



# EPA Public Access

Author manuscript

*Environ Toxicol Chem.* Author manuscript; available in PMC 2021 January 01.

About author manuscripts

Submit a manuscript

Published in final edited form as:

*Environ Toxicol Chem.* 2020 January ; 39(1): 118–130. doi:10.1002/etc.4559.

## Best Practices for Derivation and Application of Thresholds for Metals Using Bioavailability-Based Approaches

Eric Van Genderen<sup>a,\*</sup>, Jenny L. Stauber<sup>b</sup>, Charles Delos<sup>c</sup>, Diana Eignor<sup>d</sup>, Robert W. Gensemer<sup>e</sup>, James McGeer<sup>f</sup>, Graham Merrington<sup>g</sup>, Paul Whitehouse<sup>h</sup>

<sup>a</sup>International Zinc Association, Durham, North Carolina, USA

<sup>b</sup>CSIRO Land and Water, Lucas Heights, Australia

<sup>c</sup>GLEC, Traverse City, Michigan, USA

<sup>d</sup>US Environmental Protection Agency, Washington, DC

<sup>e</sup>GEI Consultants, Fort Collins, Colorado, USA

<sup>f</sup>Wilfred Laurier University, Waterloo, Ontario, Canada

<sup>g</sup>wca environment, Faringdon, Oxfordshire, UK

<sup>h</sup>Environment Agency, Howbery Park, Wallingford, UK

### Abstract

The primary goal of the present study is to provide a broad view of best practices for evaluating bioavailability models for metals for use in the protection of aquatic life. We describe the state of the science regarding 1) the evaluation and selection of ecotoxicity data, 2) the selection of bioavailability models for use in normalization, and 3) subsequent application of bioavailability models. Although many examples of normalization steps exist worldwide, a scheme is proposed to evaluate and select a model that takes account of its representativeness (water chemistry and taxonomic coverage of the ecotoxicity data set) and validation performance. important considerations for a suitable model are the quantity of inputs needed, accuracy, and ease of use, all of which are needed to set protective values for aquatic life and to use these values to evaluate potential risks to organisms in receiving waters. Although the end results of different model application approaches may be broadly similar, the differences in these application frameworks ultimately come down to a series of trade-offs between who needs to collect the data and use the bioavailability model, the different requirements of spatial scales involved (e.g., regional vs site-specific values), and model predictiveness and protectiveness. Ultimately, understanding the limits and consequences of these trade-offs allows for selection of the most appropriate model and application framework to best provide the intended levels of aquatic life protection.

\*Address correspondence to [evangenderen@zinc.org](mailto:evangenderen@zinc.org).

*Data Availability Statement*—Data and associated metadata and calculation tools are available as Supplemental Data files or from the corresponding author ([evangenderen@zinc.org](mailto:evangenderen@zinc.org)).

This article contains online-only Supplemental Data.

*Supplemental Data*—The Supplemental Data are available on the Wiley Online Library at DOI: 10.1002/etc.4559.

## Keywords

Water quality guidelines; Metals; Bioavailability; Chemical regulation

---

## INTRODUCTION

Many jurisdictions have recognized the need to include consideration of bioavailability in the derivation of protective values for aquatic life (PVALs) for metals. (When describing established regulatory precedents, the relevant term is used, for example, “water quality criteria.” When referring to recommendations in terms of future regulatory use of bioavailability models, we use broad terminology regarding PVALs, which is synonymous with guideline/water quality criteria/environmental quality standard [EQS]/benchmark as used in various jurisdictions.) In the early 1980s, the United States demonstrated a progressive movement away from single-value criteria by recognizing the importance of hardness as a toxicity-modifying factor (TMF) for metals (Adams et al. 2020). However, it was not until nearly 30 yr later that a more comprehensive approach was incorporated into the derivation process using the biotic ligand model (BLM; US Environmental Protection Agency 2007). Europe demonstrated the biggest departure from single-value standards in the early 2000s by adopting full BLM corrections while performing regional risk assessments for some metals within the framework of the European Union Existing Chemicals Regulation and implementing European Union-wide bioavailability-based standards in 2010 for nickel under the Water Framework Directive. Australia/New Zealand and Canada began updating their national guidelines for metals, including incorporation of other TMFs beyond just hardness, after 2010.

Although there is common ground in the way different jurisdictions derive PVALs, there can be marked differences in the way they are subsequently used to assess risk or to set permits. Two distinct regulatory options for the application of bioavailability-based approaches can be recognized using either site-specific or regional default values.

First, in some jurisdictions, especially those where regulatory duties are delegated to local administrations, there is a tendency toward using local or even site-specific PVALs. Essentially, the influence of water chemistry is dealt with in the effects assessment (i.e., derivation or adjustment up or down of the value itself). In the United States, “site-specific” values can be developed for a specific water body or segment receiving more than one permitted discharge or developed as a single discharge-specific criterion based on water chemistry characteristics collected at the regulatory point of compliance (i.e., traditional application for hardness-based values; Gensemer et al. 2016).

Second, in other jurisdictions (e.g., many European countries), where some PVALs for metals (e.g., Ni, Pb, Cd) are developed and promulgated centrally across the whole of Europe, there is a single regional default standard that is protective of most waters. Under the Water Framework Directive which drives regulatory standard development in Europe, the European Commission publishes a single value for each standard (EQS; European Commission 2018) for some metals. For metals that are not considered European Union-wide challenges, countries can set their own specific limits using the same methods (e.g., for

Cu, Zn). In contrast to the first approach (i.e., site-specific PVALs), the approach under the Water Framework Directive is to account for the influence of water chemistry through the exposure (or monitoring data) side of the risk assessment. That is, when assessing risk, the monitoring data must be adjusted to reflect the bioavailability that would occur at the site under consideration, and this is compared to the  $EQS_{\text{bioavailable}}$  to determine potential risk (i.e., “pass” or “fail”).

The science of metal bioavailability has advanced to the point where a number of bioavailability models are available with varying levels of complexity, sophistication, and ease of use. In addition, the incorporation of TMFs into model development has placed emphasis on nuances between the protectiveness (i.e., levels of conservatism, or the extent to which conclusions err on the side of protecting the environment) versus predictiveness (i.e., the extent to which toxicity is accurately predicted) of bioavailability models and, hence, the PVAL developed based on those models. However, the basic process (i.e., framework) for use of bioavailability models has remained relatively consistent, even as the science of bioavailability has progressed.

The present study describes the state of the science regarding 1) the evaluation and selection of ecotoxicity data with particular emphasis on measured metal concentrations and water chemistry parameters, 2) the selection of bioavailability models for use in normalization, and 3) considerations for model application based on data availability, spatial scale, and model predictiveness and protectiveness. We build on existing frameworks to incorporate bioavailability models that enable users to normalize ecotoxicity data for the purpose of derivation and application of PVALs. Although the primary focus of the present study is on the framework in the United States (Stephen et al. 1985), best practices internationally from other jurisdictions are also included to provide the end user with a broad view of the options for regulatory application of PVALs for aquatic life for metals that are derived using bioavailability models.

## CONCEPTUAL FRAMEWORK FOR USE OF BIOAVAILABILITY MODELS

Figure 1 illustrates a conceptual framework with respect to how bioavailability models could be used in both the derivation and application of acute and chronic PVALs for metals. This framework is based on the derivation process set forth in the US Environmental Protection Agency’s (USEPA’s) guidelines (Stephen et al. 1985). However, the proliferation of other, broadly similar frameworks around the world suggests that there are many aspects in common, so this framework has value beyond the United States.

This framework assumes that all bioavailability models adjust aquatic toxicity on the basis of external water chemistry characteristics known to control metal bioavailability (Adams et al. 2020). The regulatory use of these models occurs in 2 steps: 1) normalization, and 2) application (Figure 1). “Normalization” is defined as the use of the bioavailability model to predict and adjust toxicity values used to populate the species (genus) sensitivity distributions (SSDs) from which the 5% hazardous concentration (HC5; 5th percentile of the SSD) is calculated. This HC5 forms the technical basis of the regulatory PVAL. The normalization process is an important component of PVAL development, to 1) reduce

intraspecies variability when multiple studies exist that represent a spectrum of water quality conditions, 2) order/rank species sensitivity in the SSD relative to differences in bioavailability among species, and 3) adjust the magnitude of the HC5 relative to the index condition chosen (e.g., user-defined protection goals, expressed as a combination of water chemistry parameters that determine bioavailability). Application is the way in which PVALs are modified relative to water chemistry conditions specified for regional (Europe), national, or site-specific interests.

