Hazard/Risk Assessment

Data and Diversity in the Development of Acute Water Quality Criteria in the United States

Alice L. Coleman and Suzanne Edmands*

University of Southern California, Los Angeles, California, USA

Abstract: The US Environmental Protection Agency (USEPA) is responsible for the development of water quality criteria, regulatory standards that protect aquatic organisms from harmful chemical exposure. Although these criteria are intended to be broadly protective of aquatic life, the data used to derive criteria do not necessarily reflect the actual diversity of natural communities nor are they available for most chemicals. In addition, although the USEPA's current procedures emphasize using toxicity data with a certain minimum amount of biological diversity, the quantitative impact of such diversity on criteria is unclear. In the present study we assessed the changes to acute toxicity data over time, determined the prevalence of significant taxonomic differences in sensitivity, and investigated the effect of biological diversity on criteria. We found major gaps in existing toxicity data that we hypothesize have contributed to the absence of acute criteria for the majority of chemical pollutants. Taxonomic patterns of sensitivity in these data are abundant, although the resolution of the patterns is relatively poor. In addition, we found that the amount of biological diversity in a toxicity data set and the data set's taxonomic composition does not quantitatively affect criteria in most cases. Because the USEPA has published acute criteria for fewer than 20% of priority pollutants and the persistence of major gaps in toxicity data over the last 37 years, we recommend that the USEPA consider revisions to their water quality criteria guidelines that will expedite the criteria development process and advance the responsible management of pollutants in the aquatic environment. *Environ Toxicol Chem* 2022;00:1–11. © 2022 SETAC

Keywords: Ecological risk assessment; Species sensitivity distribution; Aquatic toxicology

INTRODUCTION

The Clean Water Act requires the US Environmental Protection Agency (USEPA) to derive and publish numeric water quality criteria (WQC) that are intended to protect aquatic organisms and their uses from the adverse effects of chemical pollutants. These criteria are recommended national standards that describe the maximum concentration of a pollutant in surface water that is expected not to harm the majority (95%) of aquatic life, but they do not serve as legally binding requirements. The USEPA's WQC methodology (the "National Guidelines") directs the development of these values by defining the "materials of concern" for criteria, issuing data collection and quality requirements, and establishing calculation procedures for acute and chronic criteria (Stephan et al., 1985).

The first document of its kind, the National Guidelines have been repeatedly adopted by groups outside the USEPA for WQC development (Buchwalter et al., 2017; Liu et al., 2019; Wu et al., 2015). However, several alternate criteria methodologies have since been established by other regulatory agencies that incorporate the advances in aquatic toxicology and additional toxicity data that emerged after the publication of the National Guidelines in 1985 (Canadian Council of Ministers of the Environment, 2007; Tenbrook et al., 2010; Warne et al., 2015). Given the advances in the field, the rise of divergent methodologies, and the perpetual need for effective criteria, it is necessary to consider whether revisions to the National Guidelines are warranted.

Review of the National Guidelines by the USEPA began as early as 1990 and remains an active priority for the agency (Ankley et al., 2017; Wilcut et al., 2015). Although the aim of the modernization efforts is to produce a comprehensive update of the National Guidelines, most proposed revisions have thus far maintained the original minimum data requirements for WQC set by Stephan et al. (1985). These requirements describe the minimum set of taxa that must have available toxicity

data to derive a criterion. For example, the minimum data requirements for acute saltwater criteria call for acceptable acute tests with at least one species from at least eight different families, specifically including:

- Two families from the phylum Chordata
- One family in a phylum other than Arthropoda or Chordata
- One species from either the Mysidae or Penaeidae family
- Three species from families not in the phylum Chordata (may include either Mysidae or Penaeidae if not used previously)
- One species from any other family

Similarly, the minimum data requirements for acute freshwater criteria call for acceptable tests from at least one species in at least eight different families, specifically including:

- One species from the family Salmonidae
- One additional family from Osteichthyes
- A third family from Chordata (may be in Osteichthyes)
- One planktonic crustacean
- One benthic crustacean
- One insect
- One family in a phylum other than Arthropoda or Chordata
- One family in any order of insect or phylum not already represented

Although the purpose of WQC is to protect the majority of aquatic life, the minimum data requirements do not impose specific diversity requirements aside from ensuring the representation of at least eight families and three phyla in a data set. Because the sample size for these minimum data requirements was based on the fact that most chemicals did not have more than eight acceptable toxicity values (Stephan, 1984), it is reasonable to question whether data from just eight families can confer the desired level of protection. In addition, by 1985 it had been established that the majority of available toxicity data at the time were obtained from a limited group of species and did not adequately sample from the full range of sensitivities expected in a natural species assemblage (Seegert et al., 1985). The pool of toxicity data that can be used to develop criteria has grown over the past 37 years, so it is important to assess how the taxonomic composition of these data has changed over time and to determine how many chemicals are now able to fulfill either set of minimum data requirements.

Aside from these concerns over the impact of data availability on criteria, it is also unclear whether the amount of biological diversity or specific taxonomic groups represented in a toxicity data set significantly affects criterion value. Previous studies in aquatic toxicology have identified generalized taxonomic patterns of sensitivity for some pollutants. For example, phytoplankton are considered to be the most sensitive group to the herbicide atrazine, followed by benthic invertebrates, planktonic invertebrates, and then fish (Solomon et al., 1996). Other studies report a different pattern for ammonia, wherein freshwater mussels (family Unionidae) are extremely sensitive to ammonia, fish are moderately sensitive, and crustaceans are

generally more tolerant than both groups (Arthur et al., 1987; Augspurger et al., 2003). These taxonomic sensitivity patterns could affect criteria development if biological diversity is limited during toxicity testing, resulting in a multimodal data set. For example, a toxicity data set for ammonia composed primarily of values from freshwater molluscs and crustaceans would likely have two significantly different modes. The value of a criterion from such multimodal data sets would be influenced by the number of species from each mode, and thus may not reflect the sensitivity of a natural aquatic community (Giddings et al., 2019).

In the present study we assessed the impact of data availability, biological diversity, and the minimum data requirements on the development of WQC with the aim of informing the USEPA's revisions of the National Guidelines. Using data from 12 aquatic pollutants, we analyzed the shift in the abundance and taxonomic composition of acute toxicity data since 1985, quantified the relationship between the amount of biological diversity in a toxicity data set and criterion value, and tested for taxonomic differences in acute tolerance. In addition, we estimated the proportion of chemicals with sufficient data to satisfy the USEPA's minimum data requirements for both saltwater and freshwater acute criteria and provide a brief overview of relevant potential changes to the National Guidelines.

MATERIALS AND METHODS

Data collection

We collected and analyzed acute toxicity test results from saltwater and freshwater species for 12 aquatic pollutants that span a range of chemical classes and toxic modes of action (Table 1). Candidate chemicals were drawn from the literature and the Clean Water Act's Priority Pollutant List, the set of chemicals for which the USEPA is required to develop WQC. We conducted a preliminary survey of the USEPA's (2021a) ECOTOXicology (ECOTOX) Knowledgebase to identify which chemicals potentially had enough data to fulfill the saltwater and freshwater minimum data requirements; we then used the results to narrow our list down to 12. The primary source of data for the present study, ECOTOX is an online public database comprised of more than 1,000,000 toxicity test results that cover approximately 12,300 chemicals and approximately 13,600 species.

