An Australian Government Initiative

Australian & New Zealand
GUIDELINES FOR
FRESH & MARINE
WATER QUALITY

Revised method for deriving Australian and New Zealand water quality guideline values for toxicants

Prepared for the revision of the Australian and New Zealand guidelines for fresh and marine water quality

Report

October 2018

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 and 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, 48 pp.

Contact
Australian Government Department of Agriculture and Water Resources
GPO Box 858 Canberra ACT 2601
Switchboard +61 2 6272 3933 or 1800 900 090
Email waterquality@agriculture.gov.au

Liability
The authors of this publication, all other entities associated with funding this publication or preparing and compiling this publication, and the publisher of this publication, and their employees and advisers, disclaim all liability, including liability for negligence and for any loss, damage, injury, expense or cost incurred by any person as a result of accessing, using or relying on any of the information or data in this publication to the maximum extent permitted by law.

Contents

Modifications made since August 2015

There have been a number of changes to the report since August 2015. Only those modifications to the method are described below, minor edits are not included.

The citation of the report has been changed.

Section 1.1: the final sentence has been modified.

Section 1.2: a sentence has been added to clarify the difference between Guideline Values (GVs) and Default Guideline Values (DGVs).

Section 2: a new final paragraph has been inserted.

Section 3.1, first paragraph:

Text has been added to clarify that chronic toxicity data includes data generated from both single- and multi-generation toxicity tests.

A sentence has been added to clarify when toxicity data related to uptake from water and from food are appropriate to use.

Text has been added to clarify what type of toxicity data can be used to derive GVs for chemicals that are normally released in mixtures, such as pesticides.

Text has been 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 GVs for estuarine, freshwater and marine waters has been 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 three 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 has been modified to accommodate both single- and multi-generation toxicity data and definitions of these terms are provided in the Glossary.

Section 3.2:

First paragraph: text has been 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 has been changed and a footnote added to the caption of the table.

An example has been provided in row two.

The definition in row three has been expanded.

Section 3.3.2: the text for the second dot point on the degree of replication required has been modified.

Section 3.4: two new paragraphs have been added (the last two in this section). These provide guidance on the use of field, mesocosm and microcosm data, how much replication is required, appropriate experimental design and introduce the concept of using a weight of evidence approach to derive site-specific GVs.

Section 3.4.2: text relating to the use of acute to chronic ratios (ACRs) to convert acute NOEC, LOEC and MATC values to chronic NOEC/EC10 values has been added. The ACR for essential elements has been deleted.

Section 3.4.3: the title of this section has been changed and the text has been 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 GVs.

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 four 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 has been replaced. The new section contains a weight of evidence approach for assessing if data are unimodal or multimodal.

Section 3.7: this section has been modified to explain the links between grades of ecosystem protection and the levels of protection provided by GVs. Text about rounding-off GVs has been added.

Section 3.8:

The first sentence has been modified to clarify that it refers to GVs derived using a species sensitivity distribution (SSD) method.

The first sentence of the second paragraph (after the dot points) has been modified to include the reliability allocated to GVs derived using the assessment factor method.

After the text describing how the reliability of GVs is determined, new material was added on how other factors can affect the accuracy of GVs and how this should be addressed.

The paragraph following Table 7 has been expanded considerably to provide additional context regarding the visual assessment of how well the distribution fits the toxicity data.

New text has been added after Figure 2. This explains how the reliability classification of DGVs can be improved and provides a link to a site that states how DGVs should be used.

Section 3.9: a paragraph was added on the appropriateness of single- and multi-generation toxicity data for bioaccumulating and 'persistent, bioaccumulative and toxic' (PBT) 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 GVs derived using the assessment factor method should be used.

Section 3.12:

The second-last paragraph has had some clarifying text added.

Also a new final paragraph has been 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*.

Summary

As part of the revision of the *Australian and New Zealand Water Quality Guidelines* (ANZECC/ARMCANZ 2000a, b; referred to herein as the 2000 Guidelines), a number of working groups were established to review particular sections of the guidelines. The Toxicants and Sediments Working Group was asked to investigate necessary revisions for the toxicant section. This was 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 (SCEW) in February 2013. This 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 outputs of the Toxicants and Sediments Working Group were two reports, now updated. The first report originally published in 2014 (Batley et al. 2018) describes the technical rationale for the key changes made to the method in the 2000 Guidelines. The second report is the current report, (Warne et al. 2018), which presents a revised method for deriving water quality guideline values (GVs) for metal, non-metallic inorganic and organic toxicants in Australia and New Zealand. The first version of this report was published in August 2015 and was subsequently updated in 2016 and 2017, and formally approved in 2018 (the current version). Both versions of this report were peer-reviewed by three reviewers and, subsequently, the Project Coordination Group overseeing the technical aspects of the revision.

The method has retained most of the key principles of the method described in the 2000 Guidelines (ANZECC/ARMCANZ 2000a, b) and in Warne (2001), while including the most recent advances in ecotoxicology. The updated method is a significant improvement on the method in the 2000 Guidelines. The method is focused on the derivation of default (that is national or Australian and New Zealand) GVs, but provides additional guidance, where necessary, for the derivation of regional, site-specific and short-term GVs. The preferred method for GV derivation continues to be based on the use of a species sensitivity distribution (SSD) of chronic toxicity data. The minimum data requirements for using a SSD have not changed from the 2000 Guidelines, that is, toxicity data for at least five species that belong to at least four taxonomic groups. However, using toxicity data from at least eight species is strongly encouraged, and from more than 15 species is considered optimal. Different statistical distributions are fitted to the toxicity data depending on how many species and taxa they belong to, in order to avoid over-fitting the data. The basis of the reliability classification for GVs has been expanded from the 2000 Guidelines – where the number and types of toxicity data points were considered (that is chronic or converted acute) – to also include an estimate of the fit of the distribution to the data. This report provides the rules governing the revised method for calculating toxicant GVs using the SSD method, including the collation and screening of the toxicity data and determining the reliability of these values. While the less preferred assessment factor (AF) method is also covered, it is unchanged from the 2000 Guidelines and, hence is not described in detail here.

1   Introduction

1.1   Background

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

A review of the above two guideline documents commenced in 2009. Initial investigation of technical revision requirements and some high priority revisions was 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 represents a component of the broader guidance provided in the current *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. It supersedes the guidance provided in the 2000 Guidelines. It has incorporated all the changes recommended by Batley et al. (2018) and provides a step-by-step process for deriving GVs. The method is focused on the derivation of default (that is national or Australian and New Zealand) guideline values (DGVs, refer to the Glossary), but also provides additional guidance, 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 we refer to default guideline values.

Two distinctly different methods can be used to derive GVs: the species sensitivity distribution (SSD) or assessment factor (AF) methods. However, Batley et al. (2018) only recommended changes to the SSD method, and this is reflected in the current document. For further background information on the AF method for calculating toxicant GVs, readers are referred to Warne (1998) and Warne (2001).

2   Overview of the revised method

The revised method for toxicant GV derivation is very similar to that used in the 2000 Guidelines, and has retained the following key features:

- the method is risk-based

- the method uses a hierarchical, tiered framework that recommends the use of the SSD method instead of the less reliable 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 (ANZECC/ARMCANZ 2000a, b; Warne 2001).

The revised method includes the following new key features:

- 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

- an expanded suite of statistical estimates of toxicity that are deemed acceptable to derive GVs

- the phasing out of the use of NOEC data for GV derivation. NOEC data should not be used when there are acceptable data for ≥8 species that belong to ≥4 taxonomic groups

- guidance to improve the design of toxicity tests that determine concentration-response based statistical estimates of toxicity (that is EC/IC/LC and NEC data)

- inclusion of non-traditional endpoints (for example behavioural or biochemical), provided their ecological relevance has been demonstrated

- ability to combine chronic and acute (converted to chronic) toxicity data or fresh and marine data in one dataset for GV derivation

- a revised hierarchy of dataset preferences when using SSDs to derive GVs

- updated Burrlioz software (Burrlioz 2.0, Barry and Henderson 2014) to improve its functionality and make it consistent with the revised GV derivation method

- an improved method for determining the reliability of GVs. This method considers: (i) the hierarchy of acceptable data, (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'.

The rationale for the key changes to the method is explained in Batley et al. (2018).

Only two methods can be used to derive GVs: the SSD method using the Burrlioz 2.0 software (Barry and Henderson 2014) 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 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
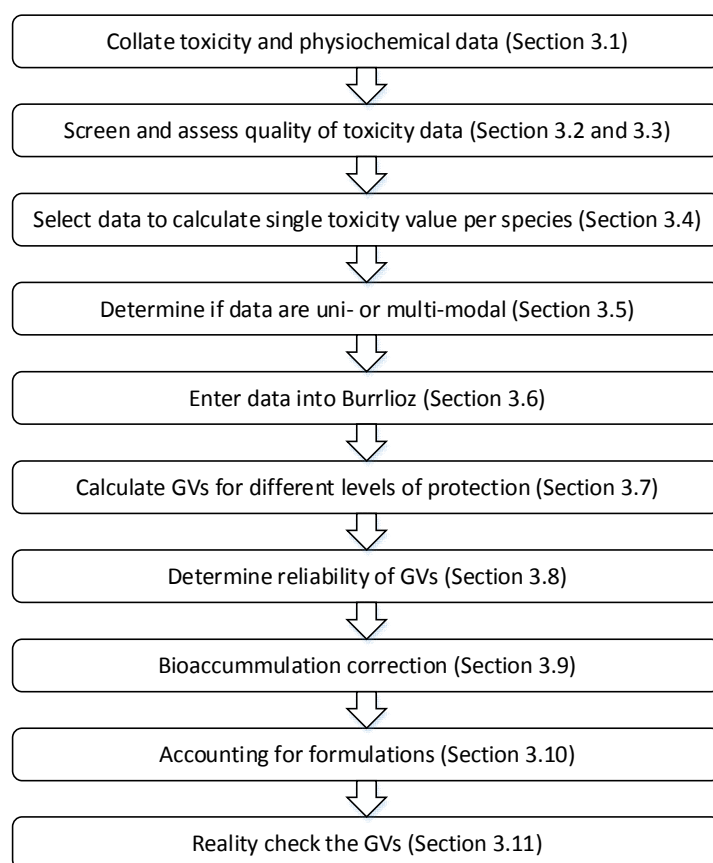
Warne 2001 for additional details). As SSD is the preferred method for calculating GVs, this report focuses on this method.

