l-norm–l-norm and its minor effect on the PC95, there is no need to consider measures to improve the modelling (i.e. by improving the dataset or removing the l-norm–l-norm from the analysis).

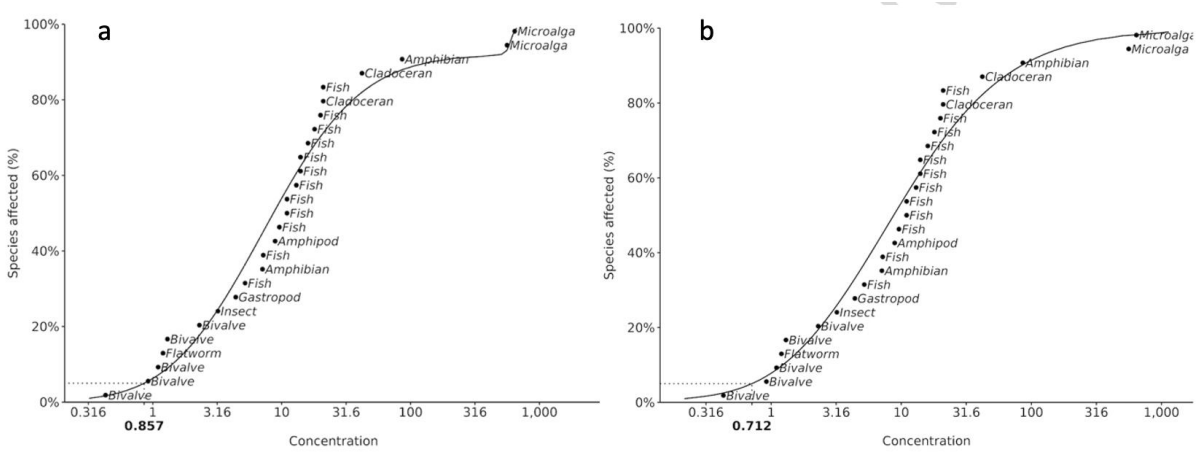
Figure C.1 Species sensitivity distribution for dataset A (n = 12) (a) using the full default model set and (b) without l-norm–l-norm in the model set.



Dataset B – bimodal dataset, high weight for l-norm–l-norm

Dataset B contained 27 chronic values from 8 taxonomic groups (Figure C.2a). The l-norm–l-norm was assigned a relatively high weight (0.661), and its influence on the SSD is most evident in the deviation at the upper end of the distribution. The PC95 was 0.86 compared to 0.71 if the l-norm–l-norm was removed from the default model set (Figure C.2b). The toxicant is also a nutrient for plants and algae. It is typically less toxic to plants and algae than to animals and, therefore, a dataset that contains data for both animals and plants/algae could be expected to be bimodal on this basis. It is somewhat surprising that this bimodality was reflected in the SSD when there were only 2 algal values compared to 25 animal values; nevertheless, the fit of the SSD, and the associated influence of the l-norm–l-norm, is consistent with what would be expected for this toxicant. Consequently, there is no need to further assess the data nor consider removal of the l-norm–l-norm distribution, and the dataset and its SSD would be considered acceptable for deriving the GVs.

Figure C.2 Species sensitivity distribution for dataset B (n = 27) (a) using the full default model set and (b) without l-norm–l-norm in the model set

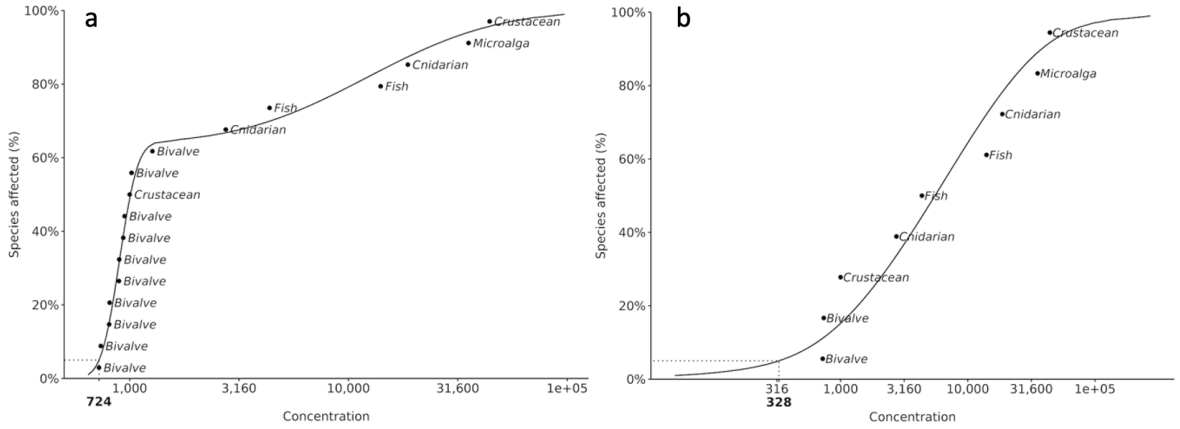


Dataset C – non-bimodal dataset, high weight for l-norm–l-norm

Dataset C(a) contained 17 chronic values from 5 taxonomic groups (Figure C.3a). The dataset contained a cluster of bivalve data, all from the same study, that show very similar sensitivity at the lower end of the SSD. The bivalve data comprise ~60% of the dataset, which is a significant over-representation for this taxonomic group. Although it could be argued that the dataset is bimodal due to bivalves being more sensitive, this could equally be an artefact due to the bivalve results being from a single study that produced very sensitive results. When combined with the fact that the toxicant does not have a specific mode of action, it is not possible to conclude that the dataset is bimodal. However, the l-norm–l-norm was assigned a very high weight (0.999), and its major influence on the SSD is evident in its dramatic change of slope at around the 60th percentile. It is evident that the l-norm–l-norm is modelling the data as if the dataset is bimodal. Although bivalves might be relatively sensitive, the over-representation of data for this group (from a single study) is disproportionately affecting the SSD. To improve the taxonomic representation across the dataset, an upper limit to the proportion of a dataset that any one taxonomic group can comprise was considered. In this case, the bivalve data were culled to ensure that they did not represent more than ~20% of the resulting dataset. The lowest bivalve values were selected when culling. This resulted in removing 8 of the 10 bivalve values, and the final dataset – dataset C(b) – comprised 9 values from 5 taxonomic groups. The SSD for this dataset is shown in Figure C.3b. Note that professional judgement

will be required for decisions on aspects such as the maximum proportion for any single taxonomic group and whether the lowest values, or a mix of values from across the range of values, for the taxonomic group are selected. In this example, a maximum taxonomic representation of 20% was selected and the lowest bivalve values retained.