We downloaded acute toxicity data for each chemical from ECOTOX that were obtained using the following search parameters:

1. CAS number
2. Kingdom: Animals
3. Endpoint: Median lethal concentration (LC50)
4. Test location: Laboratory
5. Exposure media: Freshwater or saltwater

When possible, the CAS number used for a chemical was matched to its CAS number given in the Aquatic Life Criteria Table, an online resource maintained by the USEPA that

TABLE 1: Properties of analyzed chemicals

Chemical	CAS no.	Origin	Class	Mode of action
Ammonia[a]	7664417	Natural	Inorganic	Osmoregulatory impairment
Cadmium[a,b]	7440439	Natural	Metal	Metallic iono/osmoregulatory impairment
Copper[a,b]	7440508	Natural	Metal	Metallic iono/osmoregulatory impairment
Nickel[a,b]	7440020	Natural	Metal	Metallic iono/osmoregulatory impairment
Phenol[b]	108952	Natural	Phenol	Polar narcosis
Toluene[b]	108883	Natural	Aromatic hydrocarbon	Nonpolar narcosis
4-Nitrophenol[b]	100027	Synthetic	Nitrophenol	Polar narcosis
Atrazine	1912249	Synthetic	Triazine	Narcosis
Endosulfan[a,b]	115297	Synthetic	Organochlorine	Neurotoxicity
Endrin[a,b]	72208	Synthetic	Organochlorine	Neurotoxicity
Pentachlorophenol[a,b]	87865	Synthetic	Organochlorine	Electron transport inhibition
Tributyltin oxide[a]	56359	Synthetic	Organotin	Electron transport inhibition

[a]Chemical has a criterion in the Aquatic Life Criteria Table.
[b]Chemical is a member of the Priority Pollutant List.

contains the current set of WQC (USEPA 2021b). Freshwater and saltwater data sets were downloaded separately in May 2021. Additional acute toxicity datapoints were collected with searches in Google Scholar and Web of Science using the chemical name and keywords "acute toxicity" and either "marine" or "freshwater" and then appended to the ECOTOX data sets. The combined data sets were then trimmed according to the following quality control parameters to emulate the data collection rules listed in the National Guidelines:

1. Test organism was a resident North American species.
2. Test organism was not a brine shrimp.
3. Test species was not a single-celled organism.
4. Tests performed with cladocerans and midges were 48 h in length.
5. Tests performed with all other freshwater and saltwater species were 96 h in length.

When a species had more than one LC50 value available for a chemical, the data were condensed into a species mean acute value by calculating the geometric mean of all data points. For each genus with more than one species available, the genus mean acute value was calculated as the geometric mean of all data points.

In addition, we used the USEPA's (2021b) Aquatic Life Criteria Table to identify which of the 12 chemicals had official acute criteria and to examine available criteria documents. We also performed a second survey of ECOTOX in July 2021 to estimate the proportion of chemicals from the Priority Pollutant List that could satisfy the saltwater and freshwater minimum data requirements using the same quality control parameters just listed.

The trimmed toxicity data sets were used to determine which chemicals had sufficient data to satisfy the minimum data requirements for acute saltwater and freshwater criteria. We further sampled from the trimmed data sets to create the following data set subtypes for each chemical:

1. Eight randomly selected genera that satisfy the saltwater minimum data requirements (TS8)

2. Eight randomly selected genera that satisfy the freshwater minimum data requirements (TF8)
3. Eight randomly selected saltwater genera (RS8)
4. Eight randomly selected freshwater genera (RF8)
5. Twenty randomly selected saltwater genera (RS20)
6. Twenty randomly selected freshwater genera (RF20)

A sample size of eight genera was chosen for data set types 1–4, to be consistent with the minimum data requirements. Twenty genera were used for the RF20 and RS20 data sets to approach the mean sample size ($\overline{x} \approx 29$) of the data sets used by the USEPA to develop criteria for the chemicals in our set while accounting for the sample size limitations of the data sets we assembled (Supporting Information, Table S1).

Criteria derivation

Criteria were derived using the method defined by Stephan et al. (1985) in the National Guidelines. Briefly, genus mean acute values were ordered by decreasing value and assigned a rank; then a cumulative probability was calculated for each one. The four genus mean acute values with cumulative probabilities closest to 0.05 were utilized in a series of four equations stated in the National Guidelines to calculate the final acute value. The final acute value was then divided by two to reach the criterion maximum concentration (CMC), the USEPA's acute criterion.

We also calculated criteria using species sensitivity distributions (SSDs), a relatively new derivation technique employed by other criteria methodologies that is under consideration by the USEPA for inclusion in the National Guidelines (Tenbrook et al., 2010; Warne et al., 2015; Wilcut et al., 2015). The SSDs consist of a plot of toxicity data from multiple species that are ranked and assigned a percentile. A cumulative distribution is then fitted to the data and used to calculate the hazardous concentration for 5% of taxa (HC5), which is considered equivalent to an final acute value derived by the USEPA method (Wilcut et al., 2015). To approximate a CMC from an SSD, we calculated SSDs using genus mean

acute values and divided all HC5s by 2. All SSDs were generated in R with the package ssdtools (R Core Team, 2019; Thorley & Schwarz, 2018). This package allows the user to fit multiple distributions to each data set, so we used the Akaike Information Criterion corrected for sample size to select the best-fit distributions (Burnham & Anderson, 2002).

Taxonomic analyses

We assessed the change in our trimmed data sets over time by quantifying the taxonomic composition of the data that were available in 1985 and comparing it with the makeup of the complete data sets compiled in May 2021. We further compared the data sets by calculating the average taxonomic distinctness index ($\Delta+$) of the 1985 and 2021 data sets for both saltwater and freshwater data with the R package vegan (Clarke & Warwick, 1998; Oksanen et al., 2014). Change in $\Delta+$ between 1985 and 2021 was tested with a randomization test for which we generated 1000 random subsets of the species data with the same sample size as the 1985 data sets (saltwater $n = 52$; freshwater $n = 83$) and calculated the corresponding $\Delta+$ values. We then defined the intervals that contained 95% of the simulated $\Delta+$ values and compared the intervals with the actual $\Delta+$ values from the 1985 and 2021 data sets, with the latter treated as the "true" $\Delta+$ for the data. For a broad-scale assessment of diversity, we also used Fisher's exact tests with Monte Carlo simulations ($B = 1,000,000$) to determine whether the taxonomic composition of the data available in 1985 and those data that were published in 1986 or later (up to May 2021) were different at the level of phylum. In addition, we used two-proportion z-tests to assess whether the proportion of the data that met our quality control parameters had changed between 1985 and 2021.

To evaluate the impact of the minimum data requirements on criterion value, we compared the criteria calculated from data sets that were assembled randomly (RS8/RF8) with those assembled according to the minimum data requirements (TS8/TF8). Fifty versions of each data set type were generated for each chemical, except for those that could not satisfy either of the minimum data requirements; we then calculated the CMC for each individual data set. We next compiled the individual criteria for each group, calculated the mean CMCs, and used t-tests to determine whether the means of the TS8/RS8 and TF81/RF8 data sets were significantly different. Mann–Whitney U-tests were used in place of t-tests when the data sets were not normally distributed, as indicated by a Shapiro–Wilk test. We conducted this analysis twice for each chemical, comparing criteria calculated by either the USEPA or the SSD method.

