EPA 600/R-01/020
March 2001

Method for Assessing the Chronic Toxicity of Marine and Estuarine Sediment-associated Contaminants with the Amphipod *Leptocheirus plumulosus*

First Edition

Office of Research and Development
Western Ecology Division
U.S. Environmental Protection Agency
Newport, OR 97365

Office of Science and Technology
Office of Water
U.S. Environmental Protection Agency
Washington, D.C. 20460

Engineer Research and Development Center
Waterways Experiment Station
U.S. Army Corps of Engineers
Vicksburg, MS 39180

Foreword

Sediment contamination is a widespread environmental problem that can potentially pose a threat to a variety of aquatic ecosystems. Sediment functions as a reservoir for common chemicals such as pesticides, herbicides, polychlorinated biphenyls (PCBs), polycyclic aromatic hydrocarbons (PAHs), and metals, such as lead, mercury, and arsenic. In-place contaminated sediment can result in depauperate benthic communities. Because relationships between bioavailability and concentrations of chemicals in sediment are not fully understood, determination of contaminated sediment effects on aquatic organisms requires the use of controlled toxicity and bioaccumulation tests.

As part of USEPA's Contaminated Sediment Management Strategy, Agency programs have agreed to use consistent methods to determine whether sediments have the potential to affect aquatic ecosystems. More than ten federal statutes provide authority to many USEPA program offices to address the problem of sediment contamination. The use of this uniform sediment testing procedure is expected to increase data accuracy and precision, facilitate test replication, and increase the comparative value of test results. The sediment test method in this manual may be useful in assessing sediment contamination, registration of pesticides, assessment of new and existing industrial chemicals, Superfund site assessment, and assessment and cleanup of hazardous waste treatment, storage, and disposal facilities. Each EPA Program will, however, retain the flexibility of deciding when and how to use this test and whether identified risks would trigger actions.

A chronic sediment toxicity test (which is used to study the effects of continuous, long-term exposure of a toxicant on an organism) using the estuarine benthic amphipod*, Leptocheirus plumulosus*, was developed by DeWitt et al. (1992a) for the USEPA. McGee et al. (1993) and Emery et al. (1997) independently developed chronic test methods with *L. plumulosus* that measured similar endpoints. Subsequent to these method development efforts, the USEPA and the U.S. Army Corps of Engineers (USACE) have funded further research to refine this chronic method. Findings from studies at both organizations have been incorporated into the chronic testing method described in this document. The protocol for the *L. plumulosus* 28-d sediment toxicity test will be revised periodically, as such, users of this manual are encouraged to contribute to this effort by sending to the USEPA the results of experiments that could bring to light any deficiencies or improvements to the *L. plumulosus* 28-d sediment toxicity test. Send these results and all supporting information (i.e., experimental conditions and procedures) to the U. S. Environmental Protection Agency, Office of Science and Technology/Standards and Health Protection Division (mail code 4305), ATTN: Contaminated Sediment Program, 1200 Pennsylvania Avenue, NW., Washington, D.C. 20460. Contributors to the improvement of the methodology will be acknowledged in future revisions to this manual.

This document is supplementary to USEPA (1994d), but does not replace it. The approaches described in this manual were developed from DeWitt et al. (1992; 1997a; 1997b), McGee et al. (1993), Emery et al. (1997), Scott and Redmond (1989), DeWitt et al. (1989), Schlekat et al. (1992), American Society for Testing and Materials (ASTM, 2000a; 2000f), U.S. Army Corps of Engineers (Emery and Moore, 1996), and U.S. Environmental Protection Agency (USEPA 1994d, 2000).

For additional guidance on the technical considerations in the manual, please contact Ted DeWitt, USEPA, Newport, OR (541/867-4029, fax –4049, email dewitt.ted@epamail.epa.gov) or Todd Bridges, USACE, Vicksburg, MS (601/634-3626, fax –3713, email Todd.S.Bridges@erdc.usace.army.mil).

The cover art is an illustration of *Leptocheirus plumulosus*, by E.L. Bousfield, reproduced with permission of the Canadian Museum of Nature, Ottawa, Canada.

Disclaimer

This technical manual describes recommended procedures for testing an estuarine organism in the laboratory to evaluate the potential toxicity of contaminants in whole sediments. This manual has no immediate or direct regulatory consequence. It does not impose legally binding requirements on the U. S. Environmental Protection Agency (EPA), the U.S. Army Corps of Engineers (USACE), states, tribes, other regulatory authorities, or the regulated community, and may not apply to a particular situation based upon the circumstances. EPA, USACE, state, tribal, and other decision makers retain the discretion to adopt approaches on a case-by-case basis that differ from those in this manual where appropriate. EPA or USACE may change this manual in the future.

The information in this document has been funded in part by EPA and USACE. It has been subjected to review by EPA's National Health and Environmental Effects Research Laboratory and Office of Science and Technology and approved for publication. Mention of trade names or commercial products does not constitute endorsement by either Agency or recommendation for use.

Abstract

A laboratory method is described for determining the chronic toxicity of contaminants associated with whole sediments. Sediments may be collected from estuarine or marine environments or spiked with compounds in the laboratory. The toxicity method outlined uses an estuarine crustacean, the amphipod *Leptocheirus plumulosus*. The toxicity test is conducted for 28 d in 1-L glass chambers containing 175 mL of sediment and about 725 mL of overlying water. Test temperature is 25° ±2°C, and the recommended overlying water salinity is 5‰ ±2‰ (for test sediment with pore water at 1‰ to 10‰) or 20‰ ±2‰ (for test sediment with pore water >10‰). Four hundred milliliters of overlying water is renewed three times per week, at which times test organisms are fed. The endpoints in the toxicity test are survival, growth, and reproduction of amphipods. Performance criteria established for this test include the average survival of amphipods in negative control treatment must be greater than or equal to 80% and there must measurable growth and reproduction in all replicates of the negative control treatment. This test is applicable for use with sediments from oligohaline to fully marine environments, with a silt content greater than 5% and a clay content less than 85%.

Acknowledgments

This document is a general purpose toxicity testing manual for estuarine and marine sediment. The approaches described in this manual were developed primarily from DeWitt et al. (1992, 1997a, 1997b), McGee et al. (1993), and Emery et al. (1997). That work and the impetus for this manual were derived from previous key papers and reports, including Swartz et al. (1985), Scott and Redmond (1989), DeWitt et al. (1989), Schlekat et al. (1992), American Society for Testing and Materials (ASTM, 1998b; 1998e), U.S. Army Corps of Engineers (Emery and Moore, 1996), and U.S. Environmental Protection Agency (USEPA, 1994d; 2000). This manual incorporates general guidelines that reflect the consensus of the USEPA Program Offices and the USACE.

The principal authors of this manual are Theodore H. DeWitt (USEPA, Office of Research and Development, National Health and Environmental Effects Research Laboratory, Western Ecology Division), Todd S. Bridges (USACE, Engineer Research and Development Center), D. Scott Ireland and Leanne L. Stahl (USEPA, Office of Water, Office of Science and Technology), and Margaret R. Pinza and Liam D. Antrim (Battelle Marine Science Laboratory).

Some of the material in this manual was excerpted from USEPA (1994a, 2000) for the purpose of consistency among sediment toxicity test methods manuals. Contributors to specific sections of the manual, including contributors to relevant sections of USEPA (1994a and 2000), are:

1. Section 1-9; General Guidelines
   - G.T. Ankley, USEPA, Duluth, MN
   - T.S. Bridges, USACE, Vicksburg, MS
   - G.A. Burton, Wright State University, Dayton, OH
   - T.D. Dawson, ILS, Duluth, MN
   - T.H. DeWitt, USEPA, Newport, OR
   - F.J. Dwyer, USGS, Columbia, MO
   - R.A. Hoke, DuPont, Newark, DE
   - C.G. Ingersoll, USGS, Columbia, MO
   - D.S. Ireland, USEPA, Washington, DC
   - N.E. Kemble, USGS, Columbia, MO
   - J.O. Lamberson, USEPA, Newport, OR
   - D.R. Mount, USEPA, Duluth, MN
   - T.J. Norberg-King, USEPA, Duluth, MN
   - M.S. Redmond, Northwestern Aquatic Sciences, Newport, OR
   - C.E. Schlekat, USGS, Menlo Park, CA
   - K.J. Scott, SAIC, Narragansett, RI
   - L. Stahl, USEPA, Washington, DC
   - R.C. Swartz, USEPA, Newport, OR(retired)

2. Sections 10-11; Culture and Test Methods
   - L.D. Antrim, NOAA Olympic Marine Sanctuary, Port Angeles, WA
   - T.S. Bridges, USACE, Vicksburg, MS
   - T.H. DeWitt, USEPA, Newport, OR
   - V.L. Emery, USACE, Vicksburg, MS
   - B.D. Gruendell, Battelle, Sequim, WA
   - B.L.. McGee, USFWS, Annapolis, MD

D.W. Moore, MEC Analytical Systems, Carlsbad, CA
L.A. Niewolny, Battelle, Sequim, WA
M.R. Pinza, Battelle, Sequim, WA

3. Sections 12; Statistical Analysis
   V.I. Cullinan, Battelle, Sequim, WA
   J.D. Farrar, AScI, Vicksburg, MS
   B.R. Gray, AScI, Vicksburg, MS
   J. Heltshe, SAIC, Narragansett, RI
   R.A. Hoke, DuPont, Newark, DE
   H. Lee, USEPA, Newport, OR
   T.J. Norberg-King, USEPA, Duluth, MN
   C.E. Schlekat, USGS, Menlo Park, CA

4. Section 13; Precision and Accuracy
   V.I. Cullinan, Battelle, Sequim, WA
   T.H. DeWitt, USEPA, Newport, OR
   C.G. Ingersoll, USGS, Columbia, MO
   T.J. Norberg-King, USEPA, Duluth, MN
   M.R. Pinza, Battelle, Sequim, WA

Review comments from the following individuals led to substantial improvements in the manual for which we are grateful: B.A. Barbo, Battelle, Sequim, WA; Walter Berry, USEPA, Narragansett, RI; G.A. Burton, Wright State Univ., Dayton, OH; P. Crocker, USEPA, Dallas, TX; V.L. Emery, Vicksburg, MS; Daniel Farrar, AScI, Vicksburg, MS; C.G. Ingersoll, USGS, Columbia, MO; Laura Johnson, USEPA, Washington, DC; N.P. Kohn, Battelle, Sequim, WA; J. Lazorchak, USEPA, Cincinnati, OH; M. Lewis, USEPA, Gulf Breeze, FL; B.L. McGee, USFWS, Annapolis, MD; D.W. Moore, MEC, Carlsbad, CA; C.E. Schlekat, USGS, Menlo Park, CA; J. Serbst, USEPA, Narragansett, RI; J. Smrchek, USEPA, Washington, DC; and Jeff Stevens, USACE, Vicksburg, MS.

Participation by the following laboratories in the round-robin testing is greatly appreciated: Battelle Marine Sciences Laboratory, Sequim, WA; Environment Canada, Dartmouth, Nova Scotia, Canada; EVS Consultants, Inc., Vancouver, British Columbia, Canada; Science Applications International Corporation (SAIC), Narragansett, RI; University of California at Santa Cruz, Marine Pollution Studies Laboratory, Monterey, CA; University of Maryland, Wye Research and Education Center, Queenstown, MD; University of South Carolina, Environmental Health Sciences Department, Columbia, SC; USACE Waterways Experiment Station, Vicksburg, MS; USEPA Atlantic Ecology Division Laboratory, Narragansett, RI; and USEPA Gulf of Mexico Ecology Division Laboratory, Gulf Breeze, FL.

We are very grateful to USEPA's Office of Water, Office of Science and Technology, USEPA's Office of Research and Development, and USACE's Long-term Effects of Dredging Operations research program for supporting the development of this manual and for supporting much of the underlying research.

Contents

Foreword . . . iii
Disclaimer . . . iv
Abstract . . . v
Acknowledgments . . . vi
1 Introduction . . . 1
1.1 Significance of Use . . . 1
1.2 Program Applicability . . . 3
1.3 Scope and Application . . . 5
1.4 Performance-based Criteria . . . 10
2 Summary of Method . . . 11
2.1 Method Description and Experimental Design . . . 11
2.2 Types of Tests . . . 13
2.3 Test Endpoints . . . 13
3 Definitions . . . 14
3.1 Terms . . . 14
4 Interferences . . . 16
4.1 General Introduction . . . 16
4.2 Noncontaminant Factors . . . 17
4.3 Changes in Bioavailability . . . 19
4.4 Presence of Indigenous Organisms . . . 20
5 Health, Safety, and Waste Management . . . 22
5.1 General Precautions . . . 22
5.2 Safety Equipment . . . 22
5.3 General Laboratory and Field Operations . . . 22
5.4 Disease Prevention . . . 23
5.5 Safety Manuals . . . 23
5.6 Pollution Prevention, Waste Management, and Sample Disposal . . . 23
6 Facilities, Equipment, and Supplies . . . 24
6.1 General . . . 24
6.2 Facilities . . . 24

6.3 Equipment and Supplies 24
7 Water, Reagents, and Standards 28
7.1 Water 28
7.2 Reagents 29
7.3 Standards 29
8 Sample Collection, Storage, Manipulation, and Characterization 30
8.1 Collection 30
8.2 Storage 30
8.3 Manipulation 31
8.4 Characterization 33
9 Quality Assurance and Quality Control 36
9.1 Introduction 36
9.2 Performance-based Criteria 36
9.3 Facilities, Equipment, and Test Chambers 36
9.4 Test Organisms 37
9.5 Water 37
9.6 Sample Collection and Storage 37
9.7 Test Conditions 37
9.8 Quality of Test Organisms 37
9.9 Quality of Food 38
9.10 Test Acceptability 38
9.11 Analytical Methods 38
9.12 Calibration and Standardization 38
9.13 Replication and Test Sensitivity 38
9.14 Demonstrating Acceptable Performance 39
9.15 Documenting Ongoing Laboratory Performance 39
9.16 Reference Toxicants 39
9.17 Record Keeping 40
10 Collection, Culture, and Maintaining of Test Organisms 42
10.1 Life History 42
10.2 General Culturing Procedures 43
10.3 Culturing Procedure for *Leptocheirus plumulosus* 44
10.4 Field Collection 47

10.5 Obtaining Amphipods for a Test 49
10.6 Minimization of Risk of Release of Nonindigenous Organisms 49
11 *Leptocheirus plumulosus* 28-d Chronic Test for Sediment 50
11.1 Introduction 50
11.2 Procedure for Conducting a *Leptocheirus plumulosus* 28-d Test for Measuring Sublethal Effects of Sediment-associated Contaminants 51
11.3 General Procedures 51
11.4 Interpretation of Results 60
12 Data Recording, Data Analysis and Calculations, and Reporting 63
12.1 Data Recording 63
12.2 Data Analysis 63
12.3 Data Interpretation 81
12.4 Reporting 82
13 Precision and Accuracy 83
13.1 Determining Precision and Accuracy 83
13.2 Accuracy 84
13.3 Replication and Test Sensitivity 84
13.4 Demonstrating Acceptable Laboratory Performance 84
13.5 Precision of the 28-d Chronic Sediment Toxicity Test Method 86
14 References 91

Tables

1.1 Sediment Quality Assessment Procedures 4
1.2 Statutory Needs for Sediment Quality Assessment 6
4.1 Advantages and Disadvantages for Use of Sediment Tests 16
6.1 Equipment and Supplies for Culturing and Testing *Leptocheirus plumulosus* 26
9.1 Recommended Test Conditions for Conducting Reference-toxicity Tests 41
11.1 Test Conditions for Conducting a 28-d Chronic Sediment Toxicity Test with *Leptocheirus plumulosus* 52
11.2 General Activity Schedule for Conducting a 28-d Chronic Sediment Toxicity Test with *Leptocheirus plumulosus* 54
11.3 Test Acceptability Requirements for a 28-d Chronic Sediment Toxicity Test with *Leptocheirus plumulosus* 55

12.1 Suggested alpha Levels to Use for Tests of Assumptions 73

13.1 Intralaboratory Precision Distribution of the Coefficient of Variation for Each Test Endpoint 86

13.2 Ranges of the Black Rock Harbor Sediment Dilution Series Chemical Concentrations 87

13.3 Interlaboratory Precision for Each Endpoint for *Leptocheirus plumulosus* in a 28-d Long-term Sediment Toxicity Test using Black Rock Harbor Spiked Sediments 88

13.4 Summary of Intralaboratory Precision at Five Laboratories for the 28-d *Leptocheirus plumulosus* Chronic Test Using Five Dilutions of Black Rock Harbor Sediment 89

Figures

10.1 *Leptocheirus plumulosus* morphology (A) and characteristics used to determine the gender (B-C) of the amphipod 42

12.1 Treatment response for a Type I and Type II error 66

12.2 Power of the test vs. percentage reduction in treatment response relative to the control mean at various CVs (8 replicates, " = 0.05 [one-tailed]) 68

12.3 Power of test vs. percentage reduction in treatment response relative to the control mean at various CVs (5 replicates, " = 0.05 [one-tailed]) 68

12.4 Power of the test vs. percentage reduction in treatment response relative to the control mean at various CVs (5 replicates, " = 0.10 [one-tailed]) 69

12.5 Effect of CV and number of replicates on the power to detect a 20% decrease in treatment response relative to the control mean (" = 0.05 [one-tailed]) 69

12.6 Effect of alpha and beta on the number of replicates at various CVs 71

12.7 Decision tree for analysis of survival, growth rate, and reproduction data subjected to hypothesis testing 71

12.8 Decision tree for analysis of point estimate data 78

13.1 Control charts: (A) hypothesis testing results; and (B) point estimates (LC, EC, or IC) 85

Appendix A Example Data Forms for Use with the 28-d Chronic Test 104

Section 1
Introduction

1.1 Significance of Use

1.1.1 Sediment provides habitat for many estuarine and marine organisms and is a major repository for many of the more persistent chemicals that are introduced into surface waters. In the aquatic environment, most anthropogenic chemicals and waste materials, including toxic organic and inorganic chemicals, eventually accumulate in sediment. Mounting evidence exists of environmental degradation in areas where U.S. Environmental Protection Agency (USEPA, or the Agency) Water Quality Criteria (WQC; Stephan et al., 1985) are not exceeded, yet organisms in or near sediment are adversely affected (Chapman, 1989). The WQC were developed to protect organisms in the water column and were not intended to address issues to protect organisms in sediment. Concentrations of contaminants in sediment might be several orders of magnitude higher than in the overlying water; however, bulk sediment concentrations have not been strongly correlated to bioavailability (Burton, 1991). Partitioning or sorption of a compound between water and sediment may depend on many factors, including aqueous solubility, pH, redox, affinity for sediment organic carbon and dissolved organic carbon, grain size of the sediment, sediment mineral constituents (oxides of iron, manganese, and aluminum), and the quantity of acid volatile sulfides in sediment (Di Toro et al., 1990; 1991). Although certain chemicals are highly sorbed to sediment, these compounds may still be available to the biota. Contaminated sediment may be directly toxic to aquatic life and can also be a source of contaminants for bioaccumulation in the food chain.

1.1.2 Assessments of sediment quality have commonly included sediment chemical analyses, and surveys of benthic community structure. Determination of sediment chemical concentrations on a dry weight basis alone offers little insight into predicting adverse biological effects because bioavailability may be limited by the intricate partitioning factors mentioned above. Likewise, benthic community surveys may be inadequate, because they sometimes fail to discriminate between effects of contaminants and those that result from unrelated noncontaminant factors, including water quality fluctuations, physical parameters, and biotic interactions. To obtain a direct measure of sediment toxicity, or bioaccumulation, laboratory tests have been developed in which surrogate organisms are exposed to sediments under controlled conditions. Sediment toxicity tests have evolved into effective tools that provide direct, quantifiable evidence of biological consequences of sediment contamination that can only be inferred from chemical or benthic community analyses. To evaluate sediment quality nationwide, USEPA developed the National Sediment Inventory (NSI), which is a compilation of existing sediment quality data and protocols used to evaluate the data. The NSI was used to produce the first biennial report to Congress on sediment quality in the United States as required under the Water Resources Development Act of 1992 (USEPA, 1997a;1997b; 1997c). USEPA's evaluation of the data shows that sediment contamination exists in every region and state of the country and various waters throughout the United States contain sediment that is sufficiently contaminated with toxic pollutants to pose potential risks to fish and to humans and wildlife who eat fish. The use of consistent sediment testing methods described in this manual will provide high-quality data needed for the NSI, future reports to Congress, and regulatory programs to prevent, remediate, and manage contaminated sediments (USEPA, 1998).

1.1.3 The objective of a sediment test is to determine whether contaminants in sediment are harmful to or are bioaccumulated by benthic organisms. The tests can be used to measure interactive toxic effects of complex contaminant mixtures in sediment. Furthermore, knowledge of specific interactions among sediments and test organisms is not necessary in order to conduct the tests (Kemp and Swartz, 1988). However, such knowledge can be useful to interpret toxicity data. Sediment tests can be used to (1) determine the relationship between toxic effects and bioavailability, (2) investigate interactions among contaminants, (3) compare the sensitivities of different organisms, (4) determine spatial and temporal distribution of contamination, (5) evaluate dredged material disposal suitability, (6) measure toxicity as part of product licensing or safety testing or chemical approval, (7) rank areas for cleanup, and (8) develop cleanup goals and estimate the effectiveness of remediation or management practices for marine or estuarine environments.