Figure C.3 Species sensitivity distribution for dataset C (a) original dataset (n = 17) and (b) dataset modified to reduce the taxonomic representation of bivalves (n = 9)

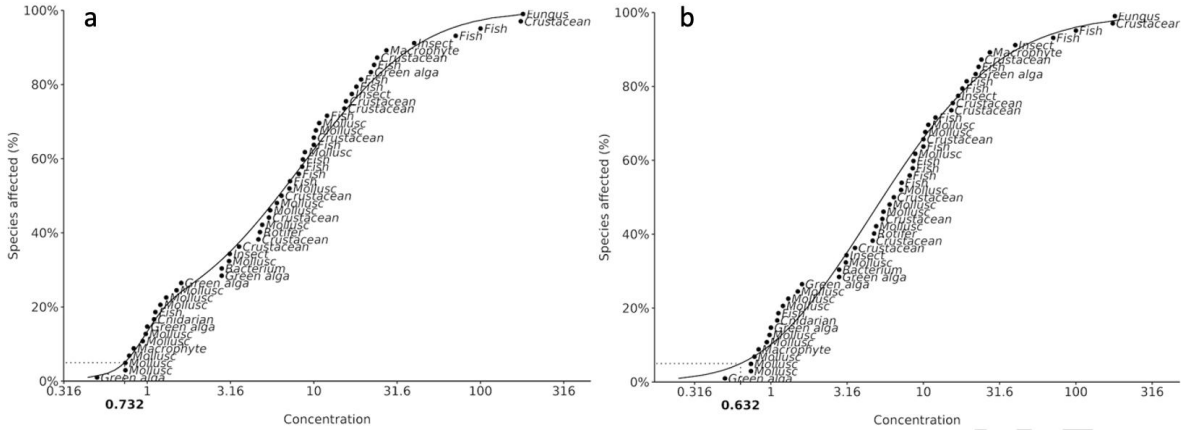


For dataset C(b), the l-norm–l-norm weight for the SSD was zero, probably as a result of both the smaller sample size and the removal of the tight (steep) cluster of bivalve data from the SSD. The resulting PC95 value was 328, compared to a markedly higher 724 for the original dataset. The fit of the SSD for dataset C(b) (Figure C.3b) would probably be classified as poor, resulting in GVs of moderate reliability as per Section 3.8 (i.e. between 8 and 14 chronic data with a poor fit of the SSD). Although the dataset was reduced by 8 values, from 17 to 9, the use of GVs based on the full dataset that was over-represented by bivalve data and for which the l-norm–l-norm was highly weighted, was not defensible. Notably, the SSD for dataset C(a) (Figure C.3a), while closely fitting the data, could not realistically be classified as being good, because the l-norm–l-norm was modelling the data artefact associated with the over-represented bivalve data.

Dataset D – non-bimodal dataset, high weight for l-norm–l-norm

Dataset D contained 51 chronic values from 10 taxonomic groups (Figure C.4a). The l-norm–l-norm was assigned a relatively high weight (0.753), and its influence on the SSD is evident in the small deviation at around the 30th percentile, where the SSD joins the lower and upper subsets of data. The PC95 was 0.73, compared to 0.63 if the l-norm–l-norm was removed from the default model set (Figure C.4b). The toxicant's mode of action is not specific to any subset of taxonomic groups, and the dataset was determined to not be bimodal.

Figure C.4 Species sensitivity distribution for dataset D (n = 51) (a) using the full default model set and (b) without l-norm–l-norm in the model set



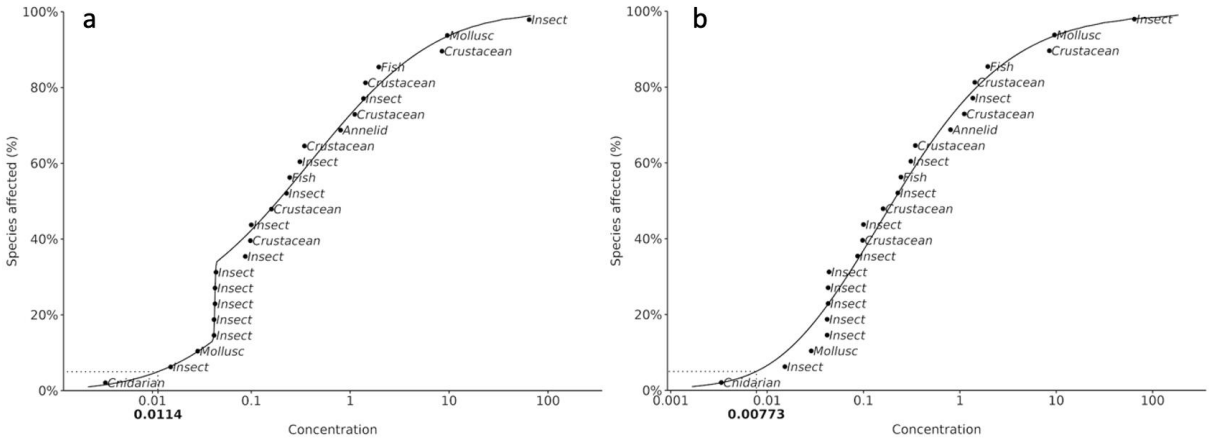
This dataset appears to be potentially bimodal, due to a gap in the data between 2 green algae at around 1.8–2.6. However, this gap is small and is almost certainly due to a lack of values in the literature within this range, not to any reasons related to mode of action. Additional analysis of the dataset revealed that it was not bimodal, and that the l-norm–l-norm was assigned a high weight because of the large difference in variance between the upper and lower subsets of data (the lower subset having lower variance than the upper subset).

Ultimately, there was no justifiable reason to remove the l-norm–l-norm distribution from the model set, and the dataset and SSD as shown were retained for deriving the GVs.

Dataset E – non-bimodal dataset, high weight for l-norm–l-norm

Dataset E(a) contained 24 values from 7 taxonomic groups (Figure C.5a). The values are a mix of chronic and converted-acute values. The l-norm–l-norm was assigned a relatively high weight (0.998), and its major influence on the SSD is evident in the almost vertical section of the SSD between around the 10th and 30th percentiles, representing a very tight cluster of values at a concentration of ~0.04. The PC95 was 0.011, compared to 0.008 if the l-norm–l-norm was removed from the default model set – dataset E(b) (see Figure C.5b). The toxicant's mode of action is not specific to any subset of taxonomic groups, and the dataset was determined to not be bimodal.