In addition, we performed a linear regression analysis to model the relationship between the amount of biological diversity in a data set and criterion value. As in the criteria comparison, we generated 50 versions of the RS20 and RF20 data sets for the chemicals with data from at least 20 genera (Supporting Information, Table S1) and then calculated the CMC for each data set using both the USEPA and SSD methods. The diversity of each data set was measured as the Shannon diversity index and calculated using the package vegan (Shannon, 1948). We then fit a linear model to the data frame containing the criteria and diversity indices, treating diversity index as the independent variable and CMC as the dependent variable. This analysis was performed twice for each chemical, using criteria derived with either the USEPA or SSD method to compare the effect of diversity on the different techniques.

Finally, Kruskal–Wallis tests, which are the nonparametric analog to the one-way analysis of variance, were used to determine whether the mean sensitivities of any phyla in a toxicity data set significantly differed from each other. Pairwise comparisons of phyla were conducted with post hoc Dunn's tests to then identify those significantly different phyla, the results of which were then used to outline taxonomic sensitivity patterns. These tests were performed twice/chemical, using each saltwater and freshwater data set separately. In our analysis, the phyla with only one datapoint available were removed from a chemical's data set because the mean is an uninformative statistic for a group with a sample size of one.

RESULTS

Data set composition

In May 2021 we found a total of 1662 saltwater and 5359 freshwater acute toxicity values (LC50s) for our set of 12 pollutants. Of these totals, 558 of the saltwater and 1662 of the freshwater data points were published in 1985 or earlier. After the application of quality control parameters, the saltwater and freshwater data sets shrank to 535 and 1626 values, respectively, and most chemicals saw a reduction in the number of species and genera represented in their individual data sets (Supporting Information, Table S1). Of the trimmed data set totals, 224 saltwater and 668 freshwater values were published in 1985 or earlier. The trimmed saltwater data set consisted of 104 species (84 genera) in eight phyla, whereas the trimmed freshwater data set contained data from 172 species (115 genera) in eight phyla. Species from Chordata and Arthropoda contributed the most data to both the saltwater and freshwater data sets, followed distantly by species from the phyla Mollusca and Rotifera (Figure 1). The number of phyla and species represented in each data set increased slightly over time.

The $\Delta+$ calculations indicated that the average taxonomic distinctness of the saltwater data significantly decreased over time (1985 $\Delta+ = 87.23$; 2021 $\Delta+ = 85.097$; Supporting Information, Figure S2), whereas the average taxonomic distinctness of the freshwater data remained relatively unchanged (1985 $\Delta+ = 84.34$; 2021 $\Delta+ = 85.28$; Supporting Information, Figure S2). Fisher's exact tests indicated that the taxonomic compositions of the data that were available in 1985 compared with the data that were published after 1985 (up to May 2021) were significantly different ($p < 5E-4$) at the phylum level for both the saltwater and freshwater data sets. In addition, the two-proportion z-tests indicated that the proportions of both the saltwater (from 40% in 1985 to 32% in 2021; $p = 0.00059$) and freshwater data sets (from 40% in 1985 to 31% in 2021;
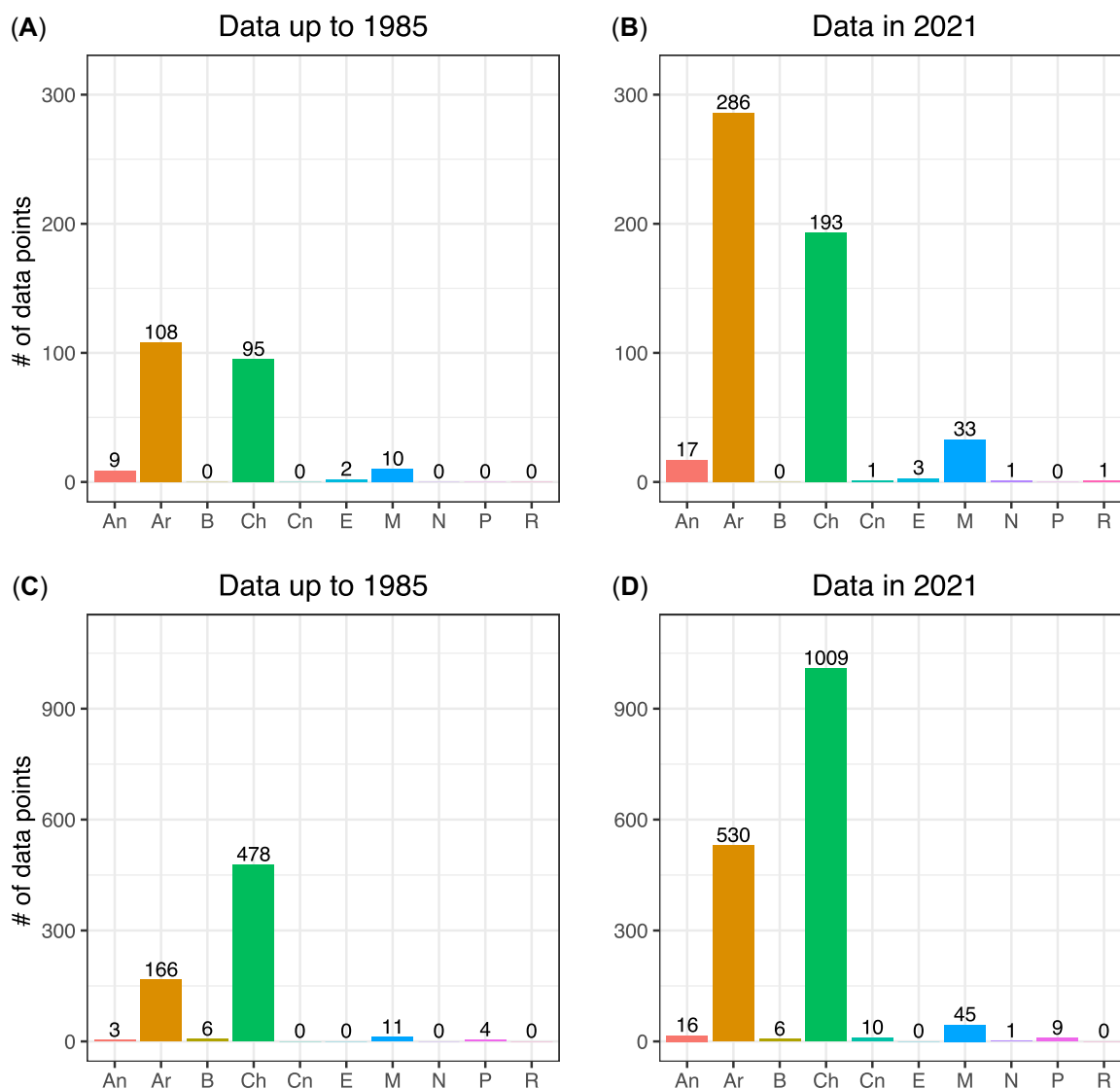
FIGURE 1: Taxonomic composition of the trimmed acute toxicity data sets from marine (**A** and **B**) and freshwater species (**C** and **D**) for 12 chemicals that were available in 1985 (**A** and **C**) and 2021 (**B** and **D**). Phylum codes: An, Annelida; Ar, Arthropoda; B, Bryozoa; Ch, Chordata; Cn, Cnidaria; E, Echinodermata; M, Mollusca; N, Nematoda; P, Platyhelminthes; R, Rotifera.