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 professional judgement throughout. Therefore, it is strongly recommended that people deriving GVs should have detailed knowledge of ecotoxicology and environmental chemistry and preferably 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 (that is chronic) exposure, the method set out in this report can also be used to derive GVs for short-term (that is acute) exposure, which may be useful at regional and/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 during intermittent and transient exposures (see Batley et al. (2018) for further guidance on the derivation of short-term GVs).

Collate toxicity and physiochemical data (Section 3.1)
⇩
Screen and assess quality of toxicity data (Section 3.2 and 3.3)
⇩
Select data to calculate single toxicity value per species (Section 3.4)
⇩
Determine if data are uni- or multi-modal (Section 3.5)
⇩
Enter data into Burrlioz (Section 3.6)
⇩
Calculate GVs for different levels of protection (Section 3.7)
⇩
Determine reliability of GVs (Section 3.8)
⇩
Bioaccummulation correction (Section 3.9)
⇩
Accounting for formulations (Section 3.10)
⇩
Reality check the GVs (Section 3.11)

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

3.1 Collating toxicity and physicochemical data

Acute, chronic (single- and multi-generation), laboratory, field, mesocosm and/or microcosm toxicity data should be obtained by conducting searches of the scientific literature including water quality documents from other countries and appropriate databases including, but not limited to, the ECOTOX database (USEPA 1994) and the Australasian Ecotoxicology Database (Warne et al. 1998; Warne and Westbury 1999; Markich et al. 2002; Langdon et al. 2009). For chemicals that do not bioaccumulate (that is chemicals with an octanol-water partition coefficient (log Kow), bioconcentration (log BCF) or bioaccumulation (log BAF) factors less than four), only toxicity data related to the uptake of chemicals from water and from water and food combined should be collated (that is 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 (for example pesticides will contain the active ingredient (AI) and additives or adjuvants designed to improve the effectiveness of the AI). For such chemicals, only toxicity data generated by exposing test organisms to a relatively pure form of the AI (refer to Section 3.10) should be used to calculate DGVs. However, toxicity data for formulations (see Glossary) 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. In the revised method, any data (including from internal reports, consultancy reports and confidential registration data) can be used provided that:

- the document is publicly available

  or

- the document is made publicly available as part of the derivation process (for example documents could be hosted on a website associated with the revised guidelines).

Commercial-in-confidence data, such as that supplied by companies for assessments by the Australian Pesticides and Veterinary Medicine Authority (APVMA) or the National Industrial Chemicals Notification and Assessment Scheme (NICNAS) can be used provided the owner of the data authorises its use and makes the data publicly available. Alternatively, where the owner approves the use of the data but not its release, the data can be used provided an agreed independent assessor with expertise in GV derivation has assessed the usability of the data (refer to Section 3.3.1).

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 rule can be made with appropriate professional judgement and justification. The emphasis of the data search should be on chronic data, as these are most appropriate for deriving GVs. However, acute data should also be collated if there are insufficient chronic data to meet the minimum data requirements for the SSD approach (Section 3.4). Preference should be given to ecotoxicity data published in peer-reviewed papers.

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 after exposure to a chemical for a short period relative to the organism's life span.

Chronic toxicity

A lethal or adverse sub-lethal effect that occurs after 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 sensitive early life stage.

A substantial portion of an organism's life span would typically be greater than 10% (Newman 2010).

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 15–25°C. The duration of acute and chronic tests for Antarctic and tropical species will differ from those presented in Table 1 (typically longer for polar and shorter for tropical) (Batley et al. 2018); however, there is currently insufficient knowledge to develop a similar table for these other climatic zones.

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 two 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.

Wherever possible, the following information should be obtained from reliable sources for every chemical for which a guideline is being derived: Chemical Abstract Services number (CAS No.), IUPAC name, common name, mode or mechanism of action, aqueous solubility, boiling and melting point, chemical formula, 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.

Table 1 Classification of acute and chronic toxicity tests for temperate species

TOXICITY TEST	LIFE STAGE[a]	RELEVANT ENDPOINTS[b]	TEST DURATION
Acute			
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 and 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			
Fish and amphibians	Adults/juveniles	All[f]	≥21 d
	Embryos/larvae/eggs	All	≥7 d
Macroinvertebrates	Adults/juveniles/larvae	All (except reproduction, larval development/metamorphosis/fertilisation)	≥14 d
	Adults/juveniles/larvae	Reproduction	≥14 d (or at least 3 broods for large cladocerans)
	Embryos	Larval development/metamorphosis	≥48 h
	Gametes	Embryo fertilisation	≥1 h
Microinvertebrates	Adults/juveniles/larvae	Reproduction	≥7 d (or at least 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 (for example decapods, echinoderms, molluscs, annelids, corals, amphipods, larger cladocerans [such as *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. Examples of invertebrates that meet this criterion are some cladocerans (for example *Ceriodaphnia dubia* and *Moina australiensis*), copepods, conchostracans, rotifer, acari, bryozoa and hydra.

f For chronic tests, 'All' encompasses all ecologically relevant endpoints measured in both single- and multi-generation tests.

The following physicochemical parameters of the water used for toxicity testing should also be documented where available: pH, salinity (or conductivity), total dissolved solids (TDS), hardness, alkalinity, dissolved organic carbon, temperature and any additives to the water (for example culture medium, food).

The salinity at which toxicity tests are conducted is also important as it can be a modifier of toxicity and defines the ecosystem type (that is fresh or marine). 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‰ while marine GVs should only use toxicity data where the salinity is between 25 and 36‰. Marine GVs should be used for estuarine ecosystems (≥0.5 to <25‰); however, if there are sufficient ecotoxicity data tested under estuarine conditions then estuarine GVs should be derived using the methods set out in this report. The salinity of marine waters can exceed 36‰ under certain conditions or times of the 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.

GVs derived for toxicants in one medium (fresh, estuarine or marine) typically should not include data from other media. 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 and/or its mode of action, there is no reason to expect differences in toxicity (refer to Section 3.4.5 for further details).

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. Data with any of the characteristics presented in Table 2 should not be used, unless a strong justification is provided. Toxicity values expressed as greater than (>) or greater than or equal to (≥) can be used, subject to professional judgement being applied to determine whether they: (i) are too far outside the existing data range and/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. Less than (<) and less than or equal to (≤) 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. When using ≥, >, ≤ and < values in calculating a GV, the actual value should be used (for example a value of >20 µg/L would, for the purposes of deriving a GV, be used as 20 µg/L). The lower value of a range of values for toxicity (for example LC50 = 25–50 µg/L) can also be used, subject to professional consideration. For all the above decisions, best professional judgement should be applied and the reasoning behind all decisions should be documented.

Table 2 Circumstances and types of toxicity data for which guideline values should not be calculated[a]

TEST CHARACTERISTIC	CONDITIONS EXCLUDED
Experimental design	Where the test concentrations differ by a large amount (for example ≥10-fold differences such as 1, 10, 100, 1000 and 10 000 µg/L)
Duration of exposure	If not stated or does not conform with Table 1
Toxicological endpoint	If not stated and/or endpoints other than lethality, immobilisation, growth, population growth or the equivalent unless the endpoint has been proven to be ecologically relevant
Aqueous solubility	If toxicity values are greater than twice the aqueous solubility

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.

Source: modified from Warne, 2001

Endpoints that are considered to be ecologically relevant (for example lethality, immobilisation, growth, development, population growth, and 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 demonstrated. An endpoint is considered to have ecological relevance when it negatively affects a species' ecological competitiveness (that is 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. Non-traditional endpoints that have not had their ecological relevance unambiguously demonstrated should only be used as an additional line of evidence in weight of evidence (WoE) based risk assessments. 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 (for example polar) for which regional or site-specific GVs are to be derived. This extends to the emerging use of ecogenomic 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, with data from non-traditional endpoints evaluated only in exceptional circumstances, for example, where there are insufficient traditional data, or to address particular site-specific concerns.

The statistical estimates of chronic toxicity (that is measures of toxicity) that can be used to derive GVs have been expanded from the 2000 Guidelines and ordered in a hierarchy. The hierarchy is as follows (in order of preference):

- No effect concentrations (NEC) (van der Hoeven et al. 1997; Fox 2009; Fox and Billoir 2011)

- x% effect/inhibition/lethal concentration (in order of preference: EC/IC/LCx) where x ≤10 (wherever possible, ECx and ICx data should be used in preference to LCx data)

- 10% bounded effect concentration (BEC10) (Hoekstra and Van Ewijk 1993)

- x% effect/inhibition concentration (EC/IC/LCx[1]) where x >10 and ≤20

- No observed effect concentration (NOEC)

- NOEC estimated from a chronic maximum acceptable toxicant concentration (MATC), lowest observed effect concentration (LOEC) or median lethal/effect value (LC/EC50).

Although NECs are not regularly reported, they are considered the preferred measure of toxicity as they are more closely aligned with the objective of GVs, that is, to protect aquatic ecosystems, as they are the concentrations that have no adverse effect on species. Reporting of NECs, and their subsequent use in GV derivation, is likely to increase in the future.

The pH range for freshwaters for which toxicity data have been generated has been retained as between 6.0 and 9.0, as per the 2000 Guidelines. 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 and/or add numerous Australian and/or New Zealand species to the dataset (with all decisions needing to be transparent and appropriately justified). Moreover, it may be useful to derive DGVs for different pH ranges, as is the case for aluminium in freshwaters, but only if pH is known to significantly affect toxicant bioavailability (for example many metals) and if sufficient data exist. Site-specific GV derivations may also be undertaken for conditions within specific pH ranges for specific sites/regions.

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/ARMCANZ (2000a, b) GVs, water quality guidelines of other jurisdictions that state that the data have been assessed (for example Canada and the USA), and those in the Australasian Ecotoxicology database (Warne et al. 1998; Warne and Westbury 1999; Markich et al. 2002; Langdon et al. 2009).

The data quality assessment should be conducted using the Excel™ spreadsheet developed by Zhang et al. (2015) that was based on the method of Hobbs et al. (2005) and developed as part of the current revision of the *Australian and New Zealand Guidelines for Fresh and Marine Water Quality*. 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 1). One of six different combinations of questions is to be answered, depending on the environmental media (freshwater, marine or estuarine), type of toxicant (metal or non-metal) and type of test organism (plant or non-plant) used (Zhang et al. 2015) (Table 3 provides the table for

freshwater, metals and non-plants, with corresponding tables for other media presented in Appendix 1). The different combinations of questions account for the different media, different toxicants and different species.

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 while data with a quality score of <50% are classed as 'unacceptable' quality. Only 'high' and 'acceptable' quality data can 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 provided (Table 3 and Appendix 1).