Figure C.5 Species sensitivity distribution for dataset E (n = 24) (a) using the full default model set and (b) without l-norm–l-norm in the model set



The almost-vertical cluster of 5 values between around the 10th and 30th percentiles corresponds to 4 converted-acute EC50s for insects from the same study and an additional converted-acute EC50 for an insect from another source (ranging from 0.042 to 0.044). A similar issue can sometimes arise for NOEC data from a single study, where the same NOEC is reported for multiple species (because the same concentration ranges were used for all species). The stacking of the converted EC50 values is an artefact of the dataset and is not related in any way to bimodality or taxa-specific sensitivity; yet the l-norm–l-norm closely models this subset of the data, resulting in a very atypical SSD.

Consequently, the dataset was assessed for any ability to improve it. Although acute EC50s are not considered to be preferred data, it was not viable to remove all the acute values from the dataset because 19 of the 24 values were acute EC50s, and this would have resulted in insufficient remaining values to derive GVs using the SSD method. Removing only the subset of 5 values was also not defensible, as all values scored in excess of 70% in the quality assessment and appeared reliable. Moreover, 4 other species tested in the study all had relatively different EC50s, indicating that there was not an inherent, systematic flaw in the study itself. It was concluded that the available dataset could not be further improved.

Consequently, the next step was to consider whether removal of the l-norm–l-norm from the default model set was justified. Figure C.5b shows the SSD based on the model set without the l-norm–l-norm. Overall, the SSD fits the dataset relatively well, although there is a generally poor fit at the lower end of the SSD. The resulting GVs would be classified as either moderate reliability (if the SSD was deemed to be a good fit) or low reliability (if the SSD was deemed to be a poor fit). As with dataset C (Figure C.5a, the SSD for dataset E(a) (Figure C.5a), while closely fitting the data, could not realistically be classified as being good, because the l-norm–l-norm was modelling the data artefact associated with the stacked subset of data.

Glossary

Term	Definition
Acute toxicity	A lethal or adverse sub-lethal effect that occurs as a result of a short (relative to the organism's life span) exposure to a chemical. Refer to Warne et al. (2018) for examples of acute exposures
Acute-to-chronic ratio (ACR)	The species' mean acute value (LC/EC50) divided by the chronic value (e.g. NOEC or EC10) for the same species
Active ingredient (AI)	The chemical(s) that is biologically active and helps a pesticide achieve its objective
Alga	Chlorophyll-bearing plants, most of which are aquatic. These can be microscopic in size and can be single-celled (such as microalgae) or multi-cellular such as macroalgae/seaweeds)
Amphipod	Small crustaceans (typically < 10 mm) found in most aquatic environments
Aquatic ecosystem	Any water environment in which plants and animals interact with the chemical and physical features of the environment
Assessment factor (AF)	A unitless number applied to the lowest toxicity value for a chemical to derive a concentration that should not cause adverse environmental effects. The size of the AF varies with the type of data. Also called 'application factor' or 'safety factor' (used when there are insufficient data to use the species sensitivity distribution method)
BEC10	Bounded-effect concentration in a toxicity test that is the highest tested concentration that has an upper 95% confidence interval that causes less than a 10% effect
Benthic	Organisms living in or on the sediments of aquatic habitats
Bioaccumulation	The process by which chemical substances are accumulated by aquatic organisms by all routes of exposures (dietary and the ambient environment)
Bioaccumulation factor (BAF)	The ratio of the concentration of a chemical in tissue of an organism to the concentration in either the surrounding media (water for aquatic organisms, soil for terrestrial organisms) or in the food they consume, once equilibrium is reached
Bioavailable	Able to be taken up by organisms
Bioconcentration	Processes by which chemicals accumulate in the tissues of living organisms from the surrounding media (e.g. soil, sediment or water)
Bioconcentration factor (BCF)	The ratio of the concentration of a contaminant in an organism to its concentration in the ambient water (or sediment) at a steady state. It can be expressed on the basis of wet weight, dry weight or lipid weight
Biomagnification	The processes by which tissue concentrations of chemicals increase as the chemical passes up through 2 or more trophic levels in a food chain. The term implies an efficient transfer of chemicals from food to consumer, so that the residue concentrations increase systematically from one trophic level to the next
Bimodality coefficient	A statistical measure of whether a set of data has one or 2 modes. In this report, it is used to determine if various groups of organisms have different sensitivities to a chemical or whether fresh and marine organisms have different sensitivities to a chemical
Biotic ligand model (BLM)	Models that consider the effect of water parameters, including water hardness, pH, and dissolved organic carbon, to model the bioavailability of metals. The ligand is the gill surface
BurrliOZ and Burrlioz	A species sensitivity distribution software package developed and used in the 2000 Guidelines to derive guideline values (previously termed 'trigger values') to protect aquatic ecosystems. BurrliOZ denotes the original version of the software, while Burrlioz and Burrlioz 2.0 denote the second, improved version of the software

Term	Definition
Burr Type III	A flexible family of parametric distributions for non-negative data
Chronic toxicity	A lethal or sub-lethal adverse effect that occurs as the result of exposure to a chemical for a period that is a substantial portion of the organism's life span or an adverse sub-lethal effect on a sensitive early life stage. Refer to Warne et al. (2018) for examples of chronic exposures
Chronic estimated toxicity value(s)/data	Chronic LC50, IC50, EC50, LOEC and MATC values that have been converted to estimates of chronic NOEC/EC10 data
Community	An assemblage of organisms characterised by a distinctive combination of species occupying a common environment and interacting with one another
Concentration	The quantifiable amount of a substance in water, biota, soil or sediment
Contaminants	Biological or chemical substances or entities, not normally present in a system, capable of producing an adverse effect in a biological system, seriously injuring structure or function
Control	Part of an experimental procedure that is ideally exactly like the treated part except that it is not subject to the test treatment. It is used as a standard of comparison, to check that the outcome of the experiment is a reflection of the test conditions and not of some unknown general factor
Copepod	A small crustacean found in marine and freshwater habitats; many are planktonic (living within the water column), but more are benthic (living on or in the sediments)
Converted-acute value(s)/data	Acute toxicity data that have been converted using experimentally derived or default acute-to-chronic ratios
Default guideline value (DGV)	A guideline value recommended for generic application in the absence of a more specific guideline value (e.g. a site-specific value) in the *Australian and New Zealand Guidelines for Fresh and Marine Water Quality*. Formerly known as 'trigger values'.
DOC	Dissolved organic carbon
Ecotoxicology	The science dealing with the adverse effects of chemicals, physical agents and natural products on populations and communities of living organisms
EC50	The toxicant concentration that is expected to cause one or more specified effects in 50% of a group of organisms or a 50% effect under specified conditions
ECx	The concentration of a substance in water or sediment that is estimated to produce an x% change in the response being measured or a certain effect in x% of the test organisms, under specified conditions
Estuarine water	For the purposes of deriving default guideline values, estuarine water is defined as any waters with a salinity between 0.5‰ and 25‰
Formulation	A commercial product that is manufactured for sale; for example, pesticides. Formulations typically contain one or more active ingredients that are added to products to 'improve the storage, handling, safety, application or effectiveness of the product' (Ware 1994). As such, formulations are not technical materials (see glossary entry for 'Technical material'). The same active ingredient may be found in multiple formulations. Examples of different formulations of pesticides include emulsifiable concentrates, wettable powders, water-dispersible granules and powders
Freshwater (FW)	For the purposes of deriving default guideline values, freshwater is defined as any waters with a salinity of < 0.5‰
Guideline value (GV)	A measurable quantity (e.g. concentration) or condition of an indicator for a specific community value below which (or above which, in the case of stressors such as pH, dissolved oxygen and many biodiversity responses) there is considered to be a low risk of unacceptable effects occurring to that community value. Guideline values for more than one indicator should be used simultaneously in a multiple lines of evidence approach. Also refer to 'default guideline value' and 'site-specific guideline value'