### Normalization framework

Guidance from the USEPA describes the steps for incorporating relationships between toxicity and one or more TMFs (Figure 2). Corresponding guidance for other jurisdictions is found elsewhere (e.g., European Commission 2011; Organisation for Economic Co-operation and Development 2016). If analysis does not indicate that TMFs are important for a metal, then individual toxicity data that have passed quality assurance standards can be aggregated without normalization prior to SSD construction and HC5 derivation for a final PVAL. If analysis instead suggests that TMFs are important, then all appropriate and validated bioavailability models are considered for selection and use for normalization. If a bioavailability model has not been developed but one or more TMFs are known or suspected, the ecotoxicity data may be used to investigate modeling options (Figure 2; Brix et al. 2020; Mebane et al. 2020; Garman et al. 2020).

In addition, the procedures for evaluating acceptable ecotoxicity data must include an additional check for how well a model represents 1) the water chemistry conditions it is intended to represent, and 2) the types of taxa used in studies to populate the SSD. A selected bioavailability model(s) can then be used to normalize individual toxicity values for which sufficient water chemistry data are available or can be estimated. Because of inevitable deficiencies in water chemistry reporting needed to populate more sophisticated models (see section Evaluation of Acceptable Ecotoxicity Data for Use in a Normalized SSD) or a model's representation of species in the SSD (see section Model Evaluation and Selection), the process of data set evaluation and model selection should include regular reassessments as information regarding the importance of water chemistry on toxicity evolves.

### Application framework

In the United States, a model-based PVAL is typically expressed not as a numerical value but as the model used to modify the HC5 for a given site water, leading to application of the suitably modified PVAL in permitting or assessment as implemented at the state level (Figure 1). In other jurisdictions (European Commission 2018), the regulatory PVAL (e.g.,  $EQS_{\text{bioavailable}}$ ) may instead be expressed as a threshold metal concentration relative to a specified index condition, rather than the model. Whereas the end results can be similar, different spatial scales and data needs influence how the models are ultimately applied (discussed further in section Model Evaluation and Selection).

## EVALUATION OF ACCEPTABLE ECOTOXICITY DATA FOR USE IN A NORMALIZED SSD

Prior to any normalization, ecotoxicity data from single-species laboratory toxicity tests need to be assessed for data reliability and relevance for use in the SSD. In this respect, metal toxicity data are no different from data for any other substances, but some metal-specific issues are highlighted in the present section. In the United States, acceptable data-quality indicators (or selection criteria) identified by Stephen et al. (1985) have been used for this purpose. Since then, several new approaches have been developed, including the Klimisch system (Klimisch et al. 1997) and the criteria for reporting and evaluating ecotoxicity data (Moermond et al. 2016). Studies are evaluated on both reliability (i.e., the inherent quality of the data resulting from the method used to conduct the test) and relevance (i.e., the extent to which the test provides useful information about the hazardous properties of a chemical), with no distinction between metal and nonmetal toxicants. Metal-specific guidance is provided by the Organisation for Economic Co-operation and Development (2016). In Australia and New Zealand, different defined lists of acceptable quality assurance are applied to different types of data depending on the environmental medium (freshwater or marine/estuarine), type of toxicant (metal or nonmetal), and type of test organism (plant or nonplant; Warne et al. 2015; Gissi et al. 2016). Additional acceptable quality assurance measures apply when mesocosm or field data are used.

Field-based and micro-/mesocosm studies may provide a more direct measure of endpoints that regulators strive to protect. They may also provide a wider range of toxicity endpoints, including nontraditional endpoints for single species within the community and population/community endpoints, thereby potentially improving the environmental relevance of PVAL derivation and application for metals. There are also challenges associated with the use of micro-/mesocosm studies, including controlling test conditions and resource constraints on experimental replication, because of the greater complexity of the systems relative to single-species laboratory toxicity tests. Australia and New Zealand have long permitted the use of micro-/mesocosm data and field data, in combination with laboratory data, in SSDs to derive PVALs (Australian and New Zealand Environment and Conservation Council and Agriculture and Resource Management Council of Australia and New Zealand 2000). In Europe, Australia, New Zealand, and the United States, these studies can be used as additional lines of supporting evidence in weight-of-evidence assessments of laboratory-derived PVALs (European Commission 2011, 2018; Warne 2001), provided that the resulting data meet certain quality conditions. Data acceptability requirements specific to micro-/mesocosm data and field data are summarized in Supplemental Data, Table S1 (modified from Warne et al. 2015).

### **Metal analysis, speciation, and solubility**

Considerations regarding the chemical form of metal used and measured in ecotoxicity tests that should be recognized when determining overall data acceptability are outlined in the present section, with more detailed guidance provided in the Supplemental Data.

**Metal analysis**—Use of nominal metal concentrations in ecotoxicity tests is no longer recommended because of loss of metals to test containers and to test organisms over the duration of the test (Franklin et al. 2002). Organism loading, in particular, can change the metal speciation in solution in static tests, and hence alter metal bioavailability and toxicity. For PVAL development purposes, most jurisdictions recognize the importance of using only measured metal concentrations; however, exceptions occur for data-poor species or when there is evidence that all metal is in solution.

In the United States, PVALs for metals were originally based on total recoverable metal, with a hardness correction. In 1993, it was recognized that dissolved metal is generally a better predictor of effects, with the exception of aluminum (Prothro 1993), although neither total nor dissolved metal measurements take into account the dietary route of exposure. Some exceptions, where total metal rather than dissolved metal is used for derivation of PVALs, include aluminum (US Environmental Protection Agency 2018) and iron (Australian and New Zealand Governments 2018a). Because most jurisdictions previously allowed total metal measurements in ecotoxicity tests, approaches were needed to convert total metal to dissolved metal to enable the use of these older data in derivation or application. These approaches have included 1) estimating solubility products to calculate metal speciation, and 2) undertaking experiments to determine the fraction of total metal present as dissolved metal. However, as shown in the Supplemental Data, conversions between total and dissolved metal concentrations are of limited reliability and will introduce a level of uncertainty into derivation or any regulatory applications based on these types of data. We recommend that only measured metal concentrations be used in the derivation of PVALs, with preference for dissolved metal concentrations over total metal concentrations.

**Metal speciation and equilibration**—Metals in the dissolved phase can be in a variety of chemical forms (species), which depend on site-specific water chemistry. Therefore, the use of dissolved metal concentrations does not distinguish between the toxic effects of the different dissolved metal species. This would require additional knowledge of metal speciation, potency of different species, and availability of analytical/ computational techniques and would not account for other geochemical interactions among major ions and organic fractions. It is also now well established that the toxicity of metals to different taxa is not always attributable to the same metal species (e.g., bioavailability of  $\text{CuOH}^+$  and  $\text{CuCO}_3$ ; US Environmental Protection Agency 2007).

Bioavailability models such as BLMs assume pseudoequilibrium between bulk water species and the biotic ligand. At the very low metal concentrations in chronic studies (and in typically occurring natural waters), kinetic control may be more important (i.e., the rate-limiting step may be slow diffusion of the metal from solution, rather than uptake into the organism; US Environmental Protection Agency 2007). However, normalization by extension also assumes equilibrium has been achieved, but test solutions may have higher bioavailability than the equilibrated dilution. Preequilibration of the metal in the test medium for 24 h before adding the organisms is one way to reduce the problems of nonequilibrium (Ma et al. 1999; Organisation for Economic Co-operation and Development 2016).

**Metal solubility**—Regulatory documents usually specify what metal salts are acceptable for ecotoxicity data used to derive PVALs (e.g., US Environmental Protection Agency 1994). Although ecotoxicity data at metal concentrations above solubility limits were previously used, more recent guidelines, for example, in Australian and New Zealand and in Europe, reject concentrations more than twice the metal solubility limit (Warne et al. 2015; European Commission 2018). This is because of concerns over unknown effects of colloids and precipitates, potentially resulting in over- or underestimation of toxicity when solubility limits are exceeded. For example, using data from tests where the metal solubility limit is exceeded can distort the upper end of the SSD, leading to unreliable HC5 estimates. This may not be the appropriate approach, however, in cases where colloids or precipitated metal have toxicological significance such as for aluminum and should be addressed (Gensemer et al. 2018; US Environmental Protection Agency 2018). More information on metal solubility is provided in the Supplemental Data.

