www.csiro.au

Technical Rationale for Changes to the Method for Deriving Australian and New Zealand Water Quality Guideline Values for Toxicants

G.E. Batley, R.A. van Dam, M.St.J. Warne, J.C. Chapman, D.R. Fox, C.W. Hickey and J.L. Stauber,

June 2014

Prepared for the Council of Australian Government's Standing Council on Environment and Water (SCEW)

Executive Summary

As part of Phase 1 of the revision of the 2000 ANZECC/ARMCANZ Water Quality Guidelines (ANZECC/ARMCANZ 2000; referred to herein as the 2000 Guidelines), the Toxicants and Sediments Working Group was asked to scope out a range of topic areas where revisions were required. A workshop held at CSIRO Land and Water at Lucas Heights, NSW in April 2010 identified a range of tasks relating to methods for the derivation of guideline values (GVs) for toxicants (termed trigger values [TVs] in the 2000 Guidelines). A contract to undertake these revision tasks was issued by the Council of Australian Government's Standing Council on Environment and Water (SCEW) in February 2013. This report was prepared by CSIRO in consultation with selected expert members from the Working Group. It provides the technical rationale for key changes to the derivation method. It supplements another report (Warne et al 2014) that provides the full revised methodology for deriving toxicant GVs.

Key changes from the 2000 Guidelines that are discussed in this report include:

- Revised definitions of acute and chronic toxicity and altered classification of toxicity tests
    - Macroalgal early life stage tests (ELS), e.g. fertilisation, germination and cell division, are now considered chronic;
    - Invertebrate early life stage tests, e.g. fertilisation and larval development are now considered chronic; and
    - 7-d juvenile or adult fish and amphibian mortality tests are now considered acute.
- Advice on short-vs long-term toxicity.
- In toxicity testing, although some indication of repeatability is desirable, it is recommended that, rather than replication of concentrations, more concentrations should be obtained 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 (e.g. behavioural) allowed only if ecological relevance can be demonstrated.
- The acceptability of toxicity endpoints follows the order NEC, EC/IC/LCx where x≤10, BEC10, EC/LC15-20, NOEC, and estimated NOEC values derived from MATC, LOEC or EC/LC50 values. The use of NOECs is discouraged in favour of ECx values.
- 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 might be 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 is available that automatically uses a log-logistic distribution for <8 data points and the Burr Type III distribution for ≥8 data points.
- A new reliability classification for GVs has been developed and is based on (i) the hierarchy of acceptable data, (ii) the sample size, and (iii) a visual estimation of goodness of fit. Very high reliability GVs have a good fit for ≥15 chronic data points. A good fit and ≥8 data points is classified as high reliability.
- Guidance is provided on weighting of data in SSDs, geographical and climatic considerations and uncertainties associated with GVs.

The changes described herein detail specific revisions to the toxicant GV derivation method that incorporate national and international developments in the science since 2000, such that the

recommended procedures represent current best practice. In the areas discussed, the recommendations supersede earlier guidance, and where a conflict or inconsistency exists between this revision and earlier advice, the advice in this report and in the companion report that describes the full revised toxicant GV derivation method (Warne et al., 2015) takes precedence.

Table of Contents

Tables

Figures

1 Introduction

This document outlines changes to the methods for calculating water quality guideline values (GVs) for toxicants (termed trigger values [TVs] in the 2000 Guidelines) in the Australian and New Zealand Guidelines for Fresh and Marine Water Quality in Document 4 of the National Water Quality Management Strategy (ANZECC/ARMCANZ, 2000) and the rationale for the changes. In particular, it deals with methods for deriving GVs for toxicants in aquatic ecosystems as covered in Chapter 3 of Volume 1 and Chapter 8 of Volume 2 of the 2000 Guidelines. The changes update approaches to specific aspects of the 2000 Guidelines to incorporate national and international developments in science since the release of the 2000 Guidelines, such that the recommended procedures represent current best practice. In the areas discussed, the recommendations supersede any earlier guidance (ANZECC/ARMCANZ, 2000; Warne, 2001; Warne et al., 2013). The revised methodology is applicable for deriving both default (i.e. nationally-derived) and site-specific GVs.

This report provides technical discussion on, and, where necessary, associated revised guidance for: definitions for acute and chronic toxicity; short- and long-term GVs (previously termed trigger values); acceptable test endpoints; use of field, mesocosm and microcosm data for deriving GVs;, a hierarchy of acceptable statistical estimates; minimum datasets for GV derivation; toxicity data inclusion rules; analytical methods for constructing species sensitivity distributions (SSDs); guidance on weighting of data in SSDs; geographical and climatic considerations; and uncertainties associated with GVs.

2 Definitions for Acute and Chronic Toxicity

2.1 Introduction

Defining acute and chronic toxicity, and determining which tests are to be considered chronic, is critical to the future development of water and sediment quality GVs (WQGVs and SQGVs) in Australia and New Zealand. A recent World Health Organisation (WHO) study (Hahn et al., 2014) showed that predicted no effect concentrations (PNECs) derived for five different chemicals varied by up to four orders of magnitude, partly due to lack of agreement on what constituted acute *versus* chronic toxicity. Clear definitions are critical because:

- The magnitude of assessment factors (AF) used in deriving GVs, and in ecological risk assessments for chemicals, depends partly on the number of acute and chronic test data available, which in turn depends on how these tests are defined.
- They affect the way acute and chronic data are combined to generate WQGs and "safe" dilutions for effluents.
- It was not clear to users of the 2000 Guidelines whether certain tests should be defined as acute or chronic, and it was evident that greater clarity was required both in the definition of acute and chronic toxicity, together with tabulation of the classification of particular test endpoints in relation to these definitions that can be used for deriving GVs.

2.2 Definitions for the Australian and New Zealand Guidelines

The 2000 Guidelines (ANZECC/ARMCANZ, 2000) provided the following definitions of acute and chronic toxicity:

Acute toxicity: Rapid adverse effect (e.g. death) caused by a substance in a living organism; can be used to define either the exposure or the response to the exposure (effect).

Chronic toxicity: Lingering or continuing for a long time; often for periods of several weeks to years; can be used to define either the exposure or the response to the exposure (effect). Chronic exposure typically includes a biological response of relatively slow progress and long continuance, often affecting a life stage.

More detailed discussions of particular endpoints were given in ANZECC/ARMCANZ (2000; Volume 2 Chapter 8.3, p. 91) and Warne (2001). For the purpose of deriving GVs, acute toxicity data used were of 24 to 96 h duration for multi-cellular organisms, and 24 to 72 h for single-celled organisms. However, this combined information did not cover all possibilities and led to confusion with respect to several common tests (e.g. larval development tests) being classified as acute or chronic.

The terms acute and chronic have been retained, but re-defined as:

Acute toxicity: A lethal or adverse sub-lethal effect that occurs as the result of a short exposure period relative to the organism's life span.

Chronic toxicity: An adverse effect that occurs as the result of exposure to a chemical for a substantial proportion of the organism's life span or an adverse sub-lethal effect on a sensitive early life stage.

A substantial proportion of an organism's lifetime would typically be greater than 10% (Newman, 2010).

General guidance on what constitutes an acute *versus* a chronic endpoint and/or test is given in Table 1.

Changes were made from the 2000 Guidelines based on current, ongoing and proposed usage. Major differences include:

- Macroalgal early life stage tests (ELS), e.g. fertilisation, germination and cell division, are now considered chronic;
- Invertebrate early life stage tests, e.g. fertilisation and larval development are now considered chronic; and
- 7-d juvenile or adult fish and amphibian mortality tests are now considered acute.

As previously, a minimum exposure duration of 24-h applies to acute toxicity tests to be used to derive GVs. However, shorter duration toxicity data (e.g. short duration fertilisation tests; see Table 1) can be used under certain circumstances (see Section 3).

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 value derivation

TOXICITY TEST	LIFE STAGE	RELEVANT ENDPOINTS [a]	TEST DURATION [b]
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
Microbes	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 reproductive broods for cladocerans)
		Lethality	≥21 d
	Larvae	Development	≥48 h

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

[a] Endpoints need to be ecologically relevant – see Section 4 (Acceptable test endpoints).

[b] The test durations specified for acute and chronic tests may not be appropriate in some cases (e.g. fish embryo tests) and best professional judgement will be needed as noted below.

[c] "All" refers to all ecologically relevant endpoints for a particular life stage of a particular species.

[d] Macroinvertebrates include large invertebrates (e.g. decapods, echinoderms, molluscs, annelids, corals, amphipods, including insect larvae of similar size and life cycle) with life cycles markedly longer than most microinvertebrates,

[e] Microinvertebrates include mostly very small invertebrates (e.g. <2 mm) with relatively short life cycles (e.g. cladocerans, copepods, conchostracans, and hydra).

The recommended test durations in Table 1 apply to tests on temperate species, typically undertaken in water temperatures ranging from 15-25^{o}C. The challenges with rigid definitions such as these, is that there will be exceptions. For example, tropical organisms (tested at >25^{o}C) develop much faster than temperate species and these are faster again than polar species. Consequently, chronic toxicity tests for tropical organisms may be of considerably shorter duration than for temperate species, while such tests for polar species (typically at 0^{o}C for Antarctic and 5^{o}C for sub-Antarctic) will be considerably longer than for temperate species. For example, a test on sea urchin development to pluteus larvae takes around 21 days for a polar species (King and Riddle, 2001) compared to 72-96 h for temperate species and 24-48 h for tropical species (R. Krassoi, Ecotoxicology Services Australasia, pers. comm.). Tests on mortality and behavioural changes in polar ostracods, gastropods and amphipods, for example, can, in some species, run for up to 10 weeks (Catherine King, Australian Antarctic Division, pers. comm.). 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). Similarly in the case of some vertebrates, the test durations listed as acute and chronic may not be appropriate, e.g. some fish embryo chronic endpoints might be achieved in less than the specified 7 days. 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.

3 Short- and Long-term Guideline Values

The 2000 Guidelines chose to not define either short- or long-term GVs as used in some jurisdictions, but to protect ecosystems from long-term exposure to toxicants. Short-term (acute) exposure GVs could be derived for particular chemicals where there are sound test data that show that effects over short times are environmentally relevant. Such tests usually (but not necessarily) measure lethality and might be appropriate in cases of a spill event, or pulse exposures as can occur with pesticides in rivers, or where contaminants are short-lived and non-persistent due to dispersion, volatilisation or degradation. The minimum exposure period is generally 96 h, but there might be circumstances where a lesser exposure time is relevant. The types of tests considered to measure acute responses were detailed in Table 1.

If a short-term GV is required for a site-specific application, it can be derived using a species sensitivity distribution (SSD) applied to acute toxicity data. The preferred data order is: 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 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. Advice on minimum datasets is provided in Section 7. No acute to chronic ratio or default assessment factors are applied to the acute data to derive short-term GVs. Apart from the use of acute toxicity data, the derivation process should follow the GV derivation method described in Warne et al. (2015). If necessary, additional guidance on short-term GVs could be sought from the Canadian protocol (CCME, 2007).

4 Acceptable Test Endpoints

The 2000 Guidelines stipulated that for GV derivation, only toxicity data that measured survival (this included survival and immobilisation), growth and reproduction were acceptable. Biomarker data, such as biochemical end-points or most behavioural data were not used, nor were data from commonly used bacterial bioluminescence assays used due to their lack of proven ecological relevance.

Canadian WQGs (CCME, 2007) admit traditional endpoints (i.e. growth, reproduction, and survival), as well as non-traditional endpoints (e.g. behaviour, predator avoidance, swimming ability, swimming speed, etc.) and physiological/biochemical changes, including endocrine-disrupting ability, if their ecological relevance can be demonstrated, i.e. whether the non-traditional endpoints influence a species' ecological competitiveness and lead to an ecologically relevant negative impact (i.e. they affect traditional endpoints).