Term	Definition
Hardness-modified guideline value	These are guideline values for chemicals whose toxicity is affected by water hardness. Such guideline values are reported at a hardness of 30mg/L $CaCO_3$ but can be adjusted to the hardness in the water body being examined
IC50	A toxicant concentration that would cause a 50% reduction in a non-quantal measurement such as fecundity or growth
ICx	The concentration of a substance in water or sediment that is estimated to produce an x% inhibition of the response being measured in test organisms relative to the control response, under specified conditions
Indicator	Measurement parameter or combination of parameters that can be used to assess the quality of water
Invertebrate	An animal lacking a notochord or backbone
LC50	The toxicant concentration that is expected to be lethal to 50% of a group of organisms under specified conditions
LCx	The concentration of a substance in water or sediment that is estimated to be lethal to x% of a group of test organisms under specified conditions
Level of protection	The acceptable level of change from a defined reference condition
Lowest-observed-effect concentration (LOEC)	The lowest concentration of a material used in a toxicity test that has a statistically significant ($p \leq 0.05$) adverse effect on the exposed population of test organisms as compared with the controls. All higher concentrations should also cause statistically significant effects
Macroinvertebrate	Macroinvertebrates are invertebrate species where full-grown adults are ≥ 2 mm long. Examples of invertebrates that meet this criterion are decapods, echinoderms, molluscs, annelids, corals, amphipods, larger cladocerans (e.g. *Daphnia magna*, *Daphnia carinata* and *Daphnia pulex*, and insect species where larvae are ≥ 2 mm long
Marine water	For the purposes of deriving default guideline values, marine water is defined as any waters with a salinity between 25‰ and 36‰
Maximum allowable toxicant concentration (MATC)	The geometric mean of the lowest exposure concentration that causes a statistically significant adverse effect (LOEC) and the highest exposure concentration where no statistically significant effect is observed (NOEC) in a chronic test
Measured concentration	The concentration (mass per unit volume) of a chemical in a test solution as determined by chemical analysis
Measurement parameter	Any parameter or variable that is measured
Mechanism of action	A detailed description of how a toxicant exerts its toxicity at a molecular level. Typically, this is not known for individual species in ecotoxicology and therefore chemicals are usually classified by their mode of action
Mesocosm	Large enclosures designed to mimic field exposure conditions, taking the form of larger tanks, enclosures or artificial channels to mimic streams. They are often, but not necessarily, located in or near waterbodies
Microcosm	A laboratory-based bench-scale artificial ecosystem
Microinvertebrate	Microinvertebrates are invertebrate species where full-grown adults are typically < 2 mm long. Examples of invertebrates that meet this criterion are some cladocerans (e.g. *Ceriodaphnia dubia* and *Moina australiensis*), copepods, conchostracans, rotifer, acari, bryozoa and hydra
Mode of action	A description of how a toxicant exerts its toxicity at a sub-cellular level. This term is used far more frequently in ecotoxicology than the more detailed mechanism of action
Multiple linear regression equations (MLR equations)	Regression equations where the y parameter is modelled using 2 or more x parameters

Term	Definition
Multi-generation toxicity test	A toxicity test that exposes more than one generation of a test organism sequentially to the test chemical. The duration of multi-generation tests depends on the generation time of the test organism. For example, a typical 72-hour test for a microalga is a multi-generation test. However, a multi-generation test on a fish species could take months to years. Typically, a multi-generation test will be of shorter duration than the average life span of the test organism. Multi-generation toxicity tests can only generate chronic toxicity data
Negligible effect	Corresponding to an estimated no-effect or acceptable-effect level; negligible-effect concentrations typically include the NEC, NSEC, BEC10, ECx/ICx where $x \leq 10$, and NOEC, but if necessary, can also include the ECx/ICx where x is > 10 and ≤ 20
No-effect concentration (NEC)	The maximum concentration of a toxicant that causes no adverse effect in a target organism, based on a threshold parameter in a concentration–response model
No-observed-effect concentration (NOEC)	The highest concentration of a toxicant used in a toxicity test that does not have a statistically significant ($p \leq 0.05$) adverse effect on the exposed population of test organisms as compared with the controls
Nominal concentration	The quantity of a chemical added to a unit volume of test media. This concentration has not been confirmed by analytical measurement. It is the opposite of a measured concentration (see glossary entry for 'Measured concentration')
No-significant-effect concentration (NSEC)	The highest concentration for which the difference between the predicted response at that concentration is statistically insignificant from the predicted response at zero concentration, as estimated from a monotonically decreasing model (see Fisher and Fox 2023)
Octanol–water partition coefficient (K_{OW})	The ratio of the concentration of a chemical dissolved in octanol to that dissolved in water once equilibrium is reached. This is usually expressed as a logarithm to the base 10 (i.e. log-K_{OW}). High K_{OW} values (e.g. log-K_{OW} values between 4 and 7.5) indicate a high solubility in lipids (fats in tissue) and high bioaccumulation potential
Organism	Any living animal or plant
Persistent, bioaccumulative and toxic substances (PBT)	Substances that persist in the environment, bioaccumulate in organisms and cause toxic effects to humans or organisms from long-term exposure. The exact definitions vary in different jurisdictions
PC	Protective concentration. A PC95 is the concentration that should protect 95% of species
Pesticide	Substance or mixture of substances used to kill unwanted species of plants or animals
pH	The intensity of the acidic or basic character of a solution, defined as the negative logarithm of the hydrogen-ion concentration of a solution
Phylum	A taxonomic rank below kingdom and above class
Quality assurance	The implementation of checks on the success of quality control (e.g. replicate samples, analysis of samples of known concentration)
Quality control	The implementation of procedures to maximise the integrity of monitoring data (e.g. cleaning procedures, contamination avoidance, sample preservation methods)
Quantitative structure-activity relationship (QSAR)	A relationship between biological activity (e.g. toxicity) and a physicochemical property or measure of the structure of a group of related chemicals. These are used to predict the activity of chemicals that belong to the same group of chemicals as those used to develop the relationship but for which toxicity data do not exist. Each QSAR can only predict the toxicity to a single species
Reference toxicant	A reference chemical (toxicant) used in a toxicity test to assess the sensitivity of a test organism and to demonstrate the repeatability of a test and the laboratory's ability to perform the test consistently

