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

MStJ Warne, GE Batley, RA van Dam, JC Chapman, DR Fox, CW Hickey and JL Stauber

August 2015

Great state. Great opportunity.

Queensland Government

Disclaimer

This document has been prepared with all due diligence and care, based on the best available information at the time of publication. The department holds no responsibility for any errors or omissions within this document. Any decisions made by other parties based on this document are solely the responsibility of those parties. Information contained in this document is from a number of sources and, as such, does not necessarily represent government or departmental policy.

If you need to access this document in a language other than English, please call the Translating and Interpreting Service (TIS National) on 131 450 and ask them to telephone Library Services on +61 7 3170 5725

August 2015

Executive summary

As part of the revision of the *Australian and New Zealand Water Quality Guidelines* (ANZECC/ARMCANZ 2000 a, b, c; 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 at 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. The first report (Batley et al., 2014) describes the technical rationale for the key changes made to the method in the 2000 Guidelines. The second is the current report (Warne et al. 2015), 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 method has retained most of the key principles of the method described in the 2000 Guidelines (ANZECC/ARMCANZ 2000a) 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 (i.e. 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 species sensitivity distributions (SSDs) of chronic toxicity data. The minimum data requirements for using an 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 (i.e. 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.

Contents

1   Introduction

1.1   Background

The *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 *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 (i.e. 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 were conducted by a series of working groups, each consisting of appropriate experts. The Toxicants and Sediments Working Group (Working Group 4) was responsible for the method for deriving GVs for toxicants in surface waters. The method described in this report is the culmination of the deliberations of that working group.

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. (2014) and provides a step-by-step process of deriving GVs. The method is focused on the derivation of default (i.e. 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.

Two distinctly different methods can be used to derive GVs: the species sensitivity distribution (SSDs) and assessment factor (AF) methods. However, Batley et al. (2014) 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 et al. (1998), ANZECC/ARMCANZ (2000b) and Warne (2001).

2 Overview of the revised method for deriving default toxicant guideline values

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; 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 (this is typically for site-specific or regional applications);
- an expanded suite of statistical estimates of toxicity that are deemed acceptable to derive GVs. They are as follows, and in order of decreasing preference of use: NEC; EC/IC/LCx where x≤10; BEC10; EC/LC15–20; NOEC; and then estimated NOEC values derived from MATC, LOEC or EC//IC/LC50 values;
- the phasing out of the use of NOEC data for GV derivation. NOEC data should not be used when there are acceptable data (see above) 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 (i.e. EC/IC/LC and NEC data). More concentrations should be used at the lower end of the concentration-response relationship (i.e. below a 50% effect) to reduce uncertainties and improve the precision of statistical estimates of toxicity;
- 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 in one dataset for GV derivation. When there are insufficient chronic toxicity data, estimates of chronic toxicity derived from acute toxicity data can be added to the chronic toxicity dataset to derive GVs (using the same rules as when chronic toxicity data are used);
- a revised hierarchy of dataset preferences when using SSDs to derive GVs. These dataset preferences are: chronic data for ≥8 species (although an aspirational target is ≥15 species); chronic + converted acute data for ≥8 species; chronic data for 5–7 species; chronic + converted acute data for 5–7 species; converted acute data for ≥8 species; and converted acute data for 5–7 species;
- updated Burrlioz software (Burrlioz 2.0, Barry and Henderson, 2014) to improve its functionality and make it consistent with the revised GV derivation method. This includes automatically fitting a log-logistic distribution when there are toxicity data for <8 species, and a Burr Type III distribution when there are toxicity data for ≥8 species;

- an improved method for determining the reliability of GVs. It considers (i) the hierarchy of acceptable data, (ii) the sample size, and (iii) a visual estimation of goodness of fit. GVs are now classed as having very high, high, moderate, low or very low reliability; and

- the classification of GVs calculated using an assessment factor method as very low reliability.

The rationale for the key changes to the method is explained in Batley et al. (2014).

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), 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 the use of SSDs is the preferred method of calculating GVs this report focuses on this method.

3 The method for calculating guideline values using the SSD approach

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 default GVs are derived to protect against harmful effects from long-term (i.e. chronic) exposures, the method set out in this report can also be used to derive GVs for short-term (i.e. acute) exposures, which may be useful at regional and/or site-specific scales or for other uses such as setting licences or in prosecutions. Short-term GVs typically aim to protect most species against lethality during intermittent and transient exposures (see Batley et al., 2014 for further guidance on the derivation of short-term GVs).

