Bringing water quality benchmark derivation approaches into the 21st century

Rick A van Dam[1], GE Batley[2], R Fisher[3], DR Fox[4], AJ Harford[1], CL Humphrey[1], A Proctor[5], JL Stauber[2], M St J Warne[6]

[1]Environmental Research Institute of the Supervising Scientist, Darwin, NT, Australia; [2] CSIRO Land and Water, Lucas Heights, NSW, Australia; [3] Australian Insitute of Marine Science, Crawley, WA, Australia; [4] Environmetrics Australia, Melbourne, Vic, Australia; [5] Australian Antarctic Division, Hobart, Tas, Australia; [6] Centre for Agroecology, Water and Resilience, Coventry University, UK.
E-mail contact: rick.vandam@environment.gov.au

1. Introduction

The development and application of water quality benchmarks (WQBs) for toxicants represents an important component of water quality management for environmental protection. Approaches for WQB development and application are well established in many jurisdictions. Notwithstanding the common use and importance of WQBs, limitations still exist across a range of elements of their development and application, including: (i) the acquisition of effects data; (ii) the methods that use the effects data to derive the WQB; and (iii) the water quality management approaches within which the WQBs are applied. This paper examines how we can improve the methods used to develop WQBs, with a focus on the derivation methods and, more specifically, SSD-based methods. The key aim is to understand why research and development outcomes in this area have generally not been adopted in formal WQB derivation approaches, and how this could be overcome.

2. Current status of WQB derivation methods

The most common method for deriving WQBs for toxicants is the use of a species sensitivity distribution (SSD) to estimate a concentration that is protective of x% of species. Although variations exist in the specifics of the methods employed by jurisdictions around the world, the fundamental SSD approach is similar and, moreover, has not changed markedly over the past 20 years, despite a significant body of published research aimed at improving or developing new derivation methods (see section 3).

The recent revision of the Australian and New Zealand SSD-based derivation method [1] has re-highlighted previously published limitations [2] of the SSD approach; for example, small sample sizes, model choice and fit, and accommodating different routes of exposure (e.g. for persistent, bioaccumulative and toxic [PBT] compounds) and specific mechanisms of toxicity (i.e. bimodality). However, areas for improvement of WQB derivation methods extend beyond just refining SSD-based approaches, to the use of non-SSD approaches and even weight of evidence approaches that give consideration to both laboratory- and field-effects data. Other opportunities for improvement exist in the acquisition of data for WQBs (e.g. type and acceptability of toxicity data), as well as the application of WQBs in water quality management. Thus, it is important to identify and target the limitations that, if addressed, will yield the biggest benefits to environmental protection.

Some significant advances in WQB development approaches have been made and formally adopted, for example, bioavailability-based approaches for metals (i.e. biotic ligand models, multiple linear regression models). Also, guidance on the use of weight of evidence for deriving WQBs has recently been published by the USEPA [3] and will soon also be formalised for Australia and New Zealand [4].

3. Review of efforts to evolve SSD-based derivation methods

It was not until the early 2000s that a significant volume of publications appeared on the limitations of, and associated modifications or alternatives for, SSDs. Initial reviews focused on the fact that key assumptions of SSDs were often not met, including the assumption that the data conform to a specific underlying distribution [2]. To address the potential problems of model choice, various bootstrapping methods that do not require the use of a specific distribution have been have been proposed and evaluated [e.g. 5,6].

The other key area of research and development has been in the use of bayesian statistical methods to improve or replace standard (frequentist) concentration-response analyses and SSD methods [e.g. 5,7]. Bayesian statistics are thought to offer advantages over frequentist statistics for a range of reasons, including: (i) relevant empirical or even subjective information can be included as a prior distribution to inform the analysis; (ii) it is closely compatible with decision theory; (iii) it can treat uncertainty in an explicit and consistent way; and (iv) inferences can be updated with new data. These methods can address a number of the assumption violations exhibited by standard SSD methods.