Term	Definition
Reference condition	An environmental quality or condition that is defined from as many similar systems as possible (including historical data) and used as a benchmark for determining the environmental quality or condition to be achieved or maintained in a particular system of equivalent type
Risk	Typically defined by the joint interaction of both the likelihood and consequence of an event having a negative or adverse impact. Estimates of risk may be expressed in absolute or relative terms. Absolute risk is the excess risk due to exposure. Relative risk is the ratio of the risk in the exposed population to the risk in the unexposed population
Salinity	The presence of soluble salts in water or soils
Sediment	Unconsolidated mineral and organic particulate material that has settled to the bottom of aquatic environments
Single-generation toxicity test	A toxicity test that exposes a single generation of the test organism to the test chemical. Single-generation toxicity tests can generate either acute or chronic toxicity data, depending on the duration of the exposure (refer to glossary entries for 'Acute toxicity' and 'Chronic toxicity')
Site-specific guideline value	A guideline value that is relevant to the specific location or conditions that are the focus of a given assessment or issue
Species	A group of organisms that resemble each other to a greater degree than members of other groups and that form a reproductively isolated group that will not normally breed with members of another group. (Chemical species are differing compounds of an element.)
Standard error (SE)	A statistic that reveals how accurately sample data represents the whole population. It is calculated as the standard deviation of the sample data divided by the square root of the sample size
Stressors	The physical, chemical or biological factors that can cause an adverse effect on an aquatic ecosystem, as measured by the condition indicators
Species sensitivity distribution (SSD)	A method that plots the cumulative frequency of species' sensitivities to a toxicant and fits a statistical distribution to the data. From the distribution, the concentration that should theoretically protect a selected percentage of species can be determined
Sub-lethal	Involving an adverse effect below the level that causes death
Taxon (taxa)	Any group of organisms considered sufficiently distinct from other such groups to be treated as a separate unit (e.g. species, genera, families)
Taxonomic group	Groups of taxa. For the purposes of deriving a guideline value, taxonomic groups are generally phyla
Technical material	A form of a pesticide that consists of the active ingredient plus impurities associated with the manufacture of the active ingredient but that is free of other 'extraneous matter and added modifying agents, except stabilisers if required' (APVMA 2014). Technical materials are not formulations
Toxicant	A chemical capable of producing an adverse response (effect) in a biological system, seriously injuring structure or function or producing death. Examples include pesticides and metals
Toxicity	The inherent potential or capacity of a material to cause adverse effects in a living organism
Toxicity-modifying factor (TMF)	Aspects of water chemistry that influence the bioavailability of a compound (e.g. pH water hardness, dissolved organic carbon)
Toxicity test	The means by which the toxicity of a chemical or other test material is determined. A toxicity test is used to measure the degree of response produced by exposure to a specific level of stimulus (or concentration of chemical) for a specified test period

Term	Definition
Trigger value	Term used in the 2000 Guidelines to define a guideline value (and which is broadly equivalent to 'default guideline value' used in ANZG 2018). It was defined as 'the concentrations (or loads) of the key performance indicators measured for the ecosystem, below which there exists a low risk that adverse biological (ecological) effects will occur. They indicate a risk of impact if exceeded and should 'trigger' some action, either further ecosystem specific investigations or implementation of management/remedial actions'
Trophic level	A notional stage in the 'food chain' that transfers matter and energy through a community; primary producers, herbivores, carnivores and decomposers each occupy a different trophic level
Uptake	A process by which materials are absorbed and incorporated into a living organism
Vertebrate	An animal having a backbone
Weight of evidence (WoE)	Describes the process to collect, analyse and evaluate a combination of different qualitative, semi-quantitative or quantitative lines of evidence to make an overall assessment of water or sediment quality and its associated management. Applying a weight-of-evidence process incorporates judgements about the quality, quantity, relevance and congruence of the data contained in the different lines of evidence.

References

ANZECC and ARMCANZ (Australian and New Zealand Environment and Conservation Council and Agriculture and Resource Management Council of Australia and New Zealand) (2000a) *Australian and New Zealand guidelines for fresh and marine water quality. Volume 1, The guidelines*, Australian and New Zealand Environment and Conservation Council and Agriculture and Resource Management Council of Australia and New Zealand, Canberra, accessed 24 December 2014.

ANZECC and ARMCANZ (2000b) *Australian and New Zealand guidelines for fresh and marine water quality. Volume 2, Aquatic ecosystems*, Australian and New Zealand Environment and Conservation Council and Agriculture and Resource Management Council of Australia and New Zealand, Canberra, accessed 24 December 2014.

ANZECC and ARMCANZ (2000c) *Australian and New Zealand guidelines for fresh and marine water quality. Volume 3, Primary industries*, Australian and New Zealand Environment and Conservation Council and Agriculture and Resource Management Council of Australia and New Zealand, Canberra, accessed 24 December 2014.

ANZECC and ARMCANZ (2000d) *National Water Quality Management Strategy, Paper No. 7, Australian guidelines for water quality monitoring and reporting*, Australian and New Zealand Environment and Conservation Council and Agriculture and Resource Management Council of Australia and New Zealand, Canberra, accessed 24 December 2014.

ANZECC and ARMCANZ (2000e) *Toxicant default guideline values for aquatic ecosystem protection: sulfides in freshwater and marine water*, Australian and New Zealand Guidelines for Fresh and Marine Water Quality, Australian and New Zealand Environment and Conservation Council and Agriculture and Resource Management Council of Australia and New Zealand. Reproduced in ANZG (2018) *Australian and New Zealand Guidelines for Fresh and Marine Water Quality*, Australian and New Zealand governments and Australian state and territory governments, Canberra, accessed 26 June 2024.

ANZG (Australian and New Zealand Guidelines) (2018) *Australian and New Zealand Guidelines for Fresh and Marine Water Quality*, Australian and New Zealand governments and Australian state and territory governments.

ANZG (2020) *Toxicant default guideline values for aquatic ecosystem protection: metolachlor in freshwater*, Australian and New Zealand Guidelines for Fresh and Marine Water Quality, Australian and New Zealand Governments and Australian state and territory governments, Canberra.

ANZG (2021) *Toxicant default guideline values for aquatic ecosystem protection: zinc in marine water*, Australian and New Zealand Guidelines for Fresh and Marine Water Quality, Australian and New Zealand Governments and Australian state and territory governments, Canberra.