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'. When professional judgement is used in assessing the quality of toxicity data, a justification for the decision should be provided in the data quality assessment spreadsheet.

Table 3 Scoring system for assessing the quality of toxicity data for metals in freshwater non-plants to be used in the derivation of guideline values for toxicants. The corresponding sets of questions for other combinations of media/toxicant type/organism type are provided in Appendix 1.

QUESTION		MARK
1	Was the duration of the exposure stated (for example 48 or 96 h)?	Yes (10), No (0)
2	Was the biological endpoint (for example immobilisation or population growth) stated and defined?	Yes (10), Stated only (5), Neither (0)
3	Was the biological effect stated (for example LC or NOEC)?	Yes (5), No (0)
4	Was the biological effect quantified (for example 50% effect, 25% effect)? Note: the effect for NOEC and LOEC data must be quantified.	Yes (5), No (0)
5	Were appropriate controls (for example a no-toxicant control and/or solvent control) used?	Yes (5), No (0)
6	Was each control and chemical concentration at least duplicated?	Yes (5), No (0)
7	Were test acceptability criteria stated (for example mortality in controls must not exceed a certain percentage) or were test acceptability criteria inferred (for example test methods used were USEPA or OECD. Note: Data that fail the acceptability criteria are automatically deemed to be of unacceptable quality and must not be used.	Stated (5), Inferred (2), Neither (0)
8	Were the characteristics of the test organism (for example length, mass, age) stated?	Yes (5), No (0)
9	Was the type of test media used stated?	Yes (5), No (0)
10	Was the type of exposure (for example static, flow-through) stated?	Yes (4), No (0)
11	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 GVs; however, professional judgement can be used to include such data provided a justification for their use is provided.	Yes (4), Measured once (2), Not measured or stated (0)
12	Were parallel reference toxicant toxicity tests conducted?	Yes (4), No (0)
13	Was there a concentration-response relationship either observable or stated?	Yes (4), No (0)

QUESTION		MARK
14	Was an appropriate statistical method or model used to determine the toxicity? Note: The method should be accepted by a recognised national or international regulatory body (for example USEPA, OECD or ASTM)	Yes (4), No (0)
15	For LC/EC/NEC/BEC data, was an estimate of variability provided? OR For NOEC/LOEC/MDEC/MATC data, was the significance level 0.05 or less?	Yes (4), No (0
16	Were the following parameters measured and stated?	
16.1	pH - pH should be measured at least at the beginning and end of the toxicity test	Measured at the beginning and end of the test and stated (3), Measured once (1), Not measured or stated (0)
16.2	Hardness	Measured and stated (3), Measured only (1), Neither (0)
16.3	Alkalinity	Measured and stated (3), Measured only (1), Neither (0)
16.4	Dissolved organic carbon concentration	Measured and stated (3), Measured only (1), Neither (0)
16.5	Dissolved oxygen	Measured and stated (3), Measured only (1), Neither (0)
16.6	Conductivity	Measured and stated (3), Measured only (1), Neither (0)
17	Was the temperature measured and stated?	Measured and stated (3), Measured but not stated or temperature of the room or chamber was stated (1), Neither (0)
18	Were test solutions, blanks and/or controls tested for contamination or were analytical reagent grade chemicals or the highest possible purity chemicals used for the experiment?	Yes (3), No (0)

Total score

Total possible score for FW/metal/non-plant data = 103

Quality score: [Total score/Total possible score] x 100

Quality class:
high quality: quality score ≥ 80%
acceptable quality: quality score ≥50– <80%
unacceptable quality: quality score <50%

Source: modified from Zhang et al. 2015 (note the modifications only affect the appearance of the table)

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 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 (2011), as summarised below.

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:

- have an adequate and unambiguous experimental set-up, including a dosing regime that reflects

    - exposure in the field, and
    - measurement of chemicals

- have at least three concentration treatments, a suitable control and appropriate replication (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, then treatments should be replicated at least in triplicate. However, if a regression method is used to calculate the toxicity then no replication is required but there should be an increase in the number of treatments to at least six)

- have a realistic biological community that

    - should be 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. Ideally fish should be included; however, 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)
    - contain taxa sensitive to the mode of action of the toxicant

- be representative of potential exposure pathways in the field, for example, in the water column. This is achieved by

    - measuring contaminant concentrations throughout the course of the experiment
    - replenishing the concentrations of any rapidly dissipating compounds

- permit a sound statistical evaluation

- measure sensitive endpoints consistent with the mode of action of the toxicant

- measure chemical and physical properties that are known to, or are likely to, affect exposure to the toxicant or the bioavailability

- permit concentration-response curves for individual contaminants to be derived

- measure individual, population and/or community level endpoints

- be of sufficient duration to account for a significant proportion of the organism's life span (at least 10%)

- be of sufficient duration to reflect the persistence of the toxicant.

Although sufficient toxicity data to derive a DGV can be generated from a single field-based, microcosm or mesocosm study, this is not acceptable. Data from multiple studies are required, or the

data from a single field-based, microcosm or mesocosm study must be combined with data from laboratory toxicity tests. DGVs derived using data from field-based, microcosm or mesocosm studies should use the method detailed in this report.

When deriving site-specific GVs, constraints around the type and amount of data/studies may require alternative, but still scientifically defensible, approaches to be used. In such cases, all decisions and associated justifications need to be documented, with peer-review necessary.

3.4 Selection of data to derive guideline values

The collated data that have successfully passed the quality assurance procedures should be entered into a document, such as Excel™, that permits easy movement and grouping of the data. Data that should be included in the spreadsheet are (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 Section 3.2 and 3.1, respectively); exposure duration; exposure type (acute or chronic); statistical estimate of toxicity (for example EC10, IC10 and NOEC); endpoint (for example immobilisation, population growth); concentration at the estimate of toxicity; the factor used to convert chronic toxicity values to the equivalent of EC10/IC10/NEC and NOEC; chronic estimated EC10/IC10/NEC and NOEC; the factor to convert acute toxicity values to chronic toxicity; and converted acute toxicity values.

For particular groups of toxicants, additional information may be required. For example, for metals affected by modifying factors (see Section 3.4.3), the spreadsheet should include the values of the modifying factors at which the test was conducted.

The data should first be sorted by media type, then species, then endpoint, then measure of toxicity, then the exposure type.

In general, the more closely an 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. However, particularly in field studies, factors other than a particular toxicant may exert a toxic effect or stress on the exposed organisms and therefore may confound the experiment and interpretations that can be drawn. Provided that field, mesocosm and microcosm ecotoxicity tests meet the criteria set out in Section 3.3.2and the resulting data pass the data screening process (Table 2), then the data are suitable for the calculation of GVs. As emphasised throughout this document, chronic data are always preferred in the derivation of GVs over acute data. Very seldom will there be sufficient field, mesocosm or microcosm data of sufficient quality, even when combined, to derive GVs using only these data. However, such data can be combined with other chronic ecotoxicity data to derive GVs. Another use of such data is as a means of testing the validity of the GVs (reality checking, Section 1.1).

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. Recent guidance on how to use WoE to help derive a GV has been provided by USEPA (2016), as well as in the current *Australian and New Zealand Guidelines for Fresh and Marine Water Quality*.

3.4.1   Conversion of toxicity data to derive guideline values

The preferred order of statistical estimates of chronic toxicity to calculate default and site-specific GVs is: chronic NEC, EC/IC/LCx where x≤10, BEC10, EC/IC/LC15–20, and NOEC. While all of these acceptable statistical estimates of toxicity are not numerically the same, they are all treated as equivalent for the purposes of deriving GVs. Professional judgement should be used to assess the magnitude of the confidence limits for point estimates of toxicity but particularly EC/IC/LCx data, where x≤10, to consider whether such data are useable. The exclusion of any data on this basis needs to be appropriately justified and documented.

In many cases, only chronic NOEC data will be available, and these may be used. The use of NOEC data to derive GVs is to be phased out as recommended by Warne and van Dam (2008) and van Dam et al. (2012a, 2012b). NOEC data should no longer be used when there are EC/IC/LC/BECx (where x≤10) or NEC data for ≥8 species that belong to ≥4 taxonomic groups. However, the effect of the omission of NOEC data from a toxicity dataset on the resultant SSD and GVs needs 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 need to be appropriately justified and documented.

In cases where there are insufficient chronic EC/IC/LCx (where x≤10), NEC, BEC10, EC/IC/LC15–20, and NOEC data to derive a GV using the SSD method, chronic LC/IC/EC50, LOEC and MATC values should be divided by 5, 2.5 and 2, respectively, to provide estimates of chronic NOEC/EC10 data (ANZECC/ARMCANZ 2000b; Warne 2001). If estimated chronic values are used, this information should be recorded in the spreadsheets used for data calculation and in the Burrlioz SSD plots and/or the accompanying text and tables of toxicity values.

The same priority in the use of toxicity data to derive default (chronic) GVs applies to the derivation of short-term GVs, that is, preference is in the order: acute EC/IC/LCx (where x ≤10), NEC, BEC10, EC/IC/LC15–20, and lastly NOEC data. However, if there are insufficient of these data then acute LC/IC/EC50, LOEC and MATC values should be divided by default conversion factors of 5, 2.5 and 2, respectively, and used. If short-term GVs are to be used for regulatory purposes such as setting license conditions or in prosecutions then the data preferences may change to reflect the purpose of the GV.

3.4.2   Conversion of acute 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 acute to chronic ratio (ACR) is the ratio of the acute toxicity (LC/EC50) to the chronic toxicity data (NOEC/EC10) for a particular chemical. Limitations to the use of ACRs are discussed by Warne (1998). 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. ACRs should be calculated directly from experimental toxicity data or those used to derive the 2000 Guidelines. 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, then 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, then 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 EC10/NOEC values. Acute NOEC, LOEC and MATC values should not be converted to chronic EC10/NOEC values and subsequently used to derive GVs. 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 (for example chlorine), or whether the chemical is an essential element (for example boron, copper, iron, manganese, molybdenum, nickel, selenium or zinc). 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 Burrlioz SSD plots and/or the accompanying text and tables of toxicity values.

It is important to note that if using acute toxicity LC/IC/EC50 data to derive short-term GVs, the data should be converted to acute NOEC/LC/IC/EC10 data prior to GV derivation (see Batley et al. 2018 for further guidance).

3.4.3 Correcting metal toxicity data for modifying factors

It is well established that there are a number of abiotic factors that can modify the toxicity and bioavailability of metals and metalloids to aquatic organisms. These include, but are not limited to organic carbon, pH, temperature, alkalinity, hardness (that is the aqueous concentration of calcium and magnesium ions) and inorganic ligands (for example Wang 1987).