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

Collate toxicity and physicochemical data

Screen and assess quality of toxicity data

Select data to calculate single toxicity value per species

Determine if data are uni- or multi-modal

Enter data into Burrlioz 2.0

Calculate GVs for different levels of protection

Determine reliability of GV

Bioaccumulation correction

Ground-truth the GVs
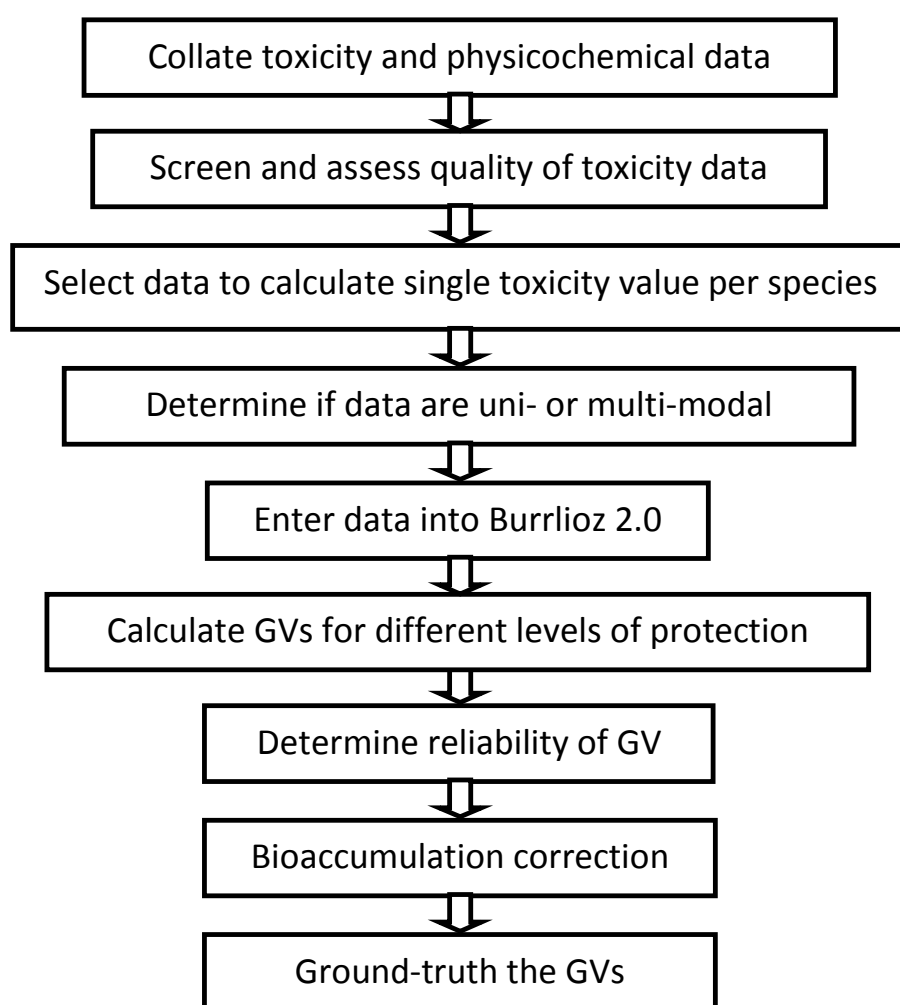
3.1 Collating toxicity and physicochemical data

Acute, chronic, 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 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). The 2000 Guidelines stipulated that only data from peer-reviewed

scientific journals be used to derive GVs. However, now any data (including from internal reports, consultancy reports and confidential registration data) can be used provided that:

- the document is publically available' or
- is made publically available as part of the derivation process (for example documents could be hosted on a web-site 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 it publically available. Alternatively, where the owner approves the use of the data but not their release, the data can be used provided an agreed independent assessor with expertise in GV derivation has assessed 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 they are considered 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 always be given to collecting 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 lifespan 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. It is acknowledged that the duration of acute and chronic tests for Antarctic and tropical species will differ from those presented in Table 1 (longer for polar and shorter for tropical) (Batley et al., 2014), however, there is currently insufficient knowledge to develop a similar table for these other climatic zones. These issues are not just restricted to climatic differences. The breadth of life histories associated with taxa represented by invertebrates means that it is not possible to make a general rule when defining chronic and acute test durations, hence invertebrates have been treated as two groups, micro- and macro-invertebrates (see Table 1). Given the above, it is likely that best professional judgement will be needed on some occasions to make a determination about whether a particular test should be regarded as acute or chronic. The basis for all professional judgement decisions must be provided and be transparent and understandable.

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, 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.

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).

Salinity in particular will categorise the test water as fresh (<0.5‰) or estuarine and marine (≥0.5‰–35‰) (currently fresh and marine are the only categories for which GVs for toxicants are derived) and other parameters provide information on factors that might modify toxicant bioavailability and toxicity.