1.1.4 Most standard whole sediment toxicity tests have been developed to produce a lethality endpoint (survival/mortality) with potential for a sublethal endpoint (reburial) in some species. Methods that measure sublethal effects have not been available or have not been routinely used to evaluate sediment toxicity in marine or estuarine sediments (Scott and Redmond, 1989; Green and Chandler, 1996; Levin et al., 1996; Ciarelli et al., 1998; Meador and Rice, 2001). Most assessments of contaminated sediment rely on short-term lethality tests (e.g., #10 d; USEPA-USACE, 1991; 1998). Short-term lethality tests are useful in identifying "hot spots" of sediment contamination, but might not be sensitive enough to evaluate moderately contaminated areas. However, sediment quality assessments using sublethal responses of benthic organisms, such as effects on growth and reproduction, have been used to successfully evaluate moderately contaminated areas (Ingersoll et al., 1998; Kemble et al., 1994; McGee et al., 1995; Scott, 1989). The 28-d toxicity test with *Leptocheirus plumulosus* has two sublethal endpoints: growth and reproduction. These sublethal endpoints have potential to exhibit a toxic response from chemicals that otherwise might not cause acute effects or significant mortality in a test. Sublethal response to chronic exposure is also valuable for population modeling of contaminant effects. This data can be used for population-level risk assessments of benthic pollutant effects.

1.1.5 Results of toxicity tests on sediments spiked at different concentrations of chemicals can be used to establish cause-and-effect relationships between chemicals and biological responses. Results of toxicity tests with test materials spiked into sediments at different concentrations may be reported in terms of a median lethal concentration(LC50), a median effect concentration (EC50), an inhibition concentration (IC50), or as a no observed effect concentration (NOEC) or lowest observed effect concentration (LOEC). However, spiked sediment might not be representative of contaminated sediment in the field. Mixing time (Stemmer et al., 1990a) and aging (Word et al., 1987; Landrum, 1989; Landrum and Faust, 1992) of spiked sediment can influence bioavailability of contaminants.

1.1.6 Evaluating effect concentrations for chemicals in sediment requires knowledge of factors controlling their bioavailability. Similar concentrations of a chemical in units of mass of chemical per mass of sediment dry weight often exhibit a range in toxicity in different sediments (Di Toro et al., 1990; 1991). Effect concentrations of chemicals in sediment have been correlated to interstitial water concentrations, and effect concentrations in interstitial water are often similar to effect concentrations in water-only exposures. The bioavailability of nonionic organic compounds in sediment is better correlated with the organic carbon normalized concentration. Whatever the route of exposure, these correlations of effect concentrations to interstitial water concentrations indicate that predicted or measured concentrations in interstitial water can be used to quantify the concentration to which an organism is exposed. Therefore, information on partitioning of chemicals between solid and liquid phases of sediment is useful for establishing effect concentrations (Di Toro et al., 1991).

1.1.7 Field surveys can be designed to provide either a qualitative reconnaissance of the distribution of sediment contamination or a quantitative statistical comparison of

contamination among sites. Surveys of sediment toxicity are usually part of more comprehensive analyses of biological, chemical, geological, and hydrographic data. Statistical correlations may be improved and sampling costs may be reduced if subsamples are taken simultaneously for sediment tests, chemical analyses, and benthic community structure.

1.1.8 Table 1.1 lists several approaches the USEPA has considered for the assessment of sediment quality (USEPA, 1992c). These approaches include (1) equilibrium partitioning, (2) tissue residues, (3) interstitial water toxicity, 4) benthic community structure, (5) whole-sediment toxicity and sediment-spiking tests, (6) Sediment Quality Triad, and (7) sediment quality guidelines (see Chapman, 1989; USEPA, 1989a; 1990a; 1990b; 1992b for a critique of these methods). The sediment assessment approaches listed in Table 1.1 can be classified as numeric (e.g., equilibrium partitioning), descriptive (e.g., whole-sediment toxicity tests), or a combination of numeric and descriptive approaches (e.g., Effects Range Median; USEPA, 1992c). Numeric methods can be used to derive chemical-specific equilibrium partitioning sediment guidelines (ESGs) or other sediment quality guidelines (SQGs). Descriptive methods, such as toxicity tests with field-collected sediment, cannot be used alone to develop numerical ESGs or other SQGs for individual chemicals. Although each approach can be used to make site-specific decisions, no single approach can adequately address sediment quality. Overall, an integration of several methods using the weight of evidence is the most desirable approach for assessing the effects of contaminants associated with sediment (Long and Morgan, 1990; MacDonald et al., 1996; Ingersoll et al., 1996; 1997). Hazard evaluations integrating data from laboratory exposures, chemical analyses, and benthic community assessments provide strong complementary evidence of the degree of pollution-induced degradation in aquatic communities (Chapman et al., 1992; 1997; Burton, 1991).

1.2 Program Applicability

1.2.1 The USEPA has authority under a variety of statutes to manage contaminated sediments (Table 1.2 and USEPA, 1990c). USEPA's Contaminated Sediment Management Strategy (USEPA, 1998) establishes the following four goals for contaminated sediments and describes actions that the Agency intends to take to accomplish these goals: (1) to prevent further contamination of sediments that may cause unacceptable ecological or human health risks; (2) when practical, to clean up existing sediment contamination that adversely affects the Nation's waterbodies or their uses, or that causes other significant effects on human health or the environment; (3) to ensure that sediment dredging and the disposal of dredged material continue to be managed in an environmentally sound manner; and (4) to develop and consistently apply methodologies for analyzing contaminated sediments. The Agency plans to employ its pollution prevention and source control programs to address the first goal. To accomplish the second goal, USEPA will consider a range of risk management alternatives to reduce the volume and effects of existing contaminated sediments, including *in-situ* containment and contaminated sediment removal. Finally, the Agency is developing tools for use in pollution prevention, source control, remediation, and dredged material management to meet the collective goals. These tools include national inventories of sediment quality and environmental releases of contaminants, numerical assessment guidelines to evaluate contaminant concentrations, and standardized bioassays to evaluate the bioaccumulation and toxicity potential of sediment samples.

1.2.2 The Clean Water Act (CWA) is the single most important law dealing with environmental quality of surface waters in the United States. Section 101 of the CWA sets forth provisions calling for the restoration an maintenance of the chemical, physical, and biological integrity of the Nation's waters. Federal and state monitoring programs traditionally have focused on evaluating water-column problems caused by point-source dischargers. Findings in the National Sediment Quality Survey, Volume I of the first biennial report to Congress on sediment quality in the United States, indicate that this focus needs to be expanded to include sediment quality impacts (Section 1.1.2 and USEPA, 1997a).

Table 1.1 Sediment Quality Assessment Procedures[1]

Method	Type			Approach
	Numeric	Descriptive	Combination	
Equilibrium Partitioning	X			A sediment quality value for a given contaminant is determined by calculating the sediment concentration of the contaminant that corresponds to an interstitial water concentration equivalent to the USEPA water quality criterion for the contaminant.
Tissue Residues	X			Safe sediment concentrations of specific chemicals are established by determining the sediment chemical concentration that results in acceptable tissue residues.
Interstitial Water Toxicity	X	X	X	Toxicity of interstitial water is quantified and identification evaluation procedures are applied to identify and quantify chemical components responsible for sediment toxicity.
Benthic Community Structure		X		Environmental degradation is measured by evaluating alterations in benthic community structure.
Whole Sediment Toxicity and Sediment Spiking	X	X	X	Test organisms are exposed to sediment that may contain known or unknown quantities of potentially toxic chemicals. At the end of a specified time period, the response of the test organisms is examined in relation to a specified endpoint. Dose-response relationships can be established by exposing test organisms to sediments that have been spiked with known amounts of chemicals or mixtures of chemicals.
Sediment Quality Triad	X	X	X	Sediment chemical contamination, sediment toxicity, and benthic community structure are measured on the same sediment sample. Correspondence between sediment chemistry, toxicity, and field effects is used to determine sediment concentrations that discriminate conditions of minimal, uncertain, and major biological effects.
Sediment Quality Guidelines	X	X	X	The sediment concentration of contaminants associated with toxic responses measured in laboratory exposures or in field assessments (i.e., Apparent Effect Threshold [AET], Effect Range Median [ERM], Probable Effect Level [PEL]).

1) Modified from USEPA (1992c).

1.2.3 The Office of Water (OW), the Office of Prevention, Pesticides, and Toxic Substances (OPPTS), the Office of Solid Waste (OSW), and the Office of Emergency and Remedial Response (OERR) are all committed to the principle of consistent tiered testing described in the Contaminated Sediment Management Strategy (USEPA, 1998). Consistent testing is desirable, because the use of uniform testing procedures is expected to increase data accuracy and precision, facilitate test replication, and increase the comparative value of test results. Each USEPA program will, however, retain the flexibility of deciding whether identified risks would trigger actions.

1.2.4 Several programs use a tiered testing approach. Tiered testing refers to a structured, hierarchical procedure for determining data needs relative to decision-making that consists of a series of tiers, or levels, of investigative intensity. Typically, increasing tiers in a tiered testing framework involve increased information and decreased uncertainty (USEPA, 1998). Each USEPA program office intends to develop guidance for interpreting the tests conducted within the tiered framework and to explain how information within each tier would trigger regulatory action. Depending on statutory and regulatory requirements, the program specific guidance will describe decisions based on a weight of evidence approach, a pass-fail approach, or comparison to a reference site. The following two approaches are currently being used by USEPA: (1) OW-U.S. Army Corps of Engineers (USACE) dredged material testing framework and (2) the OPPTS ecological risk assessment tiered testing framework. USEPA-USACE (1991; 1998) describes the dredged material testing framework, and Smrchek and Zeeman (1998) summarizes the OPPTS testing framework. A tiered testing framework has not yet been chosen for Agency-wide use, but some of the components have been identified to be standardized. These components include toxicity tests, bioaccumulation tests, sediment quality guidelines, and other measurements that may have ecological significance, including benthic community structure evaluation, colonization rate, and *in situ* sediment testing within a mesocosm (USEPA, 1992a).

1.3 Scope and Application

1.3.1 Procedures are described for laboratory testing of an estuarine amphipod to evaluate the sublethal toxicity of contaminants in whole sediments. Sediments can be collected from the field or spiked with compounds in the laboratory. The test species is *L. plumulosus*, an Atlantic coast estuarine species. The toxicity test is conducted for 28 d in 1-L glass chambers containing 175 mL of sediment and about 725 mL of overlying seawater. Four hundred milliliters of overlying water is renewed three times per week, at which time test organisms are fed. Tests are initiated with neonate amphipods that mature and reproduce during the 28-d test period. The endpoints in the 28-d toxicity test are survival, growth rate, and reproduction of amphipods. Survival is calculated as the percentage of newly born (neonate) amphipods at test initiation that survive as adults at test termination. Growth rate is calculated as the mean dry weight gain per day per adult amphipod surviving at test termination. Reproduction is calculated as the number of offspring per surviving adult. See Section 11.4 for discussion on relative sensitivity of sublethal test endpoints. This test is applicable for use with sediment having pore water salinity ranging from 1‰ to 35‰.

1.3.2 This 28-d sediment toxicity test method manual serves as a companion to the marine acute sediment test methods manual (USEPA, 1994d) and the freshwater sediment test methods manual (USEPA, 2000).

1.3.3 Procedures described in this manual are based on method refinements described in DeWitt et al. (1992a; 1997a), Emery et al. (1997), Emery and Moore (1996) and USEPA (2000). This USEPA/USACE manual outlines test methods for evaluating the chronic toxicity of sediment with *L. plumulosus*. Although standard procedures are described in the manual, further investigation of certain issues could aid in the interpretation of test results. Some of these issues include further investigation to evaluate the relative toxicological

Table 1.2 Statutory Use for Sediment Quality Assessment [1]

Law[2]	Area of Use
CERCLA	· Assessment of need for remedial action with contaminated sediments; assessment of degree of cleanup required, disposition of sediments
CWA	· National Pollutant Discharge Elimination System (NPDES) permitting, especially under Best Available Technology (BAT) in water-quality-limited water · Section 403(c) criteria for ocean discharges; mandatory additional requirements to protect marine environment · Section 301(h) waivers for publicly owned treatment works (POTWs) discharging to marine waters · Section 404 permits for dredge and fill activities (administered by the U.S. Army Corps of Engineers [USACE])
FIFRA	· Reviews of uses for new and existing chemicals · Pesticide labeling and registration
MPRSA	· Permits for ocean dumping of dredged material
NEPA	· Preparation of environmental impact statements for projects with surface water discharges
TSCA	· Section 5: Premanufacture notification reviews for new industrial chemicals · Sections 4, 6, and 8: Reviews for existing industrial chemicals
RCRA	· Assessment of suitability (and permitting of) on-land disposal or beneficial use of contaminated sediments considered "hazardous"

[1] Modified from Dickson et al. (1987) and Southerland et al. (1992).

[2] CERCLA Comprehensive Environmental Response, Compensation and Liability Act (Superfund)
CWA Clean Water Act
FIFRA Federal Insecticide, Fungicide, and Rodenticide Act
MPRSA Marine Protection, Resources and Sanctuary Act
NEPA National Environmental Policy Act
TSCA Toxic Substances Control Act
RCRA Resource Conservation and Recovery Act.

sensitivity of the lethal and sublethal endpoints to a wide variety of chemicals spiked in sediment and to mixtures of chemicals in sediments from pollution gradients in the field. Additional research is needed to evaluate the ability of the test's lethal and sublethal endpoints to estimate the responses of populations and communities of benthic invertebrates to contaminated sediments. Research is also needed to link the toxicity test's endpoints to a field-validated population model of *L. plumulosus* that would then generate estimates of population-level responses of the amphipod to test sediments and thereby provide additional ecologically relevant interpretive guidance for the toxicity test.

1.3.4 Additional sediment toxicity research and methods development are now in progress to (1) develop standard sediment bioaccumulation tests (i.e., 28-d exposures with the bivalve *Macoma nasuta*, and the polychaete *Nereis virens*) (Lee et al., 1989), (2) refine sediment spiking procedures, (3) refine sediment dilution procedures, (4) refine sediment Toxicity Identification Evaluation (TIE) procedures, (5) produce additional data on confirmation of responses in laboratory tests with natural populations of benthic organisms (i.e., field validation studies), (6) develop sediment toxicity test methods for additional species (e.g., *Neanthes*), and (7) evaluate relative sensitivity of endpoints measured in 10- and 28-d toxicity tests using marine and estuarine amphipods. This information will be described in future editions of this manual or in other USEPA or USACE manuals.

1.3.5 Altering the procedures described in this manual might affect contaminant bioavailability or organism sensitivity and produce results that are not directly comparable with results of accepted procedures. Comparison of results obtained using modified versions of these procedures might provide useful information concerning new concepts and procedures for conducting sediment tests with aquatic organisms. If tests are conducted with procedures different from those described in this manual, additional tests are required to determine comparability of results.

1.3.6 Where states have developed culturing and testing methods for indigenous species other than *L. plumulosus*, data comparing the sensitivity of the substitute species and *L. plumulosus* must be obtained with sediments or reference toxicants to ensure that the species selected are at least as sensitive and appropriate as the recommended species.

1.3.7 Selection of Test Organisms

1.3.7.1 The choice of a test organism has a major influence on the relevance, success, and interpretation of a test. Test organism selection should be based on both environmental relevance and practical concerns (DeWitt et al., 1989; Swartz, 1989). Ideally, a test organism for use in sediment tests should (1) have a toxicological database demonstrating relative sensitivity to a range of contaminants of interest in sediment; (2) have a database for interlaboratory comparisons of procedures (e.g., round-robin studies); (3) be in direct contact with sediment; (4) be readily available from culture, commercial supplier, or through field collection; (5) be easily maintained in the laboratory; (6) be easily identified; (7) be ecologically or economically important; (8) have a broad geographical distribution, be indigenous (either present or historical) to the site being evaluated, or have a niche similar to organisms of concern (e.g., similar feeding guild or behavior to the indigenous organisms); (9) be tolerant of a broad range of sediment physico-chemical characteristics (e.g., grain size); and (10) be compatible with selected exposure methods and endpoints (ASTM, 2000a). The method should also be (11) peer reviewed (e.g., journal articles, American Society of Testing and Materials [ASTM] guides) and (12) confirmed with responses with natural populations of benthic organisms.

1.3.7.2 The primary criterion used for selecting *L. plumulosus* was that it met the above criteria. Amphipods have been used extensively to test the toxicity of marine, estuarine, and freshwater sediments (Swartz et al., 1985; DeWitt et al, 1989; Scott and Redmond, 1989; DeWitt et al., 1992a; Schlekat et al., 1992; ASTM, 2000a). *L. plumulosus* is an infaunal amphipod intimately associated with sediment, due to its burrowing and sediment ingesting nature. *L. plumulosus* is found in both oligohaline and mesohaline regions of estuaries on the East Coast of the United States

and is tolerant to a wide range of sediment grain size distribution. This species is easily cultured in the laboratory and has a relatively short generation time (i.e., about 24d at 23°C, DeWitt et al. 1992a) that makes this species adaptable to chronic testing (see Section 10). Using a similar set of criteria, *L. plumulosus* was selected as one of four amphipod species recommended for short-term toxicity testing of whole sediments (USEPA, 1994d).

1.3.7.3 An important consideration in the selection of species for test method development is the organism's sensitivity to single chemicals and to complex mixtures. Studies (Schlekat 1995; DeWitt et al., 1992a) have evaluated the sensitivities in acute tests of amphipod species, including *L. plumulosus*, either relative to one another, or to other commonly tested estuarine or marine species. For example, the sensitivity of marine amphipods was compared with that of other species that were used in generating saltwater WQC. Seven amphipod genera, were among the test species used to generate saltwater WQC for 12 chemicals. Acute amphipod toxicity data from 4-d water-only tests for each of the 12 chemicals were compared with data for (1) all other species, (2) other benthic species, and (3) other infaunal species. Amphipods were generally of median sensitivity for each comparison. The average percentile rank of amphipods among all species tested was 57.2%; among all benthic species, 55.5%; and, among all infaunal species, 54.3%. Thus, amphipods are not uniquely sensitive relative to all species, benthic species, or even infaunal species (D. Hansen, USEPA, Narragansett, RI, personal communication). Additional research may be warranted to develop tests using species that are consistently more sensitive than amphipods, thereby offering protection to less sensitive groups.

1.3.7.3.1 Several studies of acute tests (10-d) have compared the sensitivity of *L. plumulosus* to other commonly used amphipod species. *L. plumulosus* was as sensitive as the freshwater amphipod *Hyalella azteca* to an artificially created gradient of sediment contamination when the latter was acclimated to oligohaline salinity (i.e., 6 ‰) (McGee et al., 1993). DeWitt et al. (1992b) compared the sensitivities of *L. plumulosus,* three other amphipod species, two molluscs, and one polychaete to highly contaminated sediment collected from Baltimore Harbor, MD, and serially diluted with clean sediment. *L. plumulosus* was more sensitive than the amphipods *H. azteca* and *Lepidactylus dytiscus* and exhibited sensitivity equal to that of *Eohaustorius estuarius*. A study using dilutions of sediment collected from Black Rock Harbor (BRH), CT, showed that *Ampelisca abdita* demonstrated greater sensitivity than *L. plumulosus* when the latter was tested at 20°C (SAIC, 1993a). However, the same study showed that *L. plumulosus* was more sensitive at 25°C (the test temperature for both the *L. plumulosus* 10- and 28-d toxicity tests) than *A. abdita* at 20°C (SAIC, 1993a).

1.3.7.3.2 The relative sensitivity and precision of 10-d acute toxicity tests with three marine and estuarine amphipod species (*A. abdita, E. haustorius*, and *L. plumulosus*) following USEPA methods (USEPA, 1994d) were evaluated in a round-robin test (Schlekat et al., 1995). All three toxicity tests consistently characterized moderate to highly contaminated sediments as toxic relative to uncontaminated control sediments. In addition, significant concordance was exhibited by all toxicity tests in ranking the toxicity of different sediments. Although there was considerable interlaboratory variability demonstrated in the round-robin, sensitivity of these three toxicity tests was similar enough to produce agreement in the categorization of sediments as toxic or nontoxic.

1.3.7.3.3 Studies have been conducted to evaluate the comparative sensitivity of the 28-d toxicity test and the 10-d toxicity test with *L. plumulosus* (DeWitt et al, 1992a; 1997b; McGee and Fisher 1999). DeWitt et al. (1992a; 1997b) found that in general, the reproductive endpoint of the 28-d test was more sensitive to chemical contaminants than the survival and growth endpoints of either the 10-d or 28-d toxicity tests. Studies conducted by the USACE demonstrated similar sensitivity amoung the lethal and sublethal endpoints of both toxicity tests. In contrast, McGee and Fisher (1999) found the sublethal endpoints less sensitive than the survival endpoint. It is possible that the different conclusions about the relative sensitivities

of the 10- and 28-d *L. plumulosus* tests resulted from either subtle differences in the testing procedures used by DeWitt et al. (1992a; 1997b) and McGee and Fisher (1999), or from response of the amphipods to different chemical contaminants in the test sediments used in the three studies. In any case, the *L. plumulosus* 28-d toxicity test provides valuable information on the impact of contaminated sediments on both lethal and sublethal endpoints, which the 10-d test does not provide.

