



ENVIRONMENTAL TOXICITY DATA FOR METALS: DEFINING THEIR UTILITY

by Dr. W. Stubblefield, Ph.D.

There are three purposes for the development and use of environmental toxicity data:

- Predicting or investigating possible adverse effects to fish, wildlife or plants resulting from exposure to environmental contaminants;
- Increasing our basic scientific understanding of the types and mechanisms of toxic effects that can occur; and
- Developing information to define environmental concentrations of contaminants for use in regulatory actions to prevent adverse effects due to environmental releases.

For all three cases, it is imperative that the highest quality data be available. For example, the evaluation and quantification of possible environmental injury resulting from the discharge of an effluent from a mining operation will likely be based on the results of toxicity tests. Decisions about the need for and selection of remedial options at contaminated sites may also rest on toxicity test results. Generating and reporting toxicity test data of the highest quality and lowest uncertainty are essential for making appropriate and responsible environmental management decisions. The societal and economic implications of using poor-quality data in decision making can be immense.

What Are “High-quality” Data?

In answering this question, two items must be considered: *data relevance* and *data reliability*. The former means that the study is consistent with the intended use of the

data, while the latter means that the study was conducted in accordance with best scientific practices and is well documented.

Data Relevance

Data relevance relates to what every investigator must consider in designing a study, and what every regulator must answer when using reported results for management purposes: *For what purpose are these data to be used, and is the test design appropriate for that purpose?*

To determine whether the data are relevant, three specific questions should be asked:

- 1. Are the test species relevant to the site in question?** For example, in evaluating the potential impacts of mine tailings to aquatic organisms in Bolivia's Lake Titicaca (a freshwater lake), freshwater species such as *Ceriodaphnia* should be used, rather than *Artemia*, a brine shrimp associated only with saline waters.
- 2. Was the test design appropriate to real-world exposures?** Addressing concerns about effects is best accomplished with toxicity tests that mimic actual exposure type and duration.

For example, there is less uncertainty in assessing long-term effects on reproduction or population stability using test data from longer-term studies, rather than relying on extrapolations from acute toxicity tests.

RELEVANCE:

The study is consistent with the intended use of the data.

RELIABILITY:

The study was conducted in accordance with best scientific practices and is well documented.



3. Were the endpoints measured in the toxicity test germane to the site/issue of concern? Tests that report endpoints that cannot be quantitatively linked to an adverse population-level effect (i.e. survival, growth and reproduction) may be of little value. For example, induction of metallothionein is a common compensatory response observed following exposure to metals; however, this response is of questionable use for estimating the effects of copper on the sustainability of a population of rainbow trout.

The more closely the test species and conditions emulate the environment of concern, the higher the likelihood that the data will be of direct value, and the lower the potential for error due to uncertainty or unmet assumptions. If a study does not meet the requirements for relevance to the issue being considered, it makes no difference whether the test is of high or low reliability.

There is, however, no standard procedure to judge whether a test is relevant or not—such decisions fall ultimately to best professional judgment. Scientific knowledge about which factors affect toxicity is increasing continuously; therefore, data-quality screening procedures must retain sufficient flexibility to permit change as new information becomes available. Data judged relevant today might be rejected in the future as our knowledge increases. For example, much of the extant metals toxicity data generated in the 1960s and 1970s was reported based on nominal (i.e., unmeasured) rather than on analytically confirmed concentrations. Today, we know the importance of measuring metals concentrations and it is now standard practice to do so.

Data Reliability

Once it has been determined that a scientific study is relevant to the question being asked, the reliability of the data must be evaluated. Reliability is related to several items:

- The test methods and conditions under which the test was conducted;
- The Quality Assurance (QA) procedures that were used;
- Whether clear exposure–response relationships were observed; and
- How well test results were reported.

RELIABILITY CRITERIA:

- **Test design**
- **Execution**
- **Reporting**

To determine reliability, questions concerning the following three stages should be asked:

1. Test design
2. Execution
3. Reporting

1. Was the study properly designed? Experimental designs and methodologies can be standardized for use in applied studies (e.g., the American Society for Testing and Materials [ASTM], United States Environmental Protection Agency [U.S. EPA], and Organization for Economic Co-operation and Development [OECD], ISO international standards and EU directives). Alternatively, non-standard methods can be used, particularly for basic research intended to understand mechanisms of toxic effects. Regardless, several basic elements of experimental design must be considered and evaluated:

- Were the experimental methods, including test setup, measuring devices/chambers, preparation of test materials and dilution water, and the physico-chemical conditions used in the study (e.g., pH, hardness, dissolved oxygen, organic carbon, etc.) appropriate?
- Was the test material (e.g., the metal salt being studied) appropriate for the experimental question being asked, and was it adequately characterized?
- Was the general experimental design, including the number of exposure concentrations, number of replicates, selection of exposure concentrations, and spacing of the intervals between exposure concentrations, adequately considered?
- Were the methods for data analysis and statistical methods appropriate for the study design?
- Were the appropriate endpoints measured (e.g., mortality, reproduction, population growth versus metallothionein or enzyme concentrations)?