Table 1. Generalised Classification of acute and chronic toxicity tests for temperate species, based on test duration and, where applicable, endpoint, for the purposes of water quality guideline derivation

TOXICITY TEST	LIFE STAGE	RELEVANT ENDPOINTS[a]	TEST DURATION
Acute			
Fish and amphibians	Adults/juveniles	All[b]	<21 d
	Embryos/larvae	All	<7 d
Macroinvertebrates[c]	Adults/juveniles	All	<14 d
	Embryos/larvae	All except fertilisation, larval development/ metamorphosis	<7 d
	Embryos/larvae	Larval development/ metamorphosis	<48 h
Microinvertebrates[d]	Adults/juveniles/larvae	All	<7 d
Macrophytes	Mature	All	<7 d
Macroalgae	Mature	Lethality, growth, photosynthesis and biochemical endpoints	<7 d
Microalgae	Not applicable	All	≤24 h
Microorganisms	Not applicable	All	≤24 h
Chronic			
Fish and amphibians	Adults/ juveniles	All	≥21 d
	Embryos/larvae/eggs	All	≥7 d
Macroinvertebrates[c]	Adults/juveniles	All	≥14 d
	Larvae	Lethality, immobilisation, growth	≥7 d
	Larvae	Larval development/ metamorphosis	≥48 h
	Embryos	Fertilisation	≥1 h
Microinvertebrates[d]	Adults/juveniles/larvae	All (except lethality)	≥7 d (or 3 broods for cladocerans)
		Lethality	≥21 d
	Larvae	Development	≥48 h
		Fertilisation	≥1 h
Macrophytes	Mature	All	≥7d
Macroalgae	Mature	All	≥7 d
	Early life stages	Lethality	≥7 d
	Early life stages	All, except lethality	≥1 h

TOXICITY TEST	LIFE STAGE	RELEVANT ENDPOINTS[a]	TEST DURATION
Microalgae	Not applicable	All	>24 h
Microorganism	Not applicable	All	>24 h

[a]Endpoints need to be ecologically relevant – see the section - Acceptable test endpoints.[b] 'All' refers to all ecologically relevant endpoints for a particular life stage of a particular species. [c]Macroinvertebrates include large invertebrates (for example decapods, echinoderms, molluscs, annelids, corals, amphipods, including insect larvae of similar size and life cycle) with life cycles markedly longer than most microinvertebrates. [d]Microinvertebrates are operationally defined here as comprising very small invertebrates (for example <2 mm) with relatively short life cycles (for example cladocerans, copepods, conchostracans, and hydra).

3.2   Screening the toxicity data

Once the toxicity data have been collated, they should be screened to determine their suitability for use in GV derivation. Data having any of the characteristics presented in Table 2 should not be used. Toxicity values expressed as a greater than (>) or greater than or equal to (≥) values 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. Providing 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 there are no other data for a species, the data point sits at the lower end of the distribution of species sensitivities, and 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 to 50 µg/L) can also be used, subject to professional consideration. For all the above types of decisions, best professional judgement should be applied and the reasoning behind all decisions should be documented.

Table 2. The types of toxicity data that should not be used to calculate guideline values (modified from Warne, 2001)

TYPE OF VARIABLE	CONDITIONS EXCLUDED
Experimental design	Where the test concentrations differ by a large amount (for example ≥10)
Duration of exposure	If not stated and/or <24 hours[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

[1] With the exception of fertilisation success tests for microinvertebrates

Endpoints that are considered to be ecologically relevant (for example lethality, immobilisation, growth, development, population growth, and reproduction or the equivalent) 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 (i.e. its ability to increase the frequency of its genes in subsequent generations). What is considered ecologically relevant will be both species and toxicant specific. An effect on a species' competitiveness can be direct or indirect in the case of a symbiotic organism such as zooxanthellae in corals. 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. In the case of 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 (for example polar) environments 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 extensive searching of the literature should 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 toxicity (i.e. 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 (EC/IC/LCx[1]) where x $\leq$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 $\leq$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 i.e., 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. 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, and if sufficient data exist and the pH is known to significantly affect toxicant bioavailability (for example many metals), it may be useful to derive default GVs for different pH ranges, as is the case for aluminium in freshwaters. Site-specific GV derivations may also be undertaken for conditions within specific pH ranges for specific sites/regions.

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 to derive GVs should be assessed, apart from those that have already been assessed. Already assessed data include those used to derive the ANZECC/ARMCANZ (2000a) GVs, water quality guidelines of other jurisdictions that state that the data have been assessed (for example Canada, USA), and those in the Australasian Ecotoxicology database (Warne et al., 1998; Warne et al., 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 have 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 are to be answered, depending on the environmental media (freshwater or marine/estuarine), type of toxicant (metal or non-metal) and type of test organism (plant or non-plant) used for the data being assessed (Zhang et al., 2015) (Table 3 and Appendix 1). This is done as different factors are important for different media, different toxicants and different species.

