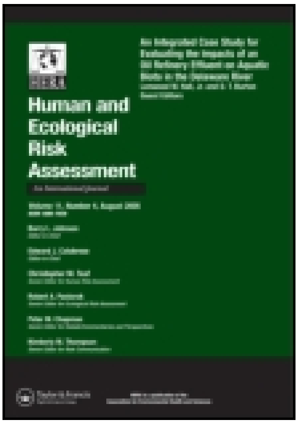


Human and Ecological Risk Assessment: An International Journal
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/bher20

Assessing the Relative Sensitivity of Aquatic Organisms to Divalent Metals and Their Representation in Toxicity Datasets Compared to Natural Aquatic Communities

Kevin V. Brix [a b], David K. DeForest [c], Mike Burger [c] & William J. Adams [d]

[a] EcoTox, 590 Ocean Drive, #2C, Key Biscayne, Florida, 33149, USA

[b] Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, Florida, 33149, USA

[c] Parametrix, Inc., 411 108th Ave. NE, Bellevue, Washington, 98004, USA

[d] Rio Tinto, 5295 South 300 West, Murray, Utah, 84107, USA

Available online: 18 Jan 2007

To cite this article: Kevin V. Brix, David K. DeForest, Mike Burger & William J. Adams (2005): Assessing the Relative Sensitivity of Aquatic Organisms to Divalent Metals and Their Representation in Toxicity Datasets Compared to Natural Aquatic Communities, Human and Ecological Risk Assessment: An International Journal, 11:6, 1139-1156

To link to this article: http://dx.doi.org/10.1080/10807030500278628

Human and Ecological Risk Assessment, 11: 1139–1156, 2005

ISSN: 1080-7039 print / 1549-7680 online
DOI: 10.1080/10807030500278628

Assessing the Relative Sensitivity of Aquatic Organisms to Divalent Metals and Their Representation in Toxicity Datasets Compared to Natural Aquatic Communities

Kevin V. Brix,[1,2] David K. DeForest,[3] Mike Burger,[3] and William J. Adams[4]
[1]EcoTox, 590 Ocean Drive, #2C, Key Biscayne, Florida 33149, USA; [2]Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, Florida 33149, USA; [3]Parametrix, Inc., 411 108th Ave. NE, Bellevue, Washington 98004, USA; [4]Rio Tinto, 5295 South 300 West, Murray, Utah 84107, USA

ABSTRACT

We compared the composition and richness of acute and chronic toxicity datasets for Cd, Cu, Ni, Pb, and Zn to several natural aquatic communities. The richness of acute datasets was reasonably representative, with the largest toxicity datasets containing a higher number of genera than some natural aquatic communities. Acute datasets also had a reasonably diverse composition compared to natural aquatic communities, although insects were under-represented and cladocerans over-represented. Given this robustness, we suggest manipulation of large acute datasets (Cd, Cu, Zn) to account for site-specific differences in aquatic community composition can be accomplished with confidence and that this will not result in under-protection of sensitive taxa. In contrast, the chronic datasets were not representative of natural aquatic communities in terms of composition or richness. Chronic dataset richness is an order of magnitude less than natural aquatic communities. Chronic datasets have minimal representation of insects, whereas cladocera and salmonids are grossly over-represented in some cases. Further, no real patterns in the relative sensitivity of genera groups can be discerned with such limited data. As a result, we conclude there is considerable uncertainty regarding how biases in genera representation may lead to under- or over-protection of aquatic communities on a chronic basis. Given this, manipulation of chronic datasets to better reflect site-specific aquatic communities is not recommended without additional chronic testing using a wider diversity of aquatic genera.

Key Words: water quality criteria, metals, composition, richness, risk assessment.

Received 24 June 2004; revised manuscript accepted 26 January 2005.
Address correspondence to Kevin V. Brix, EcoTox, 590 Ocean Drive, #2C, Key Biscayne, FL 33149, USA. E-mail: kvbrix@aol.com

INTRODUCTION

Aquatic toxicity datasets have been used for several purposes including development of water quality criteria (WQC) and species sensitivity distributions (SSDs) for risk assessment. The intent is to derive a value (concentration) that is sufficiently low to be adequately protective of ecosystem structure and function or, in the case of risk assessment, to describe the fraction of the aquatic community that might be impacted by a given exposure concentration. Water quality criteria have been derived using several types of data, including toxicity tests of single species, multiple species microcosms/mesocosm tests, and field studies. Because of the cost and complexity of field and mesocosm studies, most WQC have been derived from laboratory toxicity tests of single species. A key issue in deriving a WQC is how to ensure the distribution of taxa sensitivities on which the WQC is based is ecologically relevant and that it provides a realistic level of protection for a given substance. Characterizing the sensitivity for all taxa in an ecosystem of interest is generally not possible, necessitating the use of toxicity datasets with limited numbers of species. Various approaches have been applied to toxicity datasets to derive WQC. These approaches have used different statistical models of species sensitivity distributions (Stephan *et al.* 1985; Aldenberg and Slob 1993) and extrapolation factors such as the acute-chronic ratio (Stephan *et al.* 1985).

Extrapolation approaches were developed to avoid many of the problems inherent with use of assessment factors. The most frequently used extrapolation approaches (Stephan *et al.* 1985; Aldenberg and Slob 1993) are designed to address the issues of number of taxa represented, the variance in the SSD, and level of protection. These approaches assume that the variability in toxicity data between species can be addressed by an appropriate cumulative probability distribution (*e.g.*, SSD). This SSD and its parameters (*e.g.*, mean and standard deviation) are typically derived from laboratory toxicity tests that meet specific criteria concerning data quality and relevance. The SSD is used to estimate the concentration of a toxicant at which a specific percentage of taxa are protected. This concentration is selected as the WQC, which protects a certain percentage of the aquatic community (*e.g.*, 95%) and the specific taxa encompassed by this distribution.