$p = 2.061\text{E}{-}13$) that met our quality parameters decreased between 1985 and 2021.

In addition, based solely on the trimmed data sets we assembled, two chemicals (4-nitrophenol and toluene) could not satisfy the saltwater minimum data requirements, and three chemicals (nickel, 4-nitrophenol, and toluene) could not satisfy the freshwater minimum data requirements. We also estimated that of the 126 chemicals on the Priority Pollutant List, 21 have enough data in ECOTOX to satisfy the saltwater minimum data requirements and 37 have sufficient data to meet the freshwater minimum data requirements.

Criteria search

To date, there are 29 acute saltwater criteria and 34 acute freshwater criteria available in the USEPA's (2021b) Aquatic Life Criteria Table. These official values cover 8 of the 12 chemicals in our set, excluding atrazine, 4-nitrophenol, phenol, and toluene, which have no criteria listed (Table 1). We treated the USEPA's criteria for tributyltin and α/β-endosulfan as the criteria for tributyltin oxide and endosulfan, respectively. Importantly, the criteria for the endosulfan isomers were developed using the USEPA's (1980) precursor document to the National Guidelines, which utilized different minimum data requirements than those discussed in the present study.

In addition, despite their absence from the Aquatic Life Criteria Table, we were able to retrieve archived USEPA criteria documents for atrazine, 4-nitrophenol, phenol, and toluene. We found a complete set of draft criteria for atrazine that appear to have never been published in a finalized form (USEPA, 2003), and we also found criteria for 4-nitrophenol, phenol, and toluene that were developed using the precursor to the National Guidelines, similar to the criteria for the endosulfan isomers. It is unclear why the criteria for these three chemicals are not included in the Aquatic Life Criteria Table with the endosulfan criteria.

TABLE 2: Comparison of mean criteria from data sets that were assembled using the minimum data requirements (TS8/TF8) with the mean criteria from randomly assembled data sets (RS8/RF8)

	Saltwater		Freshwater	
Chemical	Mean TS8 criterion (µg/L)	Mean RS8 criterion (µg/L)	Mean TF8 criterion (µg/L)	Mean RF8 criterion (µg/L)
Ammonia	19	19	9.7	11
Atrazine	13[a]	12[a]	21	20
Cadmium	4.3	4.3	2.9*	4.2*
Copper	3.9*	3.5*	2.5[a]	3.2[a]
Endosulfan	0.22	0.27	0.81	0.83
Endrin	0.053	0.066	0.16	0.17
Nickel	—	—	—	—
4-Nitrophenol	—	—	—	—
Pentachlorophenol	4.1	4.3	4.1	4.1
Phenol	25	26	31	32
Tributyltin oxide	0.98[a]	0.92[a]	0.74	0.79
Toluene	—	—	—	—

*$p < 0.05$, by t-test.
[a]$p < 0.05$, by Mann–Whitney U-test.

Taxonomic analyses

We found nominal evidence of a quantitative effect of the minimum data requirements on criterion value. The comparison of criteria showed that using different methods to assemble toxicity data sets resulted in significantly different values in only 5 of 18 tests (Table 2). There appeared to be no effect of the method of data set assembly on criterion value in these significant cases, because the larger criterion was from an minimum data requirement (TS8/TF8) data set in three cases and from a random (RS8/RF8) data set in two cases.

Similarly, we found minimal evidence of a quantitative effect of the diversity in a toxicity data set on criterion value. Of the 24 models we tested, just 2 (both pentachlorophenol) described a significant relationship between a data set's Shannon diversity index and criterion (Table 3). The regression coefficient for diversity index in the significant models was negative in both cases, but the low R^2 values for these models (0.17 and 0.10) and overall low number of significant models in the set we tested suggest that diversity index is a poor predictor of criterion values. There also appeared to be no effect of condition (saltwater/freshwater) or calculation method (USEPA/SSD) on the diversity–criterion relationship, because there was one significant model in each category.

The Kruskal–Wallis tests indicated that 12 of the 24 data sets we assembled contained phyla with significantly different mean sensitivities, covering 8 of the 12 chemicals (Table 4). Eight of the data sets with significant differences were freshwater and four were saltwater. Post hoc Dunn's tests identified which specific phyla's means were significantly different from each other (Figure 2 and Supporting Information, Figures S2–S12). We also found that the phylum that was the most sensitive to a chemical varied among all 12 chemicals, although we were unable to determine whether this variation was caused by true differences in the sensitivity of phyla to chemicals or by differences in data set composition.

TABLE 3: Linear model summary for the Shannon diversity index of a quality-controlled toxicity data set as a predictor of criterion value

		USEPA criteria			SSD criteria		
Chemical	Conditions	R^2	Regression coefficient	p-value	R^2	Regression coefficient	p-value
Ammonia	Saltwater	—	—	—	—	—	—
	Freshwater	0.10	31	0.014	−0.020	−12	0.83
Atrazine	Saltwater	—	—	—	—	—	—
	Freshwater	0.019	−39	0.17	−7.5E−05	−413	0.32
Cadmium	Saltwater	0.015	3.7	0.19	−0.0061	16	0.41
	Freshwater	0.017	−1.4	0.18	−0.020	−0.46	0.90
Copper	Saltwater	—	—	—	—	—	v
	Freshwater	0.044	−1.6	0.078	0.010	−2.0	0.22
Endosulfan	Saltwater	0.028	0.028	0.13	0.031	0.011	0.11
	Freshwater	0.052	0.33	0.087	0.012	0.050	0.21
Endrin	Saltwater	−0.016	0.0068	0.63	0.0066	0.017	0.26
	Freshwater	−0.021	0.0073	0.98	−0.020	0.012	0.89
Nickel	Saltwater	—	—	—	—	—	—
	Freshwater	—	—	—	—	—	—
4–Nitrophenol	Saltwater	—	—	—	—	—	—
	Freshwater	—	—	—	—	—	—
Pentachlorophenol	Saltwater	0.17	−4.5	0.0016*	0.10	−15	0.014*
	Freshwater	−0.0017	1.0	0.34	0.0029	5.7	0.29
Phenol	Saltwater	—	—	—	—	—	—
	Freshwater	−0.016	53	0.64	−0.015	1124	0.60
Tributyltin oxide	Saltwater	—	—	—	—	—	—
	Freshwater	—	—	—	—	—	—
Toluene	Saltwater	—	—	—	—	—	—
	Freshwater	—	—	—	—	—	—

*$p < 0.05$.
USEPA = US Environmental Protection Agency; SSD = species sensitivity distribution.