In the 2000 Guidelines (ANZECC/ARMCANZ 2000a, b), GVs for cadmium, chromium (III), copper, lead, nickel and zinc were normalised to a water hardness of 30 mg $CaCO_3$/L and could subsequently be adjusted using hardness-based algorithms (adopted from USEPA (1996) and largely based on acute toxicity data for fish).

Since then considerable research has shown the limitations of these hardness algorithms. For example, Markich et al. (2005) showed that hardness has either no or a limited effect on copper toxicity for a variety of organisms. Other 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 (EQSs) concluded that they were not predictors of ecological risk posed by metals (RIVM 2004).

The above limitations contributed to the development of Biotic Ligand Models (BLMs) that consider the effect of parameters, 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 EQSs for freshwaters (European Commission 2010). At present no BLMs have been endorsed for the derivation of GVs for Australia and New Zealand, although considerable work (Peters et al. 2018) has been conducted to determine the validity of the nickel BLM, developed for European and North American waters, to

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, the use of multiple linear regression equations (MLRs), to model the toxicity of metals using water quality modifiers such as hardness, pH and DOC, has been developed and is being used by Environment Canada for zinc in freshwaters (CCME 2016). Similarly, MLRs are used in the Australian ecological investigation levels for contaminated sites (NEPC 2013). Similar to the BLM approach, toxicity data from single species can be used to develop relationships, or data for multiple species can be pooled to select the best MLR, provided that water chemistry effects on toxicity are similar across species. Such approaches may assist in the transition from simple algorithm approaches to more complex BLM approaches for metal GV derivations.

Despite their limitations, hardness-modifying algorithms, BLMs and MLRs can all be used to modify or derive default and site-specific GVs for cadmium, chromium (III), lead, nickel and zinc, particularly if recognised by other national or international regulatory authorities, while only the MLR and BLM approaches should be used for copper. The chosen method must be fully justified and must consider the following information and guidance. The method chosen to derive DGVs for cadmium, chromium (III), copper, lead, nickel and zinc will be more likely to be accepted if it has been validated using water quality conditions and species relevant for Australia and New Zealand (see further discussion in Batley et al. 2018). Additional lines of evidence supporting the decision and appropriate caveats (for example non-validation to local conditions/species) should also be documented.

Hardness-modified guideline values

It is now recommended that copper toxicity data and GVs are no longer modified for water hardness. Toxicity data for cadmium, chromium (III), lead, nickel and zinc can be modified to a standard water hardness of 30 mg $CaCO_3$/L using the algorithms presented in Table 4. Related algorithms can be used to derive hardness-modified GVs (HMGVs) from the GVs derived for 30 mg/L hardness. These would take the following form, for example for cadmium:

$$HMGV^a = GV_{(30\ mg/\ CaCO3/L)} \times (H/30)^{0.89})$$

Table 4 Hardness correction algorithms used to convert chronic toxicity data for cadmium, chromium (III), lead, nickel and zinc at a given test water hardness to a hardness of 30 mg $CaCO_3$/L

METAL	HARDNESS ALGORITHM
Cadmium	Toxicity value[a] $\div (H/30)^{0.89}$
Chromium (III)	Toxicity value $\div (H/30)^{0.82}$
Nickel and zinc	Toxicity value $\div (H/30)^{0.85}$
Lead	Toxicity value $\div (H/30)^{1.27}$

a Toxicity value = toxicity reported in the literature, H = hardness (mg/L $CaCO_3$) at which the toxicity value was determined.
Source: modified from Markich et al. 2001

Biotic Ligand Models and Multiple Linear Regression equations

Biotic Ligand Models and MLRs that are recognised by other national or international regulatory authorities may be appropriate for the derivation of both default and site-specific GVs for metals (that is cadmium, chromium (III), copper, lead, nickel and zinc), but would require case-by-case

consideration and justification. The earlier comments on justification of the chosen method will apply to other metals that may, in the future, have BLMs or MLRs developed for them.

3.4.4   Obtaining a single toxicity value for each species

Only a single toxicity value is used to represent the sensitivity of each species in an SSD. However, as there are often multiple toxicity values for each species, including data for several endpoints and exposure durations, some selection and manipulation of the toxicity data is required. An example of the application of these procedures to a dataset is presented in Table 5. The rules for data manipulation that should be applied to all toxicity data for each species are set out below:

- Determine the toxicity value for each combination of species, endpoint and duration (column 4, Table 5). If there is a single value for a combination, it is adopted (for example row 1, Table 5). If there are multiple values for a combination, the geometric mean of the values is calculated and adopted for that combination (for example rows 2 to 3 and rows 4 to 6, Table 5).

- Determine the lowest toxicity value for each combination of species and endpoint (column 5, Table 5). This will be the lowest of the values for each combination of species, endpoint and duration (column 4, Table 5). 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 (for example row 1, Table 5). If there are multiple combinations of species, endpoint and duration, the lowest geometric mean value is adopted. For example, for the combination of *Daphnia carinata* and immobilisation (rows 2 to 6, Table 5) there are geometric mean toxicity values for two durations (96 h and 144 h) of 27.4 and 5.3 µg/L, thus the value of 5.3 µg/L would be adopted as the lowest geometric mean value for this combination of species and endpoint.

- Determine the lowest value for each species (column 6, Table 5). The lowest value for all combinations of a species and endpoint is adopted as the toxicity value to represent the sensitivity of the species in the SSD calculations. For example, there are three *D. carinata* and endpoint combinations (growth, immobilisation and reproduction) with toxicity values of 7, 5.3 and 0.19 µg/L, respectively. The value of 0.19 µg/L would be adopted as the toxicity value for *D. carinata*.

Chapman (2015) provides some excellent general guidance and seven 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.

Table 5 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*)

ENDPOINT	DURATION (h)	EC10 (µg/L)	VALUE FOR EACH COMBINATION OF SPECIES, ENDPOINT AND DURATION (µg/L)	LOWEST VALUE FOR EACH COMBINATION OF SPECIES AND ENDPOINT (µg/L)	LOWEST VALUE FOR SPECIES (µg/L)
Growth	168	7	7	7	
Immobilisation	168	25	27.4		
Immobilisation	168	30			
Immobilisation	192	10		5.3	0.19
Immobilisation	192	5	5.3		
Immobilisation	192	3			

ENDPOINT	DURATION (h)	EC10 (µg/L)	VALUE FOR EACH COMBINATION OF SPECIES, ENDPOINT AND DURATION (µg/L)	LOWEST VALUE FOR EACH COMBINATION OF SPECIES AND ENDPOINT (µg/L)	LOWEST VALUE FOR SPECIES (µg/L)
Reproduction	240	1.3			
Reproduction	240	2.0	1.3		
Reproduction	240	0.9			
Reproduction	480	0.2		0.19	
Reproduction	480	0.15	0.19		
Reproduction	480	0.24			

Source: modified from Batley et al. 2018

Where water quality may have significantly varied across the tests for some reason (for example in studies specifically designed to assess the effects of physicochemical variables, such as pH, hardness or dissolved organic carbon on toxicity), then 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, then 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 (that is 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), then best professional judgement should be applied to determine whether or not the toxicity value associated with that test should be included in the dataset.

3.4.5   Do the data meet the minimum data requirements of the SSD method?

The minimum data requirements to use the SSD method are identical to those of the 2000 Guidelines. Toxicity data are required for at least five species that belong to at least four different taxonomic groups. Taxonomic groups are generally considered to be phyla (that is organisms that belong to different phyla belong to different taxonomic groups; Table 6).

Table 6 Examples of taxonomically different organism types

MAJOR TYPES OF ORGANISMS	ORGANISMS CONSIDERED TO BE TAXONOMICALLY DIFFERENT[a]
Vertebrates	Fish, amphibians
Invertebrates	Crustaceans, insects, molluscs, annelids, echinoderms, rotifers, hydra
Plants	Green algae, diatoms, brown algae, red algae, macrophytes
Others	Blue-green algae (cyanobacteria), bacteria, protozoans, coral, fungi and others

a Generally taxonomic groups are phyla.
Source: modified from Warne 2001

The minimum acceptable amount of data is not optimal, and the use of toxicity data for more species and more taxonomic groups is encouraged (see Section 3.8, Table 7). Datasets that have 5–7 species are termed 'adequate', datasets that contain data for 8–14 species that belong to at least four taxonomic groups are 'good', and datasets that contain data for at least 15 species belonging to at least four taxonomic groups are termed 'preferred'.

Toxicity datasets that meet the minimum data requirements but do not have data for eight or more species that belong to at least four taxonomic groups will have their GV calculated by fitting a 2-parameter log-logistic distribution to the data. Those toxicants that have toxicity data for at least eight species that belong to at least four taxonomic groups will have their GV calculated by fitting a 3-parameter Burr Type III distribution to the toxicity data. The rationale for this is provided by Batley et al. (2018). The selection of the type of distribution to be fitted to the toxicity data is determined automatically by Burrlioz 2.0 (Barry and Henderson 2014).

There is a strong preference to derive GVs using chronic ecotoxicity data. If, however, there are insufficient chronic ecotoxicity data to generate GVs using the SSD method or there are sufficient data but the fit of the SSD to the data is poor, then there are two potential methods to overcome these situations. The first method involves supplementing chronic data with acute data converted to chronic equivalent data as discussed in Section 3.4.2. This method can be applied to both metals and organic chemicals. It was not prescribed by the 2000 Guidelines for the derivation of site-specific GVs, but it has been used since in deriving site-specific GVs to overcome a lack of data. The value of this approach for small chronic datasets is now recognised and, as such, has been included in the revised derivation method and can be used to derive GVs at any spatial scale (for instance national, regional or site-specific). The second method is to combine chronic ecotoxicity data from more than one medium (that is fresh and/or estuarine and/or marine water). This method can only be applied to organic compounds when statistical analysis indicates there is no difference in the toxicity of the chemical in the media being combined or, based on knowledge of the chemistry of the chemical and/or its mode of action, there is no reason to expect differences in the toxicity of the chemical in the media being combined.

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

Using either method will result in a lower reliability being assigned to the resulting GVs, as described in Section 3.8.

If there are still insufficient data to derive GVs using the SSD method, then the less-favoured AF method should be used (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 most herbicides are bimodal, with the sensitivity of plants being significantly greater than that of animals due to the chemicals targeting biochemical processes that only occur in plants (for example photosynthesis). The statistical distributions typically used in the SSD method (for instance log-normal, log-logistic, Burr Type III) are unimodal. The fit of such distributions to bimodal or multimodal datasets can lead to unrepresentative results. Therefore, it is important to consider both the mode of action and data modality when deciding whether to use the whole dataset or only the most sensitive subset of the data.