Toxicity data with a quality score ≥80% are to be classed as 'high' quality, data with a quality score of ≥50 to <80% are to be classed as 'acceptable' quality while data with a quality score of <50% are to be 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.

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). Their quality 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 and a suitable control and all treatments should be conducted at least in duplicate;
- 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 logistical 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, especially in the compartment of interest, for example, water column
  - have measured contaminant concentrations throughout the course of the experiment, and
  - replenish 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; and
- be of sufficient duration to account for a significant proportion of life-history of the organisms (at least 10%) and the fate of the toxicant.

Table 3. Scoring system for assessing the quality of toxicity data for metals to freshwater non-plants to be used in the derivation of guideline values for toxicants (Zhang et al., 2015; modified from Hobbs et al., 2005). The sets of questions for other combinations of media/toxicant type/organism type for freshwater and marine species 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 (10 marks)? Award 5 marks if the endpoint is only stated	Yes (10), Stated (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) (5 marks) **OR** Were test acceptability criteria inferred (for example test methods used were USEPA or OECD, etcetera, (award 2 marks). **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 (4 marks)? Award 2 marks if they were measured only once during the test. **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), No (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 (3 marks if measured and stated, 1 mark if just measured)	
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)

	QUESTION	MARK
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 (3 marks)? Award 1 mark if the temperature was measured but not stated or if only the temperature settings of the room or chamber are stated.	Measured and stated (3), Measured only (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 (3 marks)?	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%**	

Although sufficient toxicity data to derive a default GV can be generated from a single field-based, microcosm or mesocosm study, this is not acceptable. Data from multiple such tests are required, or data from a single field-based, microcosm or mesocosm study must be combined with data from laboratory toxicity tests. Default GVs 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 being highly recommended.

3.4   Selection of data to derive guideline values

The collated data that have successfully passed the quality assurance procedures should be placed in a document, such as Excel™, that permits easy movement and grouping of the data. Data that would be appropriate to include 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, estuarine or marine); key water quality parameters such as pH and temperature (also see below); 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 IC10/EC10/IC10/NEC/ and NOEC; chronic estimated IC10/EC10/IC10/NEC and NOEC; and 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 water hardness (see Section 3.4.3 on water hardness correction), the spreadsheet should include the water hardness 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.

3.4.1   Conversion of various toxicity data to a form suitable to derive guideline values

The preferred order of statistical estimates of chronic toxicity to calculate GVs is: NEC, EC/IC/LCx where x≤10, BEC10, EC/IC/LC15–20, NOEC. While all 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,b). They 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 GV's reliability classification (see Section 3.8). Any decisions around this need to be appropriately justified and documented.

In cases where there are insufficient chronic EC/IC/LCx, NEC, BEC10, EC/IC/LC15–20, and NOEC data to derive a GV by the SSD method, chronic LC50/IC50/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). 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 and site-specific GVs, that is, preference is in the order: acute EC/IC/LCx, NEC, BEC10, EC/IC/LC15–20, and lastly NOEC data. However, if there are insufficient of those data then acute LC50/IC50/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 non-guideline 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

While, the use of chronic toxicity data is always preferred, 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. Acute to chronic ratios (ACRs) are 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 phyla; and

- if there is more than one ACR, but none for the phyla 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 LC50 values to chronic EC10/NOEC values. Use of default ACRs should be carefully considered taking into account such factors as 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 and zinc). Justification for any professional judgement decisions is required. If converted acute 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.

It is important to note that if using acute toxicity LC/IC/EC50 data to derive short-term GVs (cf. long-term GVs), the data should be converted to acute NOEC/LC/IC/EC10 data prior to GV derivation (see Batley et al., 2014 for further guidance).

3.4.3 Correcting metal toxicity data for water hardness

The toxicities of cadmium, chromium (III), lead, nickel and zinc are affected by water hardness (i.e. the aqueous concentration of calcium and magnesium ions). The toxicity data for the above metals should be modified to a standard water hardness of 30 mg $CaCO_3$/L using the algorithms presented in Table 4 and then used to derive GVs.

The hardness algorithms for these metals can be applied directly to the toxicity data or to the outcome of the SSD calculation, as both approaches will result in the same GV. The 2000 Guidelines also state that a hardness algorithm should be used for copper. However, subsequent studies that have examined the effect of true water hardness (not to be confounded with alkalinity and pH), show that hardness has either no or limited effect on copper toxicity to a variety of organisms (see Batley et al., 2014). Therefore it is recommended that copper toxicity data and GVs are no longer modified for water hardness.