ANZG (2023) *Toxicant default guideline values for aquatic ecosystem protection: dioxins in freshwater*, Australian and New Zealand Guidelines for Fresh and Marine Water Quality, Australian and New Zealand Governments and Australian state and territory governments, Canberra.

ANZG (2025a) *Toxicant default guideline values for aquatic ecosystem protection: chlorine in marine water*, Australian and New Zealand Guidelines for Fresh and Marine Water Quality, Australian and New Zealand governments and Australian state and territory governments, Canberra.

ANZG (2025b) *Toxicant default guideline values for aquatic ecosystem protection: ammonia in freshwater*, Australian and New Zealand Guidelines for Fresh and Marine Water Quality, Australian and New Zealand Governments and Australian state and territory governments, Canberra.

ANZG (2025c) *Toxicant default guideline values for aquatic ecosystem protection: nitrate in freshwater*, Australian and New Zealand Guidelines for Fresh and Marine Water Quality, Australian and New Zealand Governments and Australian state and territory governments, Canberra.

ANZG (2025d) *Toxicant default guideline values for aquatic ecosystem protection: simazine in marine water*, Australian and New Zealand Guidelines for Fresh and Marine Water Quality, Australian and New Zealand Governments and Australian state and territory governments, Canberra.

ANZG (2025e) *Toxicant default guideline values for aquatic ecosystem protection: aluminium in marine water*, Australian and New Zealand Guidelines for Fresh and Marine Water Quality, Australian and New Zealand Governments and Australian state and territory governments, Canberra.

APVMA (Australian Pesticides and Veterinary Medicines Authority) (2014) *Formulation types*, APVMA website, accessed 7 June 2018.

Australian Government (2018) *Charter: National Water Quality Management Strategy*, Department of Agriculture and Water Resources, Canberra.

Barry S and Henderson B (2014) *Burrlioz 2.0*, Commonwealth Scientific and Industrial Research Organisation website, Canberra, accessed 24 December 2014.

Batley GE, van Dam R, Warne MStJ, Chapman JC, Fox DR, Hickey CW and Stauber JL (2018) *Technical rationale for changes to the method for deriving Australian and New Zealand water quality guideline values for toxicants*, report prepared for the revision of the Australian and New Zealand guidelines for fresh and marine water quality, Australian and New Zealand governments and Australian state and territory governments.

Brix KV, DeForest DK, Tear L, Peijnenburg W, Peters A, Middleton ET and Erickson R (2020) 'Development of empirical bioavailability models for metals', *Environmental Toxicology and Chemistry*, 39:85–100, doi:10.1002/etc.4570.

Campbell E, Palmer MJ, Shao Q and Wilson D (2000) Burrlioz: a computer program for calculating toxicant trigger values for the ANZECC and ARMCANZ water quality guidelines, Perth.

CCME (Canadian Council of Ministers of the Environment) (2018) *Scientific criteria document for the development of the Canadian water quality guidelines for the protection of aquatic life: zinc*, CCME, National Guidelines and Standards Office, Water Policy and Coordination Directorate, Environment Canada, Winnipeg.

Chapman PM (2015) 'Including or excluding toxicity test data for development of a geometric mean', *Environmental Toxicology and Chemistry*, 34:1691–1692, doi:10.1002/etc.3012.

Cormier SM, Paul JF, Spehar RL, Shaw-Allen P, Berry WJ and Suter GW (2008) 'Using field data and weight of evidence to develop water quality criteria', *Integrated Environmental Assessment and Management*, 4:490–504, doi:10.1897/ieam_2008-018.1.

Cui L, Li X, Luo Y, Gao X, Wang Y, Lv X, Zhang H and Lei K (2023) 'A comprehensive review of the effects of salinity, dissolved organic carbon, pH, and temperature on copper biotoxicity: implications for setting the copper marine water quality criteria', *Science of the Total Environment*, 912:169587, doi:10.1016/j.scitotenv.2023.169587.

Dalgarno S (2018) *ssdtools: A shiny web app to analyse species sensitivity distributions*, prepared by Poisson Consulting for the Ministry of the Environment, British Columbia.

ECCC (Environment and Climate Change Canada) (2020) *Canadian Environmental Protection Act, 1999 – Federal Environmental Quality Guidelines – lead*, ECCC, Government of Canada, Ottawa.

ECCC (2022) *Canadian Environmental Protection Act, 1999 – Federal Environmental Quality Guidelines – aluminium*, ECCC, Government of Canada, Ottawa.

ECCC (2024) *Canadian Environmental Protection Act, 1999 – Federal Environmental Quality Guidelines – iron*, ECCC, Government of Canada, Ottawa.

Erickson RJ, Benoit DA and Mattson VR (1987) *A prototype toxicity factors model for site-specific copper water quality criteria*, United States Environmental Protection Agency, Environmental Research Laboratory, Duluth.

European Commission (2010) *Nickel and its compounds*, Environmental Quality Standards Sheet, prepared by Danish Environmental Protection Agency on behalf of the European Union, Copenhagen.

European Commission (2018) *Technical guidance for deriving environmental quality standards*, Guidance Document No. 27, updated version 2018, Common Implementation Strategy for the Water Framework Directive, European Commission, Brussels.

European Commission (2019) *Technical guidance for implementing Environmental Quality Standards (EQS) for metals: consideration of metal bioavailability and natural background concentrations in assessing compliance*, Guidance Document No. 38. Common Implementation Strategy for the Water Framework Directive, European Commission, Brussels.

Fisher R and Fox DR (2023) 'Introducing the no significant effect concentration', *Environmental Toxicology and Chemistry,* 42:2019–2028, doi:10.1002/etc.5610.

Fisher R, Fox DR, Negri AP, Dam JV, Flores F and Koppel D (2024) 'Methods for estimating no effect toxicity concentrations in ecotoxicology', *Integrated Environmental Assessment and Management*, 20:279–293, doi:10.1002/ieam.4809.

Fox DR (2009) 'A Bayesian approach for determining the no effect concentration and hazardous concentration in ecotoxicology', *Ecotoxicology and Environmental Safety,* 73:123–131, doi:10.1016/j.ecoenv.2009.09.012.

Fox DR and Batley GE (2022) 'Assessment factors in species sensitivity distributions for the derivation of guideline values for aquatic contaminants', *Environmental Chemistry*, 19:201–209, doi:10.1071/EN22061.

Fox DR and Billoir E (2011) 'Individual versus population effects in concentration-response modelling', *Integrated Environmental Assessment and Management*, 7(3):501–502, doi:10.1002/ieam.199.