1.3.7.3.4 Limited comparative data are available for concurrent water-only exposures of different amphipod species in single-chemical tests. Studies that have been conducted generally show that no single amphipod species is consistently the most sensitive. The relative sensitivity of four amphipod species to ammonia was determined in 10-d water-only toxicity tests to aid interpretation of results of tests on sediments in which this toxicant is present (SAIC, 1993c). These tests were static exposures that were generally conducted under conditions (e.g., salinity, photoperiod) similar to those used for standard 10-d sediment tests. Departures from standard conditions included the absence of sediment and a test temperature of 20°C for *L. plumulosus*, rather than 25°C as dictated in the acute method (USEPA, 1994d). Sensitivity to total ammonia increased with increasing pH for all four species. The rank sensitivity was *Rhepoxynius abronius* >*A. abdita* >*E. estuarius* >*L. plumulosus*. In addition, cadmium chloride has been a common reference toxicant for all four species in 4-d exposures. DeWitt et al. (1992a) reports the rank sensitivity to cadmium as *R. abronius* >*A. abdita* >*L. plumulosus* >*E. estuarius* at a common temperature of 15°C and salinity of 28‰. A series of 4-d exposures to cadmium that were conducted at species-specific temperatures and salinity values showed the following rank sensitivity: *A. abdita* >*L. plumulosus* > *R. abronius* > *E. estuarius* (SAIC, 1993a; 1993b; 1993c).

1.3.7.3.5. Ammonia is a naturally occurring compound in marine sediment that results from the degradation of organic debris. Interstitial pore water ammonia concentrations in test sediment can range from <1 mg/L to in excess of 400 mg/L (Word et al., 1997). Some benthic infauna show toxicity to ammonia at concentrations of approximately 20 mg/L (Kohn et al., 1994). Based on water-only and spiked-sediment experiments with ammonia, threshold limits for test initiation and termination have been established for the *L. plumulosus* chronic test. Smaller (younger) individuals are more sensitive to ammonia than larger (older) individuals (DeWitt et al., 1997). Results of a 28-d test indicated that neonates can tolerate very high levels of pore water ammonia (>300 mg/L total ammonia) for short periods of time with no apparent long-term effects (Moore et al., 1997). It is not surprising the L*. plumulosus* has a high tolerance for ammonia given that these amphipods are often found in organic rich sediments in which diagenesis can result in elevated pore water ammonia concentrations. Insensitivity to ammonia by *L. plumulosus* should not be construed as an indicator of the sensitivity of the *L. plumulosus* sediment toxicity test to other chemicals of concern.

1.3.7.4 The sensitivity of an organism is related to route of exposure and biochemical response to contaminants. Sediment-dwelling organisms can receive exposure from three primary sources: interstitial water, sediment particles, and overlying water. Food type, feeding rate, assimilation efficiency, and clearance rate will control the dose of contaminants from sediment. Benthic invertebrates often selectively consume different particle sizes (Harkey et al., 1994) or particles with higher organic carbon concentrations, which may have higher contaminant concentrations. Grazers and other collector-gatherers that feed on aufwuchs, or surface films, and detritus might receive most of their body burden directly from materials attached to sediment or from actual sediment ingestion. In amphipods (Landrum, 1989) and clams (Boese et al., 1990), uptake through the gut can exceed uptake across the gills for certain hydrophobic compounds. Organisms in direct contact with sediment can also accumulate contaminants by direct adsorption to the body wall or by absorption through the integument (Knezovich et al., 1987). Particle type and organic coating may affect uptake of contaminants, such as metals (Schlekat, 1998).

1.3.7.5 Despite the potential complexities in estimating the dose that an organism receives

from sediment, the toxicity and bioaccumulation of many contaminants in sediment such as Kepone®, fluoranthene, organochlorines, and metals have been correlated with either the concentration of these chemicals in interstitial water, or in the case of nonionic organic chemicals, in sediment on an organic-carbon normalized basis (Di Toro et al., 1990; 1991). The relative importance of whole sediment and interstitial water routes of exposure depends on the test organism and the specific contaminant (Knezovich et al., 1987). Because benthic communities contain a diversity of organisms, many combinations of exposure routes may be important. Therefore, behavior and feeding habits of a test organism can influence its ability to accumulate contaminants from sediment and should be considered when selecting test organisms for sediment testing.

1.3.7.6 Although no laboratory information was available at the time of publication of this document, a review of the distribution of *L. plumulosus* in Chesapeake Bay indicates that its distribution is negatively correlated with the degree of sediment contamination (Pfitzenmeyer, 1975; Reinharz, 1981). A field validation study of the 10-d and 28-d *L. plumulosus* tests by McGee et al.(1999) in Baltimore Harbor provides evidence that *L. plumulosus* mortality in 10-d toxicity tests is negatively correlated with population density of indigenous *L. plumulosus*. Protocol used by McGee et al. (1999) for the 28-d *L. plumulosus* test used a different diet than outlined in Section 11.3.6. Field validation studies with the revised 28-d *L. plumulosus* sediment toxicity test have not been conducted. The McGee and Fisher (1999) study was a field validation of the 10-d and 28-d tests; however, the feeding protocols have changed slightly from what was used in that study.

1.4 Performance-based Criteria

1.4.1 USEPA's Environmental Monitoring Management Council (EMMC) recommended the use of performance-based methods in developing chemical analytical standards (Williams, 1993). Performance-based methods were defined by EMMC as a monitoring approach that permits the use of appropriate methods that meet pre-established demonstrated performance standards (Section 9.2).

1.4.2 The key consideration for methods used to obtain test organisms, whether they are field-collected or obtained from culture, is to procure healthy organisms of known quality. A performance-based criteria approach, rather than use of control-based criteria (See Section 3 for definitions), was selected as the preferred method through which individual laboratories should evaluate culture methods or the quality of field-collected organisms. This method was chosen to allow each laboratory to optimize culture methods, determine the quality of field-collected organisms, and minimize effects of test organism health on the reliability and comparability of test results. Performance criteria used to assess the quality of cultured and field-collected amphipods and to determine the acceptability of 28-d sediment toxicity tests are listed in Table 11.3.

Section 2
Summary of Method

2.1 Method Description and Experimental Design

2.1.1 Method Description

This manual describes a laboratory method for determining the sublethal toxicity of contaminated sediment using an estuarine crustacean, the amphipod *Leptocheirus plumulosus*. Test sediments may be collected from estuarine or marine environments, or spiked with compounds in the laboratory. The toxicity test is conducted for 28 d in 1-L chambers containing 175 mL of sediment and about 725 mL of overlying water. Tests are initiated with neonate amphipods that mature and reproduce during the 28-d test period. Four hundred milliliters of overlying water is renewed three times per week, and test organisms are fed after each water renewal. The endpoints are survival, growth rate, and reproduction of test organisms. The use or choice of additional control and reference sediments depends on the nature of the test sediments or the application. This test is applicable for use with sediment from oligohaline to fully marine environments (1 ‰ to 35 ‰).

2.1.2 Experimental Design

The following section is a general summary of experimental design. See Sections 11 and 12 for additional details on actual procedures and data analysis.

2.1.2.1 Control and Reference Sediment

2.1.2.1.1 Sediment tests include a control sediment (sometimes called a negative control). A control sediment is one that is essentially free of contaminants and is used routinely to assess the acceptability of a test; it is not necessarily collected near the site of concern. Any contaminants in control sediment are thought to originate from the global spread of pollutants and do not reflect any substantial input from local or nonpoint sources (ASTM, 2000d). A control sediment provides a measure of test acceptability, evidence of test organism health, and a basis for interpreting data obtained from the test sediments. A reference sediment is typically collected near an area of concern (e.g., a disposal site) and is used to assess sediment conditions exclusive of material(s) of interest. Testing a reference sediment provides a site-specific basis for evaluating toxicity.

2.1.2.1.1.1 In general, the performance of organisms in the negative control(s) is used to judge the acceptability of a test. Either the negative control or reference sediment may be used to evaluate performance in the experimental treatments, depending on the purpose of the study. Any study in which organisms in the negative control do not meet performance criteria must be considered questionable, because it suggests that unknown adverse factors affected the test organisms. Key to avoiding this situation is using only control sediments that have a demonstrated record of performance using the same test procedure. This includes testing of new sediment collections from sources that have previously provided suitable control sediment.

2.1.2.1.1.2 Because of the uncertainties introduced by poor performance in the negative control, such studies should be repeated to insure accurate results. However, the scope or sampling associated with some studies may make it difficult or impossible to repeat a study. Some researchers have reported cases where performance in the negative control is poor, but control performance criteria are met in a reference sediment included in the study design. In these cases, it may be possible to infer that other

samples that show good performance are probably not toxic; however, any samples showing poor performance should not be judged to have shown toxicity, because it is unknown whether the adverse factors that caused poor control performance might have also caused poor performance in the test treatments.

2.1.2.1.2 Natural geomorphological and physico-chemical characteristics, such as sediment texture, may influence the response of test organisms (DeWitt et al., 1988). The physico-chemical characteristics of test sediment should be within the tolerance limits of the test organism. Ideally, the limits of a test organism should be determined in advance; however, controls for factors such as grain size and organic carbon can be evaluated if the recommended limits are approached or exceeded in a test sediment. See Section 10 and Table 11.1 for tolerance limits of *L. plumulosus* for physico-chemical characteristics. If the physico-chemical characteristic(s) of a test sediment approach or exceed the tolerance limits of the test organism, it may be desirable to include an additional control sediment that encompasses those characteristics. The effects of some sediment characteristics (e.g., grain size or total organic carbon [TOC]) on test results may be addressed with regression equations (DeWitt et al., 1988; Ankley et al., 1994a). *L. plumulosus* is relatively insensitive to a wide-range of grain sizes in test sediments (95% sand to 35% clay) (DeWitt et al., 1997a; Emery et al., 1997).

2.1.2.2 The experimental design depends on the purpose of the study. Variables that need to be considered include the number and type of control sediment(s), the source of reference sediment, the number of treatments and replicates, and water quality characteristics.

2.1.2.2.1 The purpose of the study might be to determine a specific endpoint, such as reproduction, and may include a negative control sediment, a positive control or reference toxicant, a solvent control, and several concentrations of sediment spiked with a chemical.

2.1.2.2.2 The purpose of the study might be to determine whether field-collected sediments are toxic, and may include controls, reference sediments, and test sediments. Controls are used to evaluate the acceptability of the test (Table 11.3) and a test might include one or more control sediments. Testing a reference sediment provides a site-specific basis for evaluating toxicity of the test sediments. Comparisons of test sediments to multiple reference or control sediments representative of the physical characteristics of the test sediment (i.e., grain size, organic carbon) may be useful in these evaluations. A summary of field sampling design is presented by Green (1979). See Section 12 for additional guidance on experimental design and statistics.

2.1.2.3 If the purpose of the study is to conduct a reconnaissance field survey to identify contaminated sites for further investigation, the experimental design might include only one sample from each site to allow for maximum spatial coverage. The lack of replication at a site usually precludes statistical comparisons (e.g., analysis of variance [ANOVA]) among sites), but these surveys can be used to identify contaminated sites for further study or may be evaluated using regression techniques (Sokal and Rohlf, 1981; Steel and Torrie, 1980).

2.1.2.4 In other instances, the purpose of the study might be to conduct a quantitative sediment survey of chemical contaminants and toxicity to determine statistically significant differences between effects among control and test sediments from several sites. The number of replicates per site should be based on the need for sensitivity or power (Section 12). In a quantitative survey, replicates (separate samples from different grabs collected at the same site) would need to be taken at each site. Chemical and physical characterizations of each of these grabs would be required for each of these replicates used in sediment testing. Separate subsamples might be used to determine within-sample variability or to compare test procedures (e.g., comparative sensitivity among test organisms), but these subsamples cannot be considered to be true field replicates for statistical comparisons among sites (ASTM, 2000b).

2.1.2.5 Sediments often exhibit high spatial and temporal variability (Stemmer et al., 1990a). Therefore, replicate samples may need to be collected to determine variance in sediment

characteristics. Sediments should be collected with as little disruption as possible; however, subsampling, compositing, or homogenization of sediment samples may be necessary for some experimental designs.

2.1.2.6 Site locations might be distributed along a known pollution gradient, in relation to the boundary of a disposal site, or at sites identified as being contaminated in a reconnaissance survey. Both spatial and temporal comparisons can be made. In predredging studies, a sampling design can be prepared to assess the contamination of samples representative of the project area to be dredged. Such a design should include subsampling of cores taken to the project depth.

2.1.2.7 The primary focus of the physical and experimental test design, and statistical analysis of the data, is the experimental unit. The experimental unit is defined as the smallest physical entity to which treatments can be independently assigned (Steel and Torrie, 1980) and to which air and water exchange between test chambers is kept to a minimum. As the number of test chambers per treatment increases, the number of degrees of freedom and the power of a significance test increase, and therefore, the width of the confidence interval on a point estimate, such as an LC50, decreases (Section 12). Because of factors that might affect test results, all test chambers should be treated as similarly as possible. Treatments should be randomly assigned to individual test chamber locations. Assignment of test organisms to test chambers should be nonbiased.

2.2 Types of Tests

2.2.1 A 28-d toxicity method is outlined for the estuarine amphipod *L. plumulosus*. The manual describes procedures for testing sediments from oligohaline to fully marine environments.

2.3 Test Endpoints

2.3.1 In toxicity tests, the method chosen to evaluate an endpoint has the potential to affect that bioassay's quality and cost. For example, an endpoint measure exhibiting high variance will decrease test power and increase the likelihood of false negative results (Fairweather, 1991).

Typically, endpoint selection for new bioassays is generally guided by methodologies for related bioassays (Gray et al., 1998). Sediment bioassays using macroinvertebrates often incorporate standard survival and growth endpoints (Ingersoll, 1995). Gray et al. (1998) recommend optimal endpoint measures for the *L. plumulosus* bioassay based on four criteria: relevance of each measure to its respective endpoint; signal-to-noise ratio (the ratio between the response to stressor and the normal variation in the response variable); redundancy to other measures of the same endpoint; and cost of labor, training, and equipment. Signal-to-noise ratios are independent of experiment design considerations (i.e., Type I and Type II errors, and sample size) and are positively correlated with power (Gray et al., 1998).

The recommended endpoint measures for this species in 28-d tests are survival, calculated as the percentage of neonates at test initiation that survive as adults at test termination; growth rate, calculated as the mean dry weight gain per day per adult amphipod surviving at test termination; and reproduction, calculated as the number of offspring per surviving adult. Behavior of test organisms (e.g., avoidance of sediment) should be qualitatively observed three times per week during the test, before water renewals.

Section 3
Definitions

3.1 Terms

The following terms were defined in Lee (1980), National Research Council (NRC, 1989), USEPA (1989b), USEPA-USACE (1991), USEPA-USACE (1998), Lee et al. (1994), or ASTM (2000b; 2000c; 2000i).

3.1.1 Technical Terms

3.1.1.1 Clean. Denotes a sediment or water that does not contain concentrations of test materials which cause apparent stress to the test organisms or reduce their survival.

3.1.1.2 Concentration. The ratio of weight or volume of test material(s) to the weight or volume of sediment or water.

3.1.1.3 Contaminated sediment. Sediment containing chemical substances at concentrations that pose a known or suspected threat to environmental or human health.

3.1.1.4 Control sediment. Sediment that is essentially free of contaminants and is used routinely to assess the acceptability of a test. Any contaminants in control sediment may originate from the global spread of pollutants and do not reflect any substantial input from local or nonpoint sources. Comparing test sediments to control sediment(s) is a measure of the toxicity of test sediments beyond inevitable background contamination. Control sediment is also called a **negative control** because no toxic effects are anticipated in this treatment.

3.1.1.5 Effect concentration (EC). The toxicant concentration that would cause an effect in a given percentage of the test population. Identical to LC when the observable adverse effect is death. For example, the EC50 is the concentration of toxicant that would cause a specified effect in 50% of the test population.

3.1.1.6 Inhibition concentration (IC). The toxicant concentration that would cause a given percent reduction in a nonquantal measurement for the test population. For example, the IC25 is the concentration of toxicant that would cause a 25% reduction in growth for the test population, and the IC50 is the concentration of toxicant that would cause a 50% reduction.

3.1.1.7 Interstitial water or pore water. Water occupying space between sediment or soil particles.

3.1.1.8 Lethal concentration (LC). The toxicant concentration that would cause death in a given percentage of the test population. Identical to EC when the observable adverse effect is death. For example, the LC50 is the concentration of toxicant that would cause death in 50% of the test population.

3.1.1.19 Lowest observed effect concentration (LOEC). The lowest concentration of a toxicant to which organisms are exposed in a test that causes an adverse effect on the test organisms (i.e., where a significant difference exists between the value for the observed response and that for the controls).

3.1.1.10 No observed effect concentration (NOEC). The highest concentration of a toxicant to which organisms are exposed in a test that causes no observable adverse effect on the test organisms (i.e., the highest concentration of a toxicant in which the value for the observed response is not statistically significantly different from the controls).

3.1.1.11 Overlying water. The water placed over sediment in a test chamber during a test.

3.1.1.12 Reference sediment. A whole sediment near an area of concern used to assess sediment conditions exclusive of material(s) of interest. The

reference sediment may be used as an indicator of localized sediment conditions exclusive of the specific pollutant input of concern. Such sediment would be collected near the site of concern and would represent the background conditions resulting from any localized pollutant inputs as well as global pollutant input. This is the manner in which reference sediment is used in dredged material evaluations.

3.1.1.13 Reference-toxicity test. A test conducted with reagent-grade reference chemical to assess the sensitivity of the test organisms response to a toxicant challenge. Deviations outside an established normal range may indicate a change in the sensitivity of the test organism population. Reference-toxicity tests are most often performed in the absence of sediment.

3.1.1.14 Sediment. Particulate material that usually lies below water. Formulated particulate material that is intended to lie below water in a test.

3.1.1.15 Spiked sediment. A sediment to which a material has been added for experimental purposes.

3.1.1.16 Whole sediment. Sediment and associated pore water that have had minimal manipulation. The term bulk sediment has been used synonymously with whole sediment.

3.1.2 Grammatical Terms

The words "must," "should," "may," "can," and "might" have very specific meanings in this manual.

3.1.2.1 "Must" is used to express an absolute requirement, that is, to state that a test ought to be designed specifically to satisfy the specified conditions, unless the purpose of the test requires a different design. "Must" is only used in connection with the factors that directly relate to the acceptability of a test.

3.1.2.2 "Should" is used to state that the specified condition is recommended and ought to be met if possible. Although a violation of one "should" is rarely a serious matter, violation of several will often render the results questionable.

3.1.2.3 Terms such as "is desirable," "is often desirable," and "might be desirable" are used in connection with less important factors.

3.1.2.4 "May" is used to mean "is (are) allowed to," "can" is used to mean "is (are) able to," and "might" is used to mean "could possibly." Thus, the classic distinction between "may" and "can" is preserved, and "might" is never used as a synonym for either "may" or "can."

Section 4
Interferences

4.1 General Introduction

4.1.1 Interferences are characteristics of a sediment or sediment test system, aside from those related to sediment associated chemicals of concern, that can potentially affect test organism survival, growth, or reproduction (Environment Canada, 1994; ASTM, 2000c; USEPA, 2001). These interferences can potentially confound test interpretation in two ways: (1) false-positive response (i.e., toxicity is observed in the test when contamination is not present at concentrations known to elicit a response, or there is more toxicity than expected); and (2) false-negative response (i.e., no toxicity is observed when contaminants are present at concentrations known to elicit a response, or there is less toxicity than expected).

4.1.2 There are three categories of interfering factors that can cause false-negative or false-positive responses: (1) those physical or chemical characteristics of sediments affecting survival, growth, or reproduction, independent of chemical concentration (e.g., sediment grain size), (2) changes in chemical bioavailability as a function of sediment manipulation or storage, and (3) the presence of indigenous organisms. Although test procedures and test organism selection criteria were developed to minimize these interferences, this section describes the nature of these interferences. Procedures for minimizing the effects of interfering factors are presented in Section 11.

4.1.3 Because of the heterogeneity of natural sediments, extrapolation from laboratory studies to the field can sometimes be difficult (Table 4.1; Burton, 1991). Sediment collection, handling, and storage procedures may alter contaminant bioavailability and concentration by changing the physical, chemical, or biological characteristics of the sediment. Maintaining the integrity of a field-collected sediment during removal, transport, storage, mixing, and testing is extremely difficult and may complicate the interpretation of effects (Environment Canada, 1994; USEPA, 2001). Direct comparisons of organisms exposed

Table 4.1 Advantages and Disadvantages of Use of Sediment Tests[1]

Advantages
- Sediment tests measure bioavailable fraction of contaminant(s).
- Sediment tests provide a direct measure of benthic effects, assuming no field adaptation or amelioration of effects.
- Limited special equipment is required for testing.
- Ten-day toxicity test methods are rapid and inexpensive.
- Legal and scientific precedence exists for use; ASTM standard guides are available.
- Sediment tests measure unique information relative to chemical analyses or benthic community analyses.
- Tests with spiked chemicals provide data on cause-effect relationships.
- Sediment toxicity tests can be applied to all chemicals of concern.
- Tests applied to field samples reflect cumulative effects of contaminants and contaminant interactions.
- Toxicity tests are amenable to confirmation with natural benthos populations.

Disadvantages
- Sediment collection, handling, and storage may alter bioavailability.
- Spiked sediment may not be representative of field contaminated sediment.
- Natural geochemical characteristics of sediment may affect the response of test organisms.
- Indigenous animals may be present in field-collected sediments.
- Route of exposure may be uncertain and data generated in sediment toxicity tests may be difficult to interpret if factors controlling the bioavailability of contaminants in sediment are unknown.
- Tests applied to field samples may not discriminate effects of individual chemicals.
- Few comparisons have been made of methods or species.
- Only a few chronic methods for measuring sublethal effects have been developed or extensively evaluated.
- Laboratory tests have inherent limitations in predicting ecological effects.