A key assumption in the use of SSDs to derive WQC or for risk assessment is that the toxicity data in the distribution are random and representative samples of the aquatic community intended to be protected (Stephan *et al.* 1985; OECD 1992). In an attempt to ensure this assumption is met, requirements on minimum data size and composition have been recommended in the United States and Europe (Stephan *et al.* 1985; OECD 1992). Stephan *et al.* (1985), for example, recommends that for freshwater criteria a minimum of eight genera be tested. Further, the composition of these eight genera in freshwater must include the following: (1) At least one representative from the family Salmonidae, in the class Osteichthyes; (2) a second family in the class Osteichthyes; (3) a third family in the phylum Chordata (either Osteichthyes or Amphibia); (4) a planktonic crustacean (*e.g.*, cladoceran, copepod); (5) a benthic crustacean (*e.g.*, ostracod, isopod, amphipod); (6) at least one representative from the class Insecta; (7) a family in a phylum other than Arthropoda or Chordata (*e.g.*, Rotifera, Annelida, Mollusca); and (8) a family in any order of insect or any other phylum not already represented.

The overall intent of these testing requirements is to provide a phylogenetically diverse dataset that includes taxa groups known to generally be sensitive to contaminants. Although this may be the intent of the testing requirements, since their inception questions have been raised regarding the representativeness of the datasets. For example, Seegert *et al.* (1985) evaluated a dataset used to derive draft WQC for 21 priority pollutants. They noted that of the 421 toxicity tests used to derive the criteria, 25% of the tests were on species from just 2 families, Salmonidae and Daphnidae, and 40% of the tests were conducted using just 11 species. More recently, Forbes and Calow (2002) evaluated how the composition of chronic SSDs for 11 substances (including Cd, Cu, and Zn) compared to real aquatic communities with respect to the level of representation in each trophic level. They concluded that these datasets were heavily biased toward over-representation of upper trophic levels compared to real aquatic communities. A further sensitivity analysis in which they re-sampled the distribution to artificially balance the SSDs with respect to trophic representation led them to conclude that this misrepresentation could lead to over- or under-protection depending on the substance evaluated (Forbes and Calow 2002).

As shown by Forbes and Calow, the importance of these biases in deriving WQC and conducting risk assessments can only be understood if the relative sensitivities of different taxonomic groups to a substance are also understood. Over the past 15 years, a considerable amount of additional toxicity testing has occurred that should improve the representativeness of the toxicity datasets used to derive WQC. Toxicity datasets for metals are particularly large and so should provide best-case scenarios regarding the composition and richness of toxicity datasets. The purpose of our analysis was threefold—first, to quantitatively characterize the relative sensitivity of different freshwater taxonomic groups to metals; second, to characterize the composition and richness of toxicity datasets for metals by comparing them to the composition and richness in several aquatic ecosystems across the United States; and third, to evaluate any biases in composition in the context of the relative sensitivity of various taxonomic groups to metals.

METHODS AND MATERIALS

Datasets

To accomplish these objectives, both acute and chronic toxicity datasets for Cd, Cu, Pb, Ni, and Zn were used. The datasets were compiled by searching a variety of online data sources and existing data compilations such as the U.S. Environmental Protection Agency (USEPA) ambient water quality criteria documents (USEPA 1996). Data were evaluated for quality using the guidelines described in Stephan *et al.* (1985).

For comparison to laboratory toxicity datasets, biological survey data on the composition and richness of six aquatic communities representing different ecoregions across the United States were compiled. These sites consisted of four stream systems (lotic) and two lakes (lentic). They were selected because they are known to have limited anthropogenic perturbations and thus represent "reference" conditions. These six communities by no means provide a complete representation of aquatic communities in the United States, but do provide a relatively diverse range of habitats and consequently aquatic communities.

The first lotic system, termed "Puget Sound Streams" consisted of approximately 40 second- to fifth-order streams in the Puget Sound Basin of the U.S. Pacific Northwest. This basin is well known for its salmonid-dominant ecosystem. The primary data source for this site was Black and Silkey (1997). The second lotic system was the St. Clair River, located in the U.S. Great Lakes region. The St. Clair River is a 64 km^2 connecting channel that carries Lake Huron outflow into Lake Erie (flow rate 5,200 $m^3\,s^{-1}$). Because of the relatively high quality of water in Lake Huron, the river supports a diverse and productive biological community. The primary data source for this site was Griffiths *et al.* (1991). The third lotic system was the Gila River in Arizona. This slightly saline system is representative of the relatively low species richness characteristic of the arid West ecoregion. Biological survey data for this system were obtained from Camp Dresser & McKee (CDM 2000). The final lotic system was Little Pigeon Creek in east Tennessee. This is a second- to third-order stream located in the foothills of the Great Smoky Mountains and is characteristic of streams in this region that have a high species richness (Abell *et al.* 2000). Biological survey data from this site were obtained from the Oak Ridge National Laboratory (ORNL 1999).

The first lentic system, Trout Lake, is a relatively small lake (1600 ha) located in northern Wisconsin, with limited shoreline development. This lake has been extensively studied as part of the North Temperate Lakes Long-Term Ecological Research Project conducted by the University of Wisconsin. The primary data source for Trout Lake is the university's website on the project (http://limnosun.limnology.wisc.edu/). The second lentic system, Lake Annie, is a 36 ha sinkhole in south central Florida with no development. The primary data source for Lake Annie was Layne (1979).

Data Analysis

Relative sensitivity of different taxonomic groups

To conduct the comparative analysis between toxicity datasets and natural aquatic communities, it was first necessary to determine how best to categorize the data. Brix *et al.* (2001) previously conducted such an assessment for acute and chronic Cu toxicity data by evaluating the relative sensitivity of different organism groups, using two different categorization schemes—phylogenetic and ecological/trophic position. They concluded that a combination of these two schemes was optimal for characterizing the sensitivity of different species groups. The specific groupings they identified and also used in this study were cladocera, other crustacea, insects, other invertebrates, coldwater fish, and warmwater fish. The other crustacea category includes all crustacea except cladocerans whereas the other invertebrate category includes all invertebrates that are not crustacea or insects.