2. Was the study executed properly?

Proper execution of a scientific study addresses not only how well the experimental design was followed, but also the accuracy and precision of the data obtained. In many applied studies (e.g., pesticide registration, effluent discharge permit compliance), this

involves explicit quality assurance/quality control (QA/QC) procedures. Elements of this aspect of data quality evaluation may include:

- Deviations from proposed methods or standard protocols—were any problems encountered that could impact results (e.g., lost replicates, logistic problems, etc.)?
- Organism health/performance—ensuring both proper organism acclimation to test conditions (e.g., water hardness, temperature) and appropriate organismal responses during the study. For example, in toxicity studies, it is important to report whether control mortality was within acceptable limits, or whether performance in response to a positive control (e.g., a known toxicant) was within expected limits.
- Data accuracy and precision—one should be able to evaluate not only whether the results were correct (**accuracy**), but also whether variability among replicate observations of an individual treatment (**precision**) were within acceptable limits.

3. Was the test reported adequately? One of the basic tenets for conducting scientific investigations is that it must be possible for other investigators to evaluate the results of the experiment and reach their own independent conclusions. Sufficient information must be reported for independent investigators to repeat the test and to judge the repeatability of the findings. This can only be accomplished if the original experiment was reported in sufficient detail. This is a critical—and often overlooked—aspect of data quality evaluation. Without thorough descriptions of study methods and results, it is impossible to evaluate independently whether the study met its original objectives (**relevance**), or was of sufficient quality to meet these objectives (**reliability**).

Several aspects of adequate reporting include:

- Description of experimental methods or standard protocols, including deviations (if any) and analytical methods (with detection limits). Ideally, these should be described in detail for every study, but publication length often constrains the amount of detail provided. In such cases, easily obtainable references to methods should be included;

- Description of data analyses and statistical methods used; and
- Complete data reporting, including some indication of data variability (either graphically or in tables).

To address the issue of consistency in study methods, a number of national and international authorities such as the ASTM and OECD have developed “standardized” procedures that can be used in conducting laboratory toxicity tests for fish, wildlife and plants. Standardized protocols offer the advantage of procedural consistency, data comparability, and, in some cases, a standard format for data reporting. The use of standardized methods ensures that the test results will be useful for their intended purpose. However, standardized test procedures often do not provide sufficient flexibility to allow researchers to design and conduct tests to answer specific or novel questions on a wide array of compounds. In these situations, the standardized procedure can be used to the extent possible but the investigator must document in detail how the actual methods differed. For example, conducting an acute toxicity test with fish in accordance with OECD Method 203 suggests that the test dilution water hardness be between 10 and 250 mg/L (as CaCO₃). However, if the test is being conducted to ascertain the toxicity of copper to rainbow trout in a stream with a water hardness of 500 mg/L (as CaCO₃), the test results from the standardized method will greatly overestimate the toxicity of the metal under field conditions. Copper acute toxicity to rainbow trout can vary substantially based on test water hardness, e.g., LC₅₀s of 17 and 300 µg Cu/L at a hardness of 20 and 400 mg/L as CaCO₃, respectively.

Adequate data reporting is crucial even for standard tests: simply stating that a test followed an ASTM or OECD method does not adequately characterize the procedures that were employed and does little to assure that desired test conditions were achieved. In the case of algal toxicity studies, some standard methods permit the test to be conducted with or without the addition of the chelating agent, EDTA. The presence of EDTA can drastically alter the results of algal toxicity tests, especially of metals, potentially affecting EC₅₀ concentrations by more than one order of magnitude. The investigator must report whether EDTA was added to the culture media; otherwise, it

would be difficult to determine why EC₅₀s from two tests differed and it would be impossible for an investigator to repeat the test and obtain similar results.

Criteria and Standards: A Case Study

From a regulatory point of view, laboratory toxicity test data form the basis for criteria or standards that are used to protect aquatic organisms and wildlife. The U.S. EPA's Ambient Water Quality Criteria (AWQC) are a good example of how toxicity test data are used in deriving regulatory benchmarks. AWQC are concentrations of environmental contaminants that, if not exceeded, are not anticipated to cause adverse effects in exposed organisms. Rigorous methods have been developed for deriving criteria (U.S. EPA 1985) and include requirements for both the quantity and type of data that can be used. However, when the AWQC derivation procedures were developed, few specific detailed procedures for evaluating the quality of toxicity test data were defined; more specific review criteria have been developed recently. Some of the data that were used initially to derive AWQC would not satisfy today's standards for data quality. For example, in a number of current metals criteria, toxicity test data were included in the derivation database that were based on nominal, i.e., not analytically determined exposure concentrations. By today's standards, this would not generally be acceptable.