[1]Modified from Swartz (1989).

in the laboratory and in the field would be useful to verify laboratory results. However, spiked sediment may not be representative of contaminated sediment in the field. Mixing time (Stemmer et al., 1990a), aging (Word et al., 1987; Landrum, 1989; Landrum and Faust, 1992) and the chemical form of the material can affect responses of test organisms in spiked sediment tests. Detailed recommendations for sample collection and handling are provided in Section 8 and USEPA (2001).

4.1.4 Laboratory testing with field-collected sediments may be useful in estimating cumulative effects and interactions of multiple contaminants in a sample. Tests with field samples usually cannot discriminate between effects of individual chemicals. Most sediment samples contain a complex matrix of inorganic and organic contaminants with many unidentified compounds. The use of TIE in conjunction with sediment tests with spiked chemicals may provide evidence of causal relationships and can be applied to many chemicals of concern (Ankley and Thomas, 1992; Adams et al., 1985; USEPA, 1996). Sediment spiking can also be used to investigate additive, antagonistic, or synergistic effects of specific contaminant mixtures in a sediment sample (Swartz et al., 1988).

4.1.5 Most assessments of contaminated sediment rely on short-term lethality testing methods (e.g., #10 d; USEPA-USACE, 1977; 1991). Short-term lethality tests are useful in identifying "hot spots" of sediment contamination, but may not be sensitive enough to evaluate moderately contaminated areas. Sediment quality assessments using sublethal responses of benthic organisms, such as effects on growth and reproduction, have been used to successfully evaluate moderately contaminated areas (Ingersoll et al., 1998; McGee and Fisher, 1999; Scott et al., 1996).

4.1.6 Despite the interferences discussed in this section, existing sediment testing methods can be used to provide a rapid and direct measure of effects of contaminants on benthic communities (e.g., Canfield et al., 1996; Niewolny et al. 1997; DeWitt et al. 1997c). Laboratory tests with field-collected sediment can also be used to determine temporal, horizontal, or vertical distribution of contaminants in sediment. Most tests can be completed within 2 to 4 weeks. Legal and scientific precedents exist for use of toxicity and bioaccumulation tests in regulatory decision-making (e.g., USEPA, 1990c). Furthermore, sediment tests with complex contaminant mixtures are important tools for making decisions about the extent of remedial action for contaminated aquatic sites and for evaluating the success of remediation activities.

4.2 Noncontaminant Factors

4.2.1 Noncontaminant characteristics of sediment are defined as chemical or physical characteristics that can cause reduced test organism survival, growth, or reproduction. These interferences include, but are not limited to, sediment grain size, interstitial pore water salinity, TOC, dissolved sulfides, and interstitial pore water ammonia. *L. plumulosus* is considered to be remarkably tolerant of these noncontaminant factors; however, the physico-chemical properties of each test sediment must be within the acceptable tolerance limits to ensure that a toxicological response is caused by contaminants. Tolerance limits of *L. plumulosus* for the factors listed above are summarized in Table 11.1 and defined below.

4.2.1 Sediment Grain Size

4.2.1.1 *L. plumulosus* are found in very fine muds and muddy sands and are tolerant of variable grain size. Laboratory studies have shown significant reduction in survival when clay content exceeded 84% (Emery et al., 1997). Emery et al. (1997) found an increase in growth as sediment coarseness increased up to 75% sand. However, DeWitt et al. (1997a) reported enhanced growth in finer-grained sediment as compared with more coarse-grained material, but the difference in growth was not considered to be biologically important (DeWitt et al., 1997a). *L. plumulosus* survival, growth and reproduction were significantly reduced when exposed to pure sand (Emery et al., 1997). Therefore, *L. plumulosus* should be tested with sediment with ≥ 5% silt-clay (i.e., #95% sand), but <85% clay (Table 11.1). If sediment characteristics exceed these bounds, an

appropriate clean control/reference sediment should be incorporated into the test to separate effects of sediment-associated contaminants from effects of particle size.

4.2.2 Interstitial Pore Water Salinity

4.2.2.1 *L. plumulosus* is an estuarine species tolerant of a wide range of salinity. No adverse effects have been observed in laboratory exposures to pore water salinity values ranging from 0‰ to 35‰, with overlying water salinity at 20 ‰ (DeWitt et al., 1997a). Furthermore, laboratory cultures have been successfully maintained at 5‰ (Emery and Moore, 1996) and 20‰ (DeWitt et al., 1997a) which demonstrates that reproduction had occurred in the cultures at a variety of salinity values. Although not currently recommended for testing in truly freshwater sediment, *L. plumulosus* can be used to test sediment having pore water salinity >1‰ (DeWitt et al., 1997a). Further research is required to determine whether *L. plumulosus* can be used to test sediment having pore water salinity >35‰.

4.2.3 Total Organic Carbon

4.2.3.1 Test sediment TOC content can vary greatly, ranging from near 0% to >10%. The amount of TOC can affect test organism survival, growth, and reproduction. Limited evidence suggests that the *L. plumulosus* chronic test is tolerant to most TOC concentrations; however, Scott et al. (1996) reported that growth and reproduction may be lower in uncontaminated field sediments having <2% TOC concentrations. An analysis of organism response over a wide range of sediment TOC was completed by DeWitt et al. (1997b) using reference sediment data from two studies. No effect on survival, growth, or reproduction was detected for sediments with TOC concentrations ranging from 1% to 7% TOC. There was some evidence of significantly decreased survival, growth, and reproduction in <1% TOC sediments. No data were available for test sediments with TOC >7%. Therefore until additional data are generated, if test sediment TOC concentrations are <1% or >7%, a TOC control or reference sediment with similar TOC should be tested concurrently.

4.2.4 Dissolved Sulfides

4.2.4.1 Hydrogen sulfide occurs naturally in anoxic marine sediments. Sims and Moore (1995) conducted an extensive review of the literature that focused on the effects of hydrogen sulfide on benthic organisms. Sims and Moore (1995) reported that tube-building amphipods circulate oxygenated water through their burrows, thus reducing or eliminating exposure to pore water hydrogen sulfide. In acute experiments, however, dissolved sulfides have been shown to be toxic to marine amphipods *R. abronius* and *E. estuarius* (48-h LOECs of 1.47 and 1.92 mg/L total sulfide respectively; Knezovich et al., 1992). Currently, no data exist regarding the sensitivity of *L. plumulosus* to hydrogen sulfide in 28-d exposures. Additional information on the tolerance of aquatic organisms to sulfides can be found in Bagarinao (1992).

4.2.5 Interstitial Pore Water Ammonia

4.2.5.1 Ammonia is present in sediment as a result of several independent microbial processes as well as anthropogenic sources, and ammonia concentrations may be enhanced in areas that exhibit organic enrichment. Ammonia concentrations are sometimes high in contaminated sediments. Interstitial pore water ammonia concentrations in test sediment can range from <1 mg/L to in excess of 400 mg/L (Word et al., 1997). Some benthic infauna show toxicity to ammonia at concentrations of approximately 20 mg/L (Kohn et al., 1994). Based on water-only and spiked-sediment experiments with ammonia, threshold limits for test initiation and termination have been established for the *L. plumulosus* chronic test. Smaller (younger) individuals are more sensitive to ammonia than larger (older) individuals (DeWitt et al., 1997a). Results of a 28-d test indicated that neonates can tolerate very high levels of pore water ammonia (>300 mg/L total ammonia) for short periods of time with no apparent long-term effects (Moore et al., 1997). At test initiation, pore water should not exceed 60 mg/L total ammonia (Table 11.1; DeWitt et al., 1997a; USEPA, 1994d). One study indicated that pore water ammonia levels ≥16 mg/L measured at test termination can be

associated with lethal and sublethal impacts to *L. plumulosus* (DeWitt et al. 1997a). Thus, if pore water ammonia concentrations exceed 16 mg/L at test termination, toxicity test results could be affected by ammonia.

4.2.7 If a particular sediment characteristic exceeds the tolerance of *L. plumulosus*, several measures can be taken. Suggested procedures to account for or reduce the effects of noncontaminant interferences are presented in Section 11.4.

4.3 Changes in Contaminant Bioavailability

4.3.1 Sediment toxicity tests are meant to serve as an indicator of contaminant-related toxicity that might be expected under field or natural conditions. Some studies have indicated differences between results of laboratory testing and results of field testing of sediments using *in situ* exposures (Sasson-Brickson and Burton, 1991).

4.3.2 Sediment collection, handling, and storage may alter contaminant bioavailability and concentration by changing the physical, chemical, or biological characteristics of the sediment. Manipulations such as mixing, homogenization, and sieving are generally thought to increase availability of organic compounds because of disruption of the equilibrium with organic carbon in the pore water/particle system. Similarly, oxidation of anaerobic sediment increases the availability of certain metals (Di Toro et al., 1990). Because the availability of contaminants can be a function of the degree of manipulation, this manual recommends that handling, storage, and preparation of the sediment for testing be as consistent as possible. Maintaining the integrity of a field-collected sediment during removal, transport, storage, mixing, and testing is extremely difficult. Direct comparison of organisms exposed in the laboratory and in the field would be useful to verify laboratory results. Detailed recommendations for sample collection and handling are provided in Section 8 and USEPA (2001).

4.3.2.1 Sediment Sampling. Sediment collection techniques include moderately disruptive (sediment coring and grab sampling) to highly disruptive (dredging) methods. It is impossible to collect sediment samples and remove them from samplers without altering conditions to some degree that control contaminant availability (e.g., redox potential, anaerobic environment, spatial distributions, biological activity). Oxidation, compaction, volatilization, homogenization, and exposure to light can all occur and affect contaminant distribution, speciation, partitioning, and ultimately bioavailability. It is important to select sampling techniques that not only achieve study goals, but also minimize sediment disturbance.

4.3.2.2 Sediment Storage. Sediment storage conditions can also affect contaminant availability and speciation. Type of storage container, storage time, temperature, exposure to air, and drying need to be controlled to maintain sample integrity (USEPA, 2001). It is generally recommended that sediment should be stored at 4°C, in the dark, in sealed containers with minimal headspace.

4.3.2.3 Sieving and Homogenization. Test sediments should be sieved only when there is compelling concern that indigenous predator or amphipods from the test site could accidentally be introduced into the test chamber. However, because sieving of test sediments disrupts the physical properties of the sediment, and may also affect chemical properties of the sediment, sieving should be avoided whenever possible. Press-sieving is preferable to wet-sieving because the use of water in the latter processing will dilute the pore water (and its chemical constituents) of the test sediment. To press sieve use a clean inert surface such as teflon, to help push sediment through either a nytex or stainless steel sieve (depending on project requirements). When sediments are sieved, it may be desirable to take samples before and after sieving to compare the concentration of contaminants (especially in the pore water), total organic carbon, dissolved organic carbon (in pore water), acid volatile sulfides (AVS), and sediment grain-size distribution. USEPA does not recommend unnecessary sieving of test sediments on a routine basis (see USEPA 1997d, 2000; ASTM 2000c).

4.3.2.4 Testing Conditions. Conducting sediment toxicity tests at temperatures different from those at the collection site might affect

contaminant solubility, partitioning coefficients, and other physical and chemical characteristics. Interaction between sediment and overlying water and the ratio of sediment to overlying water can influence bioavailability (Stemmer et al., 1990b). Salinity of the overlying water is another factor that can affect the bioavailability of contaminant, particularly metals. Some metals (e.g., cadmium) are more bioavailable at lower salinity values. Therefore, if a sediment sample from a low salinity location is tested with overlying waters of high salinity, there is the potential that metal toxicity may be reduced. The broad tolerance of *L. plumulosus* allows tests to be conducted over the range of pore water salinity values routinely encountered in field-collected sediments from North American estuarine and marine environments. For standardization purposes, testing should be conducted with overlying water at either 5‰ or 20‰. Sediment samples with pore water salinity values #10‰ should be tested with overlying water at 5‰, and test sediment with pore water salinity values >10‰ should be tested with overlying water at 20‰ (DeWitt et al., 1997a). Photoinduced toxicity caused by ultraviolet (UV) light may be important for some compounds associated with sediment (e.g., polycyclic aromatic hydrocarbons [PAHs]); Davenport and Spacie, 1991; Ankley et al., 1994b). However, fluorescent lighting typically used to conduct laboratory tests does not include the appropriate spectrum of UV radiation to photoactivate compounds (Oris and Giesy, 1985). Therefore, laboratory tests might not account for toxicity expressed by this mode of action.

4.3.2.5 Additions to Test Chambers*.* The addition of food, water, or solvents to the test chambers might obscure the bioavailability of contaminants in sediment or might provide a substrate for bacterial or fungal growth (Harkey et al., 1997). Without addition of food, the test organisms may starve during exposures (Ankley et al., 1994a; DeWitt et al., 1997a). However, the addition of food may alter the availability of the contaminants in the sediment (Harkey et al., 1994; Bridges et al., 1997) depending on the amount of food added, its composition (e.g., TOC), and the chemical(s) of interest.

4.3.2.6 Contaminant Uptake*.* Depletion of aqueous and sediment-sorbed contaminants resulting from uptake by an organism or absorption to test chamber can also influence contaminant availability. In most cases, the organism is a minor sink for contaminants relative to the sediment. However, within the burrow of an organism, sediment desorption kinetics might limit uptake rates. Within minutes to hours, a major portion of the total chemical can be inaccessible to the organisms because of depletion of available residues. The desorption of a particular compound from sediment can range from easily reversible (labile; within minutes) to irreversible (nonlabile; within days or months; Karickhoff and Morris, 1985). Interparticle diffusion or advection and the quality and quantity of sediment organic carbon can also affect sorption kinetics.

4.3.3 The route of exposure may be uncertain and data from sediment tests may be difficult to interpret if factors controlling the bioavailability of contaminants in sediment are unknown. Bulk-sediment chemical concentrations may be normalized to factors other than dry weight. For example, concentrations of nonionic organic compounds might be normalized to sediment organic-carbon content (USEPA, 1992c) and certain metals normalized to acid volatile sulfides (Di Toro et al., 1990). Even with the appropriate normalizing factors, determination of toxic effects from ingestion of sediment or from dissolved chemicals in the interstitial water can still be difficult (Lamberson and Swartz, 1988).

4.4 Presence of Indigenous Organisms

4.4.1 Indigenous organisms may be present in field-collected sediments. An abundance of the same organism or organisms taxonomically similar to the test organisms in the sediment sample may make interpretation of treatment effects difficult. Competing or predatory organisms can adversely affect *L. plumulosus* survival, growth, or reproduction.

4.4.2 If compelling evidence exists that indigenous organisms may be introduced into the test chamber with the test sediments, the test sediments can be sieved (see Section 4.3.2.3).

However, USEPA does not recommend unnecessary sieving of test sediments on a routine basis. Alternatively, short-term storage of test sediments may eliminate indigenous organisms in the test sediments (see Section 8.2.2).

Section 5
Health, Safety, and Waste Management

5.1 General Precautions

5.1.1 Development and maintenance of an effective health and safety program in the laboratory requires an ongoing commitment by laboratory management and includes (1) the appointment of a laboratory health and safety officer with the responsibility and authority to develop and maintain a safety program; (2) the preparation of a formal, written health and safety plan, which is provided to each laboratory staff member; (3) an ongoing training program on laboratory safety; and (4) regular safety inspections.

5.1.2 This manual addresses procedures that may involve hazardous materials, operations, and equipment, but it does not purport to address all of the safety problems associated with their use. It is the responsibility of the user to establish appropriate safety and health practices and determine the applicability of regulatory limitations before use. While some safety considerations are included in the manual, it is beyond the scope of this manual to encompass all safety requirements necessary to conduct sediment tests.

5.1.3 Collection and use of sediment may involve substantial risks to personal safety and health. Contaminants in field-collected sediment may include carcinogens, mutagens, and other potentially toxic compounds. Inasmuch as sediment testing is often begun before chemical analyses can be completed, worker contact with sediment needs to be minimized by (1) using gloves, laboratory coats, safety glasses, face shields, and respirators as appropriate; (2) manipulating sediment under a ventilated hood, in an enclosed glove box; and (3) enclosing and ventilating the exposure system. Personnel collecting sediment samples and conducting tests should take all safety precautions necessary for the prevention of bodily injury and illness that might result from ingestion or invasion of infectious agents, inhalation or absorption of corrosive or toxic substances through skin contact, and asphyxiation because of lack of oxygen or presence of noxious gases.

5.1.4 Before beginning sample collection and laboratory work, personnel should determine that all required safety equipment and materials have been obtained and are in good condition.

5.2 Safety Equipment

5.2.1 Personal Safety Gear

5.2.1.1 Personnel should use appropriate safety equipment, such as rubber aprons, laboratory coats, respirators, gloves, safety glasses, face shields, hard hats, and safety shoes.

5.2.2 Laboratory Safety Equipment

5.2.2.1 Each laboratory should be provided with safety equipment such as first aid kits, fire extinguishers, fire blankets, emergency showers, and eye wash stations.

5.2.2.2 All laboratories should be equipped with a telephone to enable personnel to summon help in case of emergency.

5.3 General Laboratory and Field Operations

5.3.1 Laboratory personnel should be trained in proper practices for handling and using chemicals that are encountered during procedures described in this manual. Routinely encountered chemicals include acids, organic solvents, and standard

materials for reference-toxicity tests. Special handling and precautionary guidance in Material Safety Data Sheets should be followed for reagents and other chemicals purchased from supply houses.

5.3.2 Work with some sediment might require compliance with rules pertaining to the handling of hazardous materials. Personnel collecting samples and performing tests should not work alone.

5.3.3 It is advisable to wash exposed parts of the body with bactericidal soap and water immediately after collecting or manipulating sediment samples.

5.3.4 Strong acids and volatile organic solvents should be used in a fume hood or under an exhaust canopy over the work area.

5.3.5 An acidic solution should not be mixed with a hypochlorite solution because hazardous vapors might be produced.

5.3.6 To prepare dilute acid solutions, concentrated acid should be added to water, not vice versa. Opening a bottle of concentrated acid and adding concentrated acid to water should be performed only under a fume hood.

5.3.7 Use of ground-fault systems and leak detectors is strongly recommended to help prevent electrical shocks. Electrical equipment or extension cords not bearing the approval of Underwriter Laboratories should not be used. Ground-fault interrupters should be installed in all "wet" laboratories where electrical equipment is used.

5.3.8 All containers should be adequately labeled to identify their contents.

5.3.9 Good housekeeping contributes to safety and reliable test results.

5.4 Disease Prevention

5.4.1 Personnel handling samples that are known or suspected to contain human wastes should be given the opportunity to be immunized against hepatitis B, tetanus, typhoid fever, and polio. Thorough washing of exposed skin with bactericidal soap should follow handling these samples.

5.5 Safety Manuals

5.5.1 For further guidance on safe practices when handling sediment samples and conducting toxicity tests, check with the permittee and consult general industrial safety manuals including USEPA (1986b) and Walters and Jameson (1984).

5.6 Pollution Prevention, Waste Management, and Sample Disposal

5.6.1 It is the laboratory's responsibility to comply with the federal, state, and local regulations governing the waste management, particularly hazardous waste identification rules and land disposal restrictions, and to protect the air, water, and land by minimizing and controlling all releases from fume hoods and bench operations. Also, compliance is required with any sewage discharge permits and regulations. For further information on waste management, consult *The Waste Management Manual for Laboratory Personnel*, available from the American Chemical Society's Department of Government Relations and Science Policy, 1155 16th Street N.W., Washington, D.C. 20036.

5.6.2 Guidelines for the handling and disposal of hazardous materials should be strictly followed. The federal government has published regulations for the management of hazardous waste and has given the states the option of either adopting those regulations or developing their own. If states develop their own regulations, they are required to be at least as stringent as the federal regulations. As a handler of hazardous materials, it is a laboratory's responsibility to know and comply with the applicable state regulations. Refer to The Bureau of National Affairs, Inc. (1986) for the citations of the federal requirements.

5.6.3 Substitution of nonhazardous chemicals and reagents should be encouraged and investigated whenever possible. For example, use of a nonhazardous compound for a positive control in reference-toxicity tests is advisable. Reference-toxicity tests with copper can provide appropriate toxicity at concentrations below regulated levels.

Section 6
Facilities, Equipment, and Supplies

6.1 General

6.1.1 Before a sediment test is conducted in any test facility, it is desirable to conduct a "nontoxicant" test with each test species in which all test chambers contain a control sediment (sometimes called the negative control) and clean overlying water. Survival, growth, or reproduction of the test organisms will demonstrate whether facilities, water, control sediment, and handling techniques are adequate to result in acceptable species-specific control numbers. Evaluations may also be made on the magnitude of between-chamber variance in a test. See Section 9.14.

6.2 Facilities

6.2.1 The facility must include separate areas for culturing test organisms and sediment testing to reduce the possibility of contamination by test materials and other substances, especially volatile compounds. Holding, acclimation, and culture chambers should not be in a room where sediment tests are conducted, stock solutions or sediments are prepared, or equipment is cleaned. Test chambers may be placed in a temperature-controlled recirculating water bath, environmental chamber, or equivalent facility with temperature control. An enclosed test system is desirable to provide ventilation during tests to limit exposure of laboratory personnel to volatile substances.

6.2.2 Light of the quality and illuminance normally obtained in the laboratory (about 500 to 1000 lux using wide-spectrum fluorescent lights; e.g., cool-white or daylight) is adequate to culture *L. plumulosus* and to conduct the chronic toxicity test. Lux is the unit selected for measuring luminance in this manual, and should be measured at the surface of the water in test or culture chambers. A uniform photoperiod of 16 h light and 8 h dark shall be maintained for cultures and during the tests, and can be achieved in the laboratory or in an environmental chamber using automatic timers.