Using these taxonomic groupings, acute and chronic toxicity data for each of the metals were binned accordingly and the relatively sensitivity of the different groups characterized. For acute datasets, this was accomplished by plotting separate SSDs for each of the species groupings. Within a metal, statistical differences between Genus Mean Acute Value (GMAV) distributions for the taxonomic groups were tested using Levene's test for homogeneity of variance and nonparametric analysis of variance (ANOVA) for differences between the means. Nonparametric assessment of differences between means was used because the distributions did not have homogeneous variances, were not normally distributed, and so did not meet the assumptions of conventional parametric analyses. To conduct the nonparametric analysis in a way that would allow for easy *post-hoc* multiple comparison of means, the data were rankit transformed and then used in one-way ANOVA and subsequent *post-hoc* multiple comparison tests with the software program SPSS (1999). To calculate the rankit, the GMAVs for all taxonomic groups were first pooled and ranked from low to high. Proportional ranks were calculated as $k/(N+1)$, where k is the rank of the GMAV and N is the total number of GMAVs for a metal. The normal scores of the proportional ranks, the rankit transformation (Conover 1999), were then used in the analyses. The rankit transformation normalizes the data and makes the variances homogeneous, allowing for a nonparametric statistical analysis to be conducted using standard parametric statistical software. If the p value result from the ANOVA was less than .05, differences between individual taxonomic groups were then assessed using a least significant difference (LSD) multiple comparison test. The LSD test evaluated all two-way pairs of distributions.

Because of the relatively small sample sizes for chronic datasets, we simply determined minimum, maximum, and mean values for each taxonomic group–metal combination. Given the relatively small sample sizes overall, developing SSDs for individual taxonomic groups and statistically testing for differences between groups did not appear warranted.

Dataset richness

For each dataset (both field and toxicity), the phylogeny (phylum, class, order, family, genus) of each of the species represented was determined. Comparisons were limited to the genus level, as this is the lowest taxonomic level used by the USEPA to derive WQC (Stephan *et al.* 1985). The richness (*i.e.*, total number of taxa) of each dataset was then characterized at the different taxonomic levels for comparison.

Dataset composition

Perhaps more important than the total number of genera represented in each dataset is the composition of each dataset. To evaluate how representative the composition of toxicity datasets of natural aquatic communities are, an analysis was conducted using the same datasets described earlier. For each dataset the genera were divided into the same groupings used to characterize their relatively sensitivity (cladocera, other crustacea, insects, other invertebrates, coldwater fish, and warmwater fish).

RESULTS

Relative Sensitivity of Different Taxonomic Groups

Evaluation of the acute toxicity datasets indicated that cladocerans were the most sensitive taxonomic group for all the metals evaluated, with the exception of Cd (Figure 1). Only in the case of Zn though, were cladocerans statistically significantly

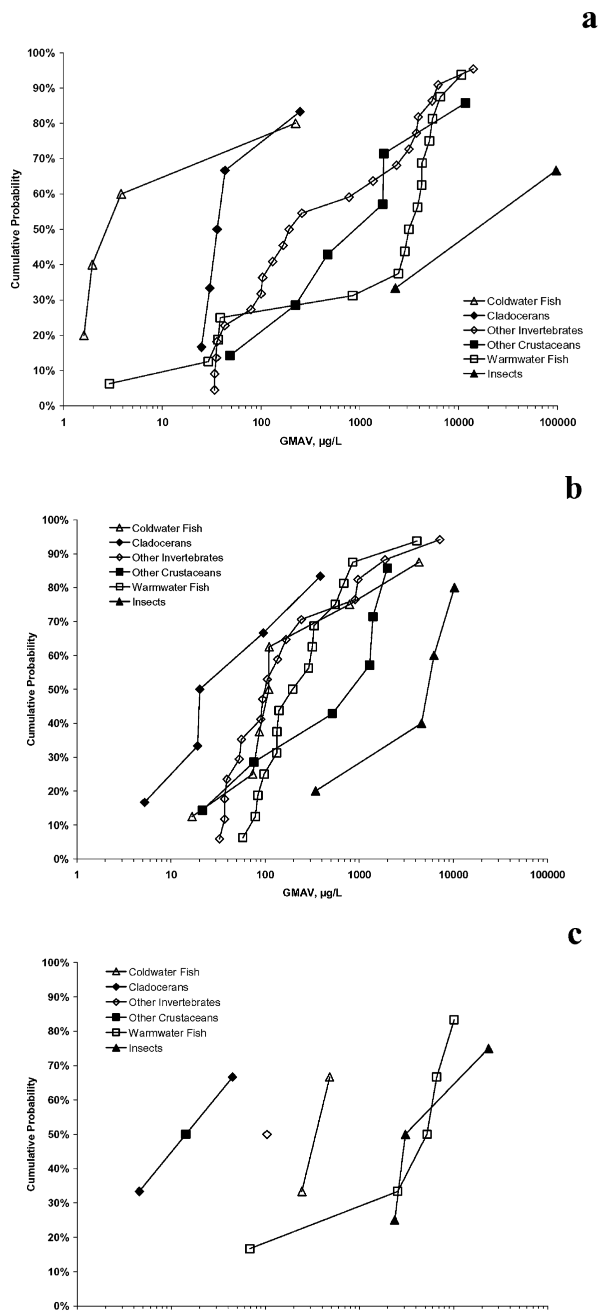
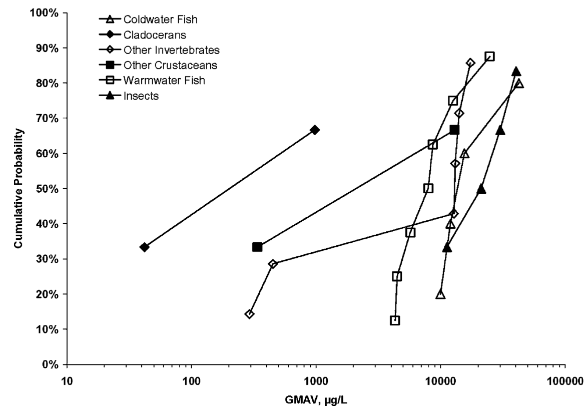