In the revised method, non-traditional endpoints such as biochemical and physiological responses are admissible for use in GV derivation, but only those based on *in vivo* testing, i.e. data from *in vitro* tests cannot be used, and where their ecological relevance can be demonstrated. Non-traditional endpoints, including those for mutagenicity and genotoxicity, 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. An argument for the inclusion of non-traditional endpoints can be made by the developer of a proposed GV, but the decision on whether it unambiguously demonstrates ecological relevance will need to be verified by through an independent review process.

Behavioural endpoints, for example, need to show demonstrated links to growth, development and reproduction endpoints to demonstrate ecological relevance. Special consideration might be given to the use of behavioural endpoints if they are the only data available for unique, e.g. polar, environments, and this is discussed further in Section 11. It is noteworthy that a recent meta-analysis of the use in aquatic toxicology of behavioural endpoints for aqueous exposures, such as swim speed, distance moved, activity levels, spatial distribution patterns, feeding rates and courting events, found that they were generally similarly sensitive, and had similar or better detectable effect sizes and statistical power to development and reproduction endpoints (Melvin and Wilson, 2013).

Some papers have demonstrated links between photosynthetic and whole organism responses in microalgae and aquatic macrophytes, especially for certain herbicides (e.g. Magnusson et al., 2008), but less so for metals. In particular, the Great Barrier Reef Marine Park Authority (GBRMPA) had tabulated these in some of its earlier GBR-specific guideline derivations (GBRMPA, 2010), based on findings such as those of Cantin et al. (2007) on the effects of diuron on spawning of corals, although they were not officially adopted. For the current methodology, photosynthetic endpoints may be admissible for GV derivation as long as their ecological relevance can be demonstrated.

A lack of effects data for single-celled non-algal microbial species (e.g. bacteria) in GV datasets in the 2000 Guidelines was noted. One way to measure microbial responses is via biochemical responses, however, biochemical responses do not, at this point, have sufficient proven ecological relevance and therefore they should not be used to derive GVs. Recent ecogenomic approaches in ecotoxicology, that quantify the functional genes of an entire biological population (micro-, meio- and macro-organisms), may in the future provide a better way of measuring ecologically relevant effects on biological communities, but at the moment they should also not be used to derive default GVs. Their use for deriving regional or site-specific GVs should be appropriately justified.

In the derivation of GVs, 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, e.g. where there are insufficient traditional data, or to address particular site-specific concerns.

5 Use of Field, Mesocosm and Microcosm Data for Deriving GVs

Mesocosms are large enclosures designed to mimic field exposure conditions. They take the form of large tanks, enclosures or artificial channels to mimic streams, often, but not necessarily, located in or near water bodies. Microcosms are smaller laboratory-based bench-scale artificial ecosystems. Compared to laboratory-based ecotoxicity testing, mesocosms enable long-term studies and permit the simultaneous evaluation of the effects of single contaminants on multiple taxonomic groups, including indirect effects such as predation and competition between species. Mesocosm studies also incorporate the effects of competing fate processes on contaminant bioavailability (Giesy et al., 1999). Because non-chemical stressors can also impact field responses, it is often difficult to establish cause-and-effect relationships.

The 2000 Guidelines indicated that the preferred data for deriving GVs come from multiple-species toxicity tests, i.e. field or model ecosystem (mesocosm, microcosm, artificial stream) tests that represent the complex interactions of species in the field. Where such data were available, and where they met OECD requirements (OECD, 1992), they were to be preferentially used in GV derivation. Many of these tests are, however, difficult to interpret and there were few such data available that met quality screening requirements. One of the challenges is that a single mesocosm study can provide data for 10-50 species (especially including ">" values), and community metrics. Thus a single study can carry significant weight compared with the rest of the individual species data. Examples include mesocosm studies of ammonia (Hickey et al., 1999) and metals (Hickey et al., 2002; Clements et al, 2013). Where contaminant mixtures (e.g. metals) are involved, the integrated toxicity response is derived so the data are not useful for deriving thresholds for individual contaminants.

Consequently, most of the GVs in the 2000 Guidelines were derived using data from single-species toxicity tests on a range of test species, because these formed the bulk of the concentration–response information available (Chapman et al., 2001). Collecting and reviewing papers containing field, mesocosm, microcosm, chronic EC10 data and those papers containing LC50 data should be part of the data collection process prior to deriving GVs.

Despite the recommendations for the use of field-based, mesocosm and microcosm data in GV derivation, there have been few examples of these reported in the literature, probably due to the costs of undertaking such experiments. Chapman et al. (2001) reported the use of mesocosm toxicity data in combination with laboratory data to develop a site-specific GV for endosulfan in northern NSW.

The EU view (European Commission, 2011) is that while field or meso/microcosm data are not necessarily considered critical for deriving GVs, they can be used to corroborate the derived GVs. It is specifically stated in that document that 'field and mesocosm data have an important role as lines of evidence in helping define the standard (guideline) (through helping reduce uncertainty)', but would not be regarded as 'higher tier' data that would replace laboratory-based ecotoxicity data except for pesticides.

If using field studies either for site-specific investigation or to assist in GV derivation, it is left by the EU to the local investigator to determine the preferred approach noting the criteria for mesocosm data acceptance (EU, 2011), which need to be considered in conjunction with the data quality assessment outlined in Table 2 (Section 8). These include, but are not restricted to:

- An adequate and unambiguous experimental set-up, including a dosing regime that reflects; exposure in the field and measurement of chemicals;
- A realistic biological community
    - Should be representative of the taxa distribution in the field
    - Should contain taxa that are sensitive to the toxicant's mode of action

- An adequate description of exposure pathways, especially in the compartment of interest, e.g. water column
    - Contaminant concentrations should be measured throughout the course of the experiment
    - Replenish the concentrations of any rapidly dissipating compounds
- A sound statistical evaluation;
- Selected sensitive endpoints should be in accordance with the mode of action of the toxicant;
- Data should meet quality requirements and enable concentration response curves for individual contaminants to be derived.

Unlike single species laboratory-derived toxicity data, field, mesocosm and microcosm data are likely to require significant additional analysis/interrogation before they are in a form ready for GV derivation. Any field, mesocosm and microcosm data should be carefully evaluated in relation to the above criteria before inclusion in any probabilistic (SSD) GV derivation, but should be acceptable provided they meet the quality criteria.

Verification of the level of protection afforded by laboratory-based toxicity testing remains an important use of field, mesocosm and microcosm data. Although the 2000 Guidelines promoted the use of field ecological data for deriving site-specific GVs, they provided little guidance on how to do this. Van Dam et al. (2013) provide several examples about how this can be done, and also identify several recent methods that may be useful for using field data to derive GVs. This is an area that is receiving increasing attention (e.g., Crane et al., 2007; Kwok et al., 2008, Hickey et al., 1999, 2002; Clements et al., 2013; Cormier et al., 2013), although there are as yet no universally accepted methods for deriving GVs from these types of data.

6 Hierarchy of Acceptable Statistical Estimates

Much has been written recently about the acceptability of the no observed effects concentration (NOEC) as a useful measure in toxicity testing (e.g. Chapman et al., 1996; Warne and van Dam, 2008; Fox et al., 2012; Landis and Chapman, 2011; Jager, 2012; van Dam et al., 2012a). The case against NOECs has been well summarised by Newman (2008) and Fox et al. (2012), namely that the values are derived from an inappropriate application of hypothesis testing.

It is an inescapable fact that NOECs make up the majority of the toxicity data used in the derivation of toxicant GVs in the 2000 Guidelines. Consequently, it is likely to be some time before derivations solely using alternative estimates or measures can be undertaken. Nevertheless, in the short term, the Australian and New Zealand GVs will move away from NOECs, to low effect concentrations, ECx, or inhibition concentrations, ICx (where x = 10-20% effect) as the preferred measure of toxicity, and for use in future GV derivations. The EC10 is defined as the estimated concentration causing a 10% effect compared to the control or an adverse effect on 10% of the organisms compared to the controls. This change is consistent with moves in other international jurisdictions (e.g. Environment Canada, 2005; CCME, 2007; USEPA, 2009) away from the use of NOEC data to the use of model-based EC/ICx data.

It may be possible to interrogate existing NOEC data and derive EC/IC10 values, although the magnitude of this task was acknowledged, and it was only likely to be undertaken on an as needs basis by specific users. A further limitation is that, in the vast majority of cases, the raw data needed to permit the calculation of ECx values from NOEC values are not presented in papers. Although recommendations have been to use EC5 and EC10 values, the latter is preferable given the greater errors and the increased likelihood that extrapolation rather than interpolation will be used to calculate EC5 values.

There have been several attempts to compare EC10 values with NOECs, despite the recognition that they are statistically not analogous (e.g. Shieh et al., 2001; Green et al., 2013), that were more out of a desire to see how the numerical values relate to one another. The results of such comparisons will depend on the statistical model used, the endpoint (growth, biomass, etc.), the number of data points used in the toxicity test, and the uncertainty associated with the endpoint. For example, Green et al. (2013), acknowledging that the results are not interchangeable because of the completely different experimental designs, reported that for 41 algal toxicity tests, a reasonable correspondence between the NOEC and the EC20 values based on biomass for 41 algal toxicity tests, but a correlation with the NOEC and EC10 values based on growth rate. It was also an illustration that a 'one size fits all' approach is inappropriate in relation to decisions over the choice of EC10 or EC20 (or other ECx).

Optimal toxicity test designs for calculating NOECs and EC10s are different. Hence, test designs need to change as the preferred statistical estimates change. Although some indication of repeatability is desirable, it is recommended that, rather than replication of concentrations, more concentrations should be obtained at the lower end of the concentration response relationship (i.e. below a 50% effect) to reduce uncertainties and improve the reliability of EC10 values.

In the longer term, a shift is foreseen to model-based toxicity measures of no-effect concentrations (NECs). The NEC is a model parameter that can be determined by statistical analysis, e.g. using Bayesian methods (van der Hoeven et al., 1997; Fox, 2009; Fox and Billoir, 2011). Ultimately, the NEC is more closely aligned with the objective of GVs to protect aquatic ecosystems, in that it is a concentration that has no adverse effect on a species. Formal guidance on whether the NEC approach is appropriate to derive GVs and suitable and readily useable approaches to estimating a NEC (including limitations, pitfalls, etc.) have yet to be fully developed, however, for the current derivation method, NECs are admissible if the tests from which they were generated meet the necessary quality requirements (see Section 8).

Although for Australian and New Zealand GV derivations, NOECs are to be replaced with ECx values, it is recognised that a combination of NEC, ECx and NOEC values will, for some time, be unavoidable. Although

in theory NOECs may be discarded once there are sufficient (≥8) NEC/ECx values, the effect of this needs be examined on a case by case basis. Such evaluations will need to consider the sizes and representativeness of the respective databases, following the reliability rankings shown in Table 4 (Section 8).

The preferred order of measures of chronic toxicity to calculate GVs is: NEC, EC/IC/LCx where x≤10, BEC10, EC/LC15-20, NOEC, and NOEC estimated from MATC, LOEC or LC50 values. The BEC10, or bounded effect concentration, is the highest tested concentration that has an upper 95% confidence interval that causes less than a 10% effect (Hoekstra and Van Ewijk, 1993). The MATC is the maximum acceptable toxicant concentration (the geometric mean of the NOEC and LOEC). In the event there are insufficient chronic toxicity data to derive a GV (see Section 7), acute toxicity data converted to chronic equivalent data, can be included. Acute EC/IC/LC50 data should be used for this purpose.

Conversion of an acute LC50 to a chronic EC10 requires division by an acute to chronic ratio or in its absence, a factor of 10 (a factor of 2 was used for essential metals). Chronic LC50/IC50 or EC50; LOEC; and MATC data were converted to chronic EC10 values by dividing by 5, 2.5 and 2, respectively. These factors are those that were used previously to convert acute EC/LC50 data to chronic NOECs (ANZECC/ARMCANZ, 2000; Warne, 2001).

With EC10 values being a preferred endpoint to NOECs, and the existing database largely based on NOECs, it may be tempting to attempt to convert these data to EC10s by reference to published concentration-response curves. Care should be taken in such attempts, as frequently the number of data points is not sufficient to derive a reliable EC10 value using the newly recommended procedure. Any such conversions should be fully transparent with appropriate justification, with reporting of the EC10 confidence interval and model goodness-of-fit. It is possible in many instances, using best professional judgement, that the NOEC estimate may be a more defensible value than the corresponding EC10 estimate based on the same dataset.