TABLE 4: Kruskal–Wallis comparison of phyla

Chemical	Conditions	Kruskal–Wallis test		
		χ^2	df	p-value
Ammonia	Saltwater	1.1	2	0.58
	Freshwater	13	3	0.0041*
Atrazine	Saltwater	1.6	1	0.20
	Freshwater	0.57	1	0.45
Cadmium	Saltwater	25	2	3.8E−06*
	Freshwater	18	4	0.0012*
Copper	Saltwater	4.6	1	0.032*
	Freshwater	79	4	2.4E−16*
Endosulfan	Saltwater	23	3	3.8E−05*
	Freshwater	59	2	1.8E−13*
Endrin	Saltwater	3.1	1	0.080
	Freshwater	9.1	1	0.0025*
Nickel	Saltwater	5.2	2	0.076
	Freshwater	6.8	2	0.054
4-Nitrophenol	Saltwater	2.3	1	0.13
	Freshwater	2.2	1	0.14
Pentachlorophenol	Saltwater	4	3	0.26
	Freshwater	122	4	<2.2E−16*
Phenol	Saltwater	5.6	2	0.062
	Freshwater	17	4	0.0024*
Tributyltin oxide	Saltwater	16	3	0.0013*
	Freshwater	23	4	0.00016*
Toluene	Saltwater	2.4	1	0.12
	Freshwater	3.7	3	0.16

*$p < 0.05$.

DISCUSSION

In the United States, the development of acute WQC relies on toxicity data that meet the specific requirements laid out in the USEPA's National Guidelines. In the present study we assembled and analyzed acute toxicity data sets for 12 chemicals and found them to lack data from saltwater and "nonstandard" laboratory test species, although the amount of biological diversity and taxonomic composition of a data set did not appear to affect criterion value.

The low availability of toxicity data from saltwater species relative to freshwater is particularly striking. Although the disparity between freshwater and saltwater data is evident from the raw counts of the cumulative data sets we assembled (freshwater $n = 1636$, saltwater $n = 535$), comparison of the single-chemical data sets indicates that marine data sets are generally smaller and contain fewer species and genera than their freshwater counterparts (Supporting Information, Table S1). Our observation of fewer saltwater values from fewer taxa is consistent with the findings of other studies in aquatic toxicology, a trend that has been attributed to the challenges of calculating the chemical speciation of toxicants in seawater and the historical focus of risk assessments on freshwater systems (Leung et al., 2001; Pavlaki et al., 2016). As a result of these differences in sample size, freshwater toxicity data sets are generally considered to be more representative of the true amount of biological diversity expected in a natural system than saltwater data sets (Leung et al., 2001). These disparities in data set sample size and representativeness pose a major challenge to the development of acute WQC for saltwater organisms.

The unbalanced taxonomic composition of our toxicity data sets hints at another significant barrier to the development of acute WQC. The overwhelming majority of the data we collected are derived from Chordata and Arthropoda (90% of the trimmed saltwater data set, 94% of the trimmed freshwater data set), with the rest of the data sets made up of small contributions from eight other phyla (Figure 1). The dominance of Chordata and Arthropoda persisted during the period between 1985 and 2021 even with the statistically significant changes in the taxonomic composition of both the saltwater and freshwater data sets during that time. In addition, the average taxonomic distinctness (Δ+) of the saltwater data decreased over that period, which we speculate was caused by the addition of chordate and arthropod species to the data set that were closely related to some of the species that had data available in 1985. Thus, we hypothesize that the minimal amount of toxicity data available from other faunal groups is the most important factor preventing the majority of chemicals
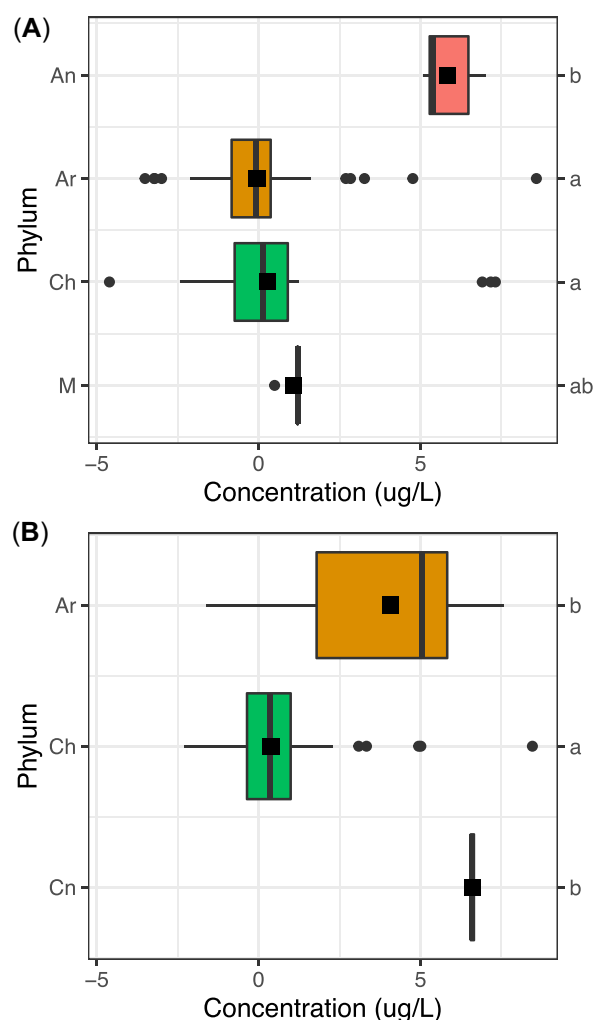
FIGURE 2: Distribution of acute toxicity data for endosulfan from saltwater (**A**) and freshwater (**B**) species. The *x*-axes of each plot have been log-transformed. The black squares in each box represent the mean acute sensitivity of a phylum; the means of phyla with a common letter are not significantly different according to Dunn's test. Phylum codes: An, Annelida; Ar, Arthropoda; Ch, Chordata; M, Mollusca.

from meeting one or both sets of the minimum data requirements for acute criteria.

Importantly, the dominance of chordate and arthropod taxa in toxicity data sets does not reflect the actual abundance of these species in the world's aquatic environments. According to the World Register of Marine Species (2021), there are 57,368 valid marine arthropod species and 24,100 valid marine chordate species. Together, the species from these phyla make up just 40% of the total number of valid marine species in the Kingdom Animalia. This persistent discrepancy between the diversity found in toxicity data sets and the actual diversity of aquatic life should be a driving force to expand the coverage of aquatic taxa included in the criteria development process.

There are multiple factors behind the reliance on chordate and arthropod species in toxicity testing. Many fish and crustacean species have historically been used in toxicity testing because they are widely available to researchers and are easily maintained in the laboratory, leading to the development of standard toxicity test procedures for certain species. As a result, common laboratory species like *Cyprinodon variegatus* and *Palaemonetes pugio* are present in the data sets of multiple chemicals. The language of the National Guidelines may also have influenced the usage of chordate and arthropod species during testing, because they explicitly emphasize protecting "commercially, recreationally, and socially important species" (Stephan, 1984), which in practice tends to translate to fishes and crustaceans.

Although the continued usage of chordates and arthropods in laboratory testing is a major determinant of the taxonomic composition of toxicity data, the makeup of the specific data sets that we assembled could also be linked to our choice of chemicals. Eight of the chemicals we analyzed have official USEPA criteria, and two (endrin and endosulfan) are banned or are being phased out in the United States, meaning that considerable toxicity testing has already been conducted for these chemicals. With these regulatory measures in place, there may be little incentive to continue testing a certain chemical, leading to minimal change in the composition of toxicity data for that chemical over time. However, when a significant issue with its criterion emerges, such as in the case of ammonia, further toxicity testing and the expansion of toxicity data sets is required.