6.2.3 During phases of rearing, holding, and testing, test organisms should be shielded from external disturbances such as rapidly changing light or pedestrian traffic.

6.2.4 The test facility should be well ventilated and free of fumes. Laboratory ventilation systems should be checked to ensure that return air from chemistry laboratories or sample handling areas is not circulated to culture or testing rooms, or that air from testing rooms does not contaminate culture rooms. Air pressure differentials between rooms should not result in a net flow of potentially contaminated air to sensitive areas through open or loose-fitting doors. Air used for aeration must be free of oil and fumes. Oil-free air pumps should be used where possible. Filters to remove oil, water, and bacteria are desirable. Particles can be removed from the air using filters such as BALSTON® Grade BX (Balston, Inc., Lexington, MA) or equivalent, and oil and other organic vapors can be removed using activated carbon filters (e.g., BALSTON® C-1 filter), or equivalent.

6.3 Equipment and Supplies

6.3.1 Equipment and supplies that contact stock solutions, sediment, or overlying water should not contain substances that can be leached or dissolved in amounts that adversely affect the test organisms. In addition, equipment and supplies that contact sediment or water should be chosen to minimize sorption of test materials from water. Glass, type 316 stainless steel, nylon, and high-density polyethylene, polypropylene, polycarbonate, and fluorocarbon plastics should be used whenever possible to minimize leaching, dissolution, and sorption. Concrete and high-density plastic containers may be used for holding

and culture chambers, and in the water-supply system. These materials should be washed in detergent, acid-rinsed, and soaked in flowing water for a week or more before use. Cast-iron pipe should not be used in water-supply systems because colloidal iron will be added to the overlying water and strainers will be needed to remove rust particles. Copper, brass, lead, galvanized metal, and natural rubber must not contact overlying water or stock solutions before or during a test. Items made of neoprene rubber and other materials not mentioned above should not be used unless it has been shown that their use will not adversely affect survival, growth, or reproduction of the test organisms.

6.3.2 New lots of plastic products should be tested for toxicity by exposing organisms to them under ordinary test conditions before general use.

6.3.3 General Equipment

6.3.3.1 Environmental chamber or equivalent facility with photoperiod and temperature control (20°C to 25°C).

6.3.3.2 Water purification system capable of producing at least 1 mega-ohm water (USEPA, 1991a).

6.3.3.3 Analytical balance capable of accurately weighing to 0.01 mg.

6.3.3.4 Reference weights, Class S—for documenting the performance of the analytical balance(s). The balance(s) should be checked with reference weights that are at the upper and lower ends of the range of the weighings made when the balance is used. A balance should be checked at the beginning of each series of weighings, periodically (such as every tenth weight) during a long series of weighings, and after taking the last weight of a series.

6.3.3.5 Volumetric flasks and graduated cylinders—Class A, borosilicate glass or nontoxic plastic labware, 10 to 1000 mL for making test solutions.

6.3.3.6 Volumetric pipets—Class A, 1 to 100 mL.

6.3.3.7 Serological pipets—1 to 10 mL, graduated.

6.3.3.8 Pipet bulbs and fillers.

6.3.3.9 Droppers, and glass tubing with fire polished edges, 4- to 6-mm ID—for transferring test organisms.

6.3.3.10 Wash bottles—for rinsing small glassware, instrument electrodes, and probes.

6.3.3.11 Electronic (digital) thermometers—for measuring water temperature. Mercury-filled glass thermometers should not be used.

6.3.3.12 National Bureau of Standards Certified thermometer (see USEPA Method 170.1; USEPA, 1979b).

6.3.3.13 Dissolved oxygen (DO), pH, and salinity meters for routine physical and chemical measurements (portable field-grade instruments are acceptable unless a test is conducted to specifically measure the effects of one of these measurements). A temperature-compensated salinity refractometer is useful for measuring salinity of water overlying field-collected sediment.

6.3.3.14 Ammonia-specific probe with a functional range between 1 and >100 mg/L total ammonia.

6.3.3.15 Table 6.1 lists additional equipment and supplies.

6.3.4 Test Chambers

6.3.4.1 Test chambers may be constructed in several ways and of various materials, depending on the experimental design and the contaminants of interest. Clear silicone adhesives, suitable for aquaria, sorb some organic compounds that might be difficult to remove. Therefore, as little adhesive as possible should be in contact with the test material. Extra beads of adhesive should be on the outside of the test chambers rather than on the inside. To leach potentially toxic compounds from the adhesive, all new test chambers constructed using silicone adhesive should be held at least 48 h in overlying water before use in a test.

Table 6.1 Equipment and Supplies for Culturing and Testing *L. plumulosus*

Biological Supplies

- Brood stock of test organisms
- TetraMin®
- Live microalgae (e.g., *Pseudoisochrysis paradoxa*, *Phaeodactylum tricornutum*, *Isochrysis galbana*, *Chaetoceros calcitrans*, *Skeletonema* sp., or *Thallasiosirus* spp. [optional food items for culturing *L. plumulosus*])

Containers and Glassware

- Culture chambers (e.g., 35-cm x 30-cm x 15-cm plastic wash bin)
- Test chambers (1-L glass jar or beaker)
- Glass bowls
- Wide-bore pipets, droppers, or glass tubing (4- to 6-mm ID) for organism transfer
- Glass disposable serological pipets or digital equivalent
- Graduated cylinders (assorted sizes, 10 mL to 2 L)

Instruments and Equipment

- Dissecting microscope
- Stainless-steel (for culture or contaminated sediment) or Nytex (for culture sediment only) sieves (U.S. Standard No. 18, 35, and 60 mesh or 1.0, 0.5 or 0.6, and 0.25 mm)
- Photoperiod timer
- Light meter
- Environmental chamber, water bath, or equivalent with photoperiod and temperature control
- Thermometer, electronic (digital)
- Continuous recording thermometer
- Dissolved oxygen meter
- pH meter
- Meter with ion-specific ammonia electrode (or functional equivalent)
- Salinity meter or temperature compensating salinity refractometer
- Drying oven
- Desiccator
- Balance (0.01 mg sensitivity)
- Refrigerator
- Freezer
- Light box
- Hemacytometer (optional)
- Mortar and pestle, blender, grain mill, coffee grinder
- Pump for water exchanges

Miscellaneous

- Ventilation system for test chambers
- Ventilation system for counts of alcohol-preserved samples
- Air supply and air stones/pipets (oil free and regulated)
- Weighing pans
- Fluorescent light bulbs
- Deionized water
- Air line tubing
- Plastic dish pan
- Sieve cups

Chemicals

- Detergent (nonphosphate)
- Acetone (reagent grade)
- Hydrochloric or nitric acid (reagent grade)
- Reagents for preparing synthetic seawater (reagent grade $CaCl_2 \bullet 2\ H_2O$, KBr, KCl, $MgCl_2 \bullet 6\ H_2O$, $Na_2B_4O_7 \bullet 10\ H_2O$, NaCl, Na HCO_3, Na_2SO_4, $SrCl_2 \bullet 6\ H_2O$ [optional])
- Alcohol (either ethyl or isopropyl)
- Rose bengal
- Reference toxicant (ammonia, copper sulfate, cadmium chloride, sodium dodecyl sulfate)

6.3.5 Cleaning

6.3.5.1 All nondisposable sample containers, test chambers, and other equipment that have come in contact with sediment should be washed after use in the manner described below to remove surface contaminants.

1. Soak 15 min in tap water and scrub with detergent or clean in an automatic dishwasher.
2. Rinse twice with tap water.
3. Carefully rinse once with fresh, dilute (10%, V:V) hydrochloric or nitric acid to remove scale, metals, and bases. To prepare a 10% solution of acid, add 10 mL of concentrated acid to 90 mL of deionized water.
4. Rinse twice with deionized water.
5. Rinse once with full-strength, pesticide-grade acetone to remove organic compounds (uses a fume hood or canopy). Hexane might also be used as a solvent for removing nonionic organic compounds. However, acetone is preferable if only one organic solvent is used to clean equipment.
6. Rinse three times with deionized water.

6.3.5.2 All test chambers and equipment should be thoroughly rinsed or soaked with the dilution water immediately before use in a test. See USEPA (2001) for information on equipment decontamination procedures with regards to collecting sediments in the field.

6.3.5.3 Many organic solvents (e.g., methylene chloride) leave a film that is insoluble in water. A dichromate-sulfuric acid cleaning solution can be used in place of both the organic solvent and the acid (see ASTM, 2000f), but the solution might attack silicone adhesive and leave chromium residues on glass. An alternative to use of dichromate-sulfuric acid could be to heat glassware for 8 h at 450°C.

Section 7

Water, Reagents, and Standards

7.1 Water

7.1.1 Requirements

7.1.1.1 Seawater used to test and culture organisms should be uniform in quality and from the same source. Acceptable seawater must allow satisfactory survival, growth, and reproduction of the test organisms. If problems are observed in the culturing or testing of organisms, the characteristics of the water should be evaluated. See USEPA (1991a) and ASTM (2000b) for a recommended list of chemical analyses of the water supply.

7.1.2 Source

7.1.2.1 Culture and testing water can be natural or synthetic seawater. The source of water will depend to some extent on the objective of the test. All natural waters should be obtained from an uncontaminated source beyond the influence of known discharges. Suitable water sources should have intakes that are positioned to (1) minimize fluctuations in quality and contamination, (2) maximize the concentration of dissolved oxygen, and (3) ensure low concentrations of sulfide and iron. Natural seawater should be collected at slack high tide, or within 1 h after high tide if taken from an semi-enclosed or urbanized area. It might be desirable or necessary to dilute full strength seawater with an appropriate freshwater source to achieve 5‰ or 20‰.

7.1.2.2 Sources of freshwater (i.e., 0‰) for dilution include distilled or deionized water, reverse osmosis water, and uncontaminated well or spring water (USEPA, 1991a). Municipal water supplies can be variable and might contain unacceptably high concentrations of materials such as copper, lead, zinc, fluoride, chlorine, or chloramines. Chlorinated water should not be used to dilute seawater used for culturing or testing because residual chlorine and chlorine-produced oxidants are toxic to many aquatic organisms. Dechlorinated water should only be used as a last resort, because dechlorination is often incomplete (ASTM, 2000d).

7.1.3 Water Treatment and Quality

7.1.3.1 Seawater and dilution water should be filtered (#5 mm) shortly before use to remove suspended particles and organisms. Water that might be contaminated with pathogens should be treated shortly before use by filtration (#0.45 mm), either alone or in combination with UV sterilization.

7.1.3.2 Water should be aerated using air stones, surface aerators, or column aerators. Adequate aeration will stabilize pH, bring concentrations of DO and other gases into equilibrium with air, and minimize oxygen demand and concentrations of volatiles. The initial concentration of DO in test water should be $6 mg/L to help ensure that DO concentrations are acceptable in test chambers.

7.1.3.3 DO, salinity, and pH should be measured on each batch of water before it is used in cultures and tests. Batches of salinity-adjusted culture water can be held for approximately 1 week; a lower holding temperature (about 4°C) helps maintain acceptable water quality. Other investigators have reported success in holding reconstituted seawater for toxicity testing for over 1 month (Ingersoll et al., 1992).

7.1.3.4 For site-specific investigations, it might be desirable to have the water-quality characteristics of the overlying water (i.e., salinity) as similar as possible to the site water. Other applications may require use of water from the site where sediment is collected. In estuarine systems, however, the pore water salinity of sediment might not be the same as the overlying water at the time of collection (Sanders et al., 1965).

7.1.4 Reconstituted/Synthetic Seawater

7.1.4.1 Although reconstituted seawater is acceptable, natural seawater is preferable, especially for tests in which the bioavailability of chemicals is affected by seawater chemistry. Reconstituted seawater is prepared by adding specified amounts of reagent-grade chemicals to high-purity distilled or deionized water (ASTM, 2000f; USEPA, 1991c). Suitable salt reagents can be reagent-grade chemicals, or commercial sea salts, such as Crystal Sea Marinemix®, Instant Ocean®, or HW Marinemix®. Preformulated brine (e.g., 60‰ to 90‰), prepared with dry ocean salts, or by heat-concentrating or freezing natural seawater, can also be used.

7.1.4.2 A synthetic sea formulation called GP2 can be prepared with reagent-grade chemicals and diluted with a suitable high-quality water to the desired salinity (Section 7.1.2.2; USEPA, 1994c).

7.1.4.3 The suitability and consistency of a particular salt formulation for use in holding and testing should be verified by laboratory tests because some formulations can produce unwanted toxic effects or sequester contaminants (Environment Canada, 1992). In controlled tests with the salt formulations mentioned above, Emery et al. (1997) found differences in survival, growth, and reproduction, and that laboratories can have acceptable performance (i.e., survival) with any of the salts evaluated. Because of higher growth rates observed in the Crystal Sea Marinemix® seasalt, they recommended its use for culturing and testing (Emery et al., 1997).

7.1.4.4 Deionized, distilled, or reverse-osmosis water should be obtained from a system capable of producing at least 1 mega-ohm water. If large quantities of high quality water are needed, it might be advisable to precondition water with a mixed-bed water treatment system. Some investigators have observed that aging of reconstituted water prepared from deionized water for several days before use in sediment tests may improve performance of test organisms. Other investigators have reported success in holding reconstituted seawater for toxicity testing for over 1 month (Ingersoll et al., 1992).

7.1.4.5 Salinity, pH, and DO should be measured on each batch of reconstituted water. The reconstituted water should be aerated before use to adjust pH and DO to the acceptable ranges (e.g., Table 11.1). Reconstituted sea water should be filtered (#5 mm) shortly before use to remove suspended particles and should be used within 24 h of filtration. USEPA (1991a) recommends holding a batch of reconstituted water for no longer than 2 weeks due to the potential for bacteriological growth. Other investigators have reported success in holding reconstituted seawater for toxicity testing for over 1 month (Ingersoll et al., 1992).

7.2 Reagents

7.2.1 Material safety data sheets should be followed for reagents and other chemicals purchased from supply houses. The test material(s) should be at least reagent grade, unless a test using a formulated commercial product, technical-grade, or use-grade material is specifically needed. Reagent containers should be dated when received from the supplier, and the shelf life of the reagent should not be exceeded. Working solutions should be dated when prepared and the recommended shelf life should not be exceeded.

7.3 Standards

7.3.1 Appropriate standard methods for chemical and physical analyses should be used when possible. For those measurements for which standards do not exist or are not sensitive enough, methods should be obtained from other reliable sources.

Section 8
Sample Collection, Storage, Manipulation, and Characterization

8.1 Collection

8.1.1 Before the preparation or collection of sediment, a procedure should be established for the handling of sediment that might contain unknown quantities of toxic contaminants (Section 5).

8.1.2 Sediments are spatially and temporally variable (Stemmer et al., 1990a). Replicate samples should be collected to determine variance in sediment characteristics. Sediment should be collected with as little disruption as possible; however, subsampling, compositing, or homogenization of sediment samples might be necessary for some experimental designs. Sampling can cause loss of sediment integrity, change in chemical speciation, or disruption of chemical equilibrium (ASTM, 2000c). Benthic grabs (i.e., Ponar, Smith-MacIntyre, Van Veen) or core samplers should be used rather than a dredge to minimize disturbance of the sediment sample. Sediment should be collected to a depth that will represent expected exposure concentration. For example, samples collected for evaluations of dredged material should include sediment cores to the depth of removal. Surveys of the toxicity of surficial sediment are often based on samples of the upper 2 cm of sediment.

8.1.3 Exposure to direct sunlight during collection should be minimized, especially if the sediment contains photolytic compounds (e.g., PAHs). Collect, manipulate, and store sediments using tools made of chemically inert materials to minimize contamination of the sample (ASTM, 2000b). Sediment samples should be cooled to 4°C as quickly as possible in the field before shipment or return to the laboratory (ASTM, 2000b). Coolers with gel packs, ice, or dry ice can be used to cool samples in the field; however, sediment should never be frozen. Continuous-recording monitors can be used to measure temperature during shipping (e.g., TempTale Temperature Monitoring and Recording System, Sensitech, Inc., Beverly, MA).

8.1.4 For additional information on sediment collection and shipment, refer to methods published by USEPA (2001) and ASTM (2000c).

8.2 Storage

8.2.1 Because the contaminants of concern and influencing sediment characteristics are not always known, it is desirable to hold the sediments after collection in the dark at 4°C. Traditional convention has held that toxicity tests should be initiated as soon as possible following collection from the field, although actual recommended storage times range from 2 weeks (ASTM, 2000c) to less than 8 weeks (USEPA-USACE, 1998). Discrepancies in recommended storage times reflected a lack of data concerning the effects of long-term storage on the physical, chemical, and toxicological characteristics of the sediment. However, numerous studies have recently been conducted to address issues related to sediment storage (Dillon et al., 1994; Becker and Ginn, 1995; Carr and Chapman, 1995; Moore et al., 1996; Sarda and Burton, 1995; Sijm et al., 1997; DeFoe and Ankley, 1998). The conclusions and recommendations offered by these studies vary substantially and appear to depend primarily upon the type or class of contaminant(s) present. Considered collectively, these studies suggest that the recommended guidance that sediments be tested sometime between the time of collection and 8 weeks storage is appropriate. Additional guidance is provided below.

8.2.2 Extended storage of sediments that contain high concentrations of labile contaminants (e.g., ammonia, volatile organics) may lead to a loss of

these contaminants and a corresponding reduction in toxicity. Under these circumstances, the sediments should be tested as soon as possible after collection, but not later than within 2 weeks (Sarda and Burton, 1995). Sediments that exhibit low-level to moderate toxicity often exhibit considerable temporal variability in toxicity, although the direction of change is often unpredictable (Carr and Chapman, 1995; Moore et al., 1996; DeFoe and Ankley, 1998). For these types of sediments, the recommended storage time of <8 weeks may be most appropriate. In some situations, a minimum storage period for low-to-moderately contaminated sediments may help reduce variability. For example, DeFoe and Ankley (1998) observed high variability in survival during early testing periods (e.g., <2 weeks) in sediments with low toxicity. DeFoe and Ankley (1998) hypothesized that this variability partially reflected the presence of indigenous predators that remained alive during this relatively short storage period. Thus, if predatory species are known to exist, and the sediment does not contain labile contaminants, it may be desirable to store the sediment for a short period before testing (e.g., 2 weeks) to reduce potential for interferences with indigenous organisms. Sediments that contain comparatively stable compounds (e.g., high-molecular-weight compounds such as polychlorinated biphenyls [PCBs]) or that exhibit a moderate-to-high level of toxicity, typically do not vary appreciably in toxicity in relation to storage duration (Moore et al., 1996; DeFoe and Ankley, 1998). For these sediments, long-term storage (e.g., >8 weeks) can be undertaken.

8.2.3 Researchers may wish to conduct additional characterizations of sediment to evaluate possible effects of storage. Concentrations of contaminants of concern could be measured periodically in pore water during the storage period and at the start of the sediment test (Kemble et al., 1994). Ingersoll et al. (1993) recommend conducting a toxicity test with pore water within 2 weeks from sediment collection and at the start of the sediment test. Freezing might further change sediment properties such as grain size or contaminant partitioning and should be avoided (ASTM, 2000c; Schuytema et al., 1989). Sediment should be stored with no air over the sealed samples (no head space) at 4°C before the start of a test (Shuba et al., 1978). Sediment may be stored in containers constructed of suitable materials as outlined in Section 6.

8.3 Manipulation

8.3.1 Homogenization and Sieving

8.3.1.1 Samples tend to settle during shipment. As a result, water above the sediment should not be discarded but should be mixed back into the sediment during homogenization. Sediment samples should only be sieved to remove indigenous organisms if there is a good reason to believe indigenous organisms may influence the response of the test organism. Sieving procedures are outlined in Section 4.3.2.3. However, large indigenous organisms and large debris can be removed using forceps. Reynoldson et al. (1994) observed reduced growth of amphipods, midges, and mayflies in sediments with elevated numbers of oligochaetes and recommended sieving sediments suspected to have high numbers of indigenous oligochaetes. If sediments must be sieved, it may be desirable to analyze samples before and after sieving (e.g., pore water metals, dissolved organic carbon [DOC], AVS, TOC) to document the influence of sieving on sediment chemistry.

8.3.1.2 If sediment is collected from multiple field samples, the sediment can be pooled and mixed by stirring or using a rolling mill, feed mixer, or other suitable apparatus (see ASTM, 2000c). Homogenization of sediment can be accomplished by hand with a teflon paddle or using a variable-speed hand-held drill outfitted with a stainless-steel auger.

8.3.2 Sediment Spiking

8.3.2.1 Test sediment can be prepared by manipulating the properties of a control sediment. Mixing time (Stemmer et al., 1990a) and aging (Word et al., 1987; Landrum, 1989; Landrum and Faust, 1992) of spiked sediment can affect bioavailability of contaminants in sediment. Many studies with spiked sediment are often started only a few days after the chemical has been added to

the sediment. This short time period may not be long enough for sediments to equilibrate with the spiked chemicals (Section 8.3.2.2.3). Consistent spiking procedures should be followed in order to make interlaboratory comparisons. See USEPA (2001) and ASTM (2000c) for additional detail regarding sediment spiking.

8.3.2.1.1 The cause of sediment toxicity and the magnitude of interactive effects of contaminants can be estimated by spiking a sediment with chemicals or complex waste mixtures (Lamberson and Swartz, 1992). Sediments spiked with a range of concentrations can be used to generate either point estimates (e.g., LC50) or a minimum concentration at which effects are observed (LOEC). The influence of sediment physico-chemical characteristics on chemical toxicity can also be determined with sediment-spiking studies (Adams et al., 1985).