Figure 1. Acute species sensitivity distributions for (a) cadmium, (b) copper, (c) lead, (d) nickel, (e) zinc. Taxonomic groups are not significantly ($p < .05$) different if they share a common letter as indicated in the figure legend. Small sample size precluded statistical analysis of lead. *(Continued)*

d



e

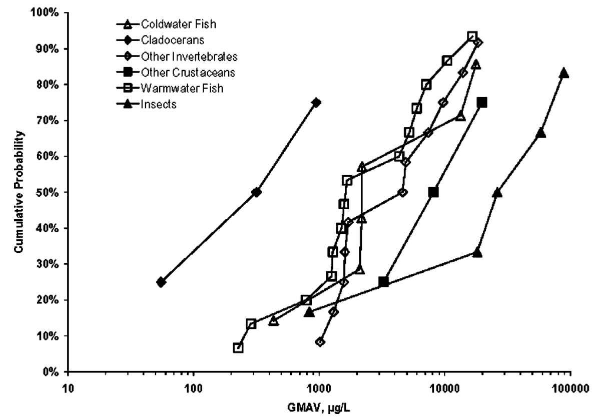


Figure 1. (Continued)

different ($p < .05$) from all other species groups (Figure 1). Warmwater fish, other crustaceans and other invertebrates were consistently of intermediate sensitivity with overlapping SSDs that were statistically similar for all metals ($p > .05$). Insects were the least sensitive taxonomic group evaluated for all five metals, and in the case of Cu were significantly less sensitive than all other species groups (Figure 1). Coldwater fish had varying sensitivity, being the most sensitive (Cd), of intermediate sensitivity (Cu, Pb, Zn), or the least sensitive (comparable to insects for Ni) taxonomic group.

It was difficult to evaluate the chronic toxicity data in terms of relative sensitivity of different taxonomic groups due to the paucity of data. In general, cladocera tended to be either the most sensitive or one of the most sensitive groups whereas insects were again consistently the least sensitive group evaluated (Table 1). These qualitative generalizations should be treated with great caution though given the extremely small sample sizes and resulting lack of statistical analysis.

Table 1. Summary of chronic values by metal and species group (μg metal/L).

Statistic	Cladocera	Other crustaceans	Insects	Other invertebrates	Warmwater fish	Coldwater fish
Cadmium						
Mean	10.0	0.3	NA	12.7	12.6	5.5
Min	0.19	0.25	NA	4.8	5.3	2.4
Max	28.3	0.25	NA	20.5	22.5	8.1
n	3	1	NA	2	6	4
Copper						
Mean	5.5	6.7	18	12	18	32
Min	5.5	6.7	18	12	5.6	13
Max	5.5	6.7	18	12	32	66
n	1	1	1	2	5	4
Lead						
Mean	26	NA	NA	6.9	NA	69
Min	26	NA	NA	6.9	NA	39
Max	26	NA	NA	6.9	NA	98
n	1	NA	NA	1	NA	2
Nickel						
Mean	9.1	19	283	NA	45	33
Min	3.6	19	120	NA	45	33
Max	15	19	445	NA	45	33
n	2	1	2	NA	1	1
Zinc						
Mean	52	NA	7861	24	140	731
Min	52	NA	7861	24	41	543
Max	52	NA	7861	24	267	919
n	1	NA	1	1	3	2

NA = Not available.

Species Richness

The field datasets were comprised of 4 to 7 phyla, 6 to 14 classes, 16 to 36 orders, 39 to 84 families, and 59 to 202 genera. Lotic systems had a higher richness than lentic systems, as would generally be expected given the greater diversity of habitats that typically occur in lotic systems (Merritt and Cummins 1996). The exception to this generalization was the Gila River in Arizona, which had relatively low richness compared to other lotic systems (approximately one half). This is characteristic of streams in the arid West where factors such as elevated water temperature and ionic strength tend to reduce diversity (Abell *et al.* 2000).

The acute toxicity datasets included data for 4 to 6 phyla, 6 to 13 classes, 10 to 26 orders, 11 to 32 families, and 14 to 53 genera. The richness of Cd, Cu, and Zn were roughly comparable whereas Ni and Pb had considerably less diversity. As would be expected, richness was considerably lower in the chronic datasets with only 2 to 4 phyla, 3 to 6 classes, 3 to 10 orders, 3 to 12 familes, and 4 to 16 genera represented. For the chronic datasets, Cd and Cu were comparable in richness, followed by Ni and Zn with comparable but lower richness, and Pb being the least rich.

Dataset Composition

Each of the acute, chronic, and natural aquatic community datasets were categorized using the same taxonomic groups used to characterize relative sensitivity. The six natural aquatic communities show that there are distinct differences between lotic and lentic systems as would be expected. In Trout Lake, warmwater fish, other invertebrates and insects each comprised one-quarter to one-third of the total community composition, with the remaining species groups (coldwater fish, other crustaceans, and cladocerans) collectively comprising approximately one quarter of

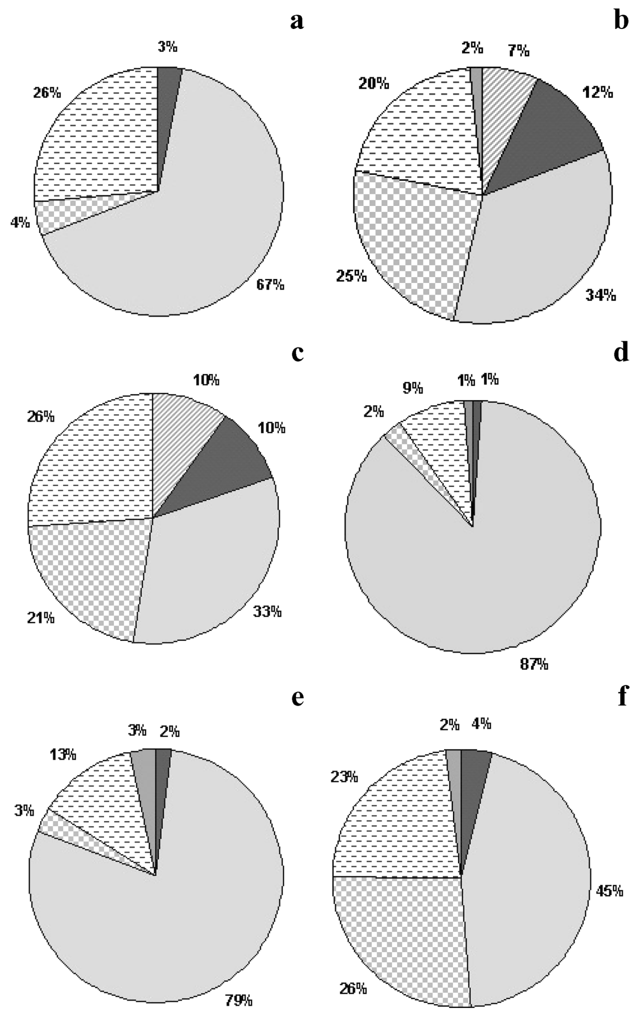


Figure 2. Composition of natural aquatic communities: (a) Lake Annie, (b) Trout Lake, (c) Gila River, (d) Little Pigeon Creek, (e) Puget Sound Streams, (f) St. Clair River. Genus used as the lowest common taxonomic level for analysis. Cladocera, Other Crustaceans, Insects, Other Invertebrates, Warmwater Fish, Coldwater Fish.

the total community (Figure 2). In contrast, Lake Annie was dominated by insects that comprised two-thirds of all genera. Warmwater fish comprised one-quarter of the total genera, with other crustaceans and other invertebrates collectively comprising about 6% of the total genera. Similar to results for species richness, species composition in the Gila River was more similar to the lentic systems than to the other lotic systems.