Alternative approaches such as Biotic Ligand Models (BLMs) accepted by recognised national or international regulatory authorities may be appropriate for derivation of both default and site-specific GVs for metals, but would require case-by-case consideration. It is likely that a BLM approach will not be adopted to derive default GVs by Australia and New Zealand unless it has been appropriately validated for a particular metal at a relevant range

of water quality conditions and species. On the other hand, existing BLMs may be of use for site-specific GVs, especially where supplemented with additional lines of evidence, and with appropriate caveats (for example non-validation to local conditions/species) documented.

Table 4. The hardness correction algorithms used to convert toxicity data for cadmium, chromium (III), lead, nickel and zinc to a water hardness of 30 mg/L CaCO$_3$ (modified from Markich et al., 2001)

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

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 a 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 the 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 *Ceriodaphnia* cf. *dubia* and immobilisation (rows 2 to 6, Table 6) there are geometric mean toxicity values for two durations (96 h and 144 h) of 27.4 µg/L 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 *C.* cf. *dubia* 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 *C.* cf. *dubia*.

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 the data manipulation rules to obtain a single toxicity value for a species - in this case the microcrustacean *Ceriodaphnia* cf. *dubia* (modified from Batley et al., 2014)

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	
Immobilisation	192	5	5.3		
Immobilisation	192	3			0.19
Reproduction	240	1.3			
Reproduction	240	2.0	1.7		
Reproduction	240	0.9		0.19	
Reproduction	480	0.2			
Reproduction	480	0.15	0.19		
Reproduction	480	0.24			

Where water quality may have significantly varied across the tests for some reason (for example in studies specifically designed to assess the effects of physico-chemical 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 physico-chemical 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 physico-chemical variable (i.e. one that affects the toxicity of the contaminant in question) in the toxicity test dilution water is well beyond the typical range of that variable in Australia and New Zealand (see Table 3 in Batley et al., 2014), then best professional judgement should again 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 not different to those of 2000 Guidelines. Toxicity data are still required for at least five species that belong to at least four different taxonomic groups. Taxonomic groups are generally considered to be phyla (thus, organisms that belong to different phyla belong to different taxonomic groups; Table 6). While data from five species is the minimum acceptable amount, it 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 just have 5–7 species are termed 'adequate', while those that contain data for 8–14 species that belong to at least four taxonomic groups are 'good', and those 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 through fitting a 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 through fitting a Burr Type III distribution to the toxicity data. 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).

Assuming that toxicity data were available for species that belonged to at least four taxonomic groups, the order of preference for data to be used in SSDs would be:

1. chronic data for ≥8 species, although an aspirational target is ≥15 species;
2. a) chronic data for 5–7 species;

   b) chronic + converted acute data for ≥8 species;
3. chronic + converted acute data for 5–7 species;
4. converted acute data for ≥8 species; and
5. converted acute data for 5–7 species.

A noticeable difference from the 2000 Guidelines is the ability to combine chronic data and acute data that have been converted to chronic equivalent data. While the combination of chronic and converted acute data was not allowed in the 2000 Guidelines to derive national guideline values, it was used, for pragmatic reasons, to derive site-specific guideline values. The value of this approach for small chronic datasets is now recognized and, as such, has been included in the revised derivation method. There is, however, a flow-on effect to the assigned reliability classification of GVs derived using such mixed datasets, as described in Section 3.8.

Table 6. Examples of types of organisms that are considered to be taxonomically different for the purpose of deriving guideline values (modified from Warne, 2001).

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.

Chemicals that do not meet the minimum data requirements for the SSD method will have their GV calculated using the less favoured assessment factor method (Warne 2001).

3.5   Check for toxicity data having multi-modal distributions

Once a single toxicity value for a toxicant has been obtained for each species, these values should be plotted in a frequency versus toxicity histogram, in order to visually identify whether the sensitivity data have a unimodal, bi-modal or other multi-modal distribution (for example plants are very sensitive to herbicides but most animals are markedly more tolerant). Preferably, statistical tests such as the parametric two sample t-test and the non-parametric Mann-Whitney test should be used to determine if the data are unimodal or not. This should be done because, while the SSD method can model such data it is breaking one of the key assumptions of SSD method and the range of the toxicity values would be very large, causing the resulting GV (for example the concentration protecting 95% of species: PC95) to be unrealistically low.

For chemicals with multi-modal distributions only the data belonging to the most sensitive group of organisms should be used to derive GVs. By using the SSD method on data for the

more sensitive group of organisms, the data can validly be modelled resulting in more environmentally realistic GVs, although technically, the resulting PC95 value will protect 95% of organisms that belong to the most sensitive group and a higher percentage of other organisms. In this case, 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 still be met for the entire dataset for the chemical (the more and less sensitive groups of organisms combined).