In 2013 the USEPA published an updated set of freshwater criteria for ammonia that represented a major decrease from the previous values set in 1999 (e.g., the CMC decreased from 24 to 17 mg/L). This change was driven by the publication of studies in 2003 that found that the existing criteria did not protect the highly sensitive Unionidae family of freshwater mussels from the harmful effects of ammonia exposure (Augspurger et al., 2003; USEPA, 2013). Although the exclusion of any major taxonomic group by a criterion is concerning, this conclusion was particularly compelling because more than half of the nearly 300 Unionidae species in North America are listed as threatened or endangered (Augspurger et al., 2003). In response, the USEPA conducted additional toxicity testing to validate these findings and ultimately drafted a new set of ammonia criteria in 2009 for bodies of water with and without

mussels present (USEPA, 2013). Following an external peer review process, the USEPA then released a finalized set of updated criteria for ammonia in 2013 that were derived from a new toxicity data set that included the previously absent unionid mussels and gill-breathing snails. This update of the ammonia criteria is significant because it demonstrated the USEPA's ability to integrate taxonomic sensitivity information specific to a single chemical into the criteria development process, an ability which may be required in the future if other conflicts between criteria protectiveness and taxonomic sensitivity patterns are exposed.

In the present study, we investigated the taxonomic patterns of sensitivity for 12 chemicals by identifying and comparing the marine and freshwater phyla with significantly different mean acute sensitivities (Figure 2 and Supporting Information, Figures S3–S13). Using these results, we can draw tentative conclusions about the relative sensitivities of different taxa to the 12 chemicals. For example, it is evident in the endosulfan data that saltwater annelids are more tolerant than both saltwater arthropods and chordates and that freshwater arthropods and molluscs are more tolerant of endosulfan than freshwater chordates (Figure 2). Unfortunately, however, our ability to resolve taxonomic sensitivity patterns to a high degree using the results of the comparisons between phyla was restricted in several instances by the low statistical power of the Dunn's test and the low amount of data available for certain phyla. For example, no conclusions can be formed about the relative sensitivities to atrazine of marine or freshwater arthropods and chordates (Supporting Information, Figure S4). Similarly, for ammonia, no distinctions can be made between the saltwater phyla, although in the freshwater data arthropods seem to be more tolerant of ammonia than chordates (Supporting Information, Figure S3) As a result of this poor degree of resolution, it is difficult to determine whether our findings are consistent with the trends in sensitivity to atrazine and ammonia documented in the literature. Despite the reoccurrence of this issue in the toxicity data sets we assembled, the frequency of significant differences between phyla in sensitivity as indicated by the results of the Kruskal–Wallis tests suggests that taxonomic relationships in sensitivity to toxicants are a common phenomenon (Table 4, Figure 1, and Supporting Information, Figures S3–S13). We anticipate that criteria developed using both the USEPA method and SSDs could be impacted by taxonomic patterns of sensitivity. The USEPA method primarily uses the four values in a data set with cumulative probabilities closest to 0.05 to calculate criteria without imposing requirements on the taxonomic makeup of those four data points. If a certain group of organisms is particularly sensitive to the target chemical, it is possible that all of the four data points used to calculate the criterion could come from a set of closely related species. The SSD approach incorporates all of the values in a toxicity data set, but if a taxonomic sensitivity pattern causes a data set to be multimodal then the cumulative distribution that is fitted to the data may be a poor fit for the left-tail of the SSD. The HC5 is derived from the left-tail of an SSD, so multimodality may result in greater uncertainty in the HC5. In light of these potential

effects, further exploration of taxonomic trends in sensitivity within phyla is required.

Ultimately, 37 years after the publication of the National Guidelines the vast majority of aquatic pollutants still do not have acute criteria. Although the overall availability of toxicity data has improved since, which in theory should benefit criteria development, there are still fewer than 60 official USEPA WQC that cover less than 30% of the chemicals on the Priority Pollutant List. Although the scarcity of criteria can be explained by multiple scientific and bureaucratic factors, we hypothesize that the cumulative lack of toxicity data from marine and non-standard test species is the most important constraint on the development of WQC in the United States. Because criteria provide a crucial measure of protection for aquatic organisms, the USEPA should consider revisions to the National Guidelines that will facilitate their development. There are several possible changes that could help achieve this goal, although none of the options we present can be considered a universal solution.

The simplest means of accelerating criteria development is to increase the amount of available toxicity data. Although the single-species toxicity test is the preferred method of data collection in aquatic toxicology, the financial and logistical constraints associated with laboratory testing make it an impractical solution for filling the large data gaps we have noted. At present, data extrapolation techniques such as quantitative structure–activity relationship (QSAR) models and interspecies correlation estimation (ICE) models are the best alternatives to traditional toxicity testing. The QSAR models use molecular descriptors and chemical properties to predict the toxicity of chemicals to organisms, whereas ICE models estimate the acute toxicity of a chemical to a data-deficient species by performing a least-squares regression with data from a related surrogate species. The SSDs can be augmented with ICE-extrapolated toxicity values, increasing data set sample size without significantly affecting HC5 uncertainty (Awkerman et al., 2014). The USEPA has developed software that can estimate toxicity values using ICE models, including an SSD module (Raimondo et al., 2010), but the Agency has not embraced ICE-derived data points in official WQC development. Thus, in light of their statistical support, we recommend that the USEPA formally integrate ICE models into the National Guidelines as an acceptable alternate method of toxicity data production. Alternatively, if the extrapolation of chemical sensitivity data is undesirable or infeasible, the USEPA could explore revising the data quality control parameters in the National Guidelines to take advantage of a greater portion of the existing pool of toxicity data.

One of the major challenges in WQC development is the conflict between data quality and quantity. Quality control parameters implemented during data collection ensure the quality and consistency in the toxicity data used to derive criteria, but may actually hinder the criteria development process by eliminating many potentially usable toxicity data points. In the present study, the imposition of relatively simple quality parameters similar to those in the National Guidelines forced an approximately 70% drop in the size of both our freshwater and saltwater data sets, and we observed a decrease in the proportion of the data that met those quality parameters over time. The majority of the data points were censored because they were either recorded from a toxicity test with an inappropriate duration or were derived from a test organism that could not be classified as a North American resident species. As a result of the removal of these data, the biological diversity in the data sets for most chemicals was severely reduced and likely made the data sets less representative of the aquatic communities that WQC are intended to protect. Data losses of similar or greater magnitude are probable regardless of chemical choice, so it may be necessary to reconsider some of the specific quality requirements in the National Guidelines. For example, it may be beneficial to allow for the use of data from species that do not have reproducing wild populations in North America during criteria derivation because the sensitivities of temperate and tropical species are relatively similar for many chemicals (Wang et al., 2014). Although the use of region-specific data is certainly preferable in theory when one is developing criteria, in practice the limitations in data availability mean that a localized approach may not be feasible for most chemicals. Alternatively, the USEPA could pivot to allow for the use of field- or mesocosm-based data during criteria development. Stephan et al. (1985) did not include these data types in the National Guidelines because of concerns over data complexity and the feasibility of field testing; however, weight-of-evidence approaches to using field data in WQC development are now available (Cormier et al., 2008; Warne et al., 2015). These approaches have allowed newer criteria methodologies to take advantage of field and mesocosm data, meaning that their integration into the National Guidelines should be possible.