7 Minimum Datasets for Guideline Value Derivation

Of critical importance is the number of data points that constitute an acceptable dataset for use in SSDs to derive GVs. The Aldenberg and Slob (1993) approach recommended that the minimum data requirement for chronic NOEC data should be for at least five different species from at least four different taxonomic groups, and this was adopted by ANZECC/ARMCANZ (2000). Clearly, as with any statistically based method, the precision of the derived GV will increase with more data points. However, in many site-specific investigations, both cost and the availability of appropriate test species often become limiting factors.

In the 2000 Guidelines, the derived GV was termed 'high reliability', provided the SSDs used chronic data and met the criterion of a minimum of 5 species from 4 taxonomic groups. The BurrliOZ software (Campbell et al., 2000) included in the 2000 Guidelines (ANZECC/ARMCANZ, 2000) did, however, warn users that fitting two and three parameter distributions on the basis of 8 or fewer toxicity values was statistically unsound. It is worth recognizing that statisticians would generally not support the notion of fitting 2- or 3-parameter distributions using fewer than about 30 data values. The limitations of this practice are acknowledged, but the reality is that such large data sets do not exist for most chemicals. It is within this ecotoxicological framework that the term 'high reliability' was used, however, no consideration was given to the goodness of fit, which in some cases was quite poor.

For the European Union, the recommended datasets to be considered 'reliable' (European Commission, 2011) are NOECs/EC10s from at least 10 species belonging to at least 8 taxonomic groups, although preferably they should include more than 15 species (European Commission, 2011). Specific guidance is provided on the taxa that would be expected to be representative. For example, for freshwater systems, where the toxicant does not have a specific mode of action, the following is expected:

- Fish (species frequently tested include salmonids, minnows, bluegill, sunfish, channel catfish, etc.);
- A second family in the phylum Chordata (e.g. fish, amphibian, etc.);
- A crustacean (e.g. cladoceran, copepod, ostracod, isopod, amphipod, crayfish etc.);
- An insect (e.g. mayfly, dragonfly, damselfly, stonefly, caddisfly, midge, etc.)
- A family in a phylum other than Arthropoda or Chordata (e.g. Rotifera, Annelida, Mollusca, etc.);
- A family in any order of insect or any phylum not already represented;
- Algae;
- Higher plants.

The Canadian Guidelines have quite specific data requirements with a required number of studies on fish, invertebrates and plants for both fresh and marine waters (CCME, 2007). From a tabulated list of minimum data required for specific taxonomic groups, it was anticipated that 'generally at least 10 to 15 data points should be available,', although it was acknowledged that on occasions less data may be acceptable to produce an adequate curve. A goodness of fit test was included as a requirement (CCME, 2007). For freshwaters the specific requirements were for:

- Three species of fish including at least one salmonoid and one non-salmonoid;
- Three aquatic or semi-aquatic invertebrates at least one of which must be a planktonic crustacean (it is desirable that one be a mayfly, caddisfly, or stonefly);
- At least one freshwater vascular plant or freshwater algal species;
- Toxicity data for amphibians are highly desirable.

The US EPA's method requires at least eight species from different families and prescribed taxa (Stephan et al., 1985). A Water Environment Research Fund (WERF) study (Parkhurst et al., 1996) claimed that eight data points were too few, while in one instance 14 values was deemed sufficient.

Rather than increasing the minimum data requirement for GVs derived by the SSD method and, thus, potentially reducing the number of chemicals for which this method can be used, it was decided to retain the minimum data requirements specified in the 2000 Guidelines. Thus, the minimum data requirements are toxicity data for at least five species that belong to at least four taxonomic groups (ANZECC/ARMCANZ, 2000). The following list can be used as a guide when considering taxonomically different organisms for the purpose of determining the minimum data requirements: fish, crustaceans, insects, molluscs, annelids, echinoderms, rotifers, hydra, diatoms, green algae, blue algae, red algae, macrophytes, blue-green algae (cyanobacteria), amphibians, bacteria (except *Photobacterium phosphoreum/Vibrio fischeri),* protozoans, corals and fungi. Note that, in most circumstances, a toxicity dataset should include data from across a range of taxonomic groups that capture aquatic vertebrates, invertebrates and plants. Exceptions to this could include datasets for toxicants such as herbicides, which are specifically targeted to being toxic to plants.

To overcome some of the limitation of small datasets, a more rigorous method for indicating the reliability of the resulting GVs was developed (Table 4, Section 8) and rules governing the types of statistical distributions that can be applied to toxicity data have been developed. It was decided not to be more prescriptive about the types of organisms that data were required for, as this violates one of the central assumptions of an SSD, that it represents a random sample of effects data that is designed to reflect the ranges of sensitivity in the receiving ecosystem (Warne, 1998).

The two-parameter log-logistic distribution should be used for datasets consisting of data for five to seven species that belong to at least four taxonomic groups, in preference to the three-parameter Burr distribution, so as to reduce the effects of 'over-fitting'. Using a two-parameter model instead of a three-parameter model reduces the uncertainty in the fitting process by not estimating more parameters than can be justified by the size of the sample. The log-logistic distribution was selected because it has a well-established role in ecotoxicology, performs well under a wide variety of conditions and is constrained to ensure it does not return negative values.

Assuming that toxicity data were available for species that belonged to at least 4 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.

In the 2000 Guidelines (ANZECC/ARMCANZ, 2000), a number of high reliability GVs were derived using eight or less toxicity data points. Examination of SSD plots showed that in a number of cases the curve fit was poor, and, in some instances two or more data points were obtained for the same contaminant concentration (usually as an artefact of using NOECs), yet the derived GV was classified as 'high reliability'. It is usual that the number of acute toxicity data far exceeds the number of chronic toxicity data. In the absence of sufficient chronic data to meet the new minimum data requirements, a combination of acute data (converted to chronic NOECs (EC10s) – refer to Section 8) and chronic data can be used to improve the robustness of the derived GVs. At the same time, the new data inclusion rules (Section 7) should be applied.

While the use of combined acute and chronic data was not allowed in the 2000 Guidelines, its value for small chronic datasets is now recognized and, as such, has been included in the revised derivation method. However, because the combination contains acute data converted to chronic, the GVs derived from the

amalgamated data should be classified as ‘moderate reliability’. These are now allowed where we have small chronic and larger acute datasets.

The conversion of acute data to chronic values remains a requirement given the greater number of acute data and their value in supplementing limited chronic data sets (when there are chronic toxicity data for less than 8 species that belong to at least four taxonomic groups) (Warne, 2001). The conversion factor from acute LC/IC/EC50 values to chronic NOEC values involved the application of an acute to chronic ratio (ACR) in the first instance or where this was not available, division by a default assessment factor of 10. This is as recommended in the 2000 Guidelines (Section 8.3.4.4). The same process is considered appropriate for the estimation of a chronic EC10, since as already stated, we are considering NOEC and IC/EC10 data as quantitatively equivalent, although this is strictly incorrect. A possible alternative method of combining chronic and acute data could involve conversion to estimates of chronic data using a Bayesian self-referential method. This method lets the data ‘speak for themselves’ whereby an optimal acute to chronic ratio (ACR) is identified that achieves that maximum harmonization of scaled acute data with chronic data (Fox, 2006). This method has potential but is not recommended here for deriving default GVs. When seeking such data, the Australasian Ecotoxicology Database should be consulted as it has compiled much of the ecotoxicity data for Australasian species or foreign species tested within Australasia up to 2009 (Warne et al., 1998, 1999; Markich et al., 2002; Langdon et al., 2009).

8 Toxicity Data Inclusion Rules and Data Quality Assessment

Data inclusion rules were documented in Sections 8.3.4.2 to 8.3.4.4 of the 2000 Guidelines (ANZECC/ARMCANZ, 2000) and in Warne (2001). There are no major changes proposed to the requirements for data quality, rather there are a number of minor refinements. The method to be used in assessing the quality of all new data (data not already assessed as part of the 2000 Guidelines or assessed by other recognised regulatory authorities) is presented in Hobbs et al. (2005). This method and the one that was used in the 2000 Guidelines are based on the method used in the AQUIRE (USEPA, 1994) database. Only 'high' (scoring 80-100) or 'acceptable' quality data sets (scoring 51-79) are deemed suitable for use in GV derivation. An electronic version of the data assessment methodology will be released as part of the revision process.

Table 2. Scoring system for assessing the quality of toxicity data for metals to freshwater non-plant organisms, to be used in the derivation of guideline values for toxicants (Zhang et al., 2015; modified from Hobbs et al., 2005)[a]

	QUESTION	MARK
1	Was the duration of the exposure stated (e.g. 48 or 96 h)?	Yes (10), No (0)
2	Was the biological endpoint (e.g. 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 (e.g. LC or NOEC)?	Yes (5), No (0)
4	Was the biological effect quantified (e.g. 50% effect, 25% effect)? **Note**: The effect for NOEC and LOEC data must be quantified.	Yes (5), No (0)
5	Were appropriate controls (e.g. 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)
8	Were the characteristics of the test organism (e.g. 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 (e.g. 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**: If the concentrations were not measured then automatically the data cannot be high quality.	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 (e.g., 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 should be measured at least at the beginning and end of the test	Measured and stated (3), measured once (1), Neither (0)

	QUESTION	MARK
16.2	Hardness	Measured and stated (3), measured (1), Neither (0)
16.3	Alkalinity	Measured and stated (3), measured (1), Neither (0)
16.4	Dissolved organic carbon concentration	Measured and stated (3), measured (1), Neither (0)
16.5	Dissolved oxygen	Measured and stated (3), measured (1), Neither (0)
16.6	Conductivity	Measured and stated (3), measured (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/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%,**	

[a] The sets of questions for other combinations of media/toxicant type/organism type for freshwater and marine species are in Warne et al. (2015)

The data quality assessment also needs to involve professional judgement, particularly where the experimental design is fundamentally flawed. For example, researchers may have measured and stated the pH, thereby scoring full marks, but the pH may have drifted by 3 units during the test period, or measured concentrations were stated (full marks) but showed a large loss of the toxicant during testing. In cases where the design is fundamentally flawed, the grade should be downgraded to 'unacceptable'.

New advice is now provided in relation to 'greater than' (>) values. For an individual species, any ">" value will be conservative for that species, but not necessarily for the ultimate GV. Greater than values are, however, permitted wherever they sit on the SSD curve, but with professional judgment being applied to evaluate whether they (i) sit too far outside the existing data range, and/or (ii) have an unacceptably large influence on the final GV, noting that this will be more likely where the value/s sit at the upper or lower end of the curve and for small datasets. Best professional judgement should be in the context of ensuring a sensible output. It is not appropriate to include ">" values in a group of species values for calculation of the species geometric mean. "Less than" (<) and "less than or equal to" (≤) values should not be included unless there are no other data for the species, the data point sits at the lower end of the SSD curve, and the

omission of the data would result in a less conservative GV. Again, best professional judgement should be applied and reasoning behind all decisions should be documented.

The pH range for freshwaters for which toxicity data have been generated at has been retained as between 6.5 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 pH is known to significantly affect toxicant bioavailability (e.g. 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. For example, some environments are naturally acidic in both New Zealand, e.g. West Coast brown-water streams (Collier and Winterbourn, 1987) and Australia, e.g. Alligator Rivers Region, NT (e.g. Klessa, 2000), in addition to those impacted by acid-sulfate soils.

General ranges of key water quality variables in fresh surface waters in Australia and New Zealand are summarised in Table 3. These are shown for indicative purposes, and may not fully represent the range of water quality in freshwaters in the two countries and, as such, have not been adopted as criteria for acceptance of toxicity data for GV derivation.