If, however, the data are unimodal, or it cannot be proved that they are bi- or multi-modal, then the dataset of single toxicity values of all species should be used to calculate GVs.

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.

3.7 Calculate different levels of protection

There are three generic levels of protection provided by the *Australian and New Zealand Guidelines for Fresh and Marine Water Quality* (ANZECC/ARMCANZ, 2000a). Four different protective concentrations (PCx values) are used to cover these levels of protection. The high conservation value and slightly to moderately disturbed ecosystems correspond to PC99 and PC95 values, respectively (protective of 99 and 95% of species, respectively). The highly disturbed ecosystems correspond to the PC90 and PC80 values, depending on the extent of the disturbance. The lower levels of protection should be calculated because they can be used where it has been agreed by stakeholders that the site is highly disturbed (degraded).

3.8 Determine the reliability of the guideline values

The classification scheme for assigning the reliability of GVs is based on three factors:

   (i) the number of species for which toxicity data are available (i.e. 5–7, 8–14 or ≥15);

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

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

There are five classes of reliability: very high, high, moderate, low and very low. The reliability of GVs associated with various combinations of these three factors is presented in Table 7.

Table 7. Classification scheme for 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	Chronic	Preferred	Good	Very high
	Chronic	Preferred	Poor	Moderate
8–14	Chronic	Good	Good	High
	Chronic	Good	Poor	Moderate
5–7	Chronic	Adequate	Good	Moderate
	Chronic	Adequate	Poor	Low
≥15	Mixed chronic and converted acute	Preferred	Good	Moderate
	Mixed chronic and converted acute	Preferred	Poor	Low
8–14	Mixed chronic and converted acute	Good	Good	Moderate
	Mixed chronic and converted acute	Good	Poor	Low
5–7	Mixed chronic and converted acute	Adequate	Good	Moderate
	Mixed chronic and converted acute	Adequate	Poor	Low
≥15	Converted acute	Preferred	Good	Moderate
	Converted acute	Preferred	Poor	Low
8–14	Converted acute	Good	Good	Moderate
	Converted acute	Good	Poor	Low
5–7	Converted acute	Adequate	Good	Low
	Converted acute	Adequate	Poor	Very low

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

For the purposes of guidance, examples of distributions that are a good and poor fit to toxicity data are presented in Figure 2. Given the level of subjectivity in determining the 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 less clear. Irrespective of how the goodness of fit is decided, a statement explaining the selected category should be provided. Moreover, for default GVs, 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 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.

Figure 2. **Examples of poor (a, b, c) and good (d and e) fits of data obtained using the Burrlioz 2.0 software (modified from Batley et al., 2014)**
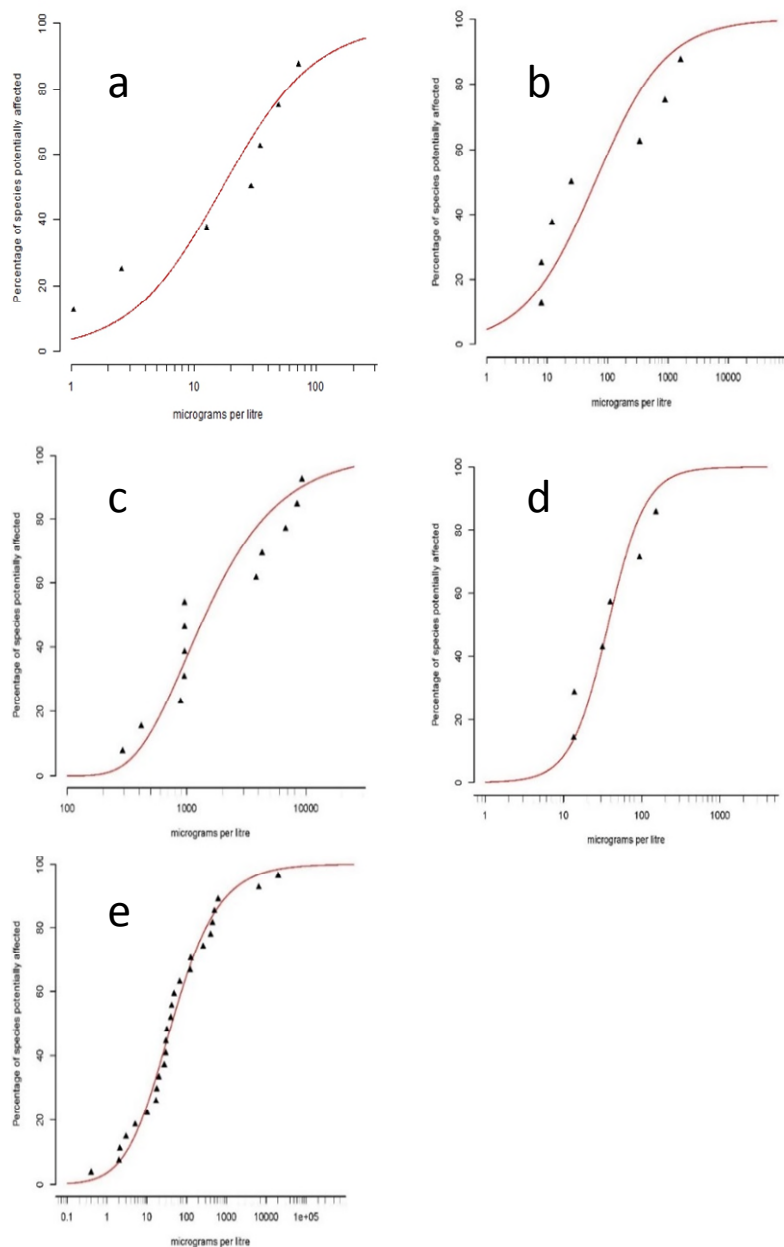
3.9   Accounting for the potential for chemicals to bioaccumulate