Fox DR and Landis WG (2016) 'Don't be fooled-a no-observed-effect concentration is no substitute for a poor concentration-response experiment', *Environmental Toxicology and Chemistry*, 35:2141–2148, doi.org/10.1002/etc.3459.

Fox DR, van Dam RA, Fisher R, Batley GE, Tillmans AR, Thorley J, Schwarz CJ, Spry DJ and McTavish K (2021) 'Recent developments in species sensitivity distribution modeling', *Environmental Toxicolology and Chemistry*, 40:293–308, doi:10.1002/etc.4925.

Fox D, van Dam R, Batley G, Fisher R and Thorley J (2023) 'Improved SSD software coming soon for the ANZG Guidelines', *Australasian Bulletin of Ecotoxicology and Environmental Chemistry,* 9:1–3.

Fox DR, Fisher R and Thorley JL (2024a) *Final report of the joint investigation into SSD modelling and ssdtools implementation for the derivation of toxicant guidelines values in Australia and New Zealand*, report prepared for the Department of Climate Change, Energy, the Environment and Water, Environmetrics Australia, Beaumaris, and the Australian Institute of Marine Science, Perth, doi:10.25845/xtvt-yc51.

Fox DR, Fisher R and Thorley JL (2024b) Sample size investigations for uni-modal and bi-modal distributions in ssdtools. Addendum to: Final report of the joint investigation into SSD modelling and ssdtools implementation for the derivation of toxicant guidelines values in Australia and New Zealand, report prepared for the Department of Climate Change, Energy, the Environment and Water, Environmetrics Australia, Beaumaris, and the Australian Institute of Marine Science, Perth.

Freeman JB and Dale R (2013) 'Assessing bimodality to detect the presence of a dual cognitive process', *Behavior Research Methods,* 45:83–97, doi:10.3758/s13428-012-0225-x.

Gadd JG, Hickey CW and Snelder AH (2024) *Acute copper and zinc guidelines for Aotearoa: technical report of the derivation including bioavailability model evaluation*, report prepared for Ministry for the Environment, New Zealand.

Garman E, Meyer J, Bergeron C, Blewett T, Clements W, Elias M, Farley K, Gissi F and Ryan A (2020) 'Validation of bioavailability-based toxicity models for metals', *Environmental Toxicology and Chemistry*, 39:101–117, doi:10.1002/etc.4563.

Gissi F, Adams MS, King CK and Jolley DF (2015) 'A robust bioassay to assess the toxicity of metals to the Antarctic marine microalga *Phaeocyctis antarctica*', *Environmental Toxicology and Chemistry*, 34:1578–1587, doi:10.1002/etc.2949.

Green JW, Springer TA and Holbech H (2018) *Statistical analysis of ecotoxicity studies*, John Wiley and Sons.

HEPA (Heads of EPA Australia and New Zealand) (2025) *PFAS National Environmental Management Plan Version 3.0*, HEPA, Canberra.

Hobbs DA, Warne MStJ and Markich SJ (2005) 'Evaluation of criteria used to assess the quality of aquatic toxicity data', *Integrated Environmental Assessment and Management*, 1:174–180, doi:10.1897/2004-003R.1.

Hoekstra JA and Van Ewijk PH (1993) 'The bounded effect concentration as an alternative to the NOEC', *Science of the Total Environment,* 134(Supplement 1):705–711, doi:10.1016/S0048-9697(05)80074-9.

Huynh T and Hobbs D (2019) *Deriving site-specific guideline values for physico-chemical parameters and toxicants*, report prepared for the Independent Expert Scientific Committee on Coal Seam Gas and Large Coal Mining Development, Department of the Environment and Energy, Canberra.

King CJ, Wasley J and Richardson J (2024) 'Derivation of site-specific environmental quality guideline values for fuel-contaminated soils on sub-Antarctic Macquarie Island', *Integrated Environmental Assessment and Management,* 20:2334–2346, doi:10.1002/ieam.4989.

Langdon K, Warne MStJ and Sunderam RIM (2009) 'A compilation of data on the toxicity of chemicals to species in Australasia, Part 4: metals (2002–2009)', *Australasian Journal of Ecotoxicology*, 15:51–184.

Markich SJ, Brown PL, Batley GE, Apte SC and Stauber JL (2001) 'Incorporating metal speciation and bioavailability into water quality guidelines for protecting aquatic ecosystems', *Australasian Journal of Ecotoxicology*, 7:109–122.

Markich SJ, Warne MStJ, Westbury A-M and Roberts CJ (2002) 'A compilation of toxicity data for chemicals to Australasian species, Part 3: metals and metalloids', *Australasian Journal of Ecotoxicology*, 8:1–137.

Markich SJ, Batley GE, Stauber JL, Rogers NJ, Apte SC, Hyne RV, Bowles KC (2005) 'Hardness corrections for copper are inappropriate for protecting sensitive freshwater biota', *Chemosphere,* 60:1–8, doi:10.1016/j.chemosphere.2004.12.005.

Moermond C, Kase R, Korkaric M and Ågerstrand M (2016) 'CRED: criteria for reporting and evaluating ecotoxicity data', *Environmental Toxicology and Chemistry,* 35:1297–1309, doi:10.1002/etc.3259.

Moore DRJ, Greer CD, Manning G, Woodling K, Beckett KJ, Brain A and Marshall G (2017) 'A weight-of-evidence approach to deriving a level of concern for atrazine that is protective of aquatic plant communities', *Integrated Environmental Assessment and Management*, 13:686–701, doi:10.1002/ieam.1865.

NEPC (National Environment Protection Council) (2013) *National Environment Protection (Assessment of Site Contamination) Measure, Schedule B5c – Guideline on soil quality guidelines for arsenic, chromium III, copper, DDT, lead, naphthalene, nickel and zinc*, NEPC, Adelaide, accessed 3 January 2017.

OECD (Organisation for Economic Co-operation and Development) (1992) *Report of the OECD workshop on extrapolation of laboratory aquatic toxicity data to the real environment*, Environment Monograph No. 59, OECD, Paris.

Peters A, Merrington G, Schlekat C, De Schamphelaere K, Stauber JL, Batley GE, Harford A, van Dam R, Pease C, Mooney T, Warne MStJ, Hickey C, Glazebrook P, Chapman JC, Smith R and Krassoi F (2018) 'Validation of the nickel ligand model for locally relevant species in Australian freshwaters', *Environmental Toxicology and Chemistry,* 37(10):2566–2574, doi:10.1002/etc.4213.

Pfister R, Schwarz KA, Janczyk M, Dale R and Freeman JB (2013) 'Good things peak in pairs: a note on the bimodality coefficient', *Frontiers in Psychology*, 4:700, doi:10.3389/fpsyg.2013.00700.