Three of the lotic systems were dominated by insects. This was particularly true for Little Pigeon Creek and the Puget Sound Streams where insects comprised 87% and 79% of the total species composition (Figure 2), respectively. Other invertebrates and other crustaceans collectively represented 30% of the species composition in the St. Clair River but only comprised 3% and 5% of the species composition in Little Pigeon Creek and Puget Sound Streams, respectively. Similarly, warmwater fish comprised a relatively large percentage (23%) of the total species composition in the St. Clair River, but a considerably lower percentage (9–13%) in the other two systems. The Gila River had a taxonomic composition much more similar to Trout Lake than the other three lotic systems. Finally, it is also important to note that in the Puget Sound Streams dataset, which is one of the most "salmonid rich" systems in the world, coldwater fish (including salmonids) only comprised 3% of the total genera for the system.

The taxonomic compositions of the acute and chronic datasets are shown in Figures 3 and 4. In the acute datasets, warmwater fish comprised roughly one third (27–37%) and coldwater fish represented 8–15% of the total composition for each metal. Insect and cladocera representation was generally low, ranging from 4–21% and 7–14% of the total tested community, respectively. Combined, the other crustacea and other invertebrate categories varied from 14–51% of the total community for each metal. In the chronic datasets, coldwater fish and cladocera were much more important groups relative to the acute datasets. Coldwater fish represented 11–50% and cladocera 7–29% of the tested community. Insect representation in the chronic datasets was roughly comparable to the acute datasets, ranging from 0–29%. The higher representation of coldwater fish and cladocera in the chronic datasets resulted in corresponding reduction in representation for warmwater fish, other crustaceans, and other invertebrates.

DISCUSSION

General

A number of observations can be made from this analysis. First, the acute datasets had a comparable level of richness in terms of phyla, classes, and orders when compared to the natural aquatic communities, but tend to have lower richness at the family and genus levels. At the genus level, real aquatic communities had 1.1 to 3.8 times the richness of the richest acute toxicity dataset (Cd) and 4.2 to 14.4 times the richness of the poorest acute toxicity dataset (Ni). In comparison, the chronic datasets are considerably less rich than the example natural aquatic communities. Even at the phylum level, the chronic toxicity datasets have only one-half to one-quarter the richness of the natural aquatic communities. This disparity continues to

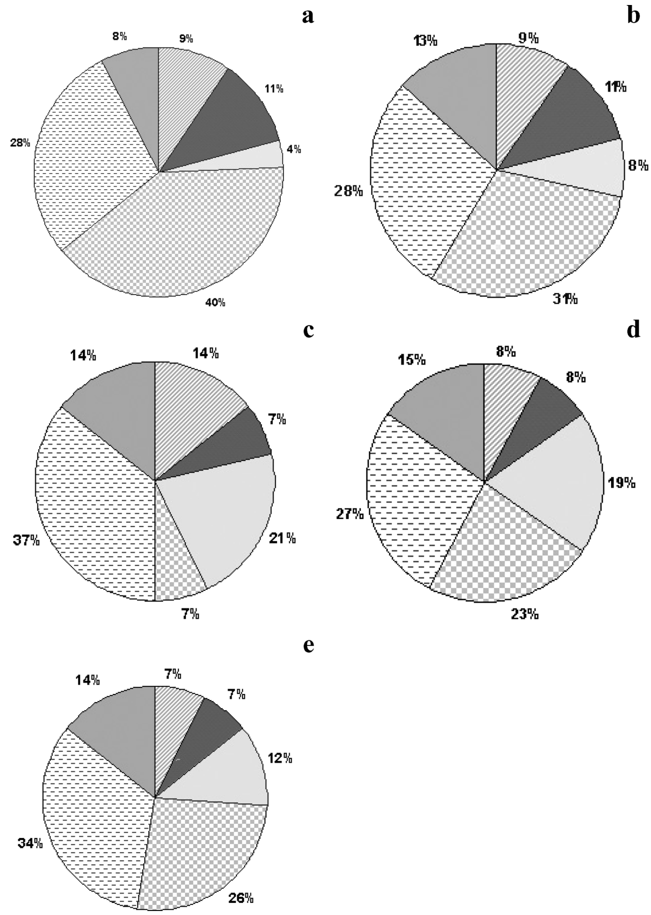


Figure 3. Composition of acute toxicity data sets: (a) cadmium, (b) copper, (c) lead, (d) nickel, (e) zinc. Genus used as the lowest common taxonomic level for analysis. ▨ Cladocera, ■ Other Crustaceans, ■ Insects, ▩ Other Invertebrates, ▒ Warmwater Fish, ■ Coldwater Fish.

increase until the genus level, where real aquatic communities are richer than the toxicity datasets by factors of 3.7 to 50.