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 are considered to have the potential to cause toxic effects to those organisms that eat organisms that have been exposed to the chemicals (i.e. secondary poisoning) (ANZECC/ARMCANZ, 2000b). 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-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.

When deriving GVs for such chemicals the following should be addressed in the text:

- it should be clearly identified that the chemical has the potential to bioaccumulate; and
- that the derivation method only considers toxic effects caused by direct aquatic exposure and as such is an indirect approach that may not provide adequate protection.

Given these limitations, users of the guidelines are encouraged to develop site-specific GVs for bioaccumulating chemicals using the methods recommended in ANZECC/ARMCANZ (2000b) or other methods that can be scientifically justified.

3.10   The assessment factor method

This method should only to 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 assumed to be very 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 very low 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, be they national or site-specific, are to be referred to only by their reliability category (i.e., very high, high, moderate, low and very low reliability GVs).

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 were available, might be affected if exposed at the guideline 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 (for instance, 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 additional toxicity data are required.

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

4 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 example toxicity data, acute to chronic ratios etcetera) 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.

5 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 from: http://waterquality.gov.au/anz-guidelines/Documents/ANZECC-ARMCANZ-2000-guidelines-vol1.pdf. 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). ANZECC and ARMCANZ, Canberra, Australia. Table 8.3.2, p. 8.3–45. Available from: http://waterquality.gov.au/anz-guidelines/Documents/ANZECC-ARMCANZ-2000-guidelines-vol2.pdf. 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. ANZECC/ARMCANZ, Canberra, Australia. Available from: http://waterquality.gov.au/anz-guidelines/Documents/ANZECC-ARMCANZ-2000-guidelines-vol3.pdf. Accessed 24 December, 2014.

——2000d. National Water Quality Management Strategy, Paper No. 7, Australian Guidelines for Water Quality Monitoring and Reporting. ANZECC/ARMCANZ, Canberra, Australia. Available from: http://waterquality.gov.au/anz-guidelines/Documents/ANZECC-ARMCANZ-monitoring-reporting.pdf. Accessed 24 December, 2014.

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. 2014. Technical rationale for changes to the method for deriving Australian and New Zealand water quality guideline values for toxicants, CSIRO Land and Water Report EP137854, Lucas Heights, NSW, Australia, 43 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.

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

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. *Integ. Environ. Assess. Manag.,* 7, 501–502.

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.

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.

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

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 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.

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

Warne, MStJ, Batley, GE, Braga, O, Chapman, JC, Fox, D, Hickey, C, Stauber, JL and Van Dam, R. 2014. Revisions to the Australian and New Zealand toxicant guidelines for fresh and marine water quality, *Environ. Sci. Pollut. Res.*, 21, 51–60.

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. Water Quality and Investigations, Department of Science, Information Technology and Innovation.

6 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 (Zhang et al., 2015, modified from Hobbs et al., 2005).

	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 (10 marks)? Award 5 marks if the endpoint is only stated	Yes (10), Stated (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) (5 marks) **OR** Were test acceptability criteria inferred (for example test methods used were USEPA or OECD etcetera (award 2 marks). **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 (4 marks)? Award 2 marks if there were measured only once during the test. **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), No (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 (3 marks if measured and stated, 1 mark if just measured)	
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 (3 marks)? Award 1 mark if the temperature was measured but not stated or if only the temperature settings of the room or chamber are stated?	Measured and stated (3), Measured only (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 (3 marks)?	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%**	

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 (Zhang et al., 2015, modified from Hobbs et al. 2005).