Table 3. Ranges for general water quality parameters in Australia and New Zealand

Parameter	Unit	Range (10th and 90th percentiles)	
		Australia[a]	New Zealand[b]
pH		6.0-7.6	7.2-8.2
Ca	mg Ca/L	0.4-45	4.2-17.4
Mg	mg Mg/L	0.5-30	0.7-3.6
Hardness	mg $CaCO_3$/L	3-236	16-56
DOC	mg/L	2-12	2.6-15[c]

[a] Data from WCA (2014; pH, n = 6418; Ca, n = 2622; Mg, n = 2653; DOC, n = 5196; hardness, calculated based on Ca and Mg concentrations). It should be noted that the data may not be fully representative of the ranges of these variables in Australian fresh surface waters.

[b] Data from Smith and Maasdam (1984).

[c] DOC for NZ estimated based on the regression relationship provided in Collier (1987).

8.1 Guideline value reliability

The 2000 Guidelines paid considerable attention to GVs derived using the assessment factor approach. In particular, GVs based on the application of an assessment factor of 10 to the lowest NOEC obtained from field or mesocosm studies were defined as high reliability. With SSDs representing a more robust method for GV derivation, coupled with the move away from NOECs as distinct from model-based estimates of ECx and ultimately NEC values, assessment factor-based derivations are not recommended. These are generally, but not necessarily, conservative and of unknown reliability.

In the past (ANZECC/ARMCANZ, 2000), there were two types of low reliability GVs: termed low reliability GVs (where there were data for a fish, an invertebrate and an alga) (Section 8.3.4.4) and low reliability environmental concern levels (ECLs) (which could be based on a single toxicity value) (Section 8.3.4.5). This was to highlight the fact that these types of GVs were likely to be very conservative and, at best, indicative of possible concern concentrations which, if exceeded, required further investigation. The terminology issue has created some confusion and it is recommended that the terms ‘environmental concern level’ and ‘interim indicative working level’ are no longer to be used in the revised Guidelines. Thus, GVs, be they default or site-specific, are to be referred to only by their reliability category.

As low and very low reliability GVs (see below) are typically calculated from insufficient datasets, they provide less confidence that aquatic ecosystems will be protected and should be recalculated once more data become available. They should not be used in the same way as high and moderate reliability GVs, and should only be used for interim guidance. The action that should result from exceedence of low and very low reliability GVs would be to search for, or generate, more data of sufficient quality. This might include data from other lines of evidence (e.g. field biological monitoring/assessment) rather than just additional toxicity data to improve the GV, which is consistent with the formalised integrated assessment approach being adopted for the revised Guidelines.

The reliability classifications to be applied to GVs derived using SSDs are listed in Table 4. In keeping with the discussions in Section 5, the minimum data requirements in Table 4 refer to single species laboratory-based testing, but these could be supplemented with data from microcosm, mesocosm or field studies where these meet the data quantity and quality requirements. This is especially appropriate for site-specific GV derivation. The hierarchy of data types considered is: all chronic data > mixed chronic and converted acute > acute data converted to chronic. The assignment of a reliability rating to the GVs is based on the data type, number of data points, and the adequacy of the fit (as 'good' or 'poor') of the SSD model. In the 2000 Guidelines, there were many instances where the GV was considered high reliability, but where the fit of the SSD model was obviously poor. A poor fit could include where a model does not appear to be a good representation of the data points either at the low end of the SSD, or of the entire dataset, but greater importance should always be placed on the fit of the distribution to the data in the lower portion of the SSD (i.e. 0 to 30% of species).

Examples of good and poor data fits are shown in Figure 1. Poor fits can be the result of a poor fit of the Burr or log-logistic distribution to the data points (Figure 1a), where the data appear not to be amenable to these distributions (e.g. potentially bimodal) (Figure 1b), or where there is an inadequate spread of the data (Figure 1c). The occurrence of the same NOEC concentration for several test organisms (Figure 1c) is most likely an artefact of the defined dilution/concentration series used for determining a NOEC value. The derivation and use of EC10 values would likely result in differing values for these organisms, especially if the determination followed the recommendation to include more measurements at concentrations the bottom end of the distribution.

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. 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.

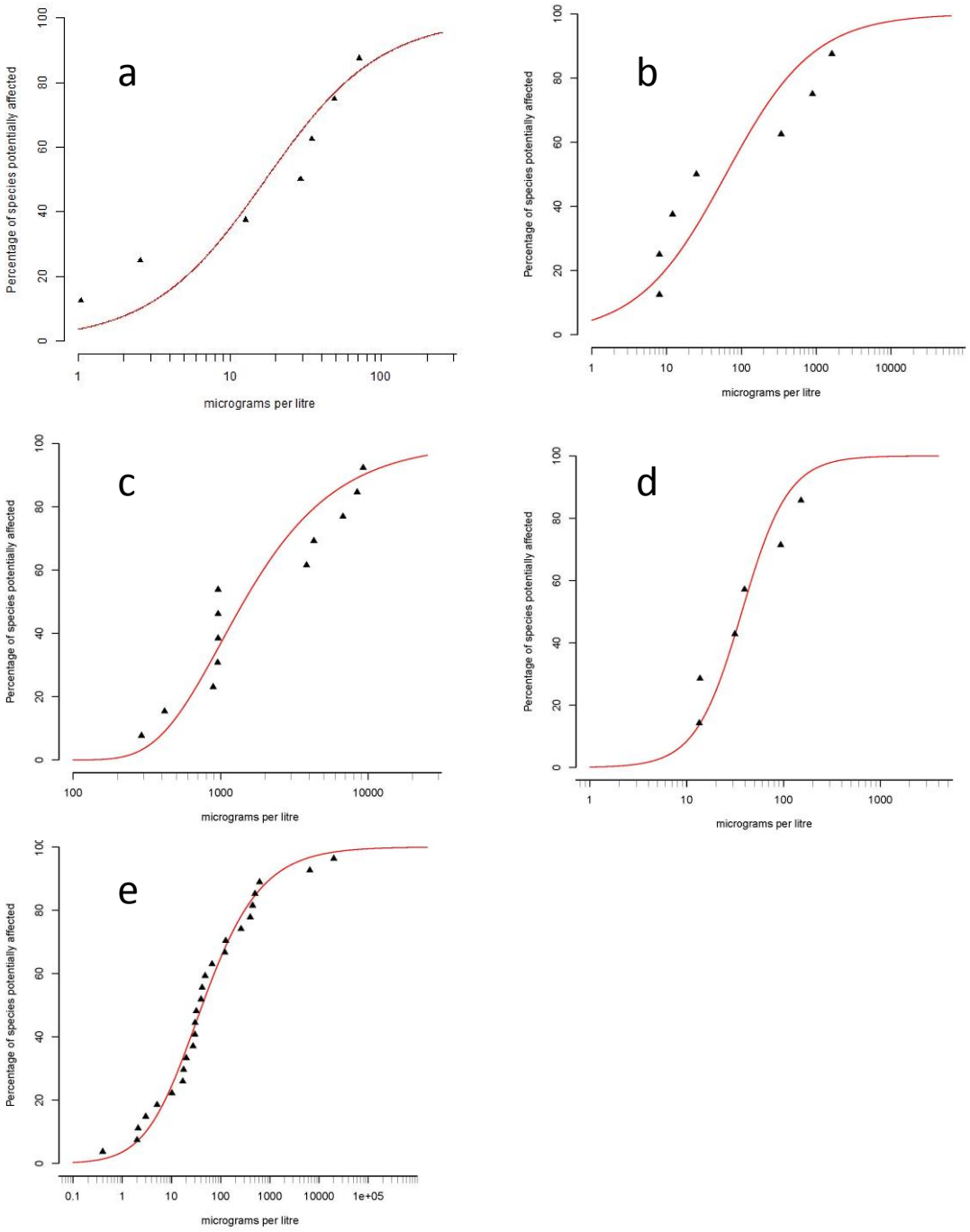

Figure 1. Examples of poor (a,b,c) and good (d,e) fits of data obtained using the revised BurrliOz software (log-logistic fits were obtained for datasets of <8, and Burr fits for ≥8)

The reliability scheme shown in Table 4 represents a technical evaluation of the level of confidence in a GV, which, when combined with other relevant knowledge (e.g. information on toxicant use and environmental concentrations), can help determine (i) the appropriateness of the GV, and (ii) priorities for future GV revisions. It is important to note that while the preferred GV derivation has a good fit to ≥15 data points, it is recognised that achieving such large datasets may not be possible in many instances, and that most derivations will use ≥8 data points. Bearing in mind that, in the 2000 Guidelines, 5 data points were sufficient for a high reliability GV, our re-evaluation recognises that this is rarely so, so these are now reclassified these as moderate reliability at best, and low reliability if the fit is poor.

Table 4. Classification of the reliability of derived guideline values

Sample size[a]	Data type	Adequacy of sample size	Adequacy of fit in SSD	Reliability
Guideline values derived using a species sensitivity distribution				
≥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	Adequate	Good	Moderate
	Mixed chronic and converted acute	Adequate	Poor	Low
5-7	Mixed chronic and converted acute	Adequate	Good	Moderate
	Mixed chronic and converted acute	Adequate	Poor	Low
≥15	Acute converted to chronic	Preferred	Good	Moderate
	Acute converted to chronic	Preferred	Poor	Low
8-14	Acute converted to chronic	Good	Good	Moderate
	Acute converted to chronic	Good	Poor	Low
5-7	Acute converted to chronic	Adequate	Good	Low
	Acute converted to chronic	Adequate	Poor	Very low
Guideline values derived using an assessment factor				
<5	Chronic, mixed chronic and acute, or converted acute	Inadequate	Not applicable	Unknown

[a] For the SSD method, the sample size is assumed to comprise data from at least 4 taxonomic groups.

8.2 Selection rules for toxicity data to be used in SSDs

The rules for manipulation of toxicity data to obtain a single value for each species have been changed from the 2000 Guidelines (ANZECC/ARMCANZ, 2000; Warne, 2001). The rules for manipulating toxicity are expressed below.

For each chemical the toxicity data that have passed the data quality assurance process should first be sorted by species, then endpoint and finally by duration of exposure. For each combination of species and endpoint the longest exposure duration is selected, unless the toxicity estimate from a shorter duration is lower, then the lower value should be chosen. If there are multiple values for the longest duration then calculate the geometric mean. Select the lowest resulting value for each combination of species and endpoint. The toxicity value for the most sensitive endpoint for each species should be adopted as the sensitivity value for that species. An example of the application of these rules to a dataset is presented in Table 5.

Table 5. Example of the application of data selection rules for a species.

SPECIES	ENDPOINT	DURATION (h)	EC10 (mg/L)	VALUE FOR EACH COMBINATION OF SPECIES, ENDPOINT AND DURATION	LOWEST VALUE FOR EACH COMBINATION OF SPECIES AND ENDPOINT	LOWEST VALUE FOR SPECIES
D. carinata	Growth	144	7	7	7	0.19
D. carinata	Immobilisation	144	10			
D. carinata	Immobilisation	144	5	5.3	5.3	
D. carinata	Immobilisation	144	3			
D. carinata	Reproduction	240	1.3			
D. carinata	Reproduction	240	2.0	1.3		
D. carinata	Reproduction	240	0.9			
D. carinata	Reproduction	480	0.2			
D. carinata	Reproduction	480	0.15	0.19	0.19	
D. carinata	Reproduction	480	0.24			

Where water quality may have significantly varied across the tests for some reason (e.g. 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 Section 8 and Table 3), then best professional judgment should again be applied to determine whether or not the toxicity value associated with that test should be included in the dataset.

It is worth noting that the European Commission (2011) does not use geometric means but rather defers to best professional judgement. However, the geometric mean in used by Canada (CCME, 2007) and the USA (Stephan et al., 1985) and in the 2000 Guidelines by Australia and New Zealand (ANZECC/ARMCANZ, 2000).

8.3 Dealing with toxicity modifying factors

In deriving some GVs, toxicity is known to be dependent on other physical and chemical factors. This is particularly so with metals, with a dependence on hardness, pH, alkalinity and/or dissolved organic carbon. Contaminants such as ammonia, cyanide, phenols etc., have a pH dependence (e.g. ANZECC/ARMCANZ, 2000). From a theoretical point of view the toxicity of many, if not most organic chemicals, are likely to be affected by a range of properties including pH, dissolved organic and total organic carbon content and the concentration of total suspended solids. However, to date, quantitative relationships between toxicity and these factors are not available for organic chemicals that could be used to calculate site-specific GVs.