Whilst SSDs have also been commonly used for field and mesocosm data, analytical techniques from other fields in ecology and biology are now recognized as being applicable for determining thresholds for measured effects [e.g. 8,9]. Moreover, there is now a growing emphasis on the ability to derive WQBs using weight of evidence approaches, whereby different types of data are drawn together to increase the rigour of, and confidence in, WQBs. Expanding the range of acceptable tools for WQB derivation beyond existing SSD approaches has the potential to provide, among other things:

- greater flexibility in capturing uncertainty in the effects data underlying the WQB derivation;
- more (and possibly simpler) options for dealing with a non-random selection of test species;
- the capacity to derive WQBs explicitly linked to operational management objectives in a formal decision science framework, where the concept of “protective of x% of species” is not always useful or valid; and
- more robust WQB derivation in the case of small sample sizes.

4. Extent of adoption

The multiple efforts to improve or replace SSD methods have seen little, if any, formal adoption in (semi)regulatory settings. There are numerous possible reasons for the lack of traction and adoption, which may include: the approaches are considered to be too complex for routine uptake; unfamiliarity with the statistical details; equivocal comparisons between existing and proposed methods; lack of a coordinated and convincing synthesis of the case; regulatory inertia; and/or a lack of effective communication by the scientific community such that the research outcomes remain largely out of sight of the decision makers. This lack of formal adoption is not unique to this issue, with a similar situation existing for the tissue residue approach to setting WQBs for PBT-type toxicants, the reason for which has been attributed to the approach being too unfamiliar or complex compared to standard (water-based) WQB approaches (Jim Meador, National Oceanic and Atmospheric Administration, pers comm).

However, good lessons can potentially be learnt from other examples. The advances made over the past 20 years with regards to incorporating bioavailability models into WQB derivation for metals is likely largely due to the investment and effective communication and lobbying of the various metals industries. Thus, the need for change can be driven from multiple interest groups.

5. Conclusions

Numerous studies have demonstrated improved methods for deriving WQBs compared to existing formal methods, yet no significant change to formal methods has been effected. In order to improve the likelihood of effecting such change, it is essential to first understand why it has been unsuccessful to date. In doing so, we hope to be able to play a role in bringing formal WQB derivation approaches into the 21st century.

6. References

[1] Warne M Batley GE, van Dam RA, Chapman JC, Fox DR, Hickey CW, Stauber JL. 2015. Revised Method for Deriving Australian and New Zealand Water Quality Guideline Values for Toxicants. Australian and New Zealand Guidelines for Fresh and Marine Water Quality.

[2] Newman MC, Ownby DR, Mezin LCA, Powell DC, Christensen TRL, Lerberg SB, Anderson BA. 2000. Applying species-sensitivity distributions in ecological risk assessment: assumptions of distribution type and sufficient numbers of species. Environ Toxicol Chem 19:508–515.

[3] US EPA. 2016. Weight of evidence in ecological assessment. Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC 20460, EPA/100/R16/001.

[4] Australian governments and the New Zealand Government. 2018. Australian and New Zealand guidelines for fresh and marine water quality. www.waterquality.gov.au/anzguidelines.

[5] Verdonck FAM, Jaworska J, Thas O, Vanrolleghem PA. 2001. Determining environmental standards using bootstrapping, bayesian and maximum likelihood techniques: A comparative study. Anal Chim Acta 446:429–438.

[6] Grist EPM, Leung KMY, Wheeler JR, Crane M. 2002. Better bootstrap estimation of hazardous concentration thresholds to protect biological assemblages. Environ Toxicol Chem 7:1515–1524.

[7] Fox D. 2010 A Bayesian approach for determining the no effect concentration and hazardous concentration in ecotoxicology. Ecotoxicol Environ Safety. 73:123–131.

[8] Kefford BJ, Schäfer RB, Liess M, Goonan P, Metzeling L, Nugegoda D. 2010. A similarity-index–based method to estimate chemical concentration limits protective for ecological communities. Environ Toxicol Chem 29: 2123–2131.

[9] van Dam RA, Humphrey CL, Harford AJ, Sinclair A, Jones DR, Davies S, Storey AW. 2014. Site-specific water quality guidelines: 1. Derivation approaches based on physicochemical, ecotoxicological and ecological data. Environ Sci Pollut Res 21:118–130.