The importance of individual data quality cannot be overemphasized. This is especially true in those cases where the results of a single toxicity test are driving the derivation of a criterion. This situation becomes extremely difficult when evaluation of the test data indicates that there may have been relevancy or reliability problems with the study or, as is often the case, the study report contains insufficient information to permit the reviewer to properly evaluate the reliability of the data. If the test result suggests a high degree of toxicity to the tested species and if no other "higher-quality" data are available for the same species, the reviewer may decide to accept the test, with reservation, resulting in a criterion that may be conservative and potentially unnecessarily restrictive. Sufficient flexibility must be incorporated into any data evaluation strategy to permit the replacement of "suspect" data with "high-quality" data as the latter become available.

Standardized Data Evaluation Methods

In the early 1980s, the U.S. EPA's Office of Pesticide Programs (OPP) developed a series of study-specific "Standard Evaluation Procedures" (SEPs) to evaluate the adequacy and acceptability of toxicity test results submitted to support the registration of pesticides in the United States. This effort provided guidance to both the reviewer and the investigator about what should be addressed in conducting and reporting toxicity test results. Although the SEPs were drafted for the evaluation of pesticide toxicity tests and are currently undergoing revision, much of the guidance is

STANDARDIZED DATA EVALUATION METHODS

- U.S. EPA's Standard Evaluation Procedures for Pesticides
- U.S. EPA's Evaluation and Interpretation of Suitable Test Results in AQUIRE (EVISTRA)
- RIVM's data quality criteria system¹
- UBA's data quality screening system²
- ETAP panel-suggested data quality system for the evaluation of metals toxicity data³

¹ RIVM: National Institute of Public Health and Environmental Protection, the Netherlands

² UBA: Umwelt Bundes Amt, Germany

³ ETAP: Ecotoxicity Technical Advisory Panel

applicable to the conduct of the tests and the reporting of metals toxicity data.

The U.S. EPA's Office of Research and Development (ORD) currently is developing guidance for evaluating results of aquatic toxicity tests to be included in the AQUIRE (**A**quatic **I**nformation **R**etrieval) database (www.epa.gov/ecotox). The AQUIRE database is a compilation of toxicity test results and related test information for individual chemicals from laboratory and field aquatic toxicity tests. Data are taken from scientific papers published

both nationally and internationally on the toxic effects of chemicals to aquatic organisms, including aquatic plants. Data currently included in AQUIRE are not evaluated for quality or utility—only the quality of the study report is provided in the “document code.” EVISTRA (Evaluation and Interpretation of Suitable Test Results in AQUIRE) is intended to remedy this deficiency and provide information about the overall *quality* of a study and the *suitability* of the data for the purpose of deriving numeric criteria and/or standards. A draft version of the EVISTRA guidance document can be obtained at www.epa.gov/med/databases/evistra.html.

Conclusions

New information and new approaches (e.g., the Biotic Ligand Model) are constantly being developed, requiring us to re-evaluate and revise our view of metals toxicity. Information currently thought to be of minimal importance may be the “key” variable in the future. Toxicity tests must be conducted and reported in a manner consistent with prevailing scientific standards and in sufficient detail to ensure future utility as those standards change. Furthermore, too often we forget to consider the variable nature of real-world exposures and the “resiliency” of natural systems; rather, we assume that laboratory tests appropriately mimic the real world.

Laboratory data are needed to characterize the effects of variable exposures and the potential interactive effects of metals mixtures; monitoring data are needed to characterize the magnitude and variable nature of field exposures; and additional ecological investigations are needed to help quantify the degree, frequency, and magnitude of metals exposures that can be tolerated by aquatic systems. The toxicity of metals, perhaps more so than that of organic chemicals, can be affected by environmental variables that alter bioavailability and thus, toxicity. Investigators must be encouraged to obtain and report such parameters, including concentrations of major cations (Ca, Mg, Na, K) and anions (bicarbonate, chloride, SO₄), pH, dissolved organic carbon, temperature, and both total and dissolved metal concentrations. Collection of these types of data will demand that even closer attention be paid to test design, data collection, and method reporting to ensure reliability and repeatability of results. Industry and regulatory scientists must work together to ensure that the data necessary to

evaluate the effects of metals on the environment are available and that the information is used for its intended purpose.