8.3.2.2 The test material(s) should be at least reagent grade, unless a test using a formulated commercial product, technical-grade, or use-grade material is specifically needed. Before a test is started, the following should be known about the test material: (1) the identity and concentration of major ingredients and impurities; (2) water solubility in test water; (3) log Kow, bioconcentration factor (BCF) from other test species, persistence, hydrolysis, and photolysis rates of the test substances; (4) estimated toxicity to the test organism and to humans; (5) if the test concentration(s) are to be measured, the precision and bias of the analytical method at the planned concentration(s) of the test material; and (6) recommended handling and disposal procedures. Addition of test material(s) to sediment may be accomplished using various methods, such as a (1) rolling mill, (2) feed mixer, or (3) hand mixing (ASTM, 2000c; USEPA, 2001). Modifications of the mixing techniques might be necessary to allow time for a test material to equilibrate with the sediment. Mixing time of spiked sediment should be limited from minutes to a few hours, and temperature should be kept low to minimize potential changes in the physico-chemical and microbial characteristics of the sediment (ASTM, 2000c). Duration of contact between the chemical and sediment can affect partitioning and bioavailability (Word et al., 1987). Care should be taken to ensure that the chemical is thoroughly and evenly distributed in the sediment. Analyses of sediment subsamples are advisable to determine the degree of mixing homogeneity (Ditsworth et al., 1990). Moreover, results from sediment-spiking studies should be compared to the response of test organisms to chemical concentrations in natural sediments (Lamberson and Swartz, 1992).

8.3.2.2.1 Organic chemicals have been added to sediments using the following procedures: (1) directly in a dry (crystalline) form; (2) coated on the inside walls of the container (Ditsworth et al., 1990); or (3) coated onto silica sand (e.g., 5% w/w of sediment) which is added to the sediment (D.R. Mount, USEPA, Duluth, MN, personal communication). In Techniques 2 and 3, the chemical is dissolved in solvent, placed in a glass spiking container (with or without sand), then the solvent is slowly evaporated. The advantage of these three approaches is that no solvent is introduced to the sediment, only the chemical being spiked. When testing spiked sediments, procedural blanks (sediments that have been handled in the same way, including solvent addition and evaporation, but that contain no added chemical) should be tested in addition to regular negative controls.

8.3.2.2.2 Metals are generally added in an aqueous solution (ASTM, 2000c; Carlson et al., 1991; Di Toro et al., 1990). Ammonia has also been successfully spiked using aqueous solutions (Moore et al., 1997; Besser et al., 1998). Inclusion of spiking blanks is recommended.

8.3.2.2.3 Sufficient time should be allowed after spiking for the spiked chemical to equilibrate with sediment components. For organic chemicals, it is recommended that the sediment be aged at least 1 month before starting a test; 2 months or more may be necessary for chemicals with a high log Kow (e.g., >6; D.R. Mount, USEPA, Duluth, MN, personal communication). For metals, shorter aging times (1 to 2 weeks) may be sufficient. Periodic monitoring of chemical concentrations in pore water during sediment aging is highly recommended as a means to

assess the equilibration of the spiked sediments. Monitoring of pore water during spiked sediment testing is also recommended.

8.3.2.3 Direct addition of a solvent (other than water) to the sediment should be avoided if possible. Addition of organic solvents may dramatically influence the concentration of dissolved organic carbon in pore water. If an organic solvent is to be used, the solvent should be at a concentration that does not affect the test organism. Further, both solvent control and negative control sediments must be included in the test. The solvent control must contain the highest concentration of solvent present and must be from the same batch used to make the stock solution (see ASTM, 2000f).

8.3.2.3.1 If direct addition of organic solvent is to be used, the same concentration of solvent should be used in all treatments. If the concentration of solvent is not the same in all treatments, a solvent test should be conducted to determine whether survival, growth, or reproduction of the test organisms is related to the concentration of the solvent.

8.3.2.4 If the test contains both a negative control and a solvent control, the survival, growth, or reproduction of the organisms tested should be compared. If a statistically significant difference is detected between the two controls, only the solvent control may be used for meeting the acceptability of the test and as the basis for calculating results. The negative control might provide additional information on the general health of the organisms tested. If no statistically significant difference is detected, the data from both controls should be used for meeting the acceptability of the test and as the basis for calculating the results (ASTM, 2000g). If performance in the solvent control is markedly different from that in the negative control, it is possible that the data are compromised by experimental artifacts and may not accurately reflect the toxicity of the chemical in natural sediments.

8.3.3 Test Concentration(s) for Laboratory Spiked Sediments

8.3.3.1 If a test is intended to generate an LC50, a toxicant concentration series (0.5 or higher) should be selected that will provide partial mortalities at two or more concentrations of the test chemical. The LC50 of a particular compound may vary depending on physical and chemical sediment characteristics. It may be desirable to conduct a range-finding test in which the organisms are exposed to a control and three or more concentrations of the test material that differ by a factor of ten. Results from water-only tests could be used to establish concentrations to be tested in a whole-sediment test based on predicted pore water concentrations (Di Toro et al., 1991).

8.3.3.2 Bulk-sediment chemical concentrations might be normalized to factors other than dry weight. For example, concentrations of nonpolar organic compounds might be normalized to sediment organic-carbon content, and simultaneously extracted metals might be normalized to acid volatile sulfides (Di Toro et al., 1990; Di Toro et al., 1991).

8.3.3.3 In some situations it might be necessary to simply determine whether a specific concentration of test material is toxic to the test organism, or whether adverse effects occur above or below a specific concentration. When there is interest in a particular concentration, it might only be necessary to test that concentration and not to determine an LC50.

8.4 Characterization

8.4.1 All sediment should be characterized and at least the following determined: salinity, pH, and ammonia of the pore water; TOC; particle-size distribution (percent sand, silt, clay); and percent water content (ASTM, 2000b; Plumb, 1981). See Section 8.4.4.7 for methods to isolate pore water.

8.4.2 Other analyses on sediment might include biological oxygen demand (BOD), chemical oxygen demand (COD), cation exchange capacity, redox potential (Eh), total inorganic carbon, total volatile solids (TVS), AVS, metals, synthetic

organic compounds, oil and grease, petroleum hydrocarbons, as well as interstitial water analyses for various physico-chemical parameters.

8.4.3 Macrobenthos can be evaluated by subsampling the field-collected sediment. If direct comparisons are to be made, subsamples for toxicity testing should be collected from the same sample to be used for analysis of physical and chemical characteristics. Qualitative descriptions of the sediment can include color, texture, presence of hydrogen sulfide, and presence of indigenous organisms. Monitoring the odor of sediment samples should be avoided because of potential hazardous volatile contaminants. It may be desirable to describe color and texture gradients that occur with sediment depth.

8.4.4 Analytical Methods

8.4.4.1 Chemical and physical data should be obtained using appropriate standard methods whenever possible. For those measurements for which standard methods do not exist or are not sensitive enough, methods should be obtained from other reliable sources.

8.4.4.2 The precision, accuracy, and bias of each analytical method used should be determined in the appropriate matrix: sediment, water, or tissue. Reagent blanks and analytical standards should be analyzed and recoveries should be calculated.

8.4.4.3 Concentration of spiked test material(s) in sediment, interstitial water, and overlying water should be measured as often as practical during a test. If possible, the concentration of the test material in overlying water, interstitial water, and sediment should be measured at the start and end of a test. Measurement of test material(s) degradation products might also be desirable.

8.4.4.4 Separate chambers should be set up at the start of a test and destructively sampled during and at the end of the test to monitor sediment chemistry. Test organisms and food might be added to these extra chambers.

8.4.4.5 Measurement of test material(s) concentrations in water can be accomplished by pipeting water samples from about 1 cm to 2 cm above the sediment surface in the test chamber. Overlying water samples should not contain any surface debris, any material from the sides of the test chamber, or any sediment.

8.4.4.6 Measurement of concentrations of test material(s) in sediment at the end of a test can be taken by siphoning most of the overlying water without disturbing the surface of the sediment, then removing appropriate aliquots of the sediment for chemical analysis.

8.4.4.7 Interstitial Water

8.4.4.7.1 Interstitial water (pore water), defined as the water occupying the spaces between sediment or soil particles, is often isolated to provide either a matrix for toxicity testing or to provide an indication of the concentration or partitioning of contaminants within the sediment matrix. Draft USEPA ESGs are based on the presumption that the concentration of chemicals in the interstitial water are correlated directly to their bioavailability and, therefore, their toxicity (Di Toro et al., 1991). Of additional importance is contaminants in interstitial waters can be transported into overlying waters through diffusion, bioturbation, and resuspension processes (Van Rees et al., 1991). The usefulness of interstitial water sampling for determining chemical contamination or toxicity will depend on the study objectives and nature of the sediments at the study site.

8.4.4.7.2 Isolation of sediment interstitial water can be accomplished by a wide variety of methods, which are based on either physical separation or on diffusion/equilibration. The common physical-isolation procedure can be categorized as (1) centrifugation, (2) compression/squeezing, or (3) suction/vacuum. Diffusion/equilibrium procedures rely on the movement (diffusion) of pore water constituents across semipermeable membranes into a collecting chamber until an equilibrium is established. A description of the materials and procedures used in the isolation of pore water is included in the reviews by Bufflap and Allen (1995a), ASTM (2000c), and USEPA (2001).

8.4.4.7.3 When relatively large volumes of water are required (>20 mL) for toxicity testing or chemical analyses, appropriate quantities of sediment are generally collected with grabs or

corers for subsequent isolation of the interstitial water. Several isolation procedures, such as centrifugation (Ankley and Scheubauer-Berigan, 1994), squeezing (Carr and Chapman, 1995) and suction (Winger and Lasier, 1991; Winger et al., 1998), have been used successfully to obtain adequate volumes for testing purposes. Peepers (dialysis) generally do not produce sufficient volumes for most analyses; however, larger sized peepers (500-mL volume) have been used for collecting interstitial water *in situ* for chemical analyses and organism exposures (Burton, 1992; Sarda and Burton, 1995).

8.4.4.7.4 There is not one superior method for the isolation of interstitial water used for toxicity testing and associated chemical analyses. Factors to consider in the selection of an isolation procedure may include (1) volume of pore water needed, (2) ease of isolation (materials, preparation time, and time required for isolation), and (3) artifacts in the pore water caused by the isolation procedure. Each approach has unique strengths and limitations (Bufflap and Allen, 1995a; 1995b; Winger et al., 1998), which vary with sediment characteristics, chemicals of concern, toxicity test methods, and desired test resolution (i.e., the data quality objectives). For suction or compression separation, which use a filter or a similar surface, there may be changes to the characteristics of the interstitial water compared with separation using centrifugation (Ankley et al., 1994; Horowitz et al., 1996). For most toxicity test procedures, relatively large volumes of interstitial water (e.g., liters) are frequently needed for static or renewal exposures with the associated water chemistry analyses. Although centrifugation can be used to generate large volumes of interstitial water, it is difficult to use centrifugation to isolate water from coarser sediment. If smaller volumes of interstitial water are adequate and logistics allow, it may be desirable to use peepers, which establish an equilibrium with the pore water through a permeable membrane. If logistics do not allow placement of peeper samplers, an alternative procedure could be to collect cores that can be sampled using side port suctioning or centrifugation (G.A. Burton, Wright State University, personal communication). However, if larger samples of interstitial water are needed, it would be necessary to collect multiple cores as quickly as possible using an inert environment and to centrifuge samples at ambient temperatures. See USEPA (2001) and ASTM (2000c) for additional detail regarding isolation of interstitial water.

Section 9
Quality Assurance and Quality Control

9.1 Introduction

9.1.1 Developing and maintaining a laboratory quality assurance (QA) program requires an ongoing commitment by laboratory management and also includes the following: (1) appointment of a laboratory quality assurance officer with the responsibility and authority to develop and maintain a QA program; (2) preparation of a Quality Assurance Project Plan with Data Quality Objectives; (3) preparation of written descriptions of laboratory Standard Operating Procedures (SOPs) for test organism culturing, testing, instrument calibration, sample chain-of-custody, laboratory sample tracking system, and other procedures, as required; and (4) provision of qualified technical staff and suitable space and equipment to assure reliable data. Additional guidance for QA can be obtained in USEPA (1989b; 1999), and Moore et al., 1994.

9.1.2 QA practices within a testing laboratory should address all activities that affect the quality of the final data, such as (1) sediment sampling and handling, (2) the source and condition of the test organisms, (3) the condition and operation of equipment, (4) test conditions, (5) instrument calibration, (6) replication, (7) use of reference toxicants, (8) record keeping, and (9) data evaluation.

9.1.3 Quality Control (QC) practices, on the other hand, consist of the more focused, routine, day-to-day activities conducted within the scope of the overall QA program. For a more detailed discussion of quality assurance and general guidance on good laboratory practices related to testing see FDA (1978), USEPA (1979a; 1980a; 1980b; 1991a; 1994b; 1995; 2001), DeWoskin (1984), and Taylor (1987).

9.2 Performance-Based Criteria

9.2.1 The USEPA EMMC recommended the use of performance-based methods in developing standards for chemical analytical methods (Williams, 1993). Performance-based methods were defined by EMMC as a monitoring approach that uses methods that meet pre-established, demonstrated performance standards. Minimum required elements of performance, such as precision, reproducibility, bias, sensitivity, and detection limits should be specified, and the method should be demonstrated to meet the performance standards.

9.2.2 In developing guidance for culturing *L. plumulosus*, it was determined that no single method has to be used to culture organisms. Success of a test relies on the health of the culture from which organisms are taken for testing. Having healthy organisms of known quality and age (i.e., size) for testing is the key consideration relative to culture methods. Therefore, a performance-based criteria approach is the preferred method by which individual laboratories should evaluate culture health, rather than a control-based criteria approach. Performance-based criteria were chosen to allow each laboratory to optimize culture methods that provide organisms that produce reliable and comparable test results. Performance criteria for culturing and testing *L. plumulosus* are listed in Table 11.3.

9.3 Facilities, Equipment, and Test Chambers

9.3.1 Separate areas for test organism culturing and testing must be provided to avoid loss of cultures from cross-contamination. Ventilation systems should be designed and operated to prevent recirculation or leakage of air from chemical analysis laboratories or sample storage

and preparation areas into test organism culturing or sediment testing areas, and from sediment testing laboratories and sample preparation areas into culture rooms.

9.3.2 Equipment for temperature control should be adequate to maintain recommended test-water temperatures. Recommended materials should be used in the fabricating of the test equipment that comes in contact with the sediment or overlying water.

9.3.3 Before a sediment test is conducted in a new facility, a "noncontaminant" test should be conducted in which all test chambers contain a control sediment and overlying water. This information is used to demonstrate that the facility, control sediment, water, and handling procedures provide acceptable responses of test organisms (Section 9.14).

9.4 Test Organisms

9.4.1 The organisms should appear healthy, behave normally, feed well, and have low mortality in test controls (#20%). The species of test organisms should be positively identified. Test organisms should not show signs of disease or apparent stress (e.g., discoloration, unusual behavior).

9.5 Water

9.5.1 The quality of water used for organism culturing and testing is extremely important. Overlying water used in culturing, holding, acclimation, and testing organisms should be uniform in quality. Acceptable water should allow satisfactory survival, growth or reproduction of the test organisms. *L. plumulosus* should not show signs of disease or apparent stress (e.g., discoloration, unusual behavior). See Section 7 for additional details.

9.6 Sample Collection and Storage

9.6.1 Sample holding times and temperatures should conform to conditions described in Section 8.

9.7 Test Conditions

9.7.1 It is desirable to measure temperature continuously in at least one chamber during each test. Temperatures should be maintained within the limits specified in Section 11. DO, temperature, salinity, ammonia, and pH should be checked as prescribed in Section 11.3.

9.8 Quality of Test Organisms

9.8.1 It may be desirable for laboratories to periodically perform 96-h water-only reference-toxicity test to assess the sensitivity of culture organisms (Section 9.16). Data from these reference-toxicity tests could be used to assess genetic strain or life-stage sensitivity to select chemicals. The previous requirement for laboratories to conduct monthly reference -toxicity tests (USEPA, 1974a; 1994d) has not been included as a requirement for sediments due to the inability of reference-toxicity tests to identify stressed populations of test organisms (McNulty et al., 1999; McGee et al., 1998). Physiological measurements such as lipid content might also provide useful information regarding the health of the cultures.

9.8.2 Test animals should only be obtained from cultures. It is likely to be impractical to obtain test-sited neonates directly from a supplier because of their sensitivity to physical disturbances and their rapid growth. Instead, test laboratories will likely want to establish their own cultures of *L. plumulosus* from which to harvest neonates. It is desirable to determine the sensitivity of *L. plumulosus* obtained from an outside source. For cultured organisms, the supplier should provide data with the shipment describing the history of the sensitivity of organisms from the same source culture. For field-collected organisms, the supplier should provide data with the shipment describing the collection location, the time and date of collection, the water salinity and temperature at the time of collection, and collection site sediment for holding and acclimation purposes. The supplier should also certify the species identification of the test organisms and provide the taxonomic references (e.g., Schoemaker, 1932; Bousfield, 1973) or name(s) of the taxonomic expert(s) consulted.

9.8.3 All organisms in a test must be from the same source (Section 10.2.2) (Table 10.1). Organisms may be obtained from laboratory cultures or from commercial or government sources. The test organisms used should be identified using an appropriate taxonomic key, and verification should be documented. The use of field-collected amphipods to start cultures is discussed in Section 10.4. Obtaining organisms from wild populations is useful for enhancement of genetic diversity of existing cultures or to establish new cultures. McGee et al. (1998) found seasonal variability in sensitivity to cadmium in field-collected *L. plumulosus*. Therefore field-collected organisms should not be used for toxicity testing unless organisms are cultured through several generations in the laboratory. In addition, the ability of the wild population of sexually reproducing organisms to cross-breed with the existing laboratory population should be determined (Duan et al., 1997). Sensitivity of the wild population to select contaminants (see Section 9.16.4) should also be documented.

9.9 Quality of Food

9.9.1 Problems with the nutritional suitability of the food will be reflected in the survival, growth, or reproduction of *L. plumulosus* in cultures or in sediment tests.

9.10 Test Acceptability

9.10.1 Test acceptability requirements related to these endpoints are provided in Table 11.3. Test acceptability requirements for the 28-d *L. plumulosus* test are as follows: (1) survival at 28-d must equal or exceed 80% in the control sediment and (2) measurable growth and reproduction must be found in all replicates of the negative control treatment. Additional requirements for acceptability of the tests are presented in Table 11.3. An individual test may be conditionally acceptable if temperature, dissolved oxygen, and other specified conditions fall outside specifications, depending on the degree of the departure and the objectives of the tests (see Table 11.1). The acceptability of a test will depend on the experience and professional judgment of the laboratory analyst and reviewing staff of the regulatory authority. Any deviation from test specifications should be noted when reporting data from a test.

9.11 Analytical Methods

9.11.1 All routine chemical and physical analyses for culture and testing water, food, and sediment should include established quality assurance practices outlined in USEPA methods manuals (1979a; 1979b; 1991c; 1994a; 1994b; 1994c; 1994d).

9.11.2 Reagent containers should be dated when received from the supplier, and the shelf life of the reagent should not be exceeded. Working solutions should be dated when prepared and the recommended shelf life should not be exceeded.

9.12 Calibration and Standardization

9.12.1 Instruments used for routine measurements of chemical and physical characteristics such as pH, DO, temperature, total ammonia, and salinity should be calibrated before use according to the instrument manufacturer's procedures as indicated in the general section on quality assurance (see USEPA Methods 150.1, 360.1, 170.1, and 120.1; USEPA, 1979a). Calibration data should be recorded in a permanent log.

9.12.2 The analytical balance(s) should be checked with reference weights, which are at the upper and lower ends of the range of weight values used. A balance should be checked at the beginning of each series of weighing, periodically (such as every tenth weight) during a long series of weighing, and after taking the last weight of series.

9.13 Replication and Test Sensitivity

9.13.1 The sensitivity of sediment tests will depend in part on the number of replicates/treatment, the significance level selected, and the type of statistical analysis. If the variability remains constant, the sensitivity of a test will increase as the number of replicates is increased. The minimum recommended number of replicates for the 28-d test with *L. plumulosus* is five, which was calculated by a cost-power analysis of test results

(see Section 13.5.1.6; DeWitt et al., 1997b). The minimum recommended number of replicates varies with the objectives of the test and the statistical method used for analysis of the data (Section 12).

9.14 Demonstrating Acceptable Performance

9.14.1 Intralaboratory precision, expressed as a coefficient of variation (CV) of the range in response for each type of test to be used in a laboratory, can be determined by performing five or more tests with different batches of test organisms, using the same reference toxicant, at the same concentrations, with the same test conditions (e.g., the same test duration, type of water, age of test organisms) and same data analysis methods. This should be done to gain experience for the toxicity tests and to serve as a point of reference for future testing. A reference-toxicity concentration series ($0.5) should be selected that will provide partial mortalities at two or more concentrations of the test chemical (Section 8.3.3). Information from previous tests can be used to improve the design of subsequent tests to optimize the dilution series selected for future testing.

9.14.2 Before conducting tests with contaminated sediment, it is strongly recommended that the laboratory conduct the tests with control sediment(s) alone. Results of these preliminary studies should be used to determine if use of the control sediment and other test conditions (i.e., water quality) result in acceptable performance in the test outlined in Table 11.1.