	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 (10 marks)? Award 5 marks if the endpoint is only stated	Yes (10), Stated (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) (5 marks) **OR** Were test acceptability criteria inferred (for example test methods used were USEPA or OECD etcetera (award 2 marks). **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 (4 marks)? Award 2 marks if there were measured only once during the test. **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), No (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? (3 marks if measured and stated, 1 mark if just measured)	
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)

	QUESTION	MARK
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? (3 marks) Award 1 mark if the temperature was measured but not stated or if only the temperature settings of the room or chamber are stated.	Measured and stated (3), Measured (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? (3 marks)	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%**	

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 (Zhang et al., 2015, modified from Hobbs et al. 2005).

	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? (10 marks) Award 5 marks if the endpoint is only stated	Yes (10), Stated (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) (5 marks) **OR** Were test acceptability criteria inferred (for example test methods used were USEPA or OECD etcetera (award 2 marks). **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? (4 marks) Award 2 marks if there were measured only once during the test. **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), No (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? (3 marks if measured and stated, 1 mark if just measured)	
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)
17	Was the temperature measured and stated? (3 marks) Award 1 mark if the temperature was measured but not stated or if only the temperature settings of the room or chamber are stated.	Measured and stated (3), Measured only (1), Neither (0)

	QUESTION	MARK
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? (3 marks)	Yes (3), No (0)
	Total score **Total possible score for FW/non-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%**	

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 (Zhang et al., 2015, modified from Hobbs et al. 2005).

	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? (10 marks) Award 5 marks if the endpoint is only stated	Yes (10), Stated (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) (5 marks) **OR** Were test acceptability criteria inferred (for example test methods used were USEPA or OECD etcetera (award 2 marks). **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? (4 marks) Award 2 marks if there were measured only once during the test. **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), No (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? (3 marks if measured and stated, 1 mark if just measured)	
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),

QUESTION	MARK
	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? (3 marks) Award 1 mark if the temperature was measured but not stated or if only the temperature settings of the room or chamber are stated.	Measured and stated (3), Measured (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? (3 marks)	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%,**	

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 (Zhang et al., 2015, modified from Hobbs et al. 2005).

	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? (10 marks) Award 5 marks if the endpoint is only stated	Yes (10), Stated (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)? (5 marks) **OR** Were test acceptability criteria inferred (for example test methods used were USEPA or OECD etcetera? (award 2 marks). **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? (4 marks) Award 2 marks if there were measured only once during the test. **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), No (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? (3 marks if measured and stated, 1 mark if just measured)	
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? (3 marks) Award 1 mark if the temperature was measured but not stated or if only the temperature settings of the room or chamber are stated.	Measured and stated (3), Measured (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? (3 marks)	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%**	

7 Glossary of Terms and Acronyms

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.

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.

Bayesian: Involving statistical methods that assign probabilities or distributions to events or parameters based on experience or best guesses before experimentation and data collection and that apply Bayes' theorem to revise the probabilities and distributions after obtaining experimental data.

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.

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.

Bioavailable: Able to be taken up by organisms.

Burrlioz: A species sensitivity distribution software package developed and used in the ANZECC/ARMCANZ (2000) guidelines to derive guideline values (previously termed trigger values) to protect aquatic ecosystems. A new version of this (Burrlioz 2.0) was developed in 2014.

Burr Type III: A flexible family of parametric distributions for non-negative data.

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.

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 x% effect under specified conditions.

Guideline value (GV): Numerical concentration limit or narrative statement to support and maintain a designated water use. If a GV is exceeded it triggers further investigation or initiates a management response.

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.

LOR: Limit of reporting.

LOEC: Lowest-observable-effect concentration; the lowest tested concentration of a material (toxicant) at which organisms were statistically significantly adversely affected compared to control organisms.

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.

Measurement parameter: Any parameter or variable that is measured to find something out about an ecosystem.

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: Laboratory-based bench-scale artificial ecosystems.

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.

NWQMS: National Water Quality Management Strategy.

Organism: Any living animal or plant; anything capable of carrying on life processes.

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).

Reference toxicant: A reference chemical (toxicant) used in a toxicity tests 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.

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.

Speciation: Measurement of different chemical forms or species of an element in a solution or solid.

Species: Generally regarded as 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.)

SSD: Species sensitivity distribution, a name used in ecotoxicology for a cumulative distribution function.

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 using BurrliOZ taxonomic groups are generally phyla.

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.

Vertebrate: An animal having a backbone.