The importance of high-quality data from studies designed to address specific research or regulatory concerns cannot be over-emphasized. Data that do not satisfy minimum scientific criteria for data quality and acceptability serve no useful purpose and may undermine the acceptance of other valid, useful data. Scientists must take care in designing, conducting and reporting the results of toxicological investigations. Keeping in mind the old question, *What does it take to un-ring a bell?*, scientists must consider the possible implications of the data they publish. Once data enter the public domain, they may be used for purposes beyond their initial intent and so must always be of the highest quality and reliability before being released.

Recommendations

- Follow strict QA procedures and obtain and report, at a minimum, parameters characterizing test conditions;
- Maintain detailed test reports and “raw” test data for studies reported in the scientific literature;
- Use test species, test designs and measured endpoints that are relevant to the question being addressed;
- Be sure the test was designed, executed, and reported according to standard data evaluation procedures; and
- Use data only for their intended purpose.

For Additional Information

Centre for Substances and Risk Assessment (CSR). 1996. QA-procedures for deriving environmental quality objectives (INS and I-values). CSR/KD/003.

ETAP. 1999. Criteria for the reliability and usefulness of ecotoxicity data for environmental risk assessment. Review Workshop on the Environmental Risk Assessment for Zinc. Ottawa, Canada. 29-30 Oct. 1998. (copies available from ICME)

RIVM. 1997. Ch. Bodar: Guidelines for the evaluation of HEDSET EU 793/93. (Internal doc. In Dutch).

TGD. 1996. Meeting notes of the IPCS task force discussions on zinc, Adelaide (AUS). U.S. Environmental Protection Agency (USEPA). 1985. Hazard Evaluation Division, Standard Evaluation Procedure: Acute toxicity test for freshwater fish. EPA-540/9-85-006. Office of Pesticide Programs. Washington, D.C. pp 18.

USEPA. 1985. Hazard Evaluation Division, Standard Evaluation Procedure: Acute toxicity test for freshwater invertebrates. EPA-540/9-85-005. Office of Pesticide Programs. Washington, D.C. pp 12.

USEPA. 1985. Hazard Evaluation Division, Standard Evaluation Procedure: Acute toxicity test for estuarine and marine organisms (Shrimp 96-h acute toxicity test). EPA-540/9-85-010. Office of Pesticide Programs. Washington, D.C. pp 17.

USEPA. 1999. EVISTRA (Evaluation and Interpretation of Suitable Test Results in AQUIRE) Draft guidance for evaluating results of aquatic toxicity tests. Office of Research and Development. Duluth, MN, USA. pp 81.

About the Author:

Dr. William Stubblefield is senior technical director of ENSR International's environmental toxicology laboratory in Fort Collins, Colorado, and serves as an affiliate faculty member in the Departments of Fish and Wildlife Biology and Environmental Health at Colorado State University. He provides consulting services to the mining industry on environmental issues, especially those associated with the evaluation of impacts to aquatic and terrestrial species from discharge of mine-associated waters and tailings. He also directs studies of the toxicity of metals and by-products of metal mining to aquatic organisms. Dr. Stubblefield received his Master's degree in toxicology/toxicodynamics at the University of Kentucky in 1979 and his Ph.D. in environmental toxicology from the University of Wyoming in 1987. He has authored more than 50 peer-reviewed publications and technical presentations in the areas of aquatic and wildlife toxicology and environmental risk assessment. Dr. Stubblefield is an active member of the Society of Environmental Toxicology and Chemistry, where he serves as Chair of the Publications Advisory Council, Associate Editor of the Society's newsletter; he was previously a member of the Society's Board of Directors.

Fact Sheet on Environmental Risk Assessment

This is the first in an occasional series of *Fact Sheets* to be produced by ICME on metal-specific issues in environmental risk assessment. Authorship selection and editorial review are coordinated by Dr. Anne Fairbrother of Parametrix, Inc. Each *Fact Sheet* is reviewed for technical merit by Dr. Erik Smolders of Katholieke Universiteit (Catholic University) Leuven, Belgium, and by a panel of experts in metal-related regulatory issues. While the *Fact Sheets* reflect the views of the authors, they are intended to provide an objective review of each topic. ICME hopes these publications provide insights into complex issues in regulatory science, and welcomes questions and comments.

These *Fact Sheets* are also available on the ICME Web site. For additional information, please contact:

The International Council on Metals and the Environment

294 Albert Street, Suite 506
Ottawa, Ontario, Canada K1P 6E6
Tel.: (+1 613) 235-4263 Fax: (+1 613) 235-2865
E-mail: info@icme.com
Web site: www.icme.com



INTERNATIONAL COUNCIL ON
METALS AND THE ENVIRONMENT

Promoting policies and practices towards sustainable development

ICME