Perhaps more important than absolute richness is the taxonomic composition of the toxicity datasets compared with natural aquatic communities. Again, the acute datasets are far more diverse and more closely approximate the composition real aquatic communities compared with the chronic datasets. In general for the acute datasets, insects are significantly under-represented, coldwater fish are significantly over-represented, and cladocerans are slightly over-represented. A similar pattern is observed for the chronic datasets, but the degree of under-representation for insects is less, whereas the degree of over-representation for coldwater fish and cladocerans is greater. The better representation of insects in the chronic datasets does not reflect more testing of insects, rather less testing of other species groups (*e.g.*, other

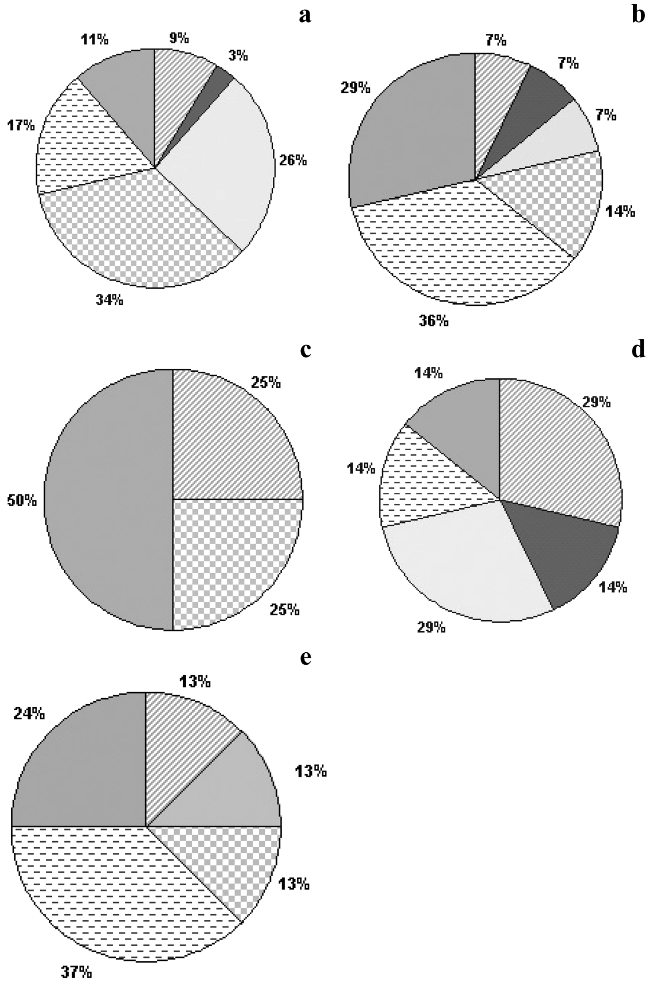


Figure 4. Composition of chronic toxicity datasets: (a) cadmium, (b) copper, (c) lead, (d) nickel, (e) zinc. Genus used as the lowest common taxonomic level for analysis. Cladocera, Other Crustaceans, Insects, Other Invertebrates, Warmwater Fish, Coldwater Fish.

invertebrates and other crustaceans) such that the few tests with insects represent a greater fraction of the total tested community. The greater over-representation for coldwater fish and cladocerans in the chronic datasets is a result of the same process, less testing of other invertebrates and other crustaceans.

The impact of these biases in taxonomic composition in the toxicity datasets on derivation of WQC and SSDs is worth consideration. The primary manner in which these biases might affect WQC and SSD derivation is if one or more of the groups is consistently sensitive or insensitive to metals relative to other species groups tested. If this were the case, under- or over-representation of the species group could result in under- or over-protection.

Acute Datasets

For acute toxicity, trends in the relative sensitivity clearly demonstrate that insects are consistently among the most insensitive groups tested (Figure 1). Under-representation of this group therefore poses an interesting scenario in terms of WQC derivation. Based on the available data, it is highly unlikely that under-representation of insects has led to the exclusion of genera sensitive near the WQC. To the contrary, it is likely that insensitive genera have been excluded and under-represented. The effect of this exclusion on WQC is expected to be minimal because of the way in which WQC are calculated in the United States. Specifically, WQC are calculated by evaluating the distribution of the four most sensitive genera and considering the total number of genera tested (Stephan *et al.* 1985). Because it is not anticipated that additional testing with insects would change the distribution of the four most sensitive genera, this bias in representation could only affect the total number of genera tested. Although the total number of genera tested can strongly influence WQC, it becomes considerably less important for large datasets such as the acute datasets for metals. For example, using the acute Cu dataset assembled for this study, a Criterion Maximum Concentration (CMC) of 7.8 μg/L at a hardness of 50 mg/L is estimated. If 30 additional insect genera were tested for acute Cu toxicity and as expected none of them were in the four most sensitive genera tested, the CMC would only increase to 10.8 μg/L Cu.

In contrast to the WQC assessment, when the entire SSD is considered as in probabilistic risk assessment, the under-representation of insects does potentially have some impact. Note, when we discuss probabilistic risk assessment we are referring to full use of the SSD and not probabilistic derivation of an HC5 (*e.g.*, hazard concentration for the 5th percentile taxa in an SSD) (Aldenberg and Slob 1993), which is then used in a hazard quotient calculation. To evaluate this scenario, we fit a distribution to the available GMAVs for insects and sampled from this distribution 30 times with replacement to artificially increase the representation of insects in the overall dataset. A comparison of the original SSD with the SSD including the 30 additional simulated insect GMAVs shows a significant shifting of the expanded SSD to the right (Figure 5a). Even in the low part of the distribution, inclusion of the additional insect data has a significant impact that could influence risk estimation results in moderately contaminated areas (Figure 5b). Hence, the under-representation of insects presents a conservative assessment of potential risks.

The effect of over-representation of coldwater fish, in particular salmonids, on the acute WQC for metals is variable. For Cd, where salmonid genera represent three of the four most sensitive taxa tested, the effect is potentially significant. For Cu, only 1 salmonid taxa is within the 4 most sensitive taxa whereas for the other three metals, salmonids are of intermediate sensitivity and so no effect on the WQC is anticipated. Specific to Cd, one might conclude that this over-representation of very sensitive taxa would lead to over-protection. However, we believe this is not the case. The presence of the three salmonid taxa tested with Cd (*Salmo, Salvelinus*, and *Oncorhynchus*) in a natural aquatic system is not particularly unusual and certainly occurs in the Puget Sound Streams system we evaluated. The reason for their over-representation in the metal datasets is not the result of testing too many salmonid taxa, but the lack of testing (*e.g.*, under-representation) of other taxa. Hence, following the earlier example