9.14.3 Laboratories should demonstrate that their personnel are able to recover an average of at least 90% of the organisms of a range of size classes (including neonates) from whole sediment. For example, test organisms could be added to control sediment or test sediment and recovery could be determined after 1 h (Tomasovic et al., 1994).

9.15 Documenting Ongoing Laboratory Performance

9.15.1 For a given reference-toxicity test, successive tests should be performed with the same reference toxicant, at the same concentrations, in the same type of water, generating LC50s using the same data analysis method (Section 12).

9.15.2 Outliers, which are data falling outside the control limits, and trends of increasing or decreasing sensitivity are readily identified. If the reference-toxicity results from a given test falls outside the "expected" range (e.g., +2 standard deviations [SD]), the sensitivity of the organisms and the credibility of the test results may be suspect. In this case, the test procedure should be examined for defects and should be repeated with a different batch of test organisms.

9.15.3 A sediment test may be acceptable if specified conditions of a reference-toxicity test fall outside the expected ranges (Section 9.10.2). Specifically, a sediment test should not be judged unacceptable if the LC50 for a given reference-toxicity test falls outside the expected range or if mortality in the control of the reference-toxicity test exceeds 10%. All the performance criteria outlined in Table 11.3 must be considered when determining the acceptability of a sediment test. The acceptability of the sediment test would depend on the experience and judgement of the investigator and the regulatory authority.

9.15.4 Performance should improve with experience, and the control limits should gradually narrow as the statistics stabilize. However, control limits of a mean +2 SD, by definition, will be exceeded 5% of the time regardless of how well a laboratory performs. For this reason, laboratories that develop very narrow control limits can be penalized if a test result that falls just outside the control limits is rejected *de facto*. The width of the control limits should be considered in decisions regarding rejection of data (Section 13).

9.16 Reference Toxicants

9.16.1 Historically, reference-toxicity testing has been thought to provide three types of information relevant to the interpretation of toxicity test data: (1) an indication of the relative "health" of the test organisms used in the test; (2) a demonstration that the laboratory can perform the test procedure in a reproducible manner; and (3) information to

indicate whether the sensitivity of the particular strain or population in use at the laboratory is comparable to those in use in other facilities. With regard to the first, recent work by McNulty et al. (1999) and McGee et al. (1998) suggests that reference-toxicity tests may not be effective in identifying stressed populations of test organisms. In addition, reference-toxicity tests recommended for use with sediment toxicity tests are short-term, water column tests, owing in part to the lack of a standard sediment for reference-toxicity testing. Because the test procedures for reference-toxicity tests are not the same as for the sediment toxicity test of interest, the applicability of reference-toxicity tests to demonstrate ability to reproducibly perform the sediment test procedures is greatly reduced. Particularly for long-term sediment toxicity tests, with *L. plumulosus* performance of control organisms over time may be a better indicator of success in handling and testing these organisms (Section 11).

9.16.2 Although the requirement for monthly testing has been removed in this manual, periodic reference-toxicity testing should be conducted as an indication of comparability of results among laboratories (minimumly one test every six months should be conducted to evaluate potential differences in genetic strain of organisms). In particular, reference-toxicity tests should be performed when organisms are obtained from outside sources, when there are changes in culture practices, or when brood stock from an outside source is incorporated into a culture.

9.16.3 In many instances, reference-toxicity tests have been conducted every time the *L. plumulosus* 28-d test was run. This may provide additional quality assurance data regarding the toxicological sensitivity of the test organism. However, the decision whether to conduct reference-toxicity tests every time the *L. plumulosus* 28-d test is run is dependent on the goal of the study (Section 9.16.2).

9.16.4 Reference toxicants such as cadmium (available as cadmium chloride [$CdCl_2$]), and ammonia, are suitable for use. Care must be taken with cadmium due to its carcinogenic nature and with ammonia because it is very labile. Use of nonhazardous alternatives for reference toxicants is recommended (Section 5.6.3). No one reference toxicant can be used to measure the sensitivity of test organisms with respect to another toxicant with a different mode of action (Lee, 1980). However, it may be unrealistic to test more than one or two reference toxicants routinely.

9.16.5 Test conditions for conducting reference-toxicity tests with *L. plumulosus* are outlined in Table 9.1.

9.16.6 Based on 96-h, water-only reference-toxicity tests at 20‰ with neonate *L. plumulosus*, one should expect a mean LC50 value for cadmium of approximately 0.5 mg/L (range: 0.2 mg/L to 0.7 mg/L) and LC50 values for total ammonia between 25 mg/L and 60 mg/L (DeWitt et al., 1997a). At 5‰, one should expect a mean LC50 value for cadmium of approximately 0.05 mg/L (range: 0.01 mg/L to 0.09 mg/L) and LC50 values for total ammonia between 37 mg/L and 53 mg/L (Emery et al., 1997; Moore et al., 1997).

9.17 Record-Keeping

9.17.1 Proper record-keeping is essential to the scientific defensibility of a testing program. A complete file should be maintained for each individual sediment test or group of tests on closely related samples. This file should contain a record of the sample chain-of-custody; a copy of the sample log sheet; the original bench sheets for the test organism responses during the sediment test(s); chemical analysis data on the sample(s); control data sheets for reference toxicants; detailed records of the test organisms used in the test(s), such as species, source, age, date of receipt, and other pertinent information relating to their history and health; information on the calibration of equipment and instruments; water quality monitoring records; test conditions used; and results of reference-toxicity tests. Laboratory data should be recorded immediately to prevent the loss of information or inadvertent introduction of errors into the record. Original data sheets should be signed and dated by the laboratory personnel performing the tests. For additional detail, see Section 12.

Table 9.1 Recommended Test Conditions for Conducting Reference-toxicity Tests

Parameter	Conditions
1. Test Type:	Static, water-only test
2. Dilution series:	Control and at least 5 test concentrations ($\geq$0.5 dilution factor)
3. Toxicant:	Cd, Ammonia
4. Temperature:	25°C $\pm$2°C
5. Salinity:	5‰ or 20‰ ($\pm$2‰), matched to salinity of 28-d sediment toxicity test (Section 11.3.6.6)
6. Light quality:	Wide-spectrum fluorescent lights
7. Illuminance	500 - 1000 lux
8. Photoperiod:	16 h light : 8 h dark
9. Renewal of water:	None
10. Age and size of test organisms:	size-selected: between 0.25 mm and 0.6 mm
11. Test chamber:	250 mL to 1-L glass beaker or jar
12. Volume of water:	80% of chamber volume (minimum)
13. Number of organisms/chamber:	n = 20 if 1 per replicate; n = 10 (minimum) if >1 replicate
14. Number of replicate chambers/ treatment:	1 minimum; 2 recommended
15. Aeration:	Not recommended; but aerate as necessary to maintain >60% DO saturation (>4.4 mg/L)
16. Dilution water:	Culture water, surface water, site water, or reconstituted water
17. Water quality monitoring frequency:	Salinity and pH, at beginning and end of test; temperature and dissolved oxygen daily
18. Test duration:	96 h
19. Endpoint:	Survival (LC50)
20. Test acceptability:	$\geq$90% control survival

Section 10
Collection, Culture, and Maintaining of Test Organisms

10.1 Life History

10.1.1 *Leptocheirus plumulosus*

10.1.1.1 *L. plumulosus* is a burrow-building member of the family Aoridae (Figure 10.1). It is an infaunal amphipod found in subtidal portions of Atlantic Coast brackish estuaries from Cape Cod, Massachusetts, to northern Florida (Bousfield, 1973; DeWitt et al., 1992a). It is common in protected embayments, but has been collected in channels of estuarine rivers at water depths up to 13 m (Schoemaker, 1932; Holland et al., 1988; Schlekat et al., 1992). In Chesapeake Bay, densities of *L. plumulosus* can reach 24,000/m^2 to 29,000m^2 (Holland et al., 1988).

10.1.1.2 *L. plumulosus* is a relatively large amphipod (adult length up to 13 mm) with a cylindrically shaped body that is brownish-grey incolor. A distinguishing feature is a series of dark bands or stripes that cross the dorsal surface of the pareons and pleons. It feeds on particles that are in suspension and on the sediment surface (DeWitt et al., 1988). Two studies have shown that *L. plumulosus* population abundance in Chesapeake Bay is negatively correlated with sediment contamination (Holland et al., 1988; McGee and Fisher, 1997). Thus, this amphipod would appear to be a good candidate to be an environmental indicator.

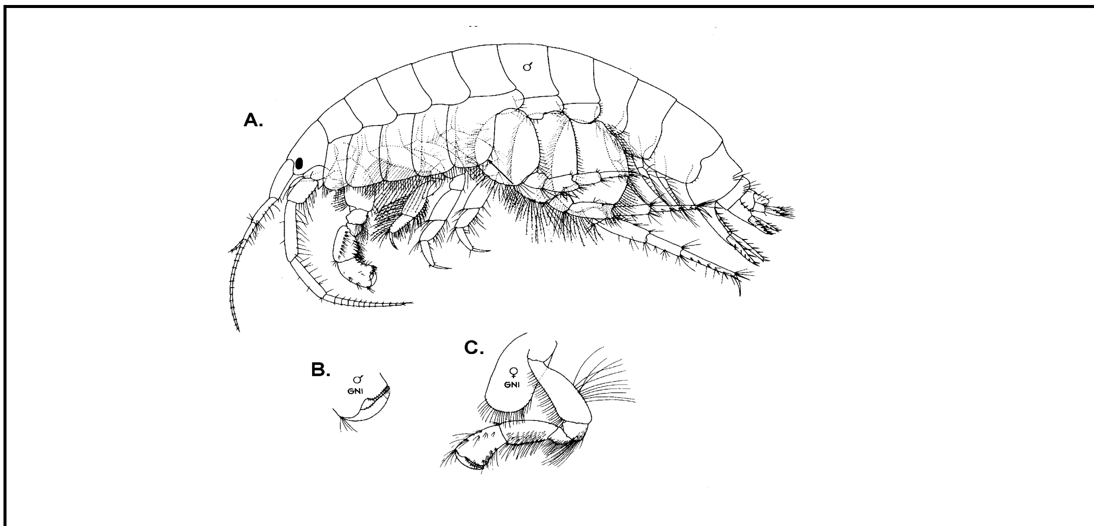


Figure 10.1 *Leptocheirus plumulosus* morphology (A) and characteristics used to determine the gender (B-C) of the amphipod. A: Adult male *L. plumulosus*. B: First gnathopod of the male (GN1), showing notched palm under dactyl. C: First gnathopod of female, showing flat palm under pactyl. Illustration of *L. plumulosus*, by E.L. Bousfield, reproduced with permission of the Canadian Museum of Nature, Ottawa, Canada.

10.1.1.3 *L. plumulosus* is found in both oligohaline and mesohaline regions of east coast estuaries; ambient water salinity at collection sites has ranged from 0‰ to 15‰ (Holland et al., 1988; DeWitt et al., 1992a; Schlekat et al., 1992; McGee et al., 1994). Laboratory studies have demonstrated that *L. plumulosus* 28-d test can be conducted at salinity values ranging from 1‰ to 35‰ (Section 11.4.4; Schlekat et al., 1992; SAIC, 1993b; DeWitt et al., 1992a, 1997a; Emery et al., 1997).

10.1.1.4 This amphipod is most often found in fine-grained sediment with a relatively high proportion of particulate organic material, although it has been collected in fine sand with low organic content (Jordan and Sutton, 1984; Holland et al., 1988; Marsh and Tenor, 1990; DeWitt et al., 1992a; Schlekat et al., 1992; McGee et al., 1994). Laboratory studies with *L. plumulosus* revealed no effect of sediment grain size on survival in control sediment containing 5% to 100% silt-clay content (DeWitt et al., 1997a). However, Emery et al. (1997) found significantly reduced survival in sediments in which clay content exceeded 84%.

10.1.1.5 Populations of *L. plumulosus* can be seasonally ephemeral with major population growth in fall and spring and large population declines in the summer (Holland et al., 1988; Marsh and Tenore, 1990; McGee, 1998). This pattern appears to be driven by changes in temperature and food availability and subsequent effects on life history traits (Marsh and Tenore, 1990; McGee, 1998). Short-term population fluctuations are also a function of the amphipod's relatively short generation time (DeWitt et al., 1992a). At 28°C, the age of the first brood release is approximately 24 d (DeWitt et al., 1992a).

10.1.1.6 *L. plumulosus* has been collected for cultures from several areas in the Maryland portion of Chesapeake Bay, including the Magothy, Chester, Corsica, and Wye Rivers. Organisms have been collected for culturing year-round from the Magothy River subestuary of Chesapeake Bay (C. Schlekat, University of South Carolina, and B. McGee, U.S. Fish and Wildlife Service, Annapolis, MD, unpublished data, personal communication).

10.2 General Culturing Procedures

10.2.1 Acceptability of a culturing procedure is based in part on performance of organisms in culture and in the sediment test (Section 1.4 and 9.2). No single technique for culturing test organisms is required. What may work well for one laboratory may not work as well for another laboratory. Although a variety of culturing procedures are outlined in Section 10.3 for *L. plumulosus,* organisms must meet the test acceptability requirements listed in Table 11.3.

10.2.2 All organisms in a test must be from the same source. Organisms may be obtained from laboratory cultures or from commercial or government sources; a partial list sources is provided in Table 10.1. The test organism used should be identified using an appropriate taxonomic key, and verification should be documented (Section 9.8.2).

Table 10.1 Sources of Starter Cultures of Test Organisms

Aquatic Biosystems, Inc.
1300 Blue Spruce Road, Suite C
Fort Collins, Colorado 80524
Scott Kellman
phone: 800/331-5916; fax: 970/484-2514
email: SRK@riverside.com

Chesapeake Cultures, Inc.
P.O. Box 507
Hays, Virginia 23702
Elizabeth Wilkins, President
phone: 804/693-4046; fax: 804/694-4703
email: growfish@c-cultures.com
website: www.c-cultures.com

Aquatic Research Organisms
P.O. Box 1271
Hampton, New Hampshire 03842-1271
Stan Sinitski or Mark Rosenqvist
phone: 800/927-1650; fax: 603/926-5278
website: www.arocentral.com

Obtaining organisms from wild populations should be avoided unless organisms are cultured through several generations in the laboratory before use in testing (Section 10.4). In addition, the ability of the wild population of sexually reproducing organisms to crossbreed with the existing laboratory population should be determined (Duan et al., 1997).

10.2.3 Test organisms obtained from commercial sources should be shipped in well-oxygenated water without sediment in insulated containers to maintain temperature during shipment. Temperature, salinity and DO of the water in the shipping containers should be measured at the time of shipment and on arrival to determine if the organisms might have been subjected to low DO, salinity change, or temperature and salinity fluctuations. The temperature and salinity of the shipped water should be gradually adjusted to the desired culture temperature and salinity at rates not exceeding 3/C or 3‰ per 24 h.

10.2.4 A group of organisms should not be used for a test if they appear to be unhealthy, discolored, or otherwise stressed (e.g., >20% mortality for 48 h before the start of a test). If the organisms fail to meet these criteria, the entire batch should be discarded and a new batch should be obtained. All organisms should be as uniform as possible in age and life stage. Test organisms should be handled as little as possible. When handling is necessary, it should be done as gently, carefully, and as quickly as possible.

10.3 Culturing Procedure for *Leptocheirus plumulosus*

10.3.1 The culturing method below is based on procedures described in DeWitt et al. (1997a). A periodic-renewal culture system is used. It consists of culture bins that contain aerated water over a thin (about 1 cm) layer of clean, fine-grained sediment in which the amphipods burrow. Culturing areas must be separate from testing areas to avoid exposing the cultures to contaminants. Before *L. plumulosus* are received at a testing facility, appropriate permits or approvals for import of live organisms should be obtained, if necessary. If culturing is to occur in an area where *L. plumulosus* are not indigenous to local waters, precautions should be taken to prevent release of living organisms to the outside environment (Section 10.6). Test animals should be destroyed at the end of toxicity test.

10.3.2 Starting a Culture

10.3.2.1 Amphipods for starting a laboratory culture of *L. plumulosus* should be obtained from a source with an established culture in which the species has been verified (see Table 10.1 for commercial sources of *L. plumulosus*). Alternatively, *L. plumulosus* can be obtained from field populations (see Section 10.4). Upon receipt of amphipods, the temperature and salinity of the water in shipping container(s) should be gradually adjusted to 20°C and desired culture salinity, at rates not exceeding 3°C or 3‰ per 24-h period. Feeding and regular maintenance should begin once the acclimation period is complete. Separate organisms into three size classes by gentle sieving: adults (retained on 1.0-mm mesh), subadults (pass through 1.0-mm mesh and retained on 0.6-mm mesh), and neonates (pass through 0.6-mm mesh and retained on 0.25-mm mesh). Seed each culture bin with approximately equal numbers of adults, subadults, and neonates to achieve a population density between 0.25/cm^2 to 0.35/cm^2 (2500/m^2 to 3500/m^2). Select only actively moving, healthy-looking organisms. Cultures should not be stocked at densities greater than 0.5/cm^2 (5000/m^2). See Section 10.3.8.4 for guidance on maintaining culture densities. Field-collected organisms should be added periodically to the culture population to maintain genetic diversity of the cultured amphipods (see also 9.8.3).

10.3.3 Culture Bins

10.3.3.1 Culture bins should be easy to maintain. Plastic wash tubs (approximately 35 cm x 30 cm x 15 cm) have been used successfully by several laboratories (DeWitt et al., 1992a). They are relatively light when filled with water and sediment, broad enough to allow for easy viewing of amphipod burrows, easily cleaned, inexpensive, and readily available. A wide variety of containers and materials may work just as well for culturing this species. New plasticware should be soaked in running water for several days prior to use in the

cultures to leach out potentially toxic compounds. Previously used culture bins usually can be satisfactorily cleaned using hot water and a scrub brush or pad, without the use of a chemical cleanser. Culture bins should not be washed with soap or detergent except in extreme conditions. If such a cleaning is deemed necessary, culture bins must be rinsed and soaked thoroughly after cleaning to remove residual cleanser.

10.3.4 Culture Sediment

10.3.4.1 Cultures should be established with a thin layer (1 cm to 1.5 cm) of sediment spread on the bottom of a culture bin. Sediment used for culture purposes should be the same as the control sediment used in sediment toxicity tests. Suitable sources for culture sediment include the amphipod collection site or an area adjacent to salt marsh vegetation. Culture sediment should be uncontaminated, organic-rich, fine-grained marine or estuarine sediment that is not anoxic. The organic carbon content (% TOC) should range between 1.5% and 4%. The sediment should be press-sieved through a 0.25-mm screen before use to facilitate the harvesting of neonates and to remove indigenous macroinvertebrates. Culture sediment can also be wet sieved. Wet-sieving involves agitating or swirling the sieve containing sediment in water so that particles smaller than the selected mesh size are washed through the sieve into a container (ASTM, 2000a). The sieve may be placed on a mechanical shaker, or the sediments on the screen can be stirred with a nylon brush to facilitate the process. Alternatively, the particles may be washed through the sieve with a small volume of running water. Culture sediment can also be frozen (>48 h) to provide additional assurance that viable macroinvertebrates are not present. Frozen sediment should be homogenized after thawing and before use. Culture sediment can be stored frozen for approximately 1 year.

10.3.5 Culture Water

10.3.5.1 Culture water used for holding and acclimating test organisms and for conducting toxicity tests should be of uniform quality and from the same source. See Section 7.1.2 for acceptable sources of water. Cultures of *L. plumulosus* are maintained at a salinity of either 5‰ or 20‰. Culture salinity will depend on the anticipated pore water salinity of test sediment and desired overlying water salinity to be used in the test (Section 11.3.6.6). To obtain these salinity values, natural or reconstituted seawater should be diluted with nonchlorinated well water, deionized water, distilled water, or reverse-osmosis water. Seawater and dilution water should be filtered (#5 m). Water that might be contaminated with pathogens should be treated shortly before use by filtration (#0.45 m), either alone or in combination with ultraviolet sterilization. DO, salinity, and pH should be checked before the water is used in cultures. Batches of salinity-adjusted culture water can be held for approximately 1 week; a lower holding temperature (<20°C) helps maintain acceptable water quality. Water depth in culture bins should be at least 10 cm. Aeration, provided through an air stone or pipet, should be moderate and constant, but not so vigorous as to resuspend sediment. Overlying water should be replaced the day after a new culture is established; thereafter, it should be renewed two or three times per week (Section 10.3.7.2).

10.3.6 Temperature and Photoperiod

10.3.6.1 Cultures should be maintained at 20°C to 25°C. The reproductive rate of *L. plumulosus* increases at temperatures greater than 20°C, necessitating more frequent culture thinning. Higher temperatures also can promote unwanted growth of nuisance organisms (such as nematodes, small worms, copepods, etc.). Temperatures below 20°C may not foster sufficiently prolific reproductive rates. Fluorescent lights should be on a 16 h light : 8 h dark photoperiod at a light intensity of 500 to 1000 lux. An efficient procedure is to maintain long-term cultures at 20°C, and increase culture temperature to about 25°C a few weeks in advance of testing.

10.3.7 Food and Feeding

10.3.7.1 This method recommends the simplest effective diet for routine use for *L. plumulosus* culture: finely milled TetraMin® provided two or three times per week. TetraMin® is a dry fish food (flake or powder) widely available in retail pet

stores. The food is prepared by milling, grinding, or chopping the flakes to a fine powder. A small flour mill, blender, or coffee grinder is useful for this. Ground powder is then sifted through a 0.25-mm mesh screen, retaining and using only the material that passes through the sieve. Use of a respirator or fume hood will minimize aspiration of dust. When establishing a new culture bin, do not add food for 3 to 4 days after amphipods are placed in new sediment. This will encourage the organisms to consume labile organic matter in sediment and to turn over the sediment by burrowing and feeding.