In the 2000 Guidelines, derivations for metals were based as much as possible on data for a fixed pH, and hardness, with hardness corrections using accepted algorithms being used for cadmium, chromium, copper, lead, nickel and zinc (ANZECC/ARMCANZ, 2000; Markich et al., 2001). However, much of the work on which the algorithm for copper was derived 'confounded the effects of true water hardness (Ca/Mg), with alkalinity (carbonate concentration) and/or pH' (Markich et al., 2006). More recent work that examined the effect of true water hardness generally found that it did not affect copper toxicity to a variety of freshwater organisms including bacteria, a unicellular green alga, a macrophyte, a hydra, a bivalve, crustaceans and fish (Lauren and McDonald 1986; Pynnönen 1995; Reithmuller et al., 2000; Grosell and Wood, 2002; De Schamphelaere and Janssen, 2004; Hyne et al., 2005; Markich et al., 2005; Markich et al., 2006) while a few have found it had a significant but small effect (Erickson et al., 1996; Perschbacher and Wurts, 1999; Heijerick et al., 2005). Therefore, for the revised GV derivation method, it is recommended that the copper GV hardness correction algorithm is not used to tailor the GV to local water hardness.

Metal speciation is specifically considered in the assessment hierarchy against GVs for laboratory-based testing. The bioavailable fraction can be measured, modelled or determined by toxicity testing and the bioavailable concentration compared to the GV. This approach is promoted in the 2000 Guidelines and continues to be adopted.

A key development in the future, as has been done with the revised Australian sediment quality guidelines (Simpson et al., 2010) and the Australian Ecological Investigation Levels for contaminated sites (NEPC, 2013), should be to develop more of these relationships with toxicity modifying factors. An example of this is the work by van Dam et al (2012b) on the influence of dissolved organic carbon on uranium toxicity.

An alternative is to develop biotic ligand models (BLMs) as has been done for copper (e.g. Santore et al., 2001; De Schamphelaere and Janssen, 2002, 2004; USEPA, 2007), nickel (Keithly et al., 2004; Hoang et al., 2004), silver (Paquin et al. 1999), zinc (Heijerick et al., 2002a, b) and manganese (Peters et al. 2011). Currently, representatives of Australian regulatory agencies, academics and consultants are working collaboratively with the Nickel Producers Environmental Research Association (NiPERA) and WCA Environment Limited (England) to determine the validity of the nickel BLM, which was initially developed for Europe, to suit Australian conditions and biota (Peters et al., 2013).

9 Analytical Methods for Deriving Species Sensitivity Distributions (SSDs)

A species sensitivity distribution (SSD) is a probabilistic model of the distribution of the toxicity of particular contaminants to a defined range of species representing a specified number of taxonomic groups. It is used to determine a contaminant concentration that is theoretically protective of a given percentage (x) of species (i.e. the protective concentration to x% of species, PCx), often expressed in risk assessments as the concentration that is hazardous to a given percentage (y) of species (i.e. HCy, where y=100-x).

In Australia and New Zealand, it has been agreed that toxicant GVs should be based on the following levels of species protection:

- High conservation value ecosystems: 99% species protection (PC99; HC1);
- Slightly to moderately disturbed ecosystems: 95% species protection (PC95; HC5); and
- Highly disturbed ecosystems: 90% or 80% species protection (PC90; HC10 or PC80; HC20).

The PC99 is often an extrapolated value (outside the range of the toxicity data) and in these cases should be interpreted cautiously.

A number of statistical models are available for describing SSDs that include both parametric and non-parametric methods (Wheeler et al., 2002a; Zajdlik, 2006; Hickey and Craig, 2012). Commonly used parametric methods that have been used include the following distributions:

(i) Logistic
(ii) Log-logistic
(iii) Log-normal
(iv) Burr Type III
(v) Gompertz
(vi) Log-triangular

Given the small data sets often used to derive GVs, it has generally been found that any reasonable candidate distribution will not be rejected by a statistical test of goodness of fit due to low statistical power. However, there are a number of reasons for rejecting the log-triangular distribution (Warne, 1998).

In Australia and New Zealand, the Burr Type III distribution was adopted in 2000 as the preferred distribution to be fitted to the toxicity data. This was incorporated into the Burrlioz software (Campbell et al., 2000). This software has generally performed well, giving acceptable data fits, although in some cases there were problems with small datasets. There are statistical reasons why the 3-parameter Burr Type III distribution is inappropriate to deal with small datasets (5-7 species). Although not solving all of the problems associated with this, it is preferable to use a simpler model such as the 2-parameter log-logistic distribution for small datasets. Adopting this approach is likely to resolve some of the existing problems with small dataset SSDs. The revised SSD analytical approach described here has been incorporated into an updated version of the Burrlioz software for deriving GVs. This software will provide a 95% confidence interval around the PC95 or other selected GV. This is designed to provide a measure of statistical confidence around the PC values, and it is important to note that it is not intended that the lower range of the confidence interval be used as the GV.

The log-logistic distribution has a long and established role in ecotoxicology (e.g. Aldenberg and Slob, 1993). Even though there is no biological/ecological/environmental/ecotoxicological basis for choosing one statistical distribution over another, there is at least a precedent for defaulting to the Burr Type III distribution, and it does perform reasonably well under a wide variety of conditions. It cannot return negative PCx values because it is constrained to lie on the interval (0 to infinity) and, if we accept that

fitting any distribution with such small samples is not best statistical practice, then we should at least be careful not to overfit, i.e. using more parameters that can be justified by the sample.

There is also the option to apply non-parametric approaches to SSD fitting. A bootstrap regression method has been described by Grist et al. (2002). Bayesian methods have been described by Fox (2009) and Hayashi and Kashiwagi (2013). Such approaches will be explored in future revisions of toxicant GV derivation methods.

Of importance is the understanding of the precision afforded by the PCx estimates. The GVs reported in the 2000 Guidelines are the median of the distribution of PC95 estimates. Use of a lower 95% confidence interval (e.g. PC95,95) is inappropriate, particularly for small data sets, given the lack of precision at the tail of the confidence limit distribution coupled with that on the tail of the SSD distribution. Most jurisdictions recommend the use of a PC95,50. To simplify notation and comprehension by users, the PCx,50 values have been simplified to PCx values (where x values are commonly 99, 95, 90 or 80).

10 Guidance on Weighting of Data in SSDs

The method of deriving toxicant GVs by SSDs requires data for at least five species that belong to at least four taxonomic groups. Where there is more than one data point for the same exposure duration and species, a single value is used to represent each species (Warne (2001) and also see Section 8.1). In determining a single sensitivity value for each species, equal importance (weighting) is given to each species in calculating the SSD and the resulting GV. The way in which data from different species are dealt with in SSDs has received criticisms from several authors. Forbes and co-workers (Forbes and Forbes, 1993; Forbes and Calow, 2002) made the point that only a fraction of the species going into the SSD determines the effects threshold. With all species being weighted equally, the loss of any species is considered to be of equal importance to the ecosystem, and keystone, foundation or other important species are assumed to be randomly distributed in the SSD (although for site-specific GV derivation, this may not be the case). For example, they suggested that while an ecologically realistic distribution of species by trophic level was 64% primary producers, 26% herbivores (invertebrates) and 10% carnivores (fish), a mean ratio from typical SSDs for different chemicals was found to be 28, 35 and 38% respectively. Thus, fish were over-represented and primary producers under-represented. Re-sampling the species sensitivity database to reflect a more environmentally realistic distribution of trophic levels yielded a PC95 value 3.4 times lower for cadmium than when the data were given equal weighting (Forbes and Calow, 2002). By contrast, the PC95 value for linear alkylbenzene sulfonate (LAS) based on an environmentally realistic distribution of trophic level was 3.2 times higher than the PC95 when the same data were given an equal weighting (Forbes and Calow, 2002). One of the underlying assumptions of all SSD methods is that the species used in the SSD should be a random selection of the species in the ecosystem, however, our selection is typically biased by the availability of toxicity tests, which may be exaggerated further for small datasets. Targeting keystone taxa in the SSD and derived GV is highly recommended if tests are available.

Duboudin et al. (2004) followed this up by questioning (i) how to deal with the various test results that exist for the same species, and (ii) how to account for the variability in the available data for a particular taxonomic class (e.g. algae, invertebrates, vertebrates). They compared SSDs that used (i) the literature ratio of data in these classes (i.e. unweighted data), (ii) one that weighted these classes equally (using the geometric mean of multiple species data (the Australian and New Zealand approach); and (iii) one using the environmentally realistic ratio for these three species classes used by Forbes and Calow (2002), and observed up to an order of magnitude difference in GVs between the three data treatments. The choice of data (intra-species variation and proportions between taxonomic groups) was found to have more effect on the value of the PC95 than the statistical method used to construct the distribution (Duboudin et al. 2004). They recommended that rather than using the geometric mean of multiple data for the same species, to weight data to retain intra-species variation while giving each species the same weight within the SSD. This uses the entire set of data, so as not to favour one species over the others, however, when the number of data points is low, the PC95 can be dominated by a few large weighted data. This highlights the desirability of having large datasets that cover a range of species and taxonomic groups as discussed in Section 8.

Rather than using SSDs for all aquatic organisms, Brix et al. (2001) applied SSDs to individual taxonomic groups, for estimating the risks of dissolved copper to specific taxonomic groups. This is possible because there is quite a large dataset available for copper. The results were compared with a site-specific food web to determine whether key food web components are potentially at risk and whether the overall aquatic community is at risk in relation to ecosystem function. In this example, substantial risk was observed for a zooplankton–planktivorous fish–piscivorous fish food web, because cladocerans were shown to be at greatest risk. While such an approach is ideal, data limitations make it impractical for most applications and not appropriate for deriving default (i.e. national level) GVs.

Grist et al. (2006) discussed ways in which expert judgment can be incorporated into SSDs using a Bayesian approach that may be applicable to weighting approaches to data.

For toxicant GVs, where possible, input data should cover important species, but giving weighting based on natural abundance or other such factors and the validity of doing this requires more research before it can be recommended. The major limitation is usually one of insufficient data. There is generally insufficient information on which to make decisions on how to weight the data, so the current system of giving each species equal weighting is to be retained.

Implicit in the statistical distribution-fitting exercise is the untested, but surely false assumption that the data represents a truly random sample from the population of all species. How we accommodate our 'selection bias' in SSD-fitting, is another unresolved issue that requires further research. Weighting might be usefully considered for site-specific assessments and the calculation of site-specific GVs where the information needed to determine the weightings may be available, but as with all deviations from the national approach, the method must be transparent and scientifically valid.

11 Geographical/Climate-Specific Considerations for Species/Data Inclusion

Guideline values for Australia and New Zealand are typically derived from North American and European data because of the limited local database, particularly of chronic toxicity data. Australia, however, encompasses environments ranging from tropical to polar and whether there is a need for different GVs for each environment or climate is an issue of concern. There have been several attempts to compare sensitivities of temperate and tropical species (Leung et al., 2001; Wheeler et al., 2002b; Maltby et al., 2005; Kwok et al., 2007; Daam and van den Brink, 2009), and temperate, tropical and polar species (Chapman and Riddle 2005), using species sensitivity distributions. These have largely been confined to acute rather than chronic endpoints. Overall, the findings suggest that any differences in acute toxicity are dependent on both the chemical and the species being compared.

For freshwater species, Kwok et al. (2007) generalized that for most metals tropical species were less sensitive (for zinc the reverse was true), while for ammonia, phenol and some pesticides, tropical species were more sensitive (based on ratios of temperate to tropical EC10 values). Based on applying various SSD methods to the ratios of temperate to tropical EC10 data they recommended that temperate GVs be divided by 10 if being used for tropical and sub-tropical regions. Based on the latest most comprehensive data for marine species, PC90 temperate/tropical ratios of 0.5, 1.7, 0.8, 1.5, and 1.6 for cadmium, copper, zinc, pentachlorophenol and phenol were reported, while the ratio for tributyltin was 8.5 (Wang et al., 2013). The variable data serve to point out that there is no consistent relationship between tropical and temperate GVs, largely because the species sample sizes are so small.