A WoE approach should be adopted to determine if a dataset should be split due to bimodality or multimodality. 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 four 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 (for example non-polar and polar-narcosis) are more likely to have unimodal toxicity datasets, while those with highly specific modes of action (for example acetylcholinesterase (AChE) and photosystem II (PSII) 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, if necessary, statistically. It is recommended that this is done using log-transformed data because concentration-based data are often positively right skewed and generally well-modelled by 'log-type' distributions such as the log-logistic or Burr (which is a generalisation of the log-logistic) distributions (D. Fox, Environmetrics Australia, *pers. comm.* 1 September 2017).

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

There are various statistical tests available to assess for bimodality (for instance Freeman and Dale 2013, Pfister et al. 2013). Of these, the bimodality coefficient (BC) is a useful and easily interpreted method. It is based on an empirical relationship between bimodality and the skewness and kurtosis of a dataset. The values of BC range from 0 to 1, with those exceeding 0.555 (the value representing a uniform distribution) suggesting bimodality (Freeman and Dale 2013). BC is calculated as follows:

$$BC = \frac{\gamma^2 + 1}{\kappa + \dfrac{3(n-1)^2}{(n-2)(n-3)}}$$

where γ = skewness, κ = excess kurtosis and n = sample size.

Where a specific mode of action or bimodality or multimodality are suspected, the BC should be computed for the final (log-transformed) toxicity dataset. The BC can be calculated and used as a guide of bimodality or multimodality for datasets of any sample size (noting the minimum sample size for an SSD is 5). With respect to the three parameters above, Freeman and Dale (2013) found that bimodality detection was least influenced by sample size.

It is important to note that there is no formal test of statistical significance associated with the BC statistic and as such it should form part of an overall WoE approach. Furthermore, the reliability of BC can decrease under certain data conditions (Freeman and Dale 2013, Pfister et al. 2013). Consequently, the BC provides only an indication, rather than a definitive test, of bimodality or multimodality. The results of the graphical inspection (frequency histogram) and BC (that is statistical) examination are given equal importance and are considered to be different types of evidence.

Question 3 - Do the data show taxa-specific sensitivity (that is through distinct grouping of different taxa types)?
The BC does not identify which sub-groups within a dataset exhibit the different statistical modes. Therefore, 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 (that is 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 taxa 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 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, is the answer an artefact associated with data selection (for example the data screening process, through inclusion and exclusion of certain data), small sample size (that is insufficient data to indicate biologically meaningful differences in sensitivity between taxa groups), or test procedures (for example differences in sensitivity reported from chronic short-term versus long-term tests)? Other factors to consider include:

- whether data for the same chemical in the alternative medium (for example 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 taxa types are biologically meaningful and potentially associated with a specific mode of action, or are artefacts of the above types of issues.

If the WoE associated with the above four 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, then the dataset should be split and only the data belonging to the most sensitive group of species should be used to derive the GV. If the WoE indicates otherwise, then the dataset of single toxicity values of all species should be used to calculate the GV. While this

process is likely to be straightforward in many cases, in others it may result in equivocal outcomes, for which professional judgement will be required.

The minimum data requirement of at least five species must still be met when deriving GVs by applying the SSD method to a subset of the species. If this is done, the reliability of the GV needs to be based on the sample size used to derive the GV (see Section 3.7). The criterion of requiring data for at least four taxonomic groups may need to be relaxed for the more sensitive group of species, but should be met for the entire dataset for the chemical (that is the more and less sensitive groups of organisms combined).

3.6   Enter toxicity data into Burrlioz

The Burrlioz 2.0 software (Barry and Henderson 2014) 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 Burrlioz 2.0 (Barry and Henderson 2014) user instructions. All GVs should be expressed to two significant figures.

3.7   Calculate guideline values for different levels of species protection

The current *Australian and New Zealand Guidelines for Fresh and Marine Water Quality* (ANZECC/ARMCANZ 2000a) specify three 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. To protect high conservation value systems and slightly-to-moderately disturbed systems, PC99 and PC95 values (that is concentrations protective of 99% and 95% of species) are recommended, respectively. For highly disturbed systems, the PC90 or PC80 values are generally recommended, depending on the extent of the disturbance and agreement amongst stakeholders about the ecosystem condition and desired level of protection. The current *Australian and New Zealand Guidelines for Fresh and Marine Water Quality* provide full guidance on the appropriate application of GVs for toxicants.

It should be noted that when the toxicity data for a toxicant are multimodal, and therefore only data for the most sensitive group of taxa are 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 sensitive 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).

3.8   Determine the reliability of the guideline values

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

1)   the number of species for which toxicity data are available (that is 5–7, 8–14 or ≥15)

2)   the type of toxicity data (chronic, a mixture of chronic and converted acute, a mixture of chronic fresh and chronic marine data, or only converted acute values)

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

There are six classes of reliability: very high, high, moderate, low and very low, with the sixth class of unknown reliability being assigned to GVs derived using the AF method (see Section 1.1). The reliability of GVs associated with various combinations of these three factors is presented in Table 7. It is recognised that there are other factors that are not considered in the assessment of GV reliability that 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: *Example 1.* 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.

Example 2. 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, then similarly appropriate caveats should be provided.

Table 7 Classification of the reliability of guideline values using the SSD method

SAMPLE SIZE[a]	DATA TYPE	ADEQUACY OF SAMPLE SIZE	ADEQUACY OF FIT IN SSD	RELIABILITY
≥15		Preferred	Good	Very high
			Poor	Moderate
8–14	Chronic[b]	Good	Good	High
			Poor	Moderate
5–7		Adequate	Good	Moderate
			Poor	Low
≥15	Combined chronic and converted acute	Preferred	Good	Moderate
			Poor	Low
8–14	or Combined chronic fresh and chronic marine	Good	Good	Moderate
			Poor	Low
5–7		Adequate	Good	Moderate
			Poor	Low
≥15		Preferred	Good	Moderate
			Poor	Low
8–14	Converted acute	Good	Good	Moderate
			Poor	Low
5–7		Adequate	Good	Low
			Poor	Very low

a The sample size is assumed to comprise data from at least four taxonomic groups.
b This includes all types of 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.

To assist in determining whether the fit of the distribution to the toxicity data is good or poor, examples are presented in Figure 2. 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.

Given the level of subjectivity in determining the functional form and fit of the SSD model, it may be preferable to have a panel of at least three 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 for a particular purpose. These review processes should ensure that the final decision on SSD model fit is appropriate and defensible.
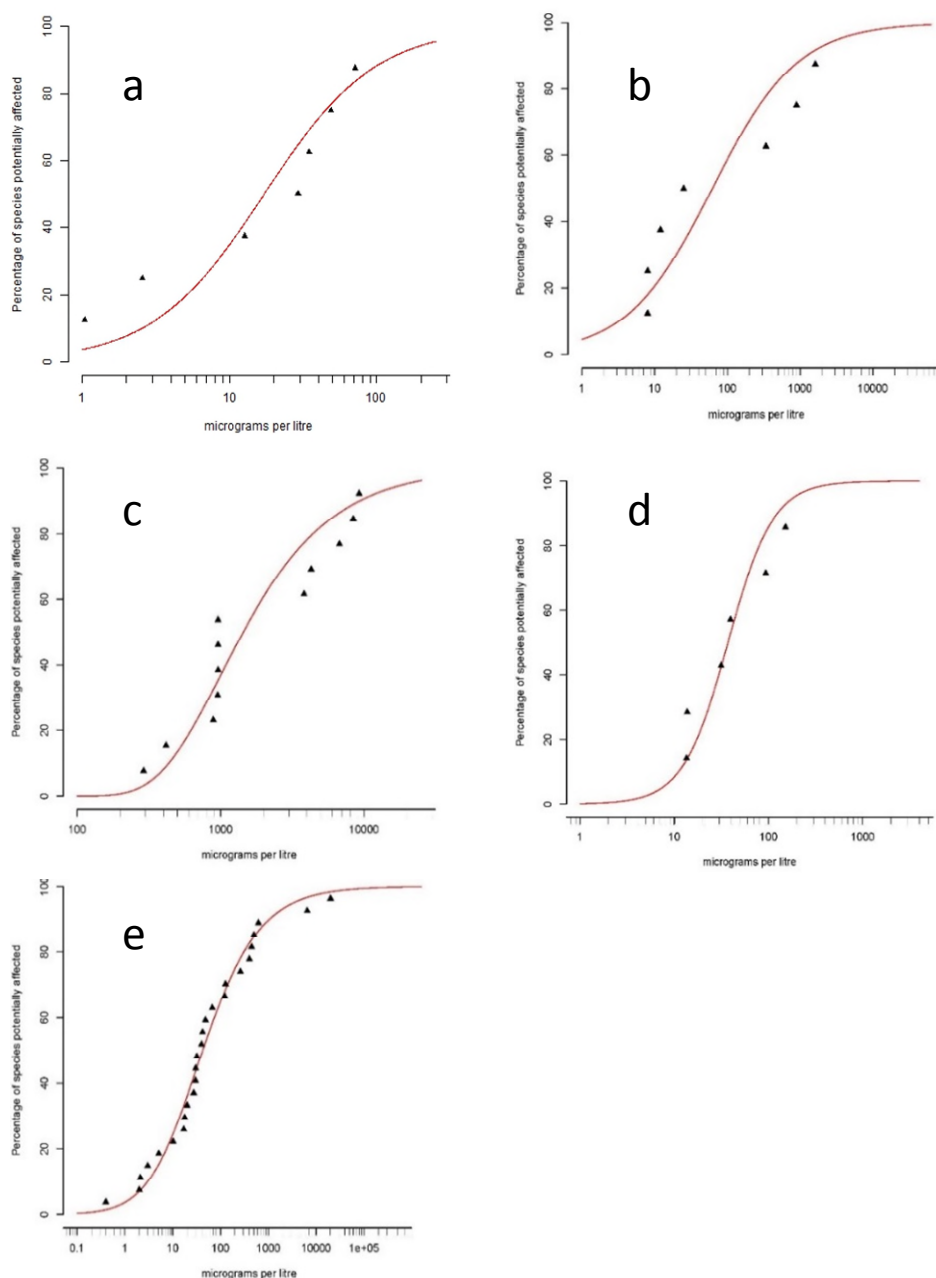
Source: Modified from Batley et al. 2018

Figure 2 Examples of poor (a, b, c) and good (d and e) fits to data obtained using the Burrlioz 2.0 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 that 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). Information on the recommended ways that DGVs should be used is provided on the [Australian and New Zealand Guidelines for Fresh and Marine Water Quality website](#).