### Water chemistry parameters

The process of normalizing toxicity test results or applying a model for derivation purposes requires that each study has reported water chemistry data suitable for applying the chosen bioavailability model. The reported water chemistry data should also be critically evaluated to determine uniformity with model requirements. Results from such an evaluation typically demonstrate that many studies/species (normalization) or monitoring data (application) may be rejected because of lack of required data inputs for modeling (e.g., dissolved organic carbon [DOC] data are typically lacking). For example, approximately 30% of otherwise acceptable studies were rejected as part of the data-screening process for developing the BLM-based criterion for copper (16 fewer genera and 18 fewer species than the previous criterion; US Environmental Protection Agency 1995, 2007).

As a result, approaches for estimating missing water chemistry parameters have been developed to address deficiencies encountered in both the normalization and application procedures. For the purpose of filling missing water chemistry data in the ecotoxicity data set, methods have included 1) personal communication with authors for newer analyses of water sources, 2) calculation of major ion concentrations from recipes for standard reconstituted laboratory waters, and 3) empirical estimation of concentrations of major ions from one or more known parameters (calcium concentrations or conductivity) for tests using natural water (e.g., Peters et al. 2011; US Environmental Protection Agency 2016). Similarly, for the purpose of filling in missing data for an otherwise complete input parameter or monitoring data set, existing national water chemistry databases can be used to develop and refine distributions. For example, the USEPA compiled water chemistry data for thousands of sites using publicly available information from the US Geological Survey's National Water Information System and National Organic Carbon Database (US Environmental Protection Agency 2016). For some critical parameters, notably pH and DOC concentration, it has been recommended that these are preferably measured, rather than estimated (Peters et al. 2011; US Environmental Protection Agency 2016).

Differences between water chemistry ranges/boundaries for the ecotoxicity data set and chosen bioavailability model must also be evaluated for uniformity. Bioavailability models

are typically calibrated and validated to cover a certain range of water chemistry, for instance, the 10th to 90th percentile of the distribution of TMFs for a region. For the purposes of normalizing an ecotoxicity data set, reported TMF values for tests with water chemistry outside the model validation range should be used and interpreted with caution (Garman et al. 2020). Water chemistry data from monitoring programs falling outside of model boundaries may suggest the need for further model refinement to account for unique water chemistry conditions.

## MODEL EVALUATION AND SELECTION

Given the availability of a variety of bioavailability models (Table 1), a process is needed for selecting the most appropriate model(s) for normalization. Whereas model evaluation and selection should be primarily guided by how well the model reflects the mechanistic understanding of chemical speciation and toxicity, the types and choice of model can also depend on how they are intended to be applied. The most acceptable model (or models) for normalization and application will depend on several considerations.

### Representation

Evaluating the representativeness of a model's water chemistry and taxonomic coverage of the ecotoxicity data set to which normalization would be applied. This should also include an evaluation of whether water chemistry input parameters are within model derivation or validation boundaries.

### Level of input

Accounting for the water chemistry data input requirements.

### Accuracy

The ability of the model to predict toxicity, as documented by the model calibration and validation studies.

### Ease of use

A critical consideration in terms of effort and resource needs, required levels of expertise in operation and interpretation of outputs, software compatibility with current systems, and dovetailing of model outputs to the regulatory use.

### Normalization procedure

The operational procedures for normalizing acceptable ecotoxicity data sets have evolved since the early 1980s as a result of 1) an improved understanding of the influences of water chemistry parameters on the aquatic toxicity of metals, and 2) the amount of additional research published for new and existing species that could be considered in the standard development process. For example, a compilation of normalization procedure summaries, used in Australia and New Zealand, Canada, Europe, and the United States over the last 40 yr, illustrates a range of priorities related to model ease of use and taxonomic representation in the SSD (Table 1).



A central challenge to the use of complex bioavailability models (e.g., BLMs and multiple linear regressions [MLRs]) is that direct application to all species in the SSD may not be possible because 1) the model was either developed using, or validated for, only a small number of widely tested species (e.g., algae, daphnids, rainbow trout) not representative of the taxa in the SSD; or 2) several studies have missing water chemistry parameters needed to run bioavailability models. Therefore, 2 approaches have been developed for normalizing a full SSD that recognize potential differences in water chemistry influences for some species and/or lack of necessary water chemistry measurements.

First, Europe (European Commission 2011, 2018) prescribes the use of trophic-level models in the normalization process. It is assumed that all species data within the SSD would best be normalized using the model with the most specificity for its taxonomic group. However, for some species that have considerable divergence from the species for which a model was developed (e.g., amphibians or higher plants), all available models may be evaluated to determine if one better describes influences on toxicity than another.

Second, hybrid normalization using both a hardness-based equation and a BLM has been used to account for the lack of necessary water chemistry measurements (DeForest and Van Genderen 2012). The intent was to retain taxonomic diversity in the SSD (to satisfy minimum data requirements) because all necessary water chemistry parameters were unavailable for some (older) data-poor species. Further, DeForest and Van Genderen (2012) focused their attention on characterizing water chemistry parameters from species in the lower 20th or 50th percentile of the acute or chronic species mean distributions, respectively, because the HC5 calculation emphasizes the lowest 4 genus mean values (Stephen et al. 1985). In concept, a hybrid normalization could be constructed using any combination of models, assuming that the procedure still conforms to the jurisdictional requirements for data selection.

**Choosing an index condition**—A final step in the normalization process that can be taken is calculating an HC5 for the ecotoxicity data set using a specified combination of water chemistry parameters (referred to as the “index condition” hereafter). The index condition is used as a simple point of reference in some jurisdictions and as a regulatory benchmark in others. In the United States, the criterion is represented by the model developed and used for normalization and subsequent calculation of the HC5. An index condition is only reported in US criteria documents as a point of reference relative to previous criteria under a common water chemistry condition. In other jurisdictions, the index condition chosen may depend on the intended management goals and level of conservatism.

As a result, the specified index condition may not always represent a known site condition, the exception being the European Union which uses a natural water composition for the index condition. To this end, the most scientifically defensible approach for defining an index condition should involve evaluating distributions of measured water chemistry conditions throughout the region of interest to help define management goals. Such an approach allows water quality managers to identify a complete water chemistry condition that is representative of natural composition and water chemistry parameter relationships for setting an index condition.

**Model selection**—The choice of models has historically been based on user decisions related to levels of conservatism and ease of use, rather than on model performance. However, although there are some correlations between model performance (i.e., accuracy as reflected in reduced uncertainty and a higher degree of confidence in predictions) and level of input or ease of use, there are also exceptions. For example, traditional single-parameter models for metals are easy to use but demonstrate marginal performance when applied (providing the motivation to move away from a hardness-based model for copper). Conversely, some simplified (user-friendly) models reduce input and computational output from full BLM simulations and streamline the user interface while preserving modeling performance (Table 2). These tools are aimed at reducing resource demands while ensuring that readily interpretable outputs are consistent with full BLM predictions (e.g., Environment Agency 2009, 2010, 2012a, 2012b, 2013; Rüdél et al. 2015). Similarly, although MLRs and BLMs have variable data needs and levels of complexity for the user, both provide increased confidence in predictions over simple hardness-based models (Table 2).

Typically, MLRs require fewer water chemistry input data than BLMs, which require relatively large data inputs and technical understanding, making them more complex and time-consuming to use in a regulatory context. Although common concerns associated with application of bioavailability-based PVALs have been to model complexity and data needs (e.g., up to 16 water chemistry parameters required) for routine use by regulators and stakeholders, the models are considered robust, scientifically advanced, and accurate (e.g., Schlekát et al. 2010). Despite attempts to simplify BLM input requirements, outputs often require some interpretation before they can be applied in regulatory frameworks.