Comparative testing of species must involve conditions of optimum health, where the organisms are unstressed by temperature. There are a number of factors that might affect species response at different temperatures. Rates of metabolism, detoxification and elimination typically increase with temperature, (Daam and van den Brink, 2009). Studies of fish species reported sensitivities to DDT (Dyer et al., 1997) in the order coldwater > temperate > tropical species, but for many other pesticides there were no significant differences (Dyer et al., 1997, Maltby et al., 2005). A comparison of the acute toxicity of oil components to polar and temperate marine species found that except for naphthalene, toxicities (as HC50s) did not differ significantly (de Hoop et al., 2011). Even the HC5 values for naphthalene were comparable. There have so far been no comparisons based on chronic endpoints.

It was noted that US and Canadian regulators had exclusion rules that did not accept data from other geographic regions, although the Canadian protocol now accepts data from non-resident species, as long as they can be considered as appropriate surrogates of local species (CCME, 2007). Excluding international data would have unacceptable consequences for our GV derivations. For Australia and New Zealand, it was therefore decided to not derive geographic or climate-specific GVs in the 2000 Guidelines or in the current revision. Where such GVs are required, particularly for tropical and Antarctic ecosystems, it is recommended that these should be derived using site-specific (or climate-specific) data.

The same data requirements that apply to the default GVs would apply to these site-specific GVs. In general, tropical and polar species are under-represented in the Australian and New Zealand species datasets, and obtaining new data for these climatic regions is encouraged.

An example of site-specific, or regional, GVs in this category are those derived to protect Australia's Great Barrier Reef (GBRMPA, 2010). In particular, GVs were derived for pesticides not covered by the 2000 Guidelines, mostly with a desired level of protection of 99%. Consideration was given to the effects of the sub-lethal endpoint of suppression of photosynthesis of corals on calculated GVs, however, at this stage, GVs were set without the sub-lethal responses included (refer to section 4).

Particular problems are experienced in deriving GVs for polar regions, where very long exposure durations are required to reach mortality in toxicity testing, or, if this is to be reached within a reasonable test duration of 2-3 weeks, require often unrealistically high concentrations of contaminants. Here, data based on behavioural endpoints might constitute a large proportion of the toxicity database, and the links to ecological relevance are not readily demonstrated. In the absence of sufficient other data, these would be admissible for region-specific guideline derivation, as the only means by which GVs could be realistically achieved.

On a related area of data availability, there are generally more freshwater toxicity data than data for marine species, and it has been advocated that in such cases, freshwater data could be included in the marine database. Leung et al. (2001) showed that for many metals, the agreement was remarkably good between SSDs based only on marine or only on freshwater data. Where the marine dataset is small (less than 8 species), the addition of data for freshwater species may be considered to improve the reliability of the GV derivation. An alternative approach, recommended in the 2000 Guidelines, would be to adopt the freshwater GV as a low reliability marine GV. However, using either of the above approaches would require consideration of speciation issues.

12 Uncertainties Associated with Guideline Values

There are a number of sources of uncertainty in GV derivations that are often overlooked in defining the precision of the derived value. These arise either from uncertainties associated with the toxicity test procedure, or from uncertainties in the model used to derive the GVs.

The variability in toxicity testing, both intra- and inter-laboratory, can come from many sources, including differing sensitivities of the test organism (within and between laboratories), differences in response of individuals of a test species; differing laboratory procedures and practices, analytical errors associated with measuring low contaminant concentrations, losses on container surfaces, analytical errors, errors in measuring the organism response, errors in replication, changes in chemical form of the toxicant, and choice of dilution series.

A comprehensive study by Moore et al. (2000), found within-laboratory coefficients of variation (CVs) on EC25 values for *Ceriodaphnia dubia* exposed to a sodium chloride reference toxicant, of between 16 and 56% irrespective of whether the endpoint was mortality or reproduction. For the fish, *Menidia beryllina* exposed to a copper reference toxicant, CVs ranged from 24-61% for mortality and 24-122% for biomass. Inter-laboratory CVs were 13-17% for the *Ceriodaphnia* and 66-177% for the *Menidia*. Such variability is consistent with other published studies (e.g. Thellen et al, 1980; Warren-Hicks et al., 2000). Derived NOEC values that make up the bulk of our existing toxicity database, will be critically dependent on the dilution series adopted, it is not meaningful to talk of errors or uncertainties in the NOEC since it is simply a numeric label derived from multiple-comparison testing procedures (Fox, 2012).

To minimize variability, good quality assurance and quality control of toxicity testing is required including: (i) analysis of all toxicant concentrations used in the toxicity test at the beginning and end of the test; (ii) use of a reference toxicant; (iii) appropriate control and measurement of experimental conditions, e.g. temperature, pH, etc.; (iv) quality control of the test containers including water changes, organism feeding etc; and (v) a reproducible control response.

The variability associated with the use of ECx values, will reduce by increasing the number of test concentrations in the concentration-response curve, particularly at the lower end.

Toxicity data with their own variability are assembled into a model to calculate an appropriate GV. Here there are uncertainties associated with the chosen models, including model parameterisation, different methods of estimation and stochastic variation. The revised BurrliOZ software allows plotting of the 95% confidence intervals around the PCx value reflecting the goodness of fit of the distribution to the species sensitivity data (but not taking into account uncertainty in the toxicity estimates themselves). The wide range of these confidence intervals will come as a surprise to many. Their magnitude is highly dependent on the size of the datasets, the nature of the statistical model used, and assumptions about the random error term.

It is important to understand the sensitivity of the derived GVs to the data values. For example, using the available 43-point database for chromium (VI) in marine waters (ANZECC/ARMCANZ, 2000) it was shown that the PC95 could vary by more than a factor of 2 by choosing the lowest data rather than the geometric mean of replicates (Batley, pers. comm.). Adding points to the low concentration end of the dataset had the expected effect of lowering the GV, while adding data points to the high concentration end significantly increased the GV in this case, although in other cases a lowering of the GV has been observed. Our choice of test organisms is often biased to cultured in the laboratory, but is this an acceptable bias? While these are usually representative of naturally occurring species, the relevance of laboratory-based tests to field conditions remains an issue.

It is important to recognise that the EC10 which we are promoting as a more reliable endpoint than the NOEC, represents a 10% chronic sub-lethal effect on the test organism (with its associated confidence limits). We are then using these values to derive a PC95. The question might be raised as to whether 10%

chronic sub-lethal toxicity to 5% of species is ecologically acceptable. This is primarily a question of population viability. Most populations can withstand a 10% decrease in productivity (at least fisheries management and most other ecosystem resource use is based on this assumption). The assumption of the GV derivation method is that the test species are adequately representative of the sensitivities of other ecosystem constituents such that at most 5% of taxa in the ecosystem might suffer some population viability impairment. If that assumption holds, it is a more conservative position than that of most resource management decisions, and implies that most aspects of ecosystem function will be minimally impaired, if at all.

Note that the uncertainties in estimating a 99% species protection (PC99) GV from an SSD are larger than that for the PC95 because there is less confidence in the SSD model fit at the extremes of the distribution compared to the middle. This is especially so for small data sets. Professional judgement should be applied in assessing the extent to which the extrapolation of the SSD to a PC99 value generates an acceptable number. Considerations might include whether the GV concentration is consistent with accepted background concentrations, and the lowest toxicity estimates. Similarly, there are generally greater uncertainties associated with EC5 values than EC10 values if they are determined using the same experimental design.

These issues are being raised in order to encourage awareness among users of the uncertainties associated with GVs. To this end, the combined uncertainties should be recognised in how GVs are reported. These should be refined to no more than two significant figures, taking into account the standard rounding-off rules, e.g. a GV of 237 µg/L should be reported as 240 µg/L.

It is important to recognise that uncertainties in GVs become less important in the context of the recommended use of multiple lines of evidence for evaluating toxicant impacts. Thus, toxicant concentrations near a low or moderate reliability GV might indicate concerns if toxicity can be demonstrated, if there is significant bioaccumulation, and there are effects on ecosystem abundance and diversity. If there is counter-evidence any observed toxicity might be attributed to another unidentified source.

13 References

Aldenberg, T. and Slob, W. (1993). Confidence limits for hazardous concentrations based on logistically distributed NOEC toxicity data. Ecotoxicol. Environ. Saf., 25, 48–63.

ANZECC/ARMCANZ (2000). Australian and New Zealand guidelines for fresh and marine water quality. Australian and New Zealand Environment and Conservation Council/Agricultural and Resource Management Council of Australia and New Zealand, Canberra, ACT, Australia.

Brix, K.V., DeForest, D.K. and Adams, W.J. (2001). Addressing acute and chronic copper risks to freshwater aquatic life using species sensitivity distributions for different taxonomic groups. Environ. Toxicol. Chem. 20, 1846–1856.

Cantin, N.E., Negri, A.P. and Willis, B.L. (2007). Photoinhibition from chronic herbicide exposure reduces reproductive output of reef-building corals. Mar. Ecol. Prog. Ser., 344, 81–93.

CCME (2007). A protocol for the derivation of water quality guidelines for the protection of aquatic life. Canadian Council of Ministers of the Environment, 37 pp.

Chapman, J.C., Warne, M., St.J. and Patra, R.W.R. (2001). Considerations when applying the revised toxicant guidelines. Australas. J. Ecotox., 7, 157–174.

Chapman, P.M. and Riddle, M.J. (2005). Toxic effects of contaminants in polar marine environments. Environ. Sci. Technol., 39, 200-206.

Chapman, P.M., Cardwell, R.S. and Chapman, P.F. (1996). A warning: NOECs are inappropriate for regulatory use. Environ. Toxicol. Chem., 15, 77–79.

Clements, W., Cadmus, P. and Brinkman, S. (2013). Responses of aquatic insects to Cu and Zn in stream microcosms: understanding differences between single species tests and field responses. Environ. Sci. Technol., 47, 7506-7513.

Collier, K.J. (1987). Spectrophotometric determination of dissolved organic carbon in some South Island streams and rivers (note). N.Z. J. Mar. Freshwater Res., 21, 349-351.

Collier, K.J. and Winterbourn, M.J. (1987). Faunal and chemical dynamics of some acid and alkaline New Zealand streams. Freshwater Biol., 18, 227-240.

Cormier, S.M. and Suter, G.W. (2013). A method for deriving water–quality benchmarks using field data. Environ. Toxicol. Chem., 32, 255–262.

Crane, M., Kwok, K.W.H., Wells, C., Whitehouse, P. and Lui, G.C.S. (2007). Use of field data to support European Water Framework Directive Quality Standards for dissolved metals. Environ. Sci. Technol., 41, 5014–5021.

Daam, M.A. and Van den Brink, P.J. (2010). Implications of differences between temperate and tropical freshwater ecosystems for the ecological risk assessment of pesticides. Ecotoxicology 19, 24–37.

de Hoop, L., Schipper, A.M., Leuven, R.S.E.W., Huijbregts, M.A.J., Olsen, G.H., Smit, M.G.D. and Hendriks, A.J. (2011). Sensitivity of polar and temperate marine organisms to oil components. Environ. Sci. Technol. 45, 9017–9023.

De Schamphelaere, K., Heijerick, D. and Janssen, C. (2002). Refinement and field validation of a biotic ligand model predicting acute copper toxicity to *Daphnia magna*. Compar. Biochem. Physiol. C, 133, 243–258.

De Schamphelaere, K.A.C. and Janssen, C.R. (2004). Development and field validation of a biotic ligand model predicting chronic copper toxicity to *Daphnia magna*. Environ. Toxicol. Chem., 23, 1365–1375.

Duboudin, C., Ciffroy, P. and Magaud, H. (2004). Effects of data manipulation and statistical methods on species sensitivity distributions. Environ. Toxicol. Chem., 23, 489–499.