3.9 Accounting for the potential for chemicals to bioaccumulate

Chemicals with $\log_{10}$ values for octanol-water partition coefficient (log Kow), bioconcentration (log BCF) or bioaccumulation (log BAF) factors greater than or equal to four have the potential to cause toxic effects to those organisms that eat organisms that have been exposed to the chemicals (that is secondary poisoning) (ANZECC/ARMCANZ 2000a, b; Warne 2001). For such chemicals, the level of protection provided should be increased to account for this potential additional form of toxicity. The 2000 Guidelines recommended that the level of protection be increased. This is not possible in high conservation value water bodies, so the PC99 is retained. However, in slightly-to-moderately disturbed ecosystems the PC level for chemicals should be increased to PC99 and the level of protection afforded to highly modified ecosystems should be increased from PC80 and PC90 to PC85 and PC95, respectively.

In the case of bioaccumulating or PBT chemicals (for example per- and polyfluoroalkyl substances and dioxins), it is particularly important that chronic field, mesocosm or microcosm, or chronic multi-generation data are used to derive the GVs. Chronic single-generation and any form of acute toxicity test are unlikely to fully characterise the long-term toxic effects of such substances in aquatic systems. If GVs are derived using only, or predominantly using, chronic, single-generation or acute data, then it should be clearly stated that the GV only considers relatively short-term effects and may not provide adequate protection.

For all potentially bioaccumulating substances, it should be clearly identified that the chemical has the potential to bioaccumulate.

Given these limitations, users of the guidelines are encouraged to develop site-specific GVs for bioaccumulating chemicals using the methods recommended in ANZECC/ARMCANZ (2000a, b) or other methods that can be scientifically justified. In the latter case, it may be appropriate, depending on the method being used, to use data from toxicity tests where the organisms were exposed to the test chemical via food.

3.10 Accounting for formulations

Many chemicals are released into the environment as part of a commercial formulation, for example, pesticides (see Glossary). As stated earlier, the DGVs for such chemicals should be based on toxicity tests where the test organism is exposed to a reasonably pure form of the technical material (≥ 70% AI). Ecotoxicity data for the test substances that occur as formulations (for example pesticides, PFOS and PFOA) can be used to derive DGV values provided that the study stated that the test substance:

- was a technical material, technical grade, technical reagent, analytical grade or analytical reagent, rather than a formulation

    or

- was not a formulation and had a stated purity of greater than 80%.

The other chemicals present in the commercial formulation may modify the toxicity of the AI. If the aim is to protect the environment from the formulation, then the toxicity of the commercial

formulation and the AI should be compared. All toxicity data used in such comparisons must have passed the screening and quality assurance process and have been tested on the same species under the same conditions (that is paired data). The ratio of formulation to AI 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 three or greater or equal to or less than 0.33 then it would be appropriate to correct the DGV by the difference – thus the resulting formulation-corrected GV might be larger or smaller than the DGV based on the AI. If, however, only qualitative information is available on the relative toxicity of the AI and the commercial formulation, then a statement of whether the DGV is likely to be providing sufficient or insufficient protection should be made. The DGVs for pesticides or other chemicals that occur as formulations should be expressed as a concentration of the AI, for example, 'x' µg AI/L.

3.11 Reality checking the guideline values

Once the GVs have been derived, their suitability should be evaluated by comparing them to the raw toxicity data used to derive them and/or to field-based, microcosm or mesocosm toxicity data. 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 any of the following conditions are met, then the GV should be considered to provide inadequate protection:

- If a GV is greater than the geometric mean of experimental chronic IC10/EC10/EC10/NEC or NOEC data for any important species (that is species that are important on the basis of commerce, rarity or ecological significance).

- If there is a discrepancy between the theoretical level of protection that should be provided and that indicated as being offered, based on experimental toxicity data. For example, if more than 5% of the experimental data are below the PC95 value.

In cases where the protection provided by SSD-derived GVs is deemed inadequate, the GV level of protection should be increased, for example, a PC95 could become a PC99 and a PC90 could be modified to a PC95. If this does not provide sufficient protection, then additional toxicity data are required.

The GVs for naturally occurring elements (for example metals) and compounds (for example some hydrocarbons and polycyclic aromatic hydrocarbons, PAHs) should be checked against background concentrations to ensure that unrealistically low GVs (lower than the background concentration) are not derived. A default set of background data for metals and metalloids is presented in the 2000 Guidelines (Table 8.3.2, ANZECC/ARMCANZ 2000b). Alternatively, site-specific or regional GVs based on background concentrations could be derived; however, this is not a trivial task.

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 (ANZG 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 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 (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 (for example low reliability environmental concern levels, ECLs), which created confusion amongst users. Thus, GVs, default or site-specific, are to be referred to only by their reliability category (that is very high, high, moderate, low, very low and unknown reliability GVs).

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 unknown reliability GVs, 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. 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 (for instance toxicity data, acute to chronic ratios) 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.

Appendix 1

Table A1 Scoring system for assessing the quality of toxicity data for non-metals to freshwater non-plants to be used in the derivation of guideline values for toxicants

QUESTION		MARK
1	Was the duration of the exposure stated (for example 48 or 96 h)?	Yes (10), No (0)
2	Was the biological endpoint (for example immobilisation or population growth) stated and defined?	Yes (10), Stated only (5), Neither (0)
3	Was the biological effect stated (for example LC or NOEC)?	Yes (5), No (0)
4	Was the biological effect quantified (for example 50% effect, 25% effect)? Note: The effect for NOEC and LOEC data must be quantified.	Yes (5), No (0)
5	Were appropriate controls (for example a no-toxicant control and/or solvent control) used?	Yes (5), No (0)
6	Was each control and chemical concentration at least duplicated?	Yes (5), No (0)
7	Were test acceptability criteria stated (for example mortality in controls must not exceed a certain percentage) or were test acceptability criteria inferred (for example test methods used were USEPA or OECD? Note: Data that fail the acceptability criteria are automatically deemed to be of unacceptable quality and must not be used.	Stated (5), Inferred (2), Neither (0)
8	Were the characteristics of the test organism (for example length, mass, age) stated?	Yes (5), No (0)
9	Was the type of test media used stated?	Yes (5), No (0)
10	Was the type of exposure (for example static, flow-through) stated?	Yes (4), No (0)
11	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 GVs; however, professional judgement can be used to include such data, provided a justification for their use is provided.	Yes (4), Measured once (2), Not measured or stated (0)
12	Were parallel reference toxicant toxicity tests conducted?	Yes (4), No (0)
13	Was there a concentration-response relationship either observable or stated?	Yes (4), No (0)
14	Was an appropriate statistical method or model used to determine the toxicity? Note: They should be accepted by a recognised national or international regulatory body (for example USEPA, OECD or ASTM)	Yes (4), No (0)
15	For LC/EC/NEC/BEC data, was an estimate of variability provided? OR For NOEC/LOEC/MDEC/MATC data, was the significance level 0.05 or less?	Yes (4), No (0)
16	Were the following parameters measured and stated?	
16.1	pH - pH should be measured at least at the beginning and end of the toxicity test	Measured at the beginning and end of the test and stated (3), Measured once (1), Not measured or stated (0)
16.2	Dissolved oxygen	Measured and stated (3), Measured only (1), Neither (0)

QUESTION		MARK
16.3	Conductivity	Measured and stated (3), Measured only (1), Neither (0)
17	Was the temperature measured and stated?	Measured and stated (3), Measured but not stated or temperature of the room or chamber was stated (1), Neither (0)
18	Were test solutions, blanks and/or controls tested for contamination or were analytical reagent grade chemicals or the highest possible purity chemicals used for the experiment?	Yes (3), No (0)
	Total score	
	Total possible score for FW/non-metal/non-plant data = 94	
	Quality score: [Total score/Total possible score] x 100	
	Quality class:	
	high quality = when quality score ≥ 80%	
	acceptable quality = when quality score ≥50–<80%	
	unacceptable quality = when quality score <50%	

Source: modified from Zhang et al. 2015 (note the modifications only affect the appearance of the table)

Table A2 Scoring system for assessing the quality of toxicity data for metals to freshwater plants to be used in the derivation of guideline values for toxicants

QUESTION		MARK
1	Was the duration of the exposure stated (for example 48 or 96 h)?	Yes (10), No (0)
2	Was the biological endpoint (for example immobilisation or population growth) stated and defined?	Yes (10), Stated only (5), Neither (0)
3	Was the biological effect stated (for example LC or NOEC)?	Yes (5), No (0)
4	Was the biological effect quantified (for example 50% effect, 25% effect)? Note: The effect for NOEC and LOEC data must be quantified.	Yes (5), No (0)
5	Were appropriate controls (for example a no-toxicant control and/or solvent control) used?	Yes (5), No (0)
6	Was each control and chemical concentration at least duplicated?	Yes (5), No (0)
7	Were test acceptability criteria stated (for example mortality in controls must not exceed a certain percentage) or were test acceptability criteria inferred (for example test methods used were USEPA or OECD)? Note: Data that fail the acceptability criteria are automatically deemed to be of unacceptable quality and must not be used.	Stated (5), Inferred (2), Neither (0)
8	Were the characteristics of the test organism (for example length, mass, age) stated?	Yes (5), No (0)
9	Was the type of test media used stated?	Yes (5), No (0)
10	Was the type of exposure (for example static, flow-through) stated?	Yes (4), No (0)
11	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 GVs; however, professional judgement can be used to include such data, provided a justification for their use is provided.	Yes (4), Measured once (2), Not measured or stated (0)

QUESTION		MARK
12	Were parallel reference toxicant toxicity tests conducted?	Yes (4), No (0)
13	Was there a concentration-response relationship either observable or stated?	Yes (4), No (0)
14	Was an appropriate statistical method or model used to determine the toxicity? Note: They should be accepted by a recognised national or international regulatory body (for example USEPA, OECD or ASTM)	Yes (4), No (0)
15	For LC/EC/NEC/BEC data, was an estimate of variability provided? OR For NOEC/LOEC/MDEC/MATC data, was the significance level 0.05 or less?	Yes (4), No (0)
16	Were the following parameters measured and stated?	
16.1	pH - pH should be measured at least at the beginning and end of the toxicity test	Measured at the beginning and end of the test and stated (3), Measured once (1), Not measured or stated (0)
16.2	Hardness	Measured and stated (3), Measured only (1), Neither (0)
16.3	Alkalinity	Measured and stated (3), Measured only (1), Neither (0)
16.4	Dissolved organic carbon concentration	Measured and stated (3), Measured only (1), Neither (0)
16.5	Conductivity	Measured and stated (3), Measured only (1), Neither (0)
17	Was the temperature measured and stated?	Measured and stated (3), Measured but not stated or temperature of the room or chamber was stated (1), Neither (0)
18	Were test solutions, blanks and/or controls tested for contamination or were analytical reagent grade chemicals or the highest possible purity chemicals used for the experiment?	Yes (3), No (0)