There could be a perceived trade-off between reduced input requirements, model complexity, mechanistic basis, and potential loss of accuracy of predictions. However, recent research indicates that the complex (BLM) and more simplified (MLR and user-friendly) models can produce very similar results across a range of water chemistry conditions, providing results within the standard plus or minus a factor of 2 of observed results (Brix et al. 2017). The relative importance of this apparent trade-off may be dependent on the regulatory framework in use, with higher levels of complexity and performance needed in some instances to support decisions. Furthermore, the acceptance of potential loss of accuracy can vary depending on the application. If a model is too complex for its intended user community, it is unlikely to be used. Simplified tools are only simple in terms of the user interface and input requirements when compared to BLMs yet retain complexity with respect to bioavailability calculations. That said, although simplified tools account for factors that impact bioavailability, the technical complexity and extent of TMFs that are taken into account are also reduced in relation to the full BLMs. For example, although MLRs and user-friendly BLMs may have reduced inputs and complexity relative to full BLMs, all require the same level of interpretation in the application process (e.g., spatial and temporal coverage of the PVAL).

In Europe, the applicability and validity ranges of the simplified tools are defined by the BLMs on which they are based and that they mimic. Peters et al. (2016) reviewed the scientific underpinning and performance of 2 simplified tools (Bio-met and PNEC-pro), with

respect to comparison to the Ni-BLM, and offered recommendations for interpreting the results of such analyses.

**Case study**—The following bioavailability model evaluation and selection procedure is presented as an example of how a user might objectively review available models for use in the normalization process. All data are provided in the Supplemental Data. This example is meant to be illustrative and admittedly uses a metal (zinc) that is data-rich and has different modeling formats for consideration. The procedure only applies to the model selection process and does not attempt to derive a zinc PVAL from the findings.

The data set selected represents only chronic ecotoxicity studies passing test acceptability in Europe for algae, invertebrates, and fish. The compilation of studies is represented in the user-friendly BIOMET tool (available at <http://bio-met.net/>). The models chosen for evaluation are represented by 1) USEPA hardness-based species equations (US Environmental Protection Agency 1987); 2) Canadian Council of Ministers of the Environment Zn-MLR equations for algae (MLR1), *Daphnia magna* (MLR2), and rainbow trout (MLR3; Canadian Council of Ministers of the Environment 2018); and 3) BLMs for algae (Van Regenmortel et al. 2017) or aquatic invertebrates and fish (DeForest and Van Genderen 2012).

For each model, the evaluation process involved 7 steps. 1) Scoring based on reported metal fractions (e.g., 2 = dissolved, 1 = total, 0 = nominal). This is not used here because of the ecotoxicity data set having measured dissolved metal concentrations reported for all tests. Scoring of endpoint or exposure duration relevance could also be considered. 2) Scoring the representativeness of water quality coverage. Relative to the range of water quality used in model calibration, a score of 1 was assigned for ecotoxicity tests with water quality reported within the model boundaries and 0 for water quality outside the model boundary range. 3) Scoring the representativeness of taxonomic coverage. Relative to the species represented by the model calibration and validation, each test was assigned a scaled score based on a within-biological organization comparison: 0 (outside kingdom), 1 (kingdom), 2, (phylum), 3 (class), 4 (order), 5 (family), 6 (genus), 7 (species). 4) Water quality and taxonomic coverage scores were summed for each test and qualified as “good” (score = 6–8), “fair” (score = 3–5), or “poor” (score = 0–2). These scores were then summed for each trophic level and summed overall. 5) Scores were developed, using residual factors (RFs), to characterize the performance of each model at the species level, using a validation data set (DeForest and Van Genderen 2012). That is, each model was validated against all currently available ecotoxicity data for species represented by the model. Although the species and number of predictions differ among models, this procedure optimizes the performance evaluation for each model. Residual factors were calculated for each test as the maximum divided by the minimum value of predicted or observed toxicity, then plotted as cumulative distributions for each model and evaluated by calculating the  $RF_{x,Factor}$  which is the percentage of predictions associated with a chosen RF. Factor = 2.0 in the example focuses on all predicted toxicity falling within a factor of 2 of the observed toxicity. Higher factors indicate poorer agreement between predicted and observed toxicity. Or the  $1/RF_{y,\%}$  or the reciprocal RF associated with a chosen prediction accuracy. Percentile = 0.84 in example as equivalent to 1 standard deviation. The higher the reciprocal score, the more accurate the

prediction. Note: Untransformed residuals (predicted toxicity divided by observed toxicity) should also be plotted and assessed to qualitatively identify tendencies for under- or overprediction of a model with the chosen data set. 6) The number of “good” scores for each trophic level and the overall scores were multiplied by  $RF_x$  or  $1/RF_y$  to obtain relative scores. 7) Finally, relative scores were ranked (rank 1 is the best) by trophic level and overall to determine if qualitative conclusions could be drawn concerning model representativeness for the ecotoxicity data set.

A summary of results from the evaluation process is shown in Table 3. The water quality evaluation did not demonstrate much differentiation among models, for all trophic levels (step 2). As a result, taxonomic coverage of a model (step 3), relative to water quality in the ecotoxicity data set, produced the greatest representative scoring differences (step 4; Table 3). The BLMs consistently scored high for all trophic levels and overall. However, the hardness model contained slightly more diversity across fish species, and the BLMs contained slightly more diversity across invertebrates. Because the MLRs were developed at the species level, overall scores are only representative of a single trophic level. In cases where reported water quality for toxicity studies was absent, missing parameter approaches could be applied (step 1 or 2). Similarly, model refinement and/or extended taxonomic coverage could be considered to improve model performance (e.g., generate pooled MLR from examples provided here).

Model performance following validation with all currently available ecotoxicity data for respective species produced a clearer comparison among models. That is, each model was used to make predictions for all available ecotoxicity data represented by the model, including data used for model calibration. Prediction accuracy at an RF of 2 ( $RF_{x,2,0}$ ) ranged from 19% for hardness, 33 to 70% for MLRs, and 87% for BLMs; and the RF associated with 84% prediction accuracy ( $1/RF_{y,0.84}$ ) ranged from 0.0078 for hardness, 0.18 to 0.41 for MLRs, and 0.53 for BLM (Figure 3). Because of the large differences between performance metrics among models, relative scores for the BLM were 1.2-fold to 93-fold higher than for other models. A final ranking of models illustrated that the BLM best represented the ecotoxicity data set and demonstrated prediction accuracy for all trophic levels and overall. Other models each have strong representation for certain trophic groups and could be limited in their application.

## REGULATORY APPLICATION OF BIOAVAILABILITY-BASED PVALS

This section focuses on the regulatory application of bioavailability-based PVALs to 1) assess receiving water quality (to help identify where management action is needed) and 2) set discharge permits on known point sources (e.g., municipal wastewater or industrial discharges). Table 1 summarized those regulatory jurisdictions which are known to have applied bioavailability-based models for regulatory purposes. The differences in these application frameworks ultimately result in a series of trade-offs between who needs to collect the data and run the bioavailability model, the different requirements of spatial scales involved, and model predictiveness and protectiveness (Figure 4). For example, selecting a regional default PVAL would require relatively little effort by the permit holder but

relatively greater effort by the regulatory agency. The opposite would be true for an individual location-specific permit.

Although many issues of application are common to other hazardous chemicals (like choice of sampling locations, sampling frequency, confidence of compliance), some issues are specific to bioavailable metals or assume a particular importance. These include 1) spatial and temporal variability in the TMFs that affect bioavailability (and hence toxicity), 2) resolving missing water chemistry data (see section Evaluation of Acceptable Ecotoxicity Data for Use in a Normalized SSD), and 3) applying tiered approaches, including screening tools, to facilitate the assessment process when there are large numbers of sites or samples.

### **Spatial and temporal variability in water chemistry**

Surface water chemistry conditions vary on both a spatial and a temporal basis in response to physical, chemical, and biological processes. The relative importance of different TMFs depends on the metal, but for most metals, DOC is the most critical TMF. However, to establish PVALs for metals, water chemistry conditions at a site should be well characterized, especially the conditions that influence metal bioavailability. This characterization often involves a robust sampling program, which should be designed to avoid bias resulting from temporal and spatial variability in water chemistry (e.g., seasonal effects on pH or spatial variation in DOC).