Dyer, S.D., Belanger, S.E. and Carr, G.J. (1997). An initial evaluation of the use of Euro/North American fish species for tropical effects assessments. Chemosphere, 35, 2767–2781.

Environment Canada (2005). Guidance document on statistical methods for environmental toxicity tests. Environment Canada Environmental Protection Series EPS 1/RM/46, Ottawa, Ontario, Canada

Erickson, R.J., Benoit, D.A., Mattson, V.R., Nelson, H.P. and Leonard, E.N. (1996). The effects of water chemistry on the toxicity of copper to fathead minnows. Environ. Toxicol. Chem., 15, 181–193.

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.

Forbes, T.L. and Forbes, V.E. (1993). A critique of the use of distribution-based extrapolation models in ecotoxicology. Funct. Ecol., 7, 249–254.

Forbes, V.E. and Calow, P. (2002). Species sensitivity distributions: a critical appraisal. Human Ecol. Risk Assess., 8, 473–492.

Fox, D.R. (2006) Statistical issues in ecological risk Assessment. Human Ecol. Risk Assess., 12, 120-129.

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

Fox, D.R. and Billoir, E. (2011). Individual versus population effects in concentration-response modeling. Integr. Environ. Assess. Manag., 7, 501–502.

Fox, D.R., Billoir, E., Charles, S., Delignette-Muller, M.L. and Lopes, C. (2012). What to do with NOECs/NOELs – prohibition or innovation? Integr. Environ. Assess. Manag., 8, 764–766.

GBRMPA (2010). Water quality guidelines for the Great Barrier Reef Marine Park, Great Barrier Reef Marine Park Authority, Townsville, Qld, 109 pp.

Giesy, J.P., Solomon, K.R., Coates, J.R., Dixon, K.R., Giddings, J.M. and Kenaga, E.E. (1999). Chlorpyrifos: ecological risk assessment in North American aquatic environments. Rev. Environ. Contam. Toxicol. 160, 1–129.

Green, J.W., Springer, T.A. and Staveley, J.P. (2013). The drive to ban the NOEC/LOEC in favor of ECx is misguided and misinformed. Integr. Environ. Assess. Manag., 9, 12–16.

Grist, E.P.M., Leung, K.M.Y., Wheeler, J.R. and Crane, M. (2002). Better bootstrap estimation of hazardous concentration thresholds to protect biological assemblages. Environ. Toxicol. Chem., 21, 1515–1524.

Grist, E.P.M., O'Hagan, A., Crane, M., Sorokin, N., Sims, I. and Whitehouse, P. (2006). Comparison of frequentist and Bayesian freshwater species sensitivity distributions for chlorpyrifos using time-to-event analysis and expert elicitation. Environ. Sci. Technol., 40, 295–301.

Grosell, M. and Wood, C.M. (2002). Copper uptake across rainbow trout gills: mechanisms of apical entry. J. Exp. Biol., 205, 1179–1188.

Hahn, T., Diamond, J., Dobson, S., Howe, P., Kielhorn, J., Koennecker, G., Lee-Steere, C., Mangelsdorf, I., Schneider, U., Sugaya, Y., Taylor, K., Van Dam R. and Stauber, J.L. (2014). Predicted no effect concentration derivation as a significant source of variability in environmental hazard assessments of chemicals in aquatic systems: An international analysis. Integr. Environ. Assess. Manag. , 10, 30-36.

Heijerick, D.G., Bossuyt, B.T.A., De Schampelaere, K.A.C., Ineherberg, M., Mingazzini, M. and Janssen, C.R. (2005). Effect of varying physicochemistry of European surface waters on the copper toxicity to the green alga *Pseudokirchneriella subcapitata*. Ecotoxicology 14, 661–670.

Heijerick, D.G., De Schamphelaere, K.A.C. and Janssen, C.R. (2002a). Biotic Ligand Model development predicting Zn toxicity to the alga *Pseudokirchneriella subcapitata*: possibilities and limitations. Compar. Biochem. Physiol. C, 133, 207–218.

Heijerick, D.G., De Schamphelaere, K.A.C. and Janssen, C.R. (2002b). Predicting acute zinc toxicity for *Daphnia magna* as a function of key water chemistry characteristics: Development and validation of a Biotic Ligand Model. Environ. Toxicol. Chem., 21, 1309–1315.

Hickey, C.W. and Golding, L.A. (2002). Response of macroinvertebrates to copper and zinc in a stream mesocosm. Environ. Toxicol. Chem., 21, 1854-1863.

Hickey, C.W., Golding, L.A., Martin, M.L. and Croker, G.C. (1999). Chronic toxicity of ammonia to New Zealand freshwater invertebrates: a mesocosm study. Arch. Environ. Contam. Toxicol. 37,338-351.

Hickey, G.L. and Craig, P.S. (2012). Competing statistical methods for the fitting of normal species sensitivity distributions: recommendations for practitioners. Risk Anal., 32, 1232–1243.

Hoang, T.C., Tomasso, J.R. and Klaine, S.J. (2004). Influence of water quality and age on nickel toxicity to fathead minnows (*Pimephales promelas*). Environ. Toxicol. Chem., 23, 86–92.

Hobbs, D.A., Warne, M.St.J. and Markich, S.J. (2005). Evaluation of criteria used to assess the quality of aquatic toxicity data. Integr. Environ. Assess. Manag., 1, 174–180.

Hoekstra, J.A. and Van Ewijk, P.H. (1993). The bounded effect concentration as an alternative to the NOEC. Sci. Total Environ., 134, Supplement 1, 705–711.

Jager, T. (2012). Bad habits die hard: the NOECs persistence reflects poorly on ecotoxicology. Environ. Toxicol. Chem., 31, 228–229.

Keithly, J., Brooker, J.A., DeForest, D.K., Wu, B.K. and Brix, K.V. (2004). Acute and chronic toxicity of nickel to a cladoceran (*Ceriodaphnia dubia*) and an amphipod (*Hyalella azteca*). Environ. Toxicol. Chem. 23, 691–696.

King C.K. and Riddle M.J. (2001). Effects of metal contaminants on the development of the common Antarctic sea urchin *Sterechinus neumayeri* and comparisons of sensitivity with tropical and temperate echinoids. Mar. Ecol. Prog. Ser., 215, 143-154.

Klessa, D.A. (2000). The chemistry of Magela Creek: A baseline for assessing change downstream of Ranger. Supervising Scientist Report 151, Supervising Scientist, Darwin.

Kwok, K.W.H., Bjorgesaeter, A., Leung, K.M.Y., Lui, G.C.S., Gray, J.S., Shin, P.K.S. and Lam, P.K.S. (2008). Deriving site-specific sediment quality guidelines for Hong Kong marine environments using field-based species sensitivity distributions. Environ. Toxicol. Chem., 27, 226–234.

Kwok, K.W.H., Leung, K.M.Y., Lui, G.S.G., Chu, V.K.H., Lam, P.K.S., Morritt, D., Maltby, L., Brock, T.C.M., Van den Brink, P.J., Warne, M.St.J. and Crane, M. (2007). Comparison of tropical and temperate freshwater animal species' acute sensitivities to chemicals: implications for deriving safe extrapolation factors. Integr. Environ. Assess. Manag., 3, 49–67.

Landis, W.G. and Chapman, P.M. (2011). Well past time to stop using NOELs and LOELs. Integr. Environ. Assess. Manag., 7, vi–viii.

Langdon, K., Warne, M.St.J. and Sunderam, R.I.M. (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.

Lauren, D.J., and McDonald, D.G. (1986). Influence of water hardness, pH, and alkalinity on the mechanisms of copper toxicity in juvenile rainbow trout, *Salmo gairdneri*. Can. J. Fish. Aquat. Sci., 43, 1488–1496.

Leung, K.M.Y., Morritt, D., Wheeler, J., Whitehouse, P., Sorokin, N., Toy, R., Holt, M. and Crane, M. (2001). Can saltwater toxicity be predicted from freshwater data? Mar. Pollut. Bull., 42, 1007–1013.

Magnusson, M., Heimann, K. and Negri, A.P. (2008). Comparative effects of herbicides on photosynthesis and growth of tropical estuarine microalgae. Mar. Pollut. Bull., 56, 1545–1552.

Maltby, L., Blake, N., Brock, T.C.M. and Van den Brink, P.J. (2005). Insecticide species sensitivity distributions: importance of test species selection and relevance to aquatic ecosystems. Environ. Toxicol. Chem., 24, 379–3 88.

Markich, S.J., Batley, G.E., Stauber, J.L., Rogers, N.J., Apte, S.C., Hyne, R.V., Bowles, K.C., Wilde, K.L. and Creighton, N.M. (2005). Hardness corrections for copper are inappropriate for protecting sensitive freshwater biota. Chemosphere, 60, 1-8.

Markich, S.J., Brown, P.L., Batley, G.E., Apte, S.C. and Stauber J.L. (2001). Incorporating metal speciation and bioavailability into water quality guidelines for protecting aquatic ecosystems. Australas. J. Ecotox. 7, 109–122.

Markich, S.J., King, A.R. and Wilson, S.P. (2006). Non-effect of water hardness on the accumulation and toxicity of copper in a freshwater macrophyte (*Ceratophyllum demersum*): How useful are hardness-modified copper guidelines for protecting freshwater biota? Chemosphere, 66, 1791–1800.

Markich, S.J., Warne, M.St.J., Westbury, A-M. and Roberts, C.J. (2002). A compilation of toxicity data for chemicals to Australasian species. Part 3: Metals and Metalloids. Australas. J. Ecotox., 8, 1–137.

Melvin, S.D. and Wilson, S.P. (2013). The utility of behavioral studies for aquatic toxicology testing: A meta-analysis. Chemosphere, 93, 2217-2223.

Moore, D.R.J., Warren-Hicks, W., Parkhurst, B.R., Teed, R.S., Baird, R.B., Berger, R., Denton, D.I. and Pletl, J.J. (2000). Intra- and intertreatment variability in reference toxicant tests: implications for whole effluent toxicity testing programs. Environ. Toxicol. Chem., 19, 105–112.

NEPC (2013). Methodology to derive Ecological Investigation Levels in Contaminated Soils. Schedule B5b. National Environment Protection Council, Adelaide, South Australia, 86 pp.

Newman M.C. (2010). Acute and chronic lethal effects to individuals. In: *Fundamentals of Ecotoxicology,* CRC Press, Boca Raton, Florida, USA, pp. 247–272.

Newman, M.C. (2008). What exactly are you inferring? A closer look at hypothesis testing. Environ. Toxicol. Chem., 27, 1013–1019.

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

Paquin, P.R., Di Toro, D.M., Santore, R.C., Trivedi, D. and Wu, K.B. (1999). A biotic ligand model of the acute toxicity of metals: III. Application to fish and *Daphnia magna* exposure to silver. US Government Printing Office: Washington DC, 1999. EPA 822-E-99-001.

Parkhurst, B.R., Warren-Hicks, W., Cardwell, R.D., Volosin J., Etchison, T., Butcher, J.B. and Covington, S.M. (1996). Methodology for aquatic ecological risk assessment. Water Environment Research Foundation Report No. RP91-AER-1, Alexandria, VA, USA.

Perschbacher, P.W. and Wurts, W.A. (1999). Effects of calcium and magnesium hardness on acute copper toxicity to juvenile channel catfish, *Ictalurus punctatus*. Aquaculture, 172, 275–280.

Peters, A., Lofts, S., Merrington, G., Brown, B., Stubblefield, W. and Harlow, K. (2011). Development of biotic ligand models for chronic manganese toxicity to fish, invertebrates, and algae. Environ. Toxicol. Chem., 30, 2407–2415.

Peters, A., Merrington, G., Schlekat,C., Stauber, J.L., Van Dam, R., Batley, G.E., Smith, R., Harford, A., Chapman, J.C., Warne, M.St.J., Hickey, C. and Glazebrook P. (2013). Implementation of a bioavailability approach for nickel in Australian surface waters. Poster presentation to SETAC, Edinburgh, Scotland.