Total score

Total possible score for FW/metal/plant data = 100

Quality score: [Total score/Total possible score] x 100

Quality class:

high quality = when quality score ≥ 80%

acceptable quality = when quality score ≥50–79%

unacceptable quality = when quality score <50%

Source: modified from Zhang et al. 2015 (note the modifications only affect the appearance of the table)

Table A3 Scoring system for assessing the quality of toxicity data for non-metals to freshwater plants to be used in the derivation of guideline values for toxicants

QUESTION		MARK
1	Was the duration of the exposure stated (for example 48 or 96 h)?	Yes (10), No (0)
2	Was the biological endpoint (for example immobilisation or population growth) stated and defined?	Yes (10), Stated only (5), Neither (0)
3	Was the biological effect stated (for example LC or NOEC)?	Yes (5), No (0)
4	Was the biological effect quantified (for example 50% effect, 25% effect)? Note: The effect for NOEC and LOEC data must be quantified.	Yes (5), No (0)
5	Were appropriate controls (for example a no-toxicant control and/or solvent control) used?	Yes (5), No (0)
6	Was each control and chemical concentration at least duplicated?	Yes (5), No (0)
7	Were test acceptability criteria stated (for example mortality in controls must not exceed a certain percentage) or were test acceptability criteria inferred (for example test methods used were USEPA or OECD)? Note: Data that fail the acceptability criteria are automatically deemed to be of unacceptable quality and must not be used.	Stated (5), Inferred (2), Neither (0)
8	Were the characteristics of the test organism (for example length, mass, age) stated?	Yes (5), No (0)
9	Was the type of test media used stated?	Yes (5), No (0)
10	Was the type of exposure (for example static, flow-through) stated?	Yes (4), No (0)
11	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 GVs; however, professional judgement can be used to include such data, provided a justification for their use is provided.	Yes (4), Measured once (2), Not measured or stated (0)
12	Were parallel reference toxicant toxicity tests conducted?	Yes (4), No (0)
13	Was there a concentration-response relationship either observable or stated?	Yes (4), No (0)
14	Was an appropriate statistical method or model used to determine the toxicity? Note: They should be accepted by a recognised national or international regulatory body (for example USEPA, OECD or ASTM)	Yes (4), No (0)
15	For LC/EC/NEC/BEC data, was an estimate of variability provided? OR For NOEC/LOEC/MDEC/MATC data, was the significance level 0.05 or less?	Yes (4), No (0)
16	Were the following parameters measured and stated?	
16.1	pH - pH should be measured at least at the beginning and end of the toxicity test	Measured at the beginning and end of the test and stated (3), Measured once (1), Not measured or stated (0)
16.2	Conductivity	Measured and stated (3), Measured only (1), Neither (0)

QUESTION		MARK
17	Was the temperature measured and stated?	Measured and stated (3), Measured but not stated or temperature of the room or chamber was stated (1), Neither (0)
18	Were test solutions, blanks and/or controls tested for contamination or were analytical reagent grade chemicals or the highest possible purity chemicals used for the experiment?	Yes (3), No (0)

Total score

Total possible score for FW/metal/plant data = 91

Quality score: [Total score/Total possible score] x 100

Quality class:
high quality = when quality score ≥ 80%
acceptable quality = when quality score ≥50–79%
unacceptable quality = when quality score <50%

Source: modified from Zhang et al. 2015 (note the modifications only affect the appearance of the table)

Table A4 Scoring system for assessing the quality of toxicity data for contaminants to marine/estuarine non-plant species to be used in the derivation of guideline values for toxicants

QUESTION		MARK
1	Was the duration of the exposure stated (for example 48 or 96 h)?	Yes (10), No (0)
2	Was the biological endpoint (for example immobilisation or population growth) stated and defined?	Yes (10), Stated only (5), Neither (0)
3	Was the biological effect stated (for example LC or NOEC)?	Yes (5), No (0)
4	Was the biological effect quantified (for example 50% effect, 25% effect)? Note: The effect for NOEC and LOEC data must be quantified.	Yes (5), No (0)
5	Were appropriate controls (for example a no-toxicant control and/or solvent control) used?	Yes (5), No (0)
6	Was each control and chemical concentration at least duplicated?	Yes (5), No (0)
7	Were test acceptability criteria stated (for example mortality in controls must not exceed a certain percentage) or were test acceptability criteria inferred (for example test methods used were USEPA or OECD)? Note: Data that fail the acceptability criteria are automatically deemed to be of unacceptable quality and must not be used.	Stated (5), Inferred (2), Neither (0)
8	Were the characteristics of the test organism (for example length, mass, age) stated?	Yes (5), No (0)
9	Was the type of test media used stated?	Yes (5), No (0)
10	Was the type of exposure (for example static, flow-through) stated?	Yes (4), No (0)
11	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 GVs; however, professional judgement can be used to include such data, provided a justification for their use is provided.	Yes (4), Measured once (2), Not measured or stated (0)
12	Were parallel reference toxicant toxicity tests conducted?	Yes (4), No (0)
13	Was there a concentration-response relationship either observable or stated?	Yes (4), No (0)

QUESTION		MARK
14	Was an appropriate statistical method or model used to determine the toxicity? Note: They should be accepted by a recognised national or international regulatory body (for example USEPA, OECD and ASTM)	Yes (4), No (0)
15	For LC/EC/NEC/BEC data, was an estimate of variability provided? OR For NOEC/LOEC/MDEC/MATC data, was the significance level 0.05 or less?	Yes (4), No (0)
16	Were the following parameters measured and stated?	
16.1	Conductivity/Salinity	Measured and stated (3), Measured only (1), Neither (0)
16.2	Dissolved oxygen	Measured and stated (3), Measured only (1), Neither (0)
16.3	Conductivity	Measured and stated (3), Measured only (1), Neither (0)
16.4	Dissolved organic carbon	Measured and stated (3), Measured only (1), Neither (0)
16.5	pH - pH should be measured at least at the beginning and end of the toxicity test	Measured at the beginning and end of the test and stated (3), Measured once (1), Not measured or stated (0)
17	Was the temperature measured and stated?	Measured and stated (3), Measured but not stated or temperature of the room or chamber was stated (1), Neither (0)
18	Were test solutions, blanks and/or controls tested for contamination or were analytical reagent grade chemicals or the highest possible purity chemicals used for the experiment?	Yes (3), No (0)

Total score

Total possible score for Marine and estuarine/contaminants/non-plant data = 100

Quality score: [Total score/Total possible score] x 100

Quality class:

high quality = when quality score ≥ 80%

acceptable quality = when quality score ≥50–79%

unacceptable quality = when quality score <50%,

Source: modified from Zhang et al. 2015 (note the modifications only affect the appearance of the table)

Table A5 Scoring system for assessing the quality of toxicity data for contaminants to marine/estuarine plant species to be used in the derivation of guideline values for toxicants

QUESTION		MARK
1	Was the duration of the exposure stated (for example 48 or 96 h)?	Yes (10), No (0)
2	Was the biological endpoint (for example immobilisation or population growth) stated and defined?	Yes (10), Stated only (5), Neither (0)
3	Was the biological effect stated (for example LC or NOEC)?	Yes (5), No (0)
4	Was the biological effect quantified (for example 50% effect, 25% effect)? Note: The effect for NOEC and LOEC data must be quantified.	Yes (5), No (0)
5	Were appropriate controls (for example a no-toxicant control and/or solvent control) used?	Yes (5), No (0)
6	Was each control and chemical concentration at least duplicated?	Yes (5), No (0)
7	Were test acceptability criteria stated (for example mortality in controls must not exceed a certain percentage) or were test acceptability criteria inferred (for example test methods used were USEPA or OECD)? Note: Data that fail the acceptability criteria are automatically deemed to be of unacceptable quality and must not be used.	Stated (5), Inferred (2), Neither (0)
8	Were the characteristics of the test organism (for example length, mass, age) stated?	Yes (5), No (0)
9	Was the type of test media used stated?	Yes (5), No (0)
10	Was the type of exposure (for example static, flow-through) stated?	Yes (4), No (0)
11	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 GVs; however, professional judgement can be used to include such data, provided a justification for their use is provided.	Yes (4), Measured once (2), Not measured or stated (0)
12	Were parallel reference toxicant toxicity tests conducted?	Yes (4), No (0)
13	Was there a concentration-response relationship either observable or stated?	Yes (4), No (0)
14	Was an appropriate statistical method or model used to determine the toxicity? Note: They should be accepted by a recognised national or international regulatory body (for example USEPA, OECD and ASTM)	Yes (4), No (0)
15	For LC/EC/NEC/BEC data, was an estimate of variability provided? OR For NOEC/LOEC/MDEC/MATC data, was the significance level 0.05 or less?	Yes (4), No (0)
16	Were the following parameters measured and stated?	
16.1	Conductivity/Salinity	Measured and stated (3), Measured only (1), Neither (0)
16.2	pH	Measured and stated (3), Measured only (1), Neither (0)
16.3	Dissolved organic carbon	Measured and stated (3), Measured only (1), Neither (0)

QUESTION		MARK
17	Was the temperature measured and stated?	Measured and stated (3), Measured but not stated or temperature of the room or chamber was stated (1), Neither (0)
18	Were test solutions, blanks and/or controls tested for contamination or were analytical reagent grade chemicals or the highest possible purity chemicals used for the experiment?	Yes (3), No (0)
	Total score	
	Total possible score for Marine and estuarine/contaminants/plant data = 94	
	Quality score: [Total score/Total possible score] x 100	
	Quality class:	
	high quality = when quality score ≥ 80%	
	acceptable quality = when quality score ≥50–79%	
	unacceptable quality = when quality score <50%	

Source: modified from Zhang et al. 2015 (note the modifications only affect the appearance of the table)