Temporal changes in metal bioavailability can be influenced by 2 factors. First, seasonal influences on water composition (DOC, Ca, Mg, and pH) may fluctuate in their magnitude (relative to average) or relationship to each other over the course of a year. Consequently, the bioavailability of a metal may be highly variable throughout the year. Second, worst-case conditions where the metal is most bioavailable are normally assumed to occur under low-flow conditions, largely because effluent flows (and concomitant pollutant loads) will be largest relative to receiving water flows. However, for some TMFs, such as DOC, low-flow events dominated by effluents would predictably include higher organic contributions (and, thus, lower metal bioavailability) than during high-flow events. Furthermore, because different TMFs may exhibit different flow-dependency patterns (e.g., DOC vs pH or hardness), most conservative or “critical” conditions may be more challenging to identify a priori.

By anticipating variability in water chemistry from established monitoring programs, the frequency of sampling may be refined to account for temporal changes. For example, if a site experiences seasonal variability and flow conditions, a sampling program that relies on more frequent sampling dispersed over the year may be indicated. However, if initial sampling shows a site has consistent water quality conditions across seasons and flow conditions, the sampling frequency may be relaxed without loss of accuracy. Until anticipated trends in water chemistry can be characterized, a minimum of monthly samples taken over a 1- to 2-yr period would help to establish variability patterns to refine future sampling needs (Gondek et al. 2018). Variability in concentrations of the metal itself can also influence probabilities of exceeding bioavailability-based criteria calculations, particularly when using a probabilistic method (Ryan et al. 2018).

Spatial considerations may also affect metal bioavailability in 2 ways. First, spatial variability in water chemistry within catchments can be common, and the extent of that variability can be analyzed from historic data and can help determine the appropriate spatial scale for regulatory monitoring. Ideally, water chemistry and metal concentration data for a given sampling location should be collected at the same time, in the same sample. Second, water chemistry, and hence metal bioavailability, can change as effluent mixes with receiving water downstream through a catchment. This is of particular relevance to the larger waterways where the geochemical characteristics may be modified locally by tributaries or multiple inputs of municipal, agricultural, or industrial wastewaters. For the purposes of application for permitting processes, both upstream and downstream sites from the discharge should be sampled.

### Tiered approaches for model application

Tiered assessment schemes can be useful when large numbers of sites require assessment and it would be helpful to develop assessment priorities. This is useful to direct resources to sites where there is a potential risk and where further analysis may be warranted. Furthermore, tiered regulatory application frameworks may be helpful in addressing the lack of consistent availability of TMF data required for more complex bioavailability models such as MLRs or BLMs.

Tiered assessment schemes are particularly useful when a single PVAL is used that is normalized to a water with characteristics that represent a conservative condition of high bioavailability (such as those used in Europe for nickel and lead) and arguably only justifiable when its conservatism has been demonstrated. The application of such a PVAL within a tiered approach is consistent with classic risk-assessment paradigms in that early tiers of assessment are most conservative but simple to perform with large numbers of sites (because data requirements are low). The intention is to screen out low-risk sites during early tiers of assessment. As progress is made through the assessment tiers, the data and calculation requirements increase; but this effort is restricted to sites where metals potentially pose the greatest risk.

An example of a tiered approach to assessment (Merrington et al. 2016), which is briefly described, represents the current state of practice in the United Kingdom. It suggests a logical process that might have value for other jurisdictions (e.g., Oregon's tiered criteria derivation approach for the copper BLM; Oregon Department of Environmental Quality 2017).

**Tier 1**—Direct comparison of the concentration from monitoring data (dissolved metal) with the standard. The standard used for comparison with monitoring data must be demonstrated as conservative, to ensure that high-bioavailability sites are not erroneously screened out at this step. Sites, or samples, exceeding the standard at this tier progress to the second tier of the assessment.

**Tier 2**—Account for bioavailability using site-specific water chemistry data. Ideally, this tier of the assessment could make use of an MLR, a full BLM, or a simplified tool for calculating the local metal bioavailability.

**Tier 3**—Provides an opportunity for “local refinement” to consider local issues that might affect the assessment of risk attributable to metals. For example, local background concentrations of metals, which may have prompted resident organisms to develop a tolerance to elevated trace metal concentrations, could be investigated as a factor requiring model calibration or refinement. Continued failure (defined impairment) of a site to meet the standard at this tier would suggest candidacy for risk-reduction measures.

## CONCLUSIONS

A conceptual framework has been developed to assist users of bioavailability models for various risk-assessment applications needed to establish PVALs and support regulatory activities for metals. Once ecotoxicity data have been selected and evaluated, a scheme is proposed to evaluate and select the appropriate validated model(s) for use in deriving the PVAL. The model selection process should include evaluations for representativeness (water chemistry and taxonomic coverage) relative to the ecotoxicity data set and model performance. Hybrid normalization may be necessary to account for potential differences in water chemistry influences for some taxa and for lack of necessary water chemistry measurements for data-poor species. Final model selection should be based on data and model input needs, model predictiveness and protectiveness, and ease of use, both in deriving PVALs for metals and subsequently for when these values are applied to assess risks in receiving waters. Ultimately, trade-offs between model complexity, data needs, and ease of use and interpretation may need to be considered to ensure that the selected model represents the appropriate balance between model performance with respect to toxicity predictions and the levels of protection afforded by any given regulatory jurisdiction.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgment

The authors thank the SETAC technical workshop steering committee and all workshop participants. The authors acknowledge the financial support of the SETAC technical workshop, including the US Environmental Protection Agency, the Metals Environmental Research Associations, Dow Chemical Company, Newmont Mining Company, Rio Tinto, Umicore, and Windward Environmental. They also thank the SETAC staff for their support in organizing the workshop.

## REFERENCES

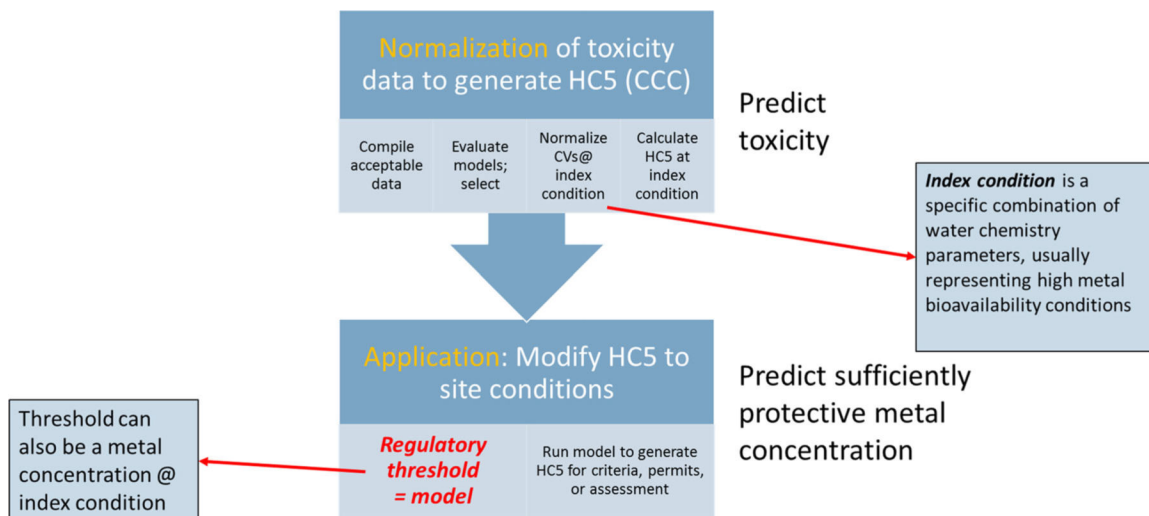
- Adams W, Blust R, Dwyer R, Mount D, Nordheim E, Rodriguez PH, Spry D. 2020 Bioavailability assessment of metals in freshwater environments: A historical review. *Environ Toxicol Chem* 39:48–59. [PubMed: 31880839]
- Australian and New Zealand Environment and Conservation Council and Agriculture and Resource Management Council of Australia and New Zealand. 2000 National water quality management strategy, paper no. 4. Australian and New Zealand guidelines for fresh and marine water quality Vol 1, The guidelines (chapters 1–7). Canberra, Australia.
- Australian and New Zealand Governments. 2018a Australian and New Zealand guidelines for fresh and marine water quality Default guidelines values for toxicants: Iron–freshwater. Canberra, Australia.