Pynnönen, K. (1995). Effect of pH, hardness and maternal pre-exposure on the toxicity of Cd, Cu and Zn to the glochidial larvae of a freshwater clam *Anodonta cygnea*. Water Res., 29, 247–254.

Riethmuller, N., Markich, S.J., van Dam, R.A. and Parry, D. (2000). The effect of true water hardness and alkalinity on the toxicity of copper and uranium to two tropical Australian freshwater organisms. SSR/155. Supervising Scientist, Canberra, Australia. <www.deh.gov.au/ssd/publications/ssr/155.html>.

Santore, R.C., Di Toro, D.M., Paquin, P.R., Allen, H.E. and Meyer, J.S. (2001). A biotic ligand model of the acute toxicity of metals. II. Application to acute copper toxicity in freshwater fish and *Daphnia*. Environ. Toxicol. Chem., 20, 2397–2402.

Shieh, J-N., Chao, M-R. and Chen, C-Y. (2001). Statistical comparisons of the no-observed-effect concentration and the effective concentration at 10% inhibition (EC10) in algal toxicity tests. Water Sci. Technol., 43, 141–146.

Simpson, S.L., Batley, G.E. and Chariton, A.A. (2010). Revision of the ANZECC/ARMCANZ Sediment Quality Guidelines. CSIRO Land and Water Science Report 08/07. August 2008, Revised July 2010. Report prepared for the Department of Environment, Water, Heritage and the Arts. 115 pp.

Smith, D.G. and Maasdam, R. (1994). New Zealand National River Water Quality Network.1. Design and physico-chemical characterisation. N.Z. J. Mar. Freshwater Res., 28, 19-35.

Stephan, C.E., Mount, D.I., Hansen, D.J., Gentile, J.R., Chapman, G.A. and Brungs, W.A. (1985). Guidelines for deriving numerical national water quality criteria for the protection of aquatic organisms and their uses (2010). US Environmental Protection Agency Office of Research and Development Report No PB85-227049, Duluth, Minnesota, USA, 59 pp.

Thellen, C., Blaise, C., Roy, Y. and Hickey, C. (1989). Round robin testing with the *Selenastrum capricornutum* microplate toxicity assay. Hydrobiologia, 188/189, 259–268.

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

USEPA. (2007). Aquatic life ambient freshwater quality criteria - copper. 2007 Revision. United States Environmental Protection Agency, Criteria and Standards Division Report No. EPA-822-R-07-001, Washington DC, USA

USEPA. (2009). 2009 update of aquatic life ambient water quality criteria for ammonia - freshwater (Draft). United States Environmental Protection Agency, Office of Water Report No EPA 822-R-99-014, Washington DC, USA.

van Dam, R.A., Harford, A.J. and Warne, M.St.J. (2012a). Time to get off the fence: the need for definitive international guidance on statistical analysis of ecotoxicity data. Integr. Environ. Assess. Manag., 8, 242–245.

van Dam, R.A., Humphrey, C.L., Harford, A.J., Sinclair, A., Jones, D.R., Davies, S. and Storey, A.W. (2013). Site-specific water quality guidelines: 1. Derivation approaches based on physico-chemical, ecotoxicological and ecological data. Environ. Sci. Poll. Res. DOI: 10.1007/s11356-013-1780-0.

van Dam, R.A., Trenfield, M.A., Markich, S.J., Harford, A.J., Humphrey, A.C. and Stauber, J.L. (2012b). Re-analysis of uranium toxicity data for freshwater organisms and the influence of dissolved organic carbon. Environ. Toxicol. Chem., 31, 2606–2614.

van der Hoeven, N., Noppert, F. and Leopold, A. (1997). How to measure no effect. Part I: towards a new measure of chronic toxicity in ecotoxicology. Introduction and workshop results. Environmetrics, 8, 241–248.

Wang, Z., Kwok, K.W.H., Lee, J-S. and Leung, K.M.Y. (2011). Acute sensitivities distribution to chemicals: implication for deriving safe extrapolation factor. Poster SETAC Asia/Pacific meeting Hong Kong.

Warne, M.St.J and Westbury, A-M. (1999). A compilation of toxicity data for chemicals to Australasian species. Part II: Organic Chemicals. Australas. J. Ecotox., 5, 21–85.

Warne, M.St.J. (1998). Critical review of methods to derive water quality guidelines for toxicants and a proposal for a new framework. Supervising Scientist Report 135, Supervising Scientist, Canberra, ACT, Australia. ISBN 0 642 24338 7. 82pp.

Warne, M.St.J. (2001). Derivation of the ANZECC and ARMCANZ water quality guidelines for toxicants. Australas. J. Ecotox., 7, 123–136.

Warne, M.St.J. and van Dam, R. (2008). NOEC and LOEC data should no longer be generated or used. Australas. J. Ecotox., 14, 1–5.

Warne, M.St.J., Batley, G.E., Braga, O., Chapman, J.C., Fox, D.R., Hickey, C.W., Stauber, J.L., and van Dam, R.A. (2013). Revisions to the derivation of the Australian and New Zealand guidelines for toxicants in fresh and marine waters. Environ. Sci. Pollut. Res., 21, 51-60.

Warne, M.St.J., Batley, G.E., Van Dam, R.A., Chapman, J., Fox, D., Hickey, C. and Stauber, J. (2014). Revised method for deriving Australian and New Zealand Water Quality Guideline Values for Toxicants. Prepared for the Council of Australian Government's Standing Council on Environment and Water (SCEW). Department of Science, Information Technology, Innovation and the Arts, Brisbane, Qld.

Warne, M.St.J., Westbury, A-M. and Sunderam, R. (1998). A compilation of toxicity data for chemicals to Australasian aquatic species. Part 1: Pesticides. Australas. J. Ecotox. 4, 93–144.

Warren-Hicks, W., Parkhurst, B.R., Moore, D.R.J., Teed, R.S., Baird, R.B., Berger, R., Denton, D.I. and Pletl, J.J. (2000). Assessment of whole effluent toxicity test variability: partitioning sources of variability. Environ. Toxicol. Chem., 19, 94–104.

WCA (2014). Assessing the geochemical variability of freshwaters in Australia. Draft Report to NiPERA. June 2014. WCA Environment Ltd, Faringdon, Oxfordshire, UK, 14 pp.

Wheeler, J.P., Grist, E.P.M., Leung, K.M.Y., Morritt, D. and Crane, M. (2002a). Species sensitivity distributions: data and model choice. Mar. Pollut. Bull. 45, 192–202.

Wheeler, J.R., Leung, K.M.Y., Morritt, D., Whitehouse, P., Sorokin, N., Toy R., Holt, M. and Crane, M. (2002). Freshwater to saltwater toxicity extrapolation using species sensitivity distributions. Environ. Toxicol. Chem., 21, 2459–2467.

Zajdlik, B.A. (2006). Potential statistical models for describing species sensitivity distributions. Prepared for Canadian Council of Ministers for the Environment (http://www.ccme.ca/assets/pdf/pn_1415_e.pdf).

Zhang Z., Warne M.St.J. and Vieritz A. 2015. Ecotoxicity data quality assessment method. Water Quality and Investigations, Department of Science, Information Technology, Innovation and the Arts, Brisbane, Qld.

14 Acknowledgement

The authors gratefully acknowledge the valuable comments provided by the reviewers and discussions held with numerous colleagues during the development of the toxicant water guideline value derivation method.

Glossary of Terms and Acronyms

ACR: Acute to chronic ratio.

Acute toxicity: A lethal or adverse sub-lethal effect that occurs as a result of a short exposure period to a chemical 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.

Benthic: Referring to organisms living in or on the sediments of aquatic habitats.

Bioaccumulation: A general term describing a process by which chemical substances are accumulated by aquatic organisms from water directly and/or through consumption of food containing the chemicals.

Bioassay: a test used to evaluate the relative toxicity of a chemical by measuring its effect on a living organism relative to a control.

Bioavailable: Able to be taken up by organisms.

Biodiversity: The variety and variability of living organisms and the ecological complexes in which they occur.

Biomagnification: The processes by which tissue concentrations of chemicals increase as the chemical passes up through two or more trophic levels in a food chain. The term implies an efficient transfer of chemicals from food to consumer so that the residue concentrations increase systematically from one trophic level to the next.

Bivalve: A mollusc with a shell in two parts, hinged together.

Bootstrapping: A statistical re-sampling technique typically used to estimate functions of statistics (for example standard errors) when analytical expressions are unavailable.

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 is being developed as part of the current revision of the 2000 Guidelines.

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

CCME: Canadian Council of Ministers of the Environment

Chronic toxicity: An adverse effect that occurs as the result of exposure to a chemical for a substantial portion of the organism's life span or a sub-lethal adverse effect on a sensitive early life stage.

Community: An assemblage of organisms characterised by a distinctive combination of species occupying a common environment and interacting with one another.

Community composition: All the types of taxa present in a community.

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.

Converted acute value(s)/data: Acute toxicity data that have been converted to chronic toxicity data using experimentally-derived or default acute to chronic ratios.

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).

CV: Coefficient of variation, defined as the ratio of the standard deviation and the mean.

Detection limit: Method detection limit is the concentration of a substance that, when processed through the complete analytical method, produces a signal that has a 99% probability of being different from the blank.

DO: Dissolved oxygen.

DOC: Dissolved organic carbon.

DTA: Direct toxicity assessment.

Ecogenomics: The examination of genetic (DNA) materials in environmental samples for the purpose of identifying the organisms present.

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.

Gompertz: A probability distribution that has been used to model species sensitivity distributions.

Guideline value: Numerical concentration limit or narrative statement to support and maintain a designated water use.

GV: Guideline value; a concentration of a toxicant that if exceeded, might potentially impair water quality and thereby trigger an investigation or initiate a management response.

HC: Hazardous concentration, usually to a given percentage of species, e.g. HC5 is the concentration hazardous to 5% of species.

IC50: A toxicant concentration that would cause a 50% 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.

Level of protection: The acceptable level of change from a defined reference condition.

LOE: Line of evidence.

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.

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: 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.

QA/QC: Quality assurance/quality control.

Quality assurance (QA): The implementation of checks on the success of quality control (e.g. replicate samples, analysis of samples of known concentration).

Quality control (QC): The implementation of procedures to maximise the integrity of monitoring data (e.g. cleaning procedures, contamination avoidance, sample preservation methods).

Redox: Simultaneous (chemical) reduction and oxidation; reduction is the transfer of electrons to an atom or molecule, whereas oxidation is the removal of electrons from an atom or molecule.

Redox potential: A measure of the oxidation-reduction potential (ORP) of sediments. The redox potential is often reported as E_h (versus the normal hydrogen electrode).

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.)

Species richness: The number of species present (generally applied to a sample or community).

Spiked sediment: A sediment to which a material has been added for experimental purposes.

Statistical power: The ability of a statistical test to detect an effect given that the effect actually exists.

Stressors: The physical, chemical or biological factors that can cause an adverse effect on an aquatic ecosystem as measured by the condition indicators.

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.

Sub-tropical: Geographic and climate zones located roughly between the Tropic of Cancer and Tropic of Capricorn and the 35th parallel in both hemispheres.

Taxon (taxa): Any group of organisms considered sufficiently distinct from other such groups to be treated as a separate unit (e.g. species, genera, families).

Taxa richness: Number of taxa present.

Taxonomic group: Groups of taxa.

Toxicant: A chemical capable of producing an adverse response (effect) in a biological system, seriously injuring structure or function or producing death. Examples include pesticides and metals.

Toxicity: The inherent potential or capacity of a material to cause adverse effects in a living organism.

Toxicity test: The means by which the toxicity of a chemical or other test material is determined. A toxicity test is used to measure the degree of response produced by exposure to a specific level of stimulus (or concentration of chemical) for a specified test period.

Trophic level: A notional stage in the `food chain' that transfers matter and energy through a community; primary producers, herbivores, carnivores and decomposers each occupy a different trophic level.

Uptake: A process by which materials are absorbed and incorporated into a living organism.

Vertebrate: An animal having a backbone.

WQG: Water quality guideline.

WOE: Weight of evidence.