RIVM (National Institute for Public Health and the Environment) (2004) European Union risk assessment report on zinc metal, zinc (II) chloride, zinc sulphate, zinc distearate, zinc oxide, trizinc bis(orthophosphate), report prepared by RIVM, Bilthoven.

Schwarz CJ and Tillmanns AR (2019) *Improving statistical methods to derive species sensitivity distributions*, Water Science Series, WSS2019-07, Province of British Columbia, Victoria.

Shao Q (2000) 'Estimation for hazardous concentrations based on NOEC toxicity data: an alternative approach', *Environmetrics,* 11:583–595, doi:10.1002/1099-095X(200009/10)11:5<583::AID-ENV456>3.0.CO;2-X.

Standards Australia (2003) *Numerical values – rounding and interpretation of limiting values (reconfirmed 2018)*, AS 2706-2003, Standards Australia.

Stauber JL, Golding L, Peters A, Merrington G, Adams MS, Binet MT, Batley GE, Gissi F, McKnight K, Garman E, Middleton E, Gadd J and Schlekat C (2021) 'Application of bioavailability models to derive guideline values for nickel in freshwaters of Australia and New Zealand', *Environmental Toxicology and Chemistry*, 40:100–112, doi:10.1002/etc.4885.

Stauber J, Price G, Evans A, Gadd J, Holland A, Albert A, Batley G, Binet M, Golding L, Hickey C, Harford A, Jolley D, Koppel D, McKnight K, Mosin M, Morais L, Ryan A, Thompson K, Van Genderen E, van Dam R and Warne M (2022) *Towards bioavailability-based guideline values for zinc in Australian and New Zealand natural waters*, report to the International Zinc Association, CSIRO Land and Water Report No. EP2022-0801, Lucas Heights.

Thorley J and Schwarz C (2018) 'Ssdtools: an R package to fit Species Sensitivity Distribution', *Journal of Open Source Software*, 3(31):1082, doi:10.21105/joss.01082.

Thorley J, Fisher R, Fox,D and Schwarz C (2025) 'Ssdtools v2: An R package to fit Species Sensitivity Distribution', *Journal of Open Source Software*, 10(105):7492, doi:10.21105/joss.07492.

Trenfield MA, Pease CJ, Walker SL, Mooney T, Tybell L, Humphrey C, van Dam RA and Harford AJ (2020) 'Standardised chronic toxicity test protocols and culturing methods for a suite of tropical freshwater species', *Australasian Bulletin of Ecotoxicology and Environmental Chemistry,* 6:1–80.

US EPA (United States Environmental Protection Agency) (1986) *Quality criteria for water*, US Department of Commerce, National Technical Information Service, US EPA, PB87-226759, EPA 440/5 86-001, Springfield.

US EPA (1994) *AQUIRE (Aquatic toxicity information retrieval)*, US EPA, Office of Research and Development, Duluth.

US EPA (1996) *Updates: water quality criteria documents for the protection of aquatic life in ambient water*, US EPA, Office of Water Report No. EPA 820-B96-001, Washington DC.

US EPA (2016) *Weight of evidence in ecological assessment*, Risk Assessment Forum, US EPA, Office of Research and Development Report No. EPA/100/R16/001, Washington DC.

US EPA (2018) *Final aquatic life ambient water quality criteria for aluminum*, US EPA, Office of Water Report No. EPA 822-R-18-001, Washington DC.

Van Dam R, Harford A and Warne MStJ (2012a) 'Time to get off the fence: the need for definitive international guidance for statistical analysis of ecotoxicity data', *Integrated Environmental Assessment and Management,* 8:242–245, doi:10.1002/ieam.1296.

Van Dam R, Harford A and Warne MStJ (2012b) 'Canada showing the lead, however, we still have a NOEC problem: response to Van der Vliet et al', *Integrated Environmental Assessment and Management*, 8:399–400, doi:10.1002/ieam.1307.

Van Dam RA, Humphrey CL, Harford AJ, Sinclair A, Jones DR, Davies S and Storey AW (2014) 'Site-specific water quality guidelines: 1. Derivation approaches based on physicochemical, ecotoxicological and ecological data', *Environmental Science and Pollution Research*, 21:118–130, doi:10.1007/s11356-013-1780-0.

Van Dam RA, Hogan AC, Humphrey CL and Harford AJ (2019) 'How specific is site-specific? A review and guidance for selecting and evaluating approaches for deriving local water quality benchmarks', *Integrated Environmental Assessment and Management,* 15:683–702, doi:10.1002/ieam.4181.

Van der Hoeven N, Noppert F and Leopold A (1997) 'How to measure no effect. Part I: towards a new measure of chronic toxicity in ecotoxicology, introduction and workshop results', *Environmetrics,* 8:241–248, doi:10.1002/(SICI)1099-095X(199705)8:3%3C241::AID-ENV244%3E3.0.CO;2-7.

Wang W (1987) 'Factors affecting metal toxicity to (and accumulation by) aquatic organisms – overview', *Environment International*, 13:437–457, doi:10.1016/0160-4120(87)90006-7.

Ware GW (1994) *The pesticide book*, 4th edition, WH Freeman, Fresno.

Warne MStJ (1998) *Critical review of methods to derive water quality guidelines for toxicants and a proposal for a new framework*, Supervising Scientist Report 135, Supervising Scientist, Canberra.

Warne MStJ (2001) 'Derivation of the ANZECC and ARMCANZ water quality guidelines for toxicants', *Australasian Journal of Ecotoxicology*, 7:123–136.

Warne MStJ and van Dam R (2008) 'NOEC and LOEC data should no longer be generated or used', *Australasian Journal of Ecotoxicology,* 14:1–5.

Warne MStJ and Westbury A-M (1999) 'A compilation of data on the toxicity of chemicals to Australasian species, Part 2: organic chemicals', *Australasian Journal of Ecotoxicology*, 5:21–86.

Warne MStJ, Westbury A-M and Sunderam RIM (1998) 'A compilation of data on the toxicity of chemicals to Australasian species, Part 1: pesticides', *Australasian Journal of Ecotoxicology*, 4:93–144.

Warne MStJ, Batley GE, van Dam RA, Chapman JC, Fox DR, Hickey CW and Stauber JL (2015) *Revised method for deriving Australian and New Zealand water quality guideline values for toxicants*, prepared for the Council of Australian Government's Standing Council on Environment and Water, Department of Science, Information Technology and Innovation, Queensland Government, Brisbane.

Warne MStJ, Batley GE, van Dam RA, Chapman JC, Fox DR, Hickey CW & Stauber JL (2018) *Revised method for deriving Australian and New Zealand water quality guideline values for toxicants – update of 2015 version*, prepared for the revision of the Australian and New Zealand Guidelines for Fresh and Marine Water Quality, Australian and New Zealand governments and Australian state and territory governments, Canberra.