- Australian and New Zealand Governments. 2018b Australian and New Zealand Guidelines for Fresh and Marine Water Quality Default guidelines values for toxicants: Zinc–freshwater. Canberra, Australia.
- Brix KV, DeForest DK, Tear L, Grosell M, Adams WJ. 2017 Use of multiple linear regression models for setting water quality criteria for copper: A complementary approach to the biotic ligand model. *Environ Sci Technol* 51:5182–5192. [PubMed: 28409924]
- Brix KV, DeForest DK, Tear L, Peijnenburg W, Peters A, Traudt E, Erikson R. 2020 Development of empirical bioavailability models for metals. *Environ Toxicol Chem* 39:85–100. [PubMed: 31880833]
- Canadian Council of Ministers of the Environment. 2018 Scientific criteria document for the development of the Canadian water quality guidelines for the protection of aquatic life: Zinc. Winnipeg, MB, Canada.
- De Forest DK, Van Genderen E. 2012 Application of U.S. EPA guidelines in a bioavailability-based assessment of ambient water quality criteria for zinc in freshwater. *Environ Toxicol Chem* 31:1264–1272. [PubMed: 22447746]
- Environment Agency. 2009 Using biotic ligand models to help implement environmental quality standards for metals under the Water Framework Directive Science Report SC080021/SR7b. Bristol, UK.
- Environment Agency. 2010 Development and use of the manganese bioavailability assessment tool (Draft) Released by the United Kingdom Technical Advisory Group (WFD-UKTAG) 2012. Bristol, UK.
- Environment Agency. 2012a Proposed EQS for Water Framework Directive Annex VIII substances: Copper (saltwater) (For consultation) Released by the United Kingdom Technical Advisory Group (WFD-UKTAG) 2012. Bristol, UK.
- Environment Agency. 2012b Development and use of the copper bioavailability assessment tool (Draft) Released by the United Kingdom Technical Advisory Group (WFD-UKTAG). Bristol, UK.
- Environment Agency. 2013 Development and use of the zinc bioavailability assessment tool (Draft) Released by the United Kingdom Technical Advisory Group (WFD-UKTAG) 2012. Bristol, UK.
- European Commission. 2010a Nickel and its compounds (final revision Oct 12 2010) EQS sheet. Prepared by Danish Environmental Protection Agency, Copenhagen, Denmark.
- European Commission. 2010b Lead and its compounds (final revision Oct 12 2010) EQS sheet. Prepared by Environment Agency, Bristol, UK.
- European Commission. 2011 Common implementation strategy for the Water Framework Directive (2000/60/EC) Guidance document no. 27. Technical guidance for deriving environmental quality standards Brussels, Belgium.
- European Commission. 2018 Technical guidance for deriving environmental quality standards Guidance document no. 27 (2000/60/EC). Brussels, Belgium.
- Franklin NM, Stauber JL, Apte SC, Lim RP. 2002 The effect of initial cell density on the bioavailability and toxicity of copper in microalgal bioassays. *Environ Toxicol Chem* 21:742–751. [PubMed: 11951947]
- Garman ER, Meyer JS, Bergeron CM, Blewett TA, Clements WH, Elias MC, Farley KJ, Gissi F, Ryan A. 2020 Validation of bioavailability-based toxicity models for metals. *Environ Toxicol Chem* 39:101–117. [PubMed: 31880834]
- Geneser RW, Gondek J, Canton SP, Kovach A, Claytor CA. 2016 Regulatory implementation of the copper biotic ligand model for aquatic life protection: What have we learned and how are we doing? Proceedings, WEFTEC 2016 Annual Conference, Water Environment Federation, New Orleans, LA, September 24–28.
- Geneser RW, Gondek JC, Rodriguez PH, Arbildua JJ, Stubblefield WA, Cardwell AS, Santore RC, Ryan AC, Adams WJ, Nordheim E. 2018 Evaluating the effects of pH, hardness, and dissolved organic carbon on the toxicity of aluminum to freshwater aquatic organisms under circumneutral conditions. *Environ Toxicol Chem* 37:49–60. [PubMed: 28833434]
- Gissi F, Stauber JL, Binet MT, Golding LA, Adams MS, Schlekot CE, Garman ER, Jolley DF. 2016 A review of nickel toxicity to marine and estuarine tropical biota with particular reference to the South East Asian and Melanesian region. *Environ Pollut* 218:1308–1323. [PubMed: 27622840]