Glossary

Term	Definition
ACR	Acute to chronic ratio
Acute toxicity	A lethal or adverse sub-lethal effect that occurs after exposure to a chemical for a short period relative to the organism's life span
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 single celled (such as microalgae) or multicellular macroalgae (such as seaweeds)
Amphipod	Small crustaceans (typically <10 mm) found in most aquatic environments
ANZECC	Australian and New Zealand Environment and Conservation Council
Aquatic ecosystem	Any water environment in which plants and animals interact with the chemical and physical features of the environment
ARMCANZ	Agriculture and Resource Management Council of Australia and New Zealand
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	A general term describing a process by which chemical substances are accumulated by aquatic organisms from water directly and/or through consumption of food containing the chemicals
Bioaccumulation factor	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 (for example soil, sediment or water).
Bioconcentration factor	The ratio of the concentration of a chemical in the organism to its concentration in the surrounding media once equilibrium is reached. For example, for an aquatic organism it is the ratio of the concentration in the organism to the concentration in the surrounding water.
Biomagnification	The processes by which tissue concentrations of chemicals increase as the chemical passes up through two 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 two 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 Models (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
Burr Type III	A flexible family of parametric distributions for non-negative data

Term	Definition
CCME	Canadian Council of Ministers of the Environment
Chronic toxicity	A lethal or sub-lethal adverse effect that occurs after 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 sensitive early life stage
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 to all Australian and New Zealand fresh or marine waterbodies in the absence of a more specific guideline value (for example site-specific) in the *Australian and New Zealand Guidelines for Fresh and Marine Water Quality*
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 toxicant concentration that is expected to cause one or more specified effects in x% of a group of organisms or a x% effect under specified conditions
Estuarine water	For the purposes of deriving GVs, estuarine water is defined as any waters with a salinity of ≥0.5 to <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). The same active ingredient can 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 DGVs, freshwater is defined as any waters with a salinity of <0.5‰
Guideline value (GV)	A measurable quantity (for example 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. There are several types of GVs: default GVs (DGVs), which are the national level GVs, regional GVs (for example the Great Barrier Reef GVs) or site-specific GVs

Term	Definition
Hardness modified Guideline Value (HMGV)	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	A toxicant concentration that would cause a x% reduction in a non-quantal measurement such as fecundity or growth
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 toxicant concentration that is expected to be lethal to x% of a group of organisms under specified conditions
Level of protection	The acceptable level of change from a defined reference condition
LOEC	Lowest-observable-effect concentration; the lowest tested concentration of a material (toxicant) that corresponds to a statistically significant difference compared to control organisms
Macroinvertebtrate	Macroinvertebrates include invertebrates where full-grown adults are ≥2 mm long (for example decapods, echinoderms, molluscs, annelids, corals, amphipods, larger cladocerans [such as *Daphnia magna*, *Daphnia carinata* and *Daphnia pulex*] and insect species where larvae are ≥2 mm long)
Marine water	For the purposes of deriving GVs, marine water is defined as any waters with a salinity from 25 to 36‰
MATC	Maximum allowable toxicant concentration: 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, often, but not necessarily, located in or near water bodies
Microcosm	A laboratory-based bench-scale artificial ecosystem
Microinvertebrate	Microinvertebrates are defined as invertebrate species where full-grown adults are typically <2 mm long. Examples of invertebrates that meet this criterion are some cladocerans (for example *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 two 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
NEC	No effect concentration
NOEC	No-observable-effect concentration; the highest tested concentration of a material (toxicant) at which the measured response is statistically indistinguishable from the control response
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 preceding)
NWQMS	The *Australian National Water Quality Management Strategy*
Octanol-water partition coefficient (Kow)	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 (that is log Kow). High Kow values (for example log Kow 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 (QA)	The implementation of checks on the success of quality control (for example replicate samples, analysis of samples of known concentration)
Quality control (QC)	The implementation of procedures to maximise the integrity of monitoring data (for example cleaning procedures, contamination avoidance, sample preservation methods)
Quantitative structure-activity relationship (QSAR)	A relationship between biological activity (for example 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
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 and/or maintained in a particular system of equivalent type

Term	Definition
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 acute and chronic toxicity)
Site-specific GVs	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)
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 cumulative distribution function that describes the variation in the sensitivity of species to a chemical
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 (for example 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 (AI) plus impurities associated with the manufacture of the AI but 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 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
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

Term	Definition
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/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/ARMCANZ 2000a. National Water Quality Management Strategy, Paper No. 4, Australian and New Zealand guidelines for fresh and marine water quality, volume 1, The guidelines (chapters 1–7), Australian and New Zealand Environment and Conservation Council and Agriculture and Resource Management Council of Australia and New Zealand, Canberra, Australia. Available at: http://www.waterquality.gov.au/anz-guidelines/resources/previous-guidelines/anzecc-armcanz-2000, accessed 24 December, 2014.

—— 2000b. National Water Quality Management Strategy, Paper No. 4, Australian and New Zealand guidelines for fresh and marine water quality, volume 2, Aquatic ecosystems – rationale and background information (chapter 8), Australian and New Zealand Environment and Conservation Council and Agriculture and Resource Management Council of Australia and New Zealand, Canberra, Australia, Table 8.3.2, p. 8.3–45. Available at: http://www.waterquality.gov.au/anz-guidelines/resources/previous-guidelines/anzecc-armcanz-2000, accessed 24 December, 2014.

—— 2000c. National Water Quality Management Strategy, Paper No. 4, Australian and New Zealand guidelines for fresh and marine water quality, Volume 3, Primary industries – rationale and background information, Australian and New Zealand Environment and Conservation Council and Agriculture and Resource Management Council of Australia and New Zealand, Canberra, Australia. Available at: http://www.waterquality.gov.au/anz-guidelines/resources/previous-guidelines/anzecc-armcanz-2000, accessed 24 December, 2014.

—— 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, Australia. Available at: http://www.waterquality.gov.au/anz-guidelines/resources/previous-guidelines/anzecc-armcanz-2000, accessed 24 December, 2014.

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 ACT, Australia.

APVMA 2014. Formulation types, Australian Pharmaceutical and Veterinary Medicine Authority, accessed 7 June 2018.

Barry, S and Henderson, B 2014. Burrlioz 2.0, Commonwealth Science and Industrial Research Organisation, Canberra, Australia, 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 – update of 2014 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, ACT, Canberra, 49 pp.

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, Western Australia, Australia.

CCME 2016, Canadian Water Quality Guidelines for the Protection of Aquatic Life: Zinc (Draft). Scientific Criteria Document, Canadian Council of Ministers of the Environment, Winnipeg, Canada.

Chapman, PM 2015. Including or excluding toxicity test data for development of a geometric mean. *Environ. Toxicol. Chem*., 34, 1691–1692.

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. *Integr. Environ. Assess. Manag.*, 4, 490–504.

European Commission 2010. Nickel and its compounds, Environmental quality standards sheet, prepared by Danish Environmental Protection Agency on behalf of the European Union, Copenhagen, Denmark.

European Commission 2011. Technical guidance for deriving environmental quality standards, Guidance Document No 27**,** Common Implementation Strategy for the Water Framework Directive, European Commission, Brussels, 204 pp.

Fox, DR 2009, A Bayesian approach for determining the no effect concentration and hazardous concentration in ecotoxicology. *Ecotox. Environ. Saf.,* 73, 123–131.

Fox, DR and Billoir, E 2011. Individual versus population effects in concentration-response modelling. *Integr. Environ. Assess. Manag.,* 7, 501–502.

Freeman, JB and Dale, R 2013. Assessing bimodality to detect the presence of a dual cognitive process. *Behav. Res. Meth.,* 45, 83–97.

Hobbs, DA, Warne, MStJ and Markich, SJ 2005. Evaluation of criteria used to assess the quality of aquatic toxicity data. *Integr. Environ. Assess. Manag.*, 1, 174–180.

Hoekstra, JA and Van Ewijk, PH 1993. The bounded effect concentration as an alternative to the NOEC. *Sci. Total Environ.,* 134*,* Supplement 1, 705–711.

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). *Australas. J. Ecotox*., 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. *Australas. J. Ecotox.*, 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, *Australas. J. Ecotox.*, 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.

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. *Integr. Environ. Assess. Manag*., 13, 686–701.

NEPC 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, National Environment Protection Council Adelaide, Australia, 177 pp., accessed 3 January, 2017.

Newman, MC 2010. Acute and chronic lethal effects to individuals, in *Fundamentals of Ecotoxicology*, CRC Press, Boca Raton, Florida, USA, pp. 247–272.

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

Peters, A, Merrington, G, Schlekat, C, De Schamphelaere, K, Stauber, JL, Batley, GE, Harford, A, Van Dam, R, Ceiwen, P, 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. *Environ. Toxicol. Chem.* 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. *Front. Psychol*., 4, 700.

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

Shao, Q 2000. Estimation for hazardous concentrations based on NOEC toxicity data: An alternative approach. *Environmetrics,* 11, 583–595.

USEPA 1994. AQUIRE (Aquatic toxicity information retrieval), United States Environmental Protection Agency Office of Research and Development, Duluth, Minnesota, USA.

USEPA 1996. Updates: Water quality criteria documents for the protection of aquatic life in ambient water, United States Environmental Protection Agency, Office of Water Report No. EPA 820-B96-001, Washington, DC, USA, 114 pp.

USEPA 2016. Weight of evidence in ecological assessment, Risk Assessment Forum, United States Environmental Protection Agency, Office of Research and Development Report No. EPA/100/R16/001, Washington, DC, USA, 104 pp.

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. *Integr. Environ. Assess. Manag.,* 8, 242–245.

—— 2012b. Canada showing the lead, however, we still have a NOEC problem: Response to Van der Vliet *et al. Integr. Environ. Assess. Manag.*, 8, 399–400.

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. *Environ. Sci. Pollut. Res.*, 21, 118–130.

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.

Wang, W 1987. Factors affecting metal toxicity to (and accumulation by) aquatic organisms. *Environ. Int.*, 13, 437–457.

Ware, GW 1994. *The pesticide book*, 4th edn, WH Freeman, Fresno, California, USA.

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, ACT, Australia, ISBN 0 642 24338 7, 92p.

Warne, MStJ 2001. Derivation of the ANZECC and ARMCANZ water quality guidelines for toxicants. *Australas. J. Ecotox.*, 7, 123–136.

Warne, MStJ and van Dam, R 2008. NOEC and LOEC data should no longer be generated or used. *Australas. J. Ecotox.,* 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. *Australas. J. Ecotox*., 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. *Australas. J. Ecotox.*, 4, 93–144.

Zhang Z, Warne MStJ and Vieritz A 2015. Ecotoxicity data quality assessment method V2.4, Excel spreadsheet, Water Quality and Investigations, Department of Science, Information Technology and Innovation, Brisbane, Queensland, Australia.