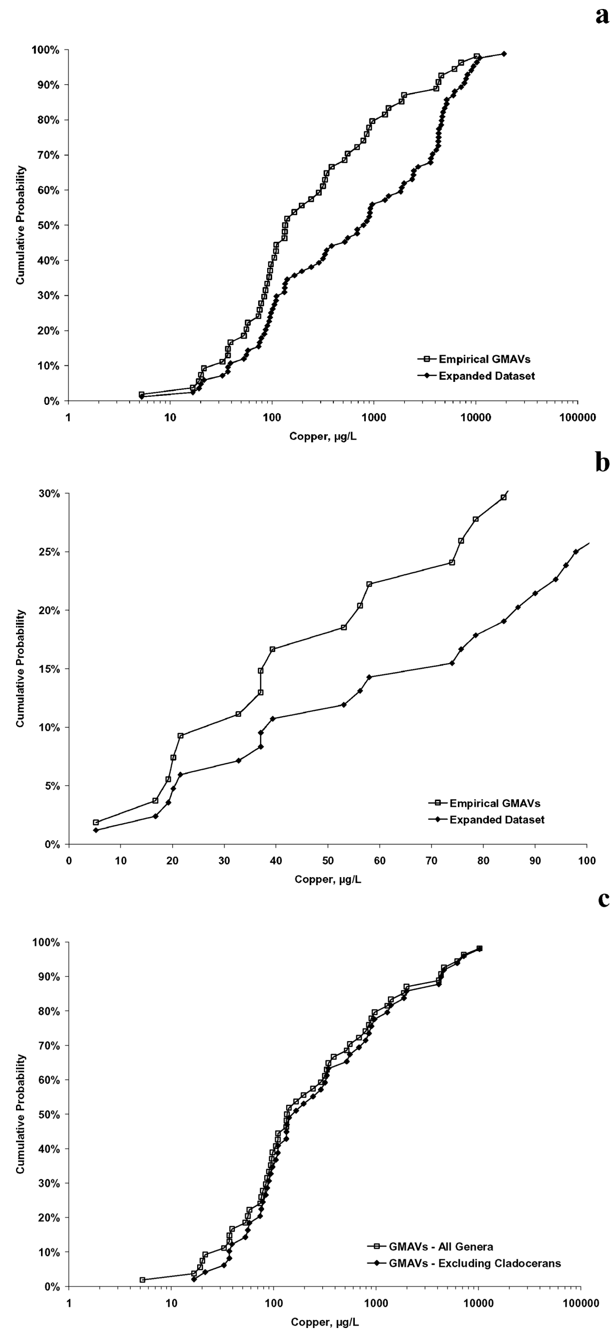


Figure 5. Acute copper SSD (a) with and without 30 additional insect taxa—entire SSD, (b) with and without 30 additional insect taxa—lower 30th percentile of SSD, (c) with and without all cladoceran taxa removed from the dataset—entire SSD, and (d) with and without all cladoceran taxa removed from the dataset—lower 30th percentile of SSD. *(Continued)*

d

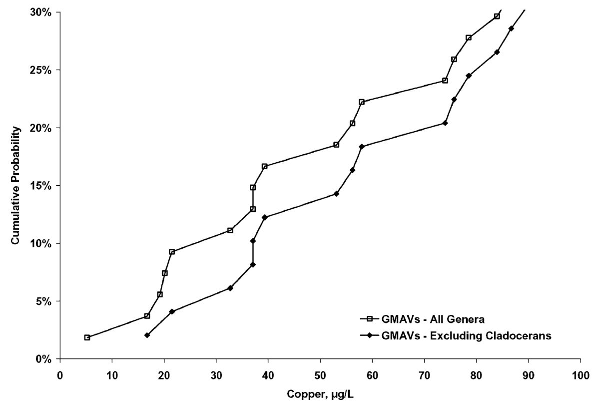


Figure 5. (Continued)

with insects, if the dataset composition was appropriately balanced by testing more of the under-represented groups, the reduced percentage of salmonids present would likely still represent the most sensitive genera and there would be the same general changes in the WQC and SSDs observed when discussing the under-representation of insects.

Cladocera were over-represented relative to some systems (Puget Sound Streams, Little Pigeon Creek, St. Clair River, Lake Annie), but representative of other systems (Trout Lake, Gila River). For the systems where they were over-represented, cladocerans do not occur at all. It is worth noting that although cladocerans are typically associated with lentic systems, this is not always the case (*e.g.*, Gila River). The issue of sensitivity is important because cladocerans are the most acutely sensitive taxonomic group for all of the metals except Cd, where they are the second most sensitive. Hence, for aquatic systems where they do not occur, criteria may indeed be over-protective. Using the full datasets compiled for this evaluation, estimated CMCs were 1.3, 7.8, 15.9, 40.7, and 80.5 μg/L for Cd, Cu, Pb, Ni, and Zn, respectively. Excluding cladocera from the datasets for each metal would result in CMCs of 1.3, 13.0, 46.8, 97.3, and 165.5 μg/L for Cd, Cu, Pb, Ni, and Zn, respectively. Hence, with the exception of Cd, excluding cladocera from the acute datasets increases the CMCs by a factor of 2 to 3 for each metal. As would be expected, removal of cladocera from SSDs for risk assessment purposes would have a relatively minor effect on the overall SSD (Figure 5c for example with Cu), but an effect roughly comparable to that observed for WQC on the lower decile of the distribution (Figure 5d).

Although removal of cladocera from the datasets is an interesting intellectual exercise, in the regulatory arena there has been considerable reluctance to selectively remove taxa that do not occur at a site from toxicity datasets and recalculate criteria accordingly. The philosophy has been that the toxicity datasets are intended to represent natural aquatic communities and although particular taxa in the toxicity dataset may not occur at a site, these taxa are representative (in terms of sensitivity) of taxa that do occur at the site but have not been tested. The U.S. Environmental Protection Agency (USEPA) does have a procedure for re-calculating AWQC based on the site-specific aquatic community (USEPA 1994). However, implementation of this procedure is often met with considerable resistance at the State level where anti-backsliding regulations frequently prevent such recalculations.