Another means of expediting criteria development would be to adjust how data are grouped prior to the calculation of criteria. One such option is to combine data from freshwater and saltwater species during criteria derivation: Leung et al. (2001) and Wheeler et al. (2002) demonstrated that freshwater data sets for ammonia and several metals could provide adequate protection for saltwater taxa. However, these same studies also indicated that freshwater data would not be protective of saltwater species for pesticides such as chlordane, endosulfan, and chlorpyrifos largely because of differences in data set sample sizes and taxonomic composition. Thus, the integration of freshwater with saltwater data may not be appropriate for all chemicals. The USEPA could also consider setting criteria for groups of chemicals with the same adverse outcome pathway instead of creating criteria for single chemicals (Elias et al., 2019; Giddings et al., 2019). Group criteria, or normalized hazardous concentrations (HC5n), are derived from SSDs populated with the toxicity data from multiple chemicals (Giddings et al., 2019). This approach does not require the individual chemical data sets to satisfy the minimum data requirements, thereby reducing the need for additional toxicity testing or data extrapolation during criteria development. In addition, the higher sample size of a combined data set can lead to greater biological diversity in an SSD and can increase the statistical precision of the HC5n estimate (Carr & Belanger, 2018). Giddings et al. (2019) utilized this

approach to derive a criterion for a set of nine pyrethroids (a class of synthetic pesticides) and calculated an HC5n for the entire group that was more statistically robust than the individual criteria would have been. Although these results are promising, the group criteria approach is limited to those chemicals that share a toxic mode of action. The modes of action have not been defined for many chemicals, making it difficult to determine when grouping chemicals for criteria derivation is appropriate. Thus, revising the National Guidelines to allow for group criteria is not yet viable.

A final option to consider is a modification of the minimum data requirements in the National Guidelines to include aquatic plant and algal species. Although aquatic plants and algae contribute considerable ecological and economic value, they have historically been excluded from WQC development in the United States because of uncertainty over how to include them in the process (Lewis & Thursby, 2018). There are now numerous studies available that use plant data in SSDs (Ding et al., 2016; Lewis & Thursby, 2018; Song et al., 2015), suggesting that this uncertainty should now be less of a factor in criteria development. In the context of our study, the inclusion of nonanimal data would increase the biological diversity of toxicity data sets and confer greater protection to primary producers from the effects of chemical contaminants. However, it is important to note that like many animal phyla, plants and algae also suffer from significant data shortages in aquatic toxicology. For example, an ECOTOX search for plant and algal data for the 12 chemicals assessed in the present study returned a total of just 60 saltwater and 4 freshwater data points that met our quality parameters. Although this low amount of data could be caused by our choice of chemicals, most of which are not typically used as herbicides, the overall low amount of sensitivity data from aquatic plants and algae is a recognized trend in the literature (Lewis & Thursby, 2018). The lack of qualified data for aquatic plants and algae is striking given their critical importance to aquatic food webs and ecosystem, and needs to be addressed in WQC development.

CONCLUSIONS

Thirty-seven years after the publication of the USEPA's National Guidelines, it has become critical to reassess the factors that influence the development of WQC. We have identified the persistence of major gaps in toxicity data since 1985 that are acting collectively to limit the USEPA's development of acute WQC. This is particularly evident when it comes to the need to generate and incorporate toxicity data for primary producers. Although there was minimal evidence to suggest that the amount of biological diversity of a toxicity data set can influence the value of a criterion, we did identify significant taxonomic differences in sensitivity in the toxicity data from a diverse group of aquatic pollutants, which suggests that it may be necessary to consider taxonomic patterns of sensitivity during future criteria development. Finally, because the current rate of criteria development dramatically lags behind the

identification of harmful pollutants in aquatic environments, we recommend that the USEPA formally incorporate interspecies correlation estimation models into the National Guidelines and reconsider their definition of acceptable toxicity data when calculating WQC. Changes such as these are necessary if the USEPA is to meet the goals of the Clean Water Act and keep pace with the growing number of chemical pollutants that threaten the aquatic environment.

REFERENCES

Ankley, G., Barron, M., Bergeron, C., Burgess, R., Cormier, S., Etterson, M., Hoff, D., Lalone, C., Mount, D., Suter, G., Villeneuve, D., Ho, K., Cantwell, M., Willming, M., Raimondo, S., Cruz, L., Jarvis, A., Eignor, D., Justice, J., … Gallagher, K. (2017). *Update on US EPA's revision to the 1985 guidelines for deriving aquatic life criteria* (pp. 1–3). Society of Environmental Toxicology and Chemistry.

Arthur, J. W., West, C. W., Allen, K. N., & Hedtke, S. F. (1987). Seasonal toxicity of ammonia to five fish and nine invertebrate species. *Bulletin of Environmental Contamination and Toxicology*, *38*, 324–331.

Augspurger, T., Keller, A., Black, M., Cope, G., & Dwyer, J. (2003). Water quality guidance for protection of freshwater mussels (Unionidae) from ammonia exposure. *Environmental Science and Technology*, *22*(11), 2569–2575.

Awkerman, J., Raimondo, S., Jackson, C., & Barron, M. (2014). Augmenting aquatic species sensitivity distributions with interspecies toxicity estimation models. *Environmental Toxicology and Chemistry*, *33*(3), 688–695. https://doi.org/10.1002/etc.2456

Buchwalter, D. B., Clements, W. H., & Luoma, S. N. (2017). Modernizing water quality criteria in the United States: A need to expand the definition of acceptable data. *Environmental Toxicology and Chemistry*, *36*(2), 285–291. https://doi.org/10.1002/etc.3654

Burnham, K., & Anderson, D. (2002). *Model selection and multimodel inference*. Springer.

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

Carr, G. J., & Belanger, S. E. (2018). SSDs revisited: Part I—A framework for sample size guidance on species sensitivity distribution analysis. *Environmental Toxicology and Chemistry*, 38(7), 1514–1525. https://doi.org/10.1002/etc.4445

Clarke, K. R., & Warwick, R. M. (1998). A taxonomic distinctness index and its statistical properties. *Journal of Applied Ecology*, 35(4), 523–531. https://doi.org/10.1046/j.1365-2664.1998.3540523.x

Cormier, S. M., Paul, J. F., Spehar, R. L., Shaw-Allen, P., Berry, W. J., & Suter, G. W. (2008). Using field data and weight of evidence to develop water quality criteria. *Integrated Environmental Assessment and Management*, 4(4), 490–504. https://doi.org/10.1897/IEAM_2008-018.1

Ding, C., Ma, Y., Li, X., Zhang, T., & Wang, X. (2016). Derivation of soil thresholds for lead applying species sensitivity distribution: A case study for root vegetables. *Journal of Hazardous Materials*, 303, 21–27. https://doi.org/10.1016/j.jhazmat.2015.10.027. https://www.sciencedirect.com/science/article/pii/S0304389415301552

Elias, M., Ankley, G., Barron, M., Eignor, D., Erickson, R., Hoff, D., Jarvis, L., Justice, J. R., Lalone, C., Mount, D., Villeneuve, D., & Gallagher, K. (2019). *Developing protective aquatic life values for data-limited PFAS chemicals criteria development and minimum data requirements (MDRs)* (pp. 1–18). SETAC North America, Toronto, ON, Canada, November 3–07.