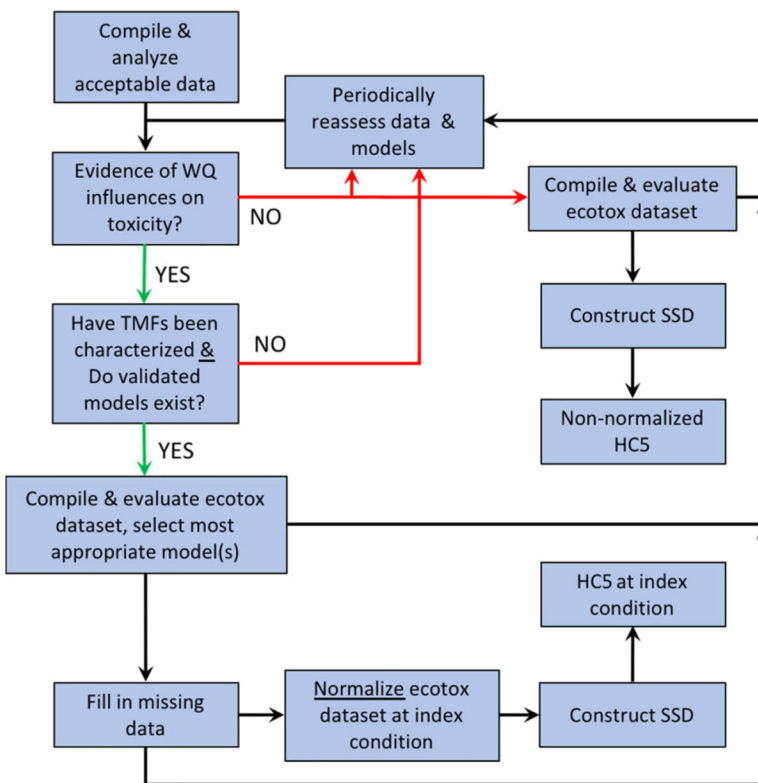


- Gondek JC, Gensemer RW, Claytor CA, Canton SP, Gorsuch JW. 2018 Framework for derivation of water quality criteria using the biotic ligand model: Copper as a case study. *Integr Environ Assess Manag* 14:736–749. [PubMed: 29858531]
- Istituto Superiore per la Protezione e la Ricerca Ambientale. 2016 Linea guida per il monitoraggio delle sostanze prioritarie (secondo D.Lgs. 172/2015). Rome, Italy.
- Klimisch H-J, Andreae M, Tillmann U. 1997 A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. *Regul Toxicol Pharmacol* 25:1–5. [PubMed: 9056496]
- Ma H, Kim SD, Cha DK, Allen HE. 1999 Effect of kinetics of complexation by humic acid on toxicity of copper to *Ceriodaphnia dubia*. *Environ Toxicol Chem* 18:828–837.
- Mebane CA, Chowdhury MJ, De Schamphelaere KAC, Lofts S, Paquin PR, Santore RC, Wood CM. 2020 Metal bioavailability models: Current status, lessons learned, considerations for regulatory use, and the path forward. *Environ Toxicol Chem* 39:60–84. [PubMed: 31880840]
- Merrington G, Peters A, Schlekot CE. 2016 Accounting for metal bioavailability in assessing water quality: A step change? *Environ Toxicol Chem* 35:257–265. [PubMed: 26808908]
- Moermond CTA, Kase R, Korkaric M, Agerstrand M. 2016 CRED: Criteria for reporting and evaluating ecotoxicity data. *Environ Toxicol Chem* 35:1297–1309. [PubMed: 26399705]
- Oregon Department of Environmental Quality. 2017 Table 30, Aquatic life water quality criteria for toxic pollutants. Oregon Administrative Rules 340–041-8033 Portland, OR, USA.
- Organisation for Economic Co-operation and Development. 2016 Guidance on the incorporation of bioavailability concepts for assessing the chemical ecological risk and/or environmental threshold values for metals and inorganic metal compounds. Series on Testing and Assessment No. 259. ENV/JM/MONO(2016)66. Paris, France.
- Peters A, Merrington G, Delbeke K, de Schamphelaere K. 2011 Regulatory consideration of bioavailability for metals: Simplification of input parameters for the chronic copper biotic ligand model. *Integr Environ Assess Manag* 7:437–444. [PubMed: 21082669]
- Peters A, Schlekot CE, Merrington G. 2016 Does the scientific underpinning of regulatory tools to estimate bioavailability of nickel in freshwaters matter? The European-wide environmental quality standard for nickel. *Environ Toxicol Chem* 35:2397–2404. [PubMed: 27253879]
- Prothro MG. 1993 Office of Water Policy and technical guidance on interpretation and implementation of aquatic life metals criteria Memorandum from US Environmental Protection Agency acting assistant administrator for water to Water Management Division directors and Environmental Services Division directors, Regions I–X
- Rüdel H, Díaz Muñoz C, Garelick H, Kandile NG, Miller BW, Pantoja Munoz L, Peijnenburg WJGM, Purchase D, Shevah Y, van Sprang P, Vijver M, Vink JPM. 2015 Consideration of the bioavailability of metal/metalloid species in freshwaters: Experiences regarding the implementation of biotic ligand model-based approaches in risk assessment frameworks. *Environ Sci Pollut Res* 22:7405–7421.
- Ryan A, Santore R, Delos C. 2018 Application of a fixed monitoring benchmark approach to evaluate attainment of time-variable water quality criteria: Copper biotic ligand model as a case study. *Integr Environ Assess Manag* 14:722–735. [PubMed: 29920938]
- Schlekot CE, DeSchamphelaere K, Van Genderen E, Antunes P, Stubblefield W, Rogevich E. 2010 Cross-species extrapolation of chronic nickel biotic ligand models. *Sci Total Environ* 408:6148–6157. [PubMed: 20920817]
- Stephen CE, Mount DI, Hansen DJ, Gentile JR, Chapman GA, Brungs WA. 1985 Guidelines for deriving numerical national water quality criteria for the protection of aquatic organisms and their uses Report PB85–227049. US Environmental Protection Agency, Washington, DC.
- Swedish Marine and Water Authority. 2016 Miljögifter i vatten—Klassificering av ytvattenstatus. Vägledning för tillämpning av HVMFS 2013:19 Gothenburg, Sweden.
- Tack K 2012 Prise en compte de la biodisponibilité des métaux selon la DCE: Guide méthodologique. INERIS report DRC-12–126834-07511A. Verneuil-en-Halatte, France.
- US Environmental Protection Agency 1987 Ambient aquatic life water quality criteria for zinc. EPA 440/5–87-003. Washington, DC.

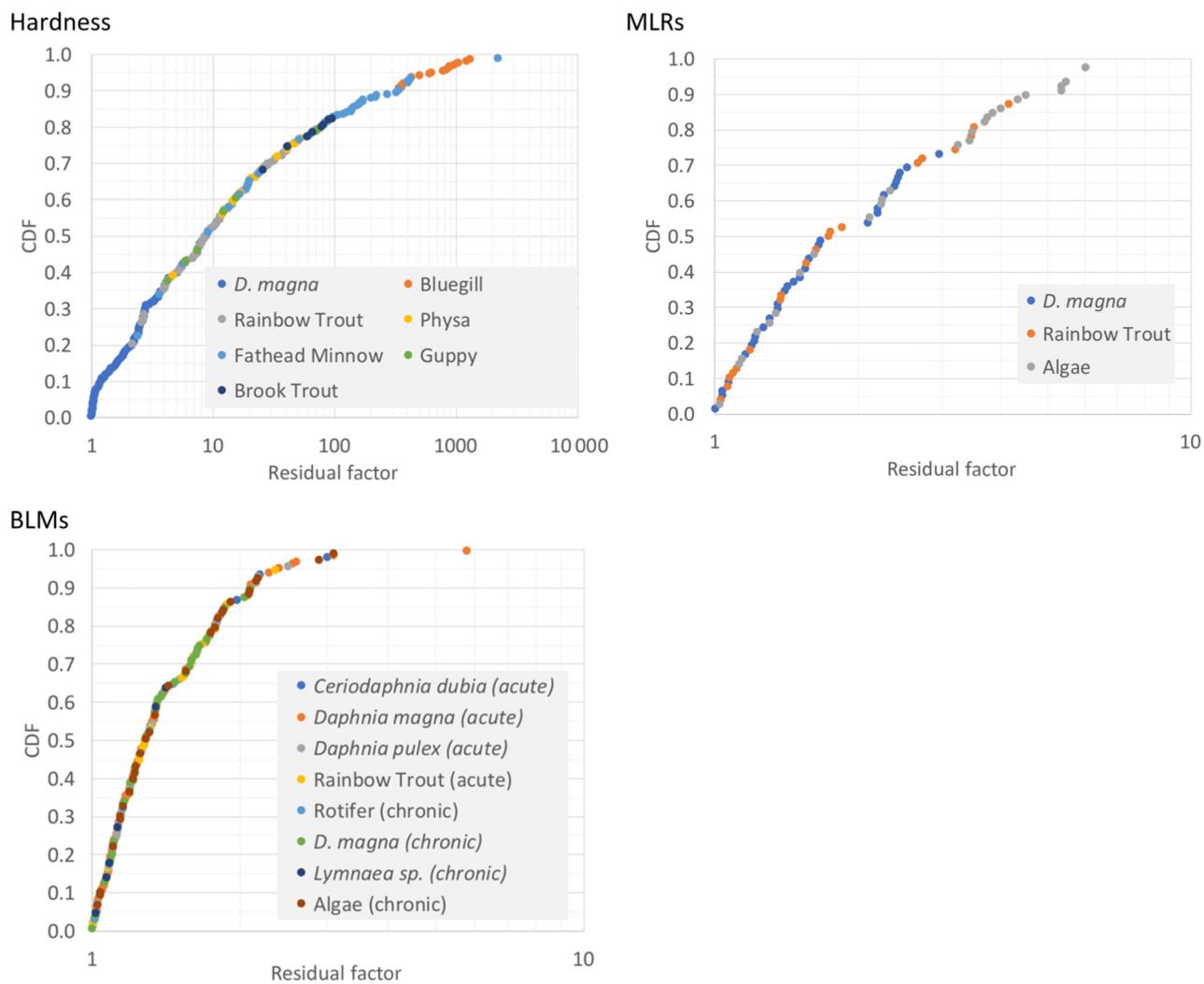
- US Environmental Protection Agency 1994 Interim guidance on determination and use of water-effect ratios for metals. EPA-823-B-94-001. Washington, DC.
- US Environmental Protection Agency 1995 Water quality criteria documents for the protection of aquatic life in ambient water. EPA-820-B-96-001. Washington, DC.
- US Environmental Protection Agency 2007 Aquatic life ambient freshwater quality criteria—Copper. EPA-822-R-07-001. Washington DC.
- US Environmental Protection Agency 2016 Recommended estimates for missing water quality parameters for application in EPA's BLM. Draft technical support document. EPA 820-R-15-106. Washington DC.
- US Environmental Protection Agency 2018 Final aquatic life ambient water quality criteria for aluminum 2018. EPA-822-R-18-001. Washington, DC.
- Van Regenmortel T, Berteloot O, Janssen CR, De Schampheleere KAC. 2017 Analyzing the capacity of the *Daphnia magna* and *Pseudokirchneriella subcapitata* bioavailability models to predict chronic toxicity at high pH and low calcium concentrations and formulation of a generalized bioavailability model for *D. magna*. *Environ Toxicol Chem* 36:2781–2798. [PubMed: 28452073]
- Warne MStJ. 2001 Derivation of the ANZECC and ARMCANZ water quality guidelines for toxicants. *Australasian Journal of Ecotoxicology* 7:123–136.
- Warne MStJ, Batley GE, van Dam RA, Chapman JC, Fox DR, Hickey CW, Stauber JL. 2015 Revised method for deriving Australian and New Zealand water quality guideline values for toxicants. Department of Science, Information Technology, Innovation and the Arts, Brisbane, QLD, Australia.