For the acute metal datasets (at least Cd, Cu, and Zn), the application of this philosophy is becoming increasingly questionable. As we have shown, these specific datasets are now generally comparable to natural aquatic communities in terms of both richness and diversity. As a result, there is now a strong understanding of the relative sensitivity of different taxonomic groups (see Figure 1). Further, studies on the mechanisms of metal toxicity are progressing sufficiently to allow reliable predictions of which taxa will be acutely sensitive to metals. For example, Grosell *et al.* (2002) recently developed a mechanistic model to predict the acute sensitivity of freshwater organisms to Ag and Cu as a function of Na turnover rate. Given the state of the science, concerns regarding the presence of an untested, but sensitive taxa are becoming increasingly unrealistic for acute metal toxicity. We suggest that the acute Cd, Cu, and Zn datasets are now sufficiently robust to be adjusted with confidence given proper consideration of site-specific or region-specific community composition (*e.g.*, removal of cladocera from the datasets when they do not occur in a given system). The smaller acute datasets of Ni and Pb, however, may not be sufficiently rich or diverse to allow for this type of adjustment.

Chronic Datasets

Unlike the earlier discussion for acute datasets, it is not possible to discuss potential biases in the chronic datasets for several reasons. First, the relatively small number of genera chronically tested makes it difficult to discern trends in relative sensitivity for the different groups, such as insects, where only one or two genera may have been tested for a given metal (Table 1). Second, we currently lack the mechanistic understanding of what causes differences in chronic sensitivity to metals and so cannot reliably predict the sensitivity of untested taxa. Third, the range in sensitivities for chronic toxicity data appears to be significantly less than the range observed for acute toxicity. For example, for Cu the range in acute sensitivities spans nearly four orders of magnitude whereas for chronic toxicity only one order of magnitude (Brix *et al.* 2001). Hence, biases in the composition of the toxicity dataset should have less of an impact on WQC or SSDs than might be observed for acute datasets. This of course assumes that the range of chronic sensitivities has been reasonably characterized by the limited testing to date.

Given these observations, a discussion of potential biases in the chronic toxicity datasets would be highly speculative at this point. As a consequence, manipulation of these datasets to reflect site-specific or region-specific community composition is not recommended until considerably more chronic toxicity testing with a diverse range of taxa has been undertaken.

CONCLUSIONS

There are no strict criteria by which to judge the extent of diversity that is required for a toxicity dataset to be considered adequate. Non-random selection of genera for testing does introduce bias into the toxicity datasets as evidenced for lotic systems where insects dominate, but are infrequently tested in the laboratory. In general, there is a need to ensure that species sensitivity is adequately assessed and sensitive taxa are protected. Our review of the metal datasets did not indicate that major groups of sensitive taxa have not been evaluated. Analysis of the composition and richness of acute and chronic toxicity datasets for metals as compared to natural aquatic communities leads us to conclude that the larger acute datasets (Cd, Cu, Zn) are generally comparable to the richness of natural aquatic communities and have a representative diversity although additional testing with insects would be useful. Overall, acute WQC derived using these datasets should be relatively representative. In contrast, the chronic datasets are typically an order of magnitude or more less rich than natural aquatic communities and may have more significant biases in terms of composition than the acute datasets. The implications of these biases are difficult to assess given the limited understanding of trends in species sensitivity to chronic metal exposure. Additional chronic toxicity testing with a greater diversity of species is clearly needed.

Finally, we suggest that for the large acute metal datasets (Cd, Cu, Zn), the high level of richness and diversity relative to natural aquatic communities and the strong understanding (both descriptive and mechanistic) of relative sensitivity of different taxonomic groups should allow for environmentally protective adjustment of these datasets to better reflect site- or region-specific aquatic community composition.

REFERENCES

Abell RA, Olson DM, Dinerstein E, *et al.* 2000. Freshwater Ecoregions of North America: A Conservation Assessment. Island Press, Washington, DC, USA

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

Black RW and Silkey M. 1997. Water Quality Assessment of the Puget Sound Basin, Washington, Summary of Stream Biological Data Through 1995. US Geological Survey

Brix KV, DeForest DK, and Adams WJ. 2001. Assessing acute and chronic copper risks to freshwater aquatic life using species sensitivity distributions for different taxonomic groups. Environ Toxicol Chem 20(8):1846–56

CDM. 2000. Arid west water quality research project: draft framework for task 2 data integration, Camp Dresser & McKee, Cambridge, MA, USA

Conover WJ. 1999. Practical Nonparametric Statistics. John Wiley and Sons, New York, NY, USA

Forbes VE and Calow P. 2002. Species sensitivity distributions revisited: a critical appraisal. Hum Ecol Risk Assess 8(3):473–92

Griffiths RW, Thornley S, and Edsall TA. 1991. Limnological aspects of the St. Clair River. Hydrobiol 219:97–123

Grosell M, Nielsen C, and Bianchini A. 2002. Sodium turnover rate determines sensitivity to acute copper and silver exposure in freshwater animals. Comp Biochem Physiol 133C(1–2): 287–303

Layne JN. 1979. Natural Features of the Lake Annie Tract, Highlands County, Florida. Archbold Biological Station, Lake Placid, New York, NY, USA

Merritt RW and Cummins KW. 1996. An Introduction to Aquatic Insects of North America. Kendall-Hunt Publishing, Dubuque, IA, USA

OECD (Organization for Economic Cooperation and Development). 1992. Report of the OECD Workshop on the Extrapolation of Laboratory Aquatic Toxicity Data to the Real Environment. Paris, France

ORNL (Oak Ridge National Laboratory). 1999. Foothills Parkway Section 8B Environmental Report. National Park Service, Denver, CO, USA

Seegert G, Fava JA, and Cumbie PM. 1985. How representative are the datasets used to derive national water quality criteria? In: Caldwell RD, Purdy R., and Bahner HC (eds), Aquatic Toxicology and Hazard Assessment: Seventh Symposium, pp 527–37. American Society for Testing and Materials, Philadelphia, PA, USA

SPSS. 1999. SPSS Software. SPSS, Inc, Chicago, IL, USA

Stephan CE, Mount DI, Hansen DJ, *et al.* 1985. Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Organisms and Their Uses. US Environmental Protection Agency, Environmental Research Laboratory, Duluth, MN, USA

USEPA (US Environmental Protection Agency). 1994. Interim Guidance on Determination and Use of Water-Effect Ratios for Metals. US Environmental Protection Agency, Office of Water, Washington, DC, USA

USEPA (US Environmental Protection Agency). 1996. 1995 Updates: Water Quality Criteria Documents for the Protection of Aquatic Life in Ambient Water. US Environmental Protection Agency, Office of Water, Washington, DC, USA