Giddings, J. M., Wirtz, J., Campana, D., & Dobbs, M. (2019). Derivation of combined species sensitivity distributions for acute toxicity of pyrethroids to aquatic animals. *Ecotoxicology*, 28, 242–250. https://doi.org/10.1007/s10646-019-02018-0

Leung, K. M. Y., Morritt, D., Wheeler, J. R., Whitehouse, P., Sorokin, N., Toy, R., Holt, M., & Crane, M. (2001). Can saltwater toxicity be predicted from freshwater data? *Marine Pollution Bulletin*, 42(11), 1007–1013. https://doi.org/10.1016/S0025-326X(01)00135-7

Lewis, M., & Thursby, G. (2018). Aquatic plants: Test species sensitivity and minimum data requirement evaluations for chemical risk assessments and aquatic life criteria development for the USA. *Environmental Pollution*, 238, 270–280. https://doi.org/10.1016/j.envpol.2018.03.003

Liu, Q., Xu, X., Zeng, J., Huang, W., Xu, X., Shou, L., & Chen, Q. (2019). Development of marine water quality criteria for inorganic mercury in China based on the retrievable toxicity data and a comparison with relevant criteria or guidelines. *Ecotoxicology*, 28(4), 412–421. https://doi.org/10.1007/s10646-019-02032-2

Oksanen, A. J., Kindt, R., Legendre, P., Hara, B. O., Simpson, G. L., Stevens, M. H. H., & Wagner, H. (2014). *Vegan: Community ecology package. R package Ver 2.2-0*. http://CRAN.Rproject.org/package=vegan

Pavlaki, M. D., Araújo, M. J., Cardoso, D. N., Silva, A. R. R., Cruz, A., Mendo, S., Soares, A. M. V. M., Calado, R., & Loureiro, S. (2016). Ecotoxicity and genotoxicity of cadmium in different marine trophic levels. *Environmental Pollution*, 215, 203–212. https://doi.org/10.1016/j.envpol.2016.05.010

R Core Team. (2019). *R: A language and environment for statistical computing*. https://www.r-project.org/

Raimondo, S., Vivian, D. N., & Barron, M. G. (2010). *Web-based Interspecies Correlation Estimation (Web-ICE) for acute toxicity: User manual Ver 3.1*. U.S. Environmental Protection Agency.

Seegert, G., Fava, J., & Cumbia, P. 1985. How representative are the data sets used to derive national water quality criteria? *Aquatic Toxicology and Hazard Assessment. Seventh Symposium*, 854, 527–537. https://doi.org/10.1520/STP36288S.

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423.

Solomon, K. R., Baker, D. B., Richards, R. P., Dixon, K. R., Klaine, S. J., La Point, T. W., Kendall, R. J., Weisskopf, C. P., Giddings, J. M., Giesy, J. P.,

Hall, L. W., & Williams, W. M. (1996). Ecological risk assessment of atrazine in North American surface waters. *Environmental Toxicology and Chemistry*, 15(1), 31–76. https://doi.org/10.1002/etc.5620150105

Song, W., Chen, S., Liu, J., Chen, L., Song, N., Li, N., & Liu, B. (2015). Variation of Cd concentration in various rice cultivars and derivation of cadmium toxicity thresholds for paddy soil by species-sensitivity distribution. *Journal of Integrative Agriculture*, 14(9), 1845–1854. https://doi.org/10.1016/S2095-3119(14)60926-6. https://www.sciencedirect.com/science/article/pii/S2095311914609266

Stephan, C. E. (1984). Are the "Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Life and Its Uses" based on sound judgements? *Aquatic Toxicology and Hazard Assessment. Seventh Symposium*, 847, 515–526.

Stephan, C. E., Mount, D. I., Hansen, D. J., Gentile, J. R., Chapman, G. A., & Brungs, W. A. (1985). Guidelines for deriving numerical national water quality criteria for the protection of aquatic organisms and their uses. U. S. Environmental Protection Agency.

Tenbrook, P. L., Palumbo, A. J., Fojut, T. L., Hann, P., Karkoski, J., & Tjeerdema, R. S. (2010). The University of California-Davis methodology for deriving aquatic life pesticide water quality criteria. *Reviews of Environmental Contamination and Toxicology*, 209(August), 1–172. https://doi.org/10.1007/978-1-4419-6883-8_1

Thorley, J., & Schwarz, C. (2018). ssdtools: An R package to fit species sensitivity distributions. *Journal of Open Source Software*, 3(31), 2–3. https://doi.org/10.1007/b97636

U.S. Environmental Protection Agency. (1980). Guidelines and methodology used in the preparation of health effect assessment chapters of the consent decree water criteria documents. *Federal Register*, 45(231), 79347.

U.S. Environmental Protection Agency. (2003). Ambient aquatic life water quality criteria for atrazine (revised draft). Office of Water.

U.S. Environmental Protection Agency. (2013). Aquatic life ambient water quality criteria for ammonia—Freshwater (Fact Sheet). Office of Water.

U.S. Environmental Protection Agency. (2021a). *ECOTOX Knowledgebase*. https://cfpub.epa.gov/ecotox/

U.S. Environmental Protection Agency. (2021b). *National recommended water quality criteria—Aquatic life criteria table*. https://www.epa.gov/wqc/national-recommended-water-quality-criteria-aquatic-life-criteria-table

Wang, Z., Kwok, K. W. H., Lui, G. C. S., Zhou, G.-J., Lee, J.-S., Lam, M. H. W., & Leung, K. M. Y. (2014). The difference between temperate and tropical saltwater species' acute sensitivity to chemicals is relatively small. *Chemosphere*, 105, 31–43. https://doi.org/10.1016/j.chemosphere.2013.10.066. https://www.sciencedirect.com/science/article/pii/S0045653513015002

Warne, M. S., Batley, G., van Dam, R., Chapman, J., Fox, D., Hickey, C., & Stauber, J. (2015). *Revised method for deriving Australian and New Zealand water quality guideline values for toxicants*. Australian Government Initiative.

Wheeler, J. R., Leung, K. M. Y., Morritt, D., Sorokin, N., Rogers, H., Toy, R., Holt, M., Whitehouse, P., & Crane, M. (2002). Freshwater to saltwater toxicity extrapolation using species sensitivity distributions. *Environmental Toxicology and Chemistry*, 21(11), 2459–2467. https://doi.org/10.1002/etc.5620211127

Wilcut, L., Lehmann, W., & Elias, M. (2015). Updating EPA's guidelines for deriving national recommended water quality criteria. Invited Expert Meeting on Revising US EPA's Guidelines for Deriving Aquatic Life Criteria, September 14–16, 2015, Washington, DC.

World Register of Marine Names. (2021). An authoritative classification and catalogue of marine names. https://doi.org/10.14284/170.

Wu, J., Yan, Z., Liu, Z., Liu, J., Liang, F., Wang, X., & Wang, W. (2015). Development of water quality criteria for phenanthrene and comparison of the sensitivity between native and non-native species. *Environmental Pollution*, 196, 141–146. https://doi.org/10.1016/j.envpol.2014.09.024