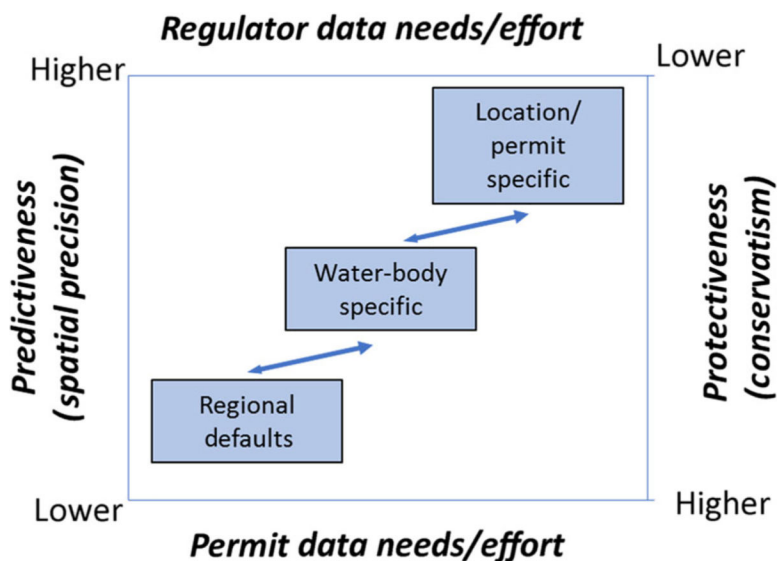
**FIGURE 1:** Conceptual model for use of bioavailability models. HC5 = 5% hazardous concentration; CCC = criterion continuous concentration; CV = chronic value.



**FIGURE 2:** General framework for normalization of toxicity data using bioavailability models. HC5 = 5% hazardous concentration; SSD = species sensitivity distribution; TMF = toxicity-modifying factor; WQ = water quality.



**FIGURE 3:** Residual factor plots (see text) for models used in case study. BLM = biotic ligand model; CDF = cumulative distribution function; MLR = multiple linear regression.



**FIGURE 4:** Example application framework in United States. “Predictiveness” refers to precision of model predictions relative to water quality at a particular location. “Protectiveness” refers to levels of conservatism. “Permit data needs/effort” refers to how much data-collection or model-calculation effort would be required by an individual permit holder. “Regulator data needs/effort” refers to how much data-collection or model-calculation effort would be required by the regulatory agency.

**TABLE 1:** Summary of normalization approaches for metal ecotoxicity data sets from selected jurisdictions

Jurisdiction (year)	Metal(s)	Toxicity-modifying factor(s)	Type of model	Procedure summary	Reference
United States (1983)	Cd, Cu, Pb, Zn	Hardness	Linear	<ul style="list-style-type: none"> <li>Model pooled for multiple species</li> <li>Species mean calculated</li> <li>Genus mean acute values used in SD</li> <li>ACR used for chronic criterion calculation</li> <li>One model for fish and invertebrates</li> <li>Species mean acute value calculated</li> <li>Genus mean acute values used in SD</li> <li>ACR used for chronic criterion calculation</li> <li>Chronic trophic level-specific models</li> <li>Species mean used in SSD</li> <li>DOC correction, driven by most sensitive species</li> <li>Chronic trophic level-specific models</li> <li>Species mean used in SSD</li> <li>DOC correction, driven by most sensitive species</li> <li>Chronic trophic level-specific models</li> <li>Species mean used in SSD</li> <li>Chronic trophic level-specific models</li> <li>Species mean used in SSD</li> <li>Chronic trophic level-specific models</li> <li>Species mean used in SSD</li> <li>Chronic trophic level-specific models</li> <li>Species mean used in SSD</li> <li>Model developed from chronic single-species (fish) data set</li> <li>Species value used in SSD</li> <li>Separate acute and chronic models for fish and invertebrates</li> <li>Species mean chronic values calculated</li> <li>Genus mean chronic values used in SD</li> <li>Chronic models applied to acute SD</li> <li>Species-specific models proposed (chronic data)</li> <li>Species mean used in SSD</li> </ul>	US Environmental Protection Agency (1987)
United States (2007)	Cu	Several water chemistry parameters	BLM		Hardness model in case study
United Kingdom (2009)	Cu, Zn, Mn, Cu	Several water chemistry parameters	BLM Marine MLR		US Environmental Protection Agency (2007)
European Union (2010)	Ni, Pb	Several water chemistry parameters	BLM MLR		Environment Agency (2010, 2012a, 2012b, 2013)
France (2012)	Cu, Zn	Several water chemistry parameters	BLM		European Commission (2010a, 2010b)
Italy (2016)	Cu, Zn, Mn	Several water chemistry parameters	BLM		Tack (2012)
Sweden (2017)	Cu, Zn	Several water chemistry parameters	BLM		Istituto Superiore per la Protezione e la Ricerca Ambientale (2016)
Canada (2018)	Zn	Hardness, pH, DOC	MLR		Swedish Marine and Water Authority (2016)
United States (2018)	Al	Hardness, pH, DOC	MLR		Canadian Council of Ministers of the Environment (2018)
Australia/New Zealand (2018)	Zn	Hardness, pH, DOC	MLR		MLR1-3 models in example US Environmental Protection Agency (2018) Australian and New Zealand Governments (2018b)

ACR = acute to chronic ratio; BLM = biotic ligand model; DOC = dissolved organic carbon; MLR = multiple linear regression; SD = sensitivity distribution; SSD = species sensitivity distribution.

**TABLE 2:**

Among-model comparison relative to accuracy, level of input, and ease of use

Type of model	Accuracy	Level of input	Ease of use
Single-parameter	Low	Low	High
MLR	High	Moderate	High
BLM	High	High	Moderate
Simplified	High	Moderate	Moderate

BLM = biotic ligand model; MLR = multiple linear regression.



**TABLE 3:**

Summary of model evaluation and selection process for a case study with zinc

Model	Scoring <sup>a</sup>	Model performance <sup>b</sup>	Relative scores <sup>c</sup>	Model rank <sup>d</sup>
Hardness	Fish = 33	$RF_{x,2.0} = 0.19$	Fish = 6.3 (0.26)	Fish = 3
	Invertebrate = 40	$1/RF_{y,0.84} = 0.0078$	Invertebrate = 7.6 (0.31)	Invertebrate = 3
	Algae = 0		Algae = 0 (0)	Algae = 3
MLR1	Overall = 73		Overall = 14 (0.57)	Overall = 4
	Fish = 0	$RF_{x,2.0} = 0.33$	Fish = 0 (0)	Fish = 4
	Invertebrate = 0	$1/RF_{y,0.84} = 0.18$	Invertebrate = 0 (0)	Invertebrate = 4
MLR2	Algae = 26		Algae = 8.6 (4.8)	Algae = 2
	Overall = 26		Overall = 8.6 (4.8)	Overall = 4
	Fish = 0	$RF_{x,2.0} = 0.64$	Fish = 0 (0)	Fish = 4
MLR3	Invertebrate = 40	$1/RF_{y,0.84} = 0.41$	Invertebrate = 25.6 (16.3)	Invertebrate = 2
	Algae = 0		Algae = 0 (0)	Algae = 3
	Overall = 40		Overall = 25.6 (16.3)	Overall = 2
BLM	Fish = 22	$RF_{x,2.0} = 0.70$	Fish = 15.4 (6.3)	Fish = 2
	Invertebrate = 0	$1/RF_{y,0.84} = 0.29$	Invertebrate = 0 (0)	Invertebrate = 4
	Algae = 0		Algae = 0 (0)	Algae = 3
Overall = 22	Overall = 22		Overall = 15.4 (6.3)	Overall = 3
	Fish = 22	$RF_{x,2.0} = 0.87$	Fish = 19 (12)	Fish = 1
	Invertebrate = 53	$1/RF_{y,0.84} = 0.53$	Invertebrate = 46 (28)	Invertebrate = 1
Overall = 101	Algae = 26		Algae = 23 (14)	Algae = 1
	Overall = 101		Overall = 88 (53)	Overall = 1

<sup>a</sup> Total of "good" scores summed for water quality and taxonomic coverage for each test, trophic level, and overall (step 4; see text).

<sup>b</sup>  $RF_{x,2.0}$  = percent of predictions within a factor of 2.0;  $1/RF_{y,0.84}$  = residual factor associated with 84% model accuracy (step 5; see text).

<sup>c</sup> Product of RF and evaluation score (step 6; see text). Results presented as  $RF_x (1/RF_y)$ .

<sup>d</sup> Among-model ranking (step 7; see text).

BLM = biotic ligand model; MLR = multiple linear regression; RF = residual factor.

EPA Author Manuscript

EPA Author Manuscript

EPA Author Manuscript