10.3.7.2 Culture bins should be provided with food in conjunction with water renewal. Two or three times a week, approximately 60% of culture water should be removed from each culture bin (by decanting, siphoning, or pumping) and replaced with the same volume of renewal water. Each culture bin is provided with approximately 0.4 g of dry food sprinkled evenly over the water surface, or as a slurry in culture water two or three times per week (e.g., Monday-Wednesday-Friday or Monday-Thursday). The amount of dry food added will depend on the density of each culture bin. Newly started culture bins should receive slightly less food (e.g., 0.3 g) than bins containing mature cultures. Excess food can decompose encouraging microbial and fungal growth on the sediment surface deteriorating water quality.

10.3.7.3 Some laboratories have experienced success in culturing *L. plumulosus* when other food is provided (i.e., live microalgae or a mixed dried food; DeWitt et al., 1992a). Modifications to the diet can be used by laboratories in order to optimize culture practices as long as performance criteria are satisfied (Table 11.3).

10.3.7.4 One feeding alternative is to supply renewal water consisting of seawater, cultured phytoplankton, and deionized water combined to the proper salinity and adjusted to an algal density of approximately 10^6 cells/mL (DeWitt et al. 1992a). Proportions will vary depending upon the salinity of the seawater and the density of the cultured phytoplankton. Live algae also can be used periodically to supplement a routine supply of dry food. The algae used can include a single or multiple species (e.g., *Pseudoisochrysis paradoxa*, *Phaeodactylum tricornutum*, *Isochrysis galbana*, *Chaetoceros calcitrans*, *Skeletonema* sp., *Dunalicella tertiolecta,* and/or *Thallasiosirus* spp.). Other algal species might be used if it can be demonstrated that they foster amphipod growth and reproductive rates equal to those of the aforementioned food alternatives. A mixture of algal species is recommended, if live algae is included in the diet.

10.3.8 Culture Maintenance

10.3.8.1 Observations and Measurements. Cultures should be observed daily to ensure that temperature is acceptable and aeration is adequate in all culture bins. Inspection for the presence of oligochaetes, polychaetes, copepods, infaunal sea anemones, or chironomids should be conducted weekly. The presence of excessive densities of these or other competing or predatory organisms should prompt renewal of culture sediment after separating *L. plumulosus* from the invasive organisms. During routine maintenance, cultures should be inspected for the presence of microbial and fungal build-up on the sediment surface. This build-up appears as a white or gray growth that may originate near uneaten food. Presence of microbial build-up may indicate that more food is being provided than is required by the amphipods. No additional food should be provided to culture bins with surficial microbial build-up until the build-up is no longer present. Sieving of sediment and renewal of culture bins can expedite removal of decaying material.

10.3.8.2 Healthy cultures are characterized by an abundance of burrow-openings on the sediment surface and turbid water from amphipod activity. Although amphipods may leave their burrows to search for food or mates, they will ordinarily remain in their burrows during the illuminated portion of the photoperiod. Amphipod density should therefore only be estimated by examining the number of burrow openings. An abundance of organisms on the sediment surface (e.g., >15 per culture bin) could indicate inadequate sediment quality, low DO concentrations, or overcrowding. A culture bin with an abundance of amphipods or unhealthy individuals on the sediment surface

should be examined closely, and the dissolved oxygen concentration should be measured in the overlying water. If the DO concentration is below 60% saturation (<4.4 mg/L), the culture bin should be sieved, and the population and culture sediment examined. If the population is too dense (i.e., >1.5/cm^2), the culture should be thinned as described in Section 10.2.8.4. If the sediment becomes an unacceptable habitat because it is anaerobic or black and sulphidic below the sediment surface, or contains an excess of competitive or predatory organisms, the healthy surviving amphipods should be placed in a new culture bin with newly prepared culture sediment.

10.3.8.3 Water temperature and DO should be measured in culture bins on a regular basis, approximately every week. Cultures should be continuously aerated. Salinity should be measured after water renewal. Ammonia and pH in overlying water should be measured with each new batch of sediment before organisms are added.

10.3.8.4 Renewal of Cultures. *L. plumulosus* can be prolific, and care must be taken to ensure that culture bins do not get overcrowded. Amphipods in overcrowded culture bins can be stressed because of food and space limitations, causing the fecundity of females to drop below five eggs/brood/female and the abundance of neonates and subadults to decline dramatically. Culture density should not exceed 1.5 amphipods/cm^2 and should ideally be maintained at approximately 0.5 amphipods/cm^2. A typical indication of overcrowding is a fairly uniform size distribution of amphipods (mostly small adults) and the presence of only two to four eggs in the brood pouches of gravid females. If sediment is not replaced occasionally, the cultures may become infested with undesirable species, such as worms or copepods. These “pests” may compete for food, bind sediment as fecal pellets, or produce mucus, thereby reducing culture productivity or increasing the effort required to harvest amphipods. Field-collected organisms should be added to the culture population periodically (approximately annually) in order to maintain genetic diversity of the culture organisms.

10.3.8.5 To avoid overcrowding, cultures should be thinned every 6 to 8 weeks by sieving through a 0.25-mm mesh screen to remove sediment. Sediment can be used for a total of 2 to 4 months before it should be replaced. Discard old sediment, prepare new culture bins and sediment, and restock each bin as described in Section 10.3.2.

10.4 Field Collection

10.4.1 Although established cultures of *L. plumulosus* are the recommended source of organisms for new cultures, it is recognized that field collection of amphipods might be necessary to enhance genetic diversity of existing cultures or to establish new cultures at a laboratory. The taxonomy of the organisms must be confirmed before they are introduced into existing laboratory populations. New organisms must be carefully inspected, and all other species of amphipods must be removed. The ability of a wild population of sexually reproducing organisms to crossbreed with existing laboratory populations of *L. plumulosus* must be confirmed through long-term culture maintenance (Duan et al., 1997). Collection areas should be relatively free of contamination. Field collection of *L. plumulosus* neonates for immediate use in a chronic toxicity test is not recommended.

10.4.2 Collection Methods

10.4.2.1 *L. plumulosus* is subtidal and can be collected with a small dredge or grab (e.g., Ponar, Smith-McIntyre, or Van Veen). In very shallow water, sediment containing *L. plumulosus* can be collected with a shovel or scoop, or using a suction dredge (DeWitt et al., 1992a).

10.4.2.2 All apparatus used for collecting, sieving, and transporting amphipods and control-site sediment should be clean and made of nontoxic material. They should be marked "live only," must never be used for working with formalin or any other toxic materials, and should be stored separately to avoid cross-contamination. The containers and other collection equipment should be cleaned and rinsed with distilled water, deionized water, dechlorinated laboratory water, reconstituted seawater, or natural seawater from

the collection site or an uncontaminated seawater source before use.

10.4.2.3 To minimize stress, amphipods should be handled carefully, gently, quickly, and only when necessary. Sieving should be performed by slow immersion in collection-site water. Once sieved, the amphipods should remain submersed in collection-site water at the ambient temperature at all times. Amphipods that are dropped or injured should be discarded. Once separated from the sediment, amphipods should not be exposed to direct sunlight.

10.4.2.4 *L. plumulosus* can be isolated easily from collection-site sediment by gentle sieving. Ideally, amphipods will be separated into adults, subadult, and neonates as described in Section 10.3.2. To reduce field processing time, 1.0-mm and 0.6-mm mesh sieves can be used to isolate adults and subadults with which to start a culture. Sediment passing through the 0.6-mm sieve could be temporarily used for holding until further processing of the sediment is practical. The final sieving of collection-site sediment through 0.25-mm mesh can be deferred until materials are returned to the laboratory. Collection-site water should be used to sieve sediment in the field.

10.4.2.5 No sediment should be placed in transport containers, with collection-site water. Detritus and predators recovered by sieving should be removed, and the collected amphipods should be gently washed into the transport containers with collection-site water. An adequate portion of collection-site sediment should be returned with the amphipods to serve as both laboratory holding sediment and control sediment in toxicity tests.

10.4.2.6 Water salinity and temperature at the surface and bottom of the collection site should be measured and recorded.

10.4.2.7 During transport to the laboratory, amphipods should be held at or slightly below the collection-site temperature. Containers of amphipods and sediment should be transported to the laboratory in coolers; ice-packs might be necessary to maintain temperature. The water in the containers of amphipods should be aerated if transport time exceeds 1 h.

10.4.2.8 Holding and acclimation procedures are the same as those described in Sections 10.3.2 through 10.3.7 for initiation of a culture.

10.4.3 Shipping Methods

10.4.3.1 It is critical that demonstrated shipping methods are used to ensure that organisms arrive in a healthy condition. Additionally, the amphipods that are received by a laboratory should meet the shipping acceptance criteria recommended in Section 10.4.4.3.

10.4.3.2 *L. plumulosus* should be shipped in water only. Care must be taken to select containers with a firm seal that is not easily broken in shipment. The containers are filled to the top with well-aerated water. No more than 100 amphipods/L should be added to each container. For shipping, sealable plastic bags, cubitainers, and other sealable plastic containers can be used. The containers should be filled with well-aerated collection-site water or culture water before they are sealed. The double packing bags should be placed in a container that has a protective layer of material (i.e., Styrofoam or newspaper) sufficiently thick to prevent excessive movement with an underlayer of ice packs. The shipping container should be marked to prevent it from being inverted.

10.4.3.3 Performance Criteria for Shipped Amphipods. The process of ensuring the availability of healthy amphipods for starting cultures begins before the organisms arrive in the laboratory from the supplier. It is desirable to assess the quality and acceptability of each batch of shipped amphipods using the criteria that follow. Biological criteria should include an exhibition of active swimming, crawling, or burrowing behavior upon placement in water, and an acceptable color. *L. plumulosus* should be brownish or orangish-gray. Mortality among the shipped organisms should not exceed 5%. The shipping containers should arrive intact, and the temperature of water in shipping containers should be between 10°C and 20°C. Information on culture conditions, including at least temperature and salinity, should be provided by the supplier. Finally, a quantity of

collection-site sediment should be included as substratum for amphipods for use during the acclimation period, in culturing, and in toxicity tests as control sediment. It might be desirable for the testing facility to specify these criteria for the supplier. If these criteria are not met, the organisms might have experienced stress during shipment, and culturing success might be jeopardized.

10.5 Obtaining Amphipods for Test

10.5.1.1 The cultures usually can be harvested approximately 4 to 5 weeks after initiation or up until the cultures are thinned and renewed (6 to 8 weeks after initiation). Neonates used for testing may be selected on the basis of size or age. For size-selected neonates, the contents of culture bins are gently sieved through 0.60-mm and 0.25-mm screens. Juveniles passing through the 0.6-mm mesh and retained on the 0.25-mm mesh are used for testing, and individual neonates typically have a dry weight of approximately 0.03 mg to 0.06 mg and body length of approximately 1.3 mm to 1.7 mm. Culture bins of about 35 cm X 30 cm typically produce at least 300 to 400 neonates with a healthy culture. Selecting neonates for testing based on size is the preferred option for method comparability. For age-selected neonates, gravid females are isolated from cultures 5 d before test initiation. Gravid females are placed in separate culture bins with sediment and are fed. Two days prior to test initiation, these females are then transferred to bins containing only water (at 25°C and 5‰ or 20‰). On the day of test initiation, the contents of these bins are gently passed through a 1-mm screen on which adults are retained. Neonates that pass through this screen are transferred to a shallow glass container for sorting. Special care must be taken to ensure that the neonates are handled gently, selecting and transferring them with wide-bore pipets only, and maintaining the water temperature and salinity within recommended test conditions.

10.5.1.2 Approximately one-third more amphipods than are needed for the test should be sieved from the sediment and transferred to a sorting tray. The additional organisms allow for the selection of healthy, active individuals. Organisms not used in toxicity tests can be used to establish new cultures.

10.5.2 Acceptability of Organisms

10.5.2.1 Amphipods placed in the holding bins should be active and healthy. Sluggish or apparently dead individuals should be discarded. If greater than 5% of the amphipods in the holding bins appear unhealthy or are dead, the entire group should be discarded and not used in tests.

10.6 Minimization of Risk of Release of Nonindigenous Organisms

10.6.1 If *L. plumulosus* is not endemic to the local estuarine environment, containment and water treatment procedures should be implemented to minimize the chance of accidental release of organisms or pathogens to local waters. The same precautions might also be required if the culture population of *L. plumulosus* is not derived from local sources. Some local or state authorities might require special permits and procedures to allow receipt and culturing of nonindigenous species. Containment and treatment policies and procedures could include the procedures described below. All test animals should be destroyed at the end of toxicity tests.

10.6.2 Culturing and holding of the amphipods should only occur in specially designated laboratory areas, separate from those used to hold, culture, or experiment with native species. These areas should have no access to drains leading directly to local surface waters. Handling of nonindigenous species should be limited to trained and authorized personnel. The amphipods should be cultured in a static-renewal manner to minimize the amount of water that must be treated. Any seawater removed from culture bins should be treated with chlorine bleach or ozonation to kill any escaping organisms and pathogens. All equipment and labware used to culture or handle the amphipods should be cleaned thoroughly. Any excess or dead amphipods should be placed in bleach or treated by ozonation or heat killed (boiling water) to ensure they are killed prior to disposal as sanitary waste.

Section 11
Leptocheirus plumulosus 28-d Chronic Test for Sediment

11.1 Introduction

11.1.1 *Leptocheirus plumulosus* has been used extensively to test the toxicity of estuarine and marine sediment. The choice of this amphipod species as a test organism is based on sensitivity to sediment-associated contaminants, availability and ease of collection and culturing, tolerance of environmental conditions (e.g., temperature, salinity, grain-size), ecological importance, ease of handling in the laboratory, and ease of measuring test endpoints. Additionally, this species is intimately associated with sediment by nature of its burrowing and feeding habits. *L. plumulosus* is tolerant of salinity values between >1‰ to 35‰ and sediment from fine- to coarse-grained. Field validation studies have shown that amphipods are absent or have reduced abundances at sites where toxicity has been demonstrated in laboratory tests. Amphipod sediment toxicity tests have been successfully performed for regulatory and research purposes by numerous laboratories, including state and federal government agencies, private corporations, and academic institutions (see Section 1 for additional details).

11.1.2 Guidance for *L. plumulosus* has been developed previously (ASTM, 2000c; USEPA, 1994d). Most standard whole sediment toxicity tests have been developed to produce a survival endpoint with potential for a sublethal endpoint (reburial) with some species. Methods that measure sublethal effects have either not been previously available or used routinely to evaluate sediment toxicity (Craig, 1984; Dillon and Gibson, 1986; Ingersoll and Nelson, 1990; Ingersoll, 1991; Burton et al., 1992). Most assessments of contaminated sediment rely on short-term lethality testing methods (e.g., #10 d; USEPA-USACE, 1991, 1998). Short-term lethality tests are useful in identifying "hot spots" of contamination, but may not be sensitive enough to evaluate moderately contaminated areas. However, sediment quality assessments using sublethal responses of benthic organisms, such as growth and reproduction, have been used to successfully evaluate moderately contaminated areas (Scott, 1989; Niewlony et al. 1997; DeWitt et al. 1997c).

11.1.3 The 28-d toxicity test with *L. plumulosus* is a test with a lethality endpoint and two sublethal endpoints: growth and reproduction. These sublethal endpoints have potential to provide a toxic response to chemicals that might not cause acute effects or significant mortality in a test. Sublethal response in 28-d exposures is also valuable for population modeling of contaminant effects. This data can be used for population-level risk assessments of benthic pollutant impacts.

11.1.4 Section 11.2 describes guidance for conducting the 28-d test with *L. plumulosus* that can be used to evaluate the effects sediment contaminants on survival, growth, and reproduction. Refinement of these methods may be described in future editions of this manual, after additional laboratories have successfully used this method (Section 13.5). These methods are based on procedures described in DeWitt et al. (1997a; 1997b) and Emery et al. (1997).

11.1.5 Results of tests using procedures different from the procedures described in Section 11.2 may not be comparable, and these different procedures may alter contaminant bioavailability. Comparisons of results obtained using modified versions of these procedures might provide useful information concerning new concepts and procedures for conducting sediment tests with estuarine and marine organisms. If tests are conducted using procedures different from those described in this manual, additional tests are required to determine comparability of results (Section 1.3).

11.2 Procedure for Conducting a *Leptocheirus plumulosus* 28-d Test for Measuring Sublethal Effects of Sediment-associated Contaminants

11.2.1 Recommended conditions for conducting a 28-d chronic sediment toxicity test with *L. plumulosus* are summarized in Table 11.1. A general activity schedule is outlined in Table 11.2. Decisions concerning the various aspects of experimental design, such as the number of treatments and water quality characteristics, should be based on the purpose of the test and the methods of data analysis (Section 12).

11.2.2 The 28-d chronic sediment toxicity test with *L. plumulosus* is conducted at 25°C and a salinity of either 5‰ or 20‰ with a 16 h light : 8 h dark photoperiod at an illuminance of about 500 to 1000 lux (Table 11.1). Test chambers are 1-L glass chambers containing 175 mL of sediment and about 725 mL of overlying seawater. Twenty neonate amphipods are added to each test chamber at the start of a test. Five replicate test containers per treatment are recommended for routine testing (Section 13.5.1.2). Exposure is static-renewal with water exchanges and feeding three times per week, on Monday, Wednesday, and Friday. The test organisms are fed after water renewals. Overlying water can be culture, surface, site, or reconstituted water adjusted to the test salinity. For site-specific evaluations, the characteristics of the overlying water should be as similar as possible to the site where sediment is collected. Requirements for test acceptability are summarized in Table 11.3.

11.3 General Procedures

11.3.1 Sediment into Test Chambers

11.3.1.1 The day before the addition of amphipods (Day-1), each test sediment, including control and reference sediment, should be homogenized among replicate beakers. This can be achieved by mixing, by stirring manually, or by using a rolling mill, feed mixer, or other apparatus (Section 8.3.1.2) or by serially spooning out small aliquots of sediment to each test chamber. If a quantitative confirmation of homogeneity is required, replicate subsamples should be taken from the sediment batch and analyzed for TOC, chemical concentrations, and particle size. Ammonia can be measured in the pore water.

11.3.1.2 A 175-mL aliquot of sediment is added to each test chamber with five replicates per sediment treatment. It is important that an identical volume be added to each replicate test chamber; the volume added should provide a sediment depth of 2 cm in the test chamber. The sediment added to the test chamber should be settled by tapping the bottom or side of the test chamber against the palm of the hand or another soft object. Alternatively, sediment can be smoothed with a nylon, fluorocarbon, glass, or polyethylene spatula. Sediment known or suspected to be contaminated should be added to test chambers in a certified laboratory fume hood.

11.3.2 Addition of Overlying Water

11.3.2.1 The procedure for addition of overlying water should not suspend significant portions of the sediment in test chambers. A turbulence reducer can be used to minimize disruption of sediment as test water is added. The turbulence reducer can be either a disk cut from polyethylene, nylon, or Teflon® sheeting (4 to 6 mil) attached to a nylon monofilament line (or nontoxic equivalent), or a glass or plastic plate attached (open face up) to a glass or plastic rod. The turbulence reducer needs to fit inside the test chamber. It is positioned just above the sediment surface and raised as water is added. It is convenient to mark each test chamber on the side at 900 mL and to fill with water to reach the mark. A turbulence reducer can be rinsed with clean water between replicates of a treatment, but a separate turbulence reducer should be used for each treatment. The test chambers should be covered, and placed in a temperature controlled water bath (or acceptable equivalent) in randomly assigned positions. Aeration is started when suspended sediment has settled (often overnight). A test begins when the test organisms are added to the test chambers (Day 0).

Table 11.1 Test Conditions for Conducting a 28-d Sediment Toxicity Test with *Leptocheirus plumulosus*

Parameter	Conditions
1. Test type:	Whole sediment toxicity test, static-renewal
2. Test sediment grain size:	>5% silt and clay to <85% clay
3. Test sediment pore water salinity:	1‰ to 35‰
4. Overlying water salinity:	Daily limits: 5‰ (+3‰) if pore water is 1‰ to 10‰, 20‰ (+3‰) if pore water is >10‰ to 35‰; 28-d mean: 5‰ (+2‰) or 20‰ (+2‰)
5. Test sediment pore water ammonia:	< 60mg/L (total mg/L, pH 7.7); < 0.8 mg/L (unionized mg/L, pH 7.7)
6. Test sediment pore water sulfides:	Not established.
7. Temperature:	Daily limits: 25/C (±3/C); 28-d mean: 25/C (±2/C)
8. Light quality:	Wide-spectrum fluorescent lights
9. Illuminance:	500 - 1000 lux
10. Photoperiod:	16 h light: 8 h dark
11. Test chamber:	1-L glass beaker or jar with 10-cm inner diameter
12. Sediment volume:	175 mL (about 2 cm depth)
13. Sediment preparation:	Press-sieved through 0.25-mm (see Section 4.3.2.3)
14. Overlying water volume:	Fill to 900 mL mark in test chamber (c.725 mL H_2O)
15. Renewal of overlying water:	3 times per week: siphon off and replace 400 mL
16. Source:	Laboratory cultures
17. Life stage and size:	Neonates: age-selected (<48 h old) or size-selected: retained between 0.25-mm and 0.6-mm mesh screens
18. Number test organisms/chamber:	20
19. Number of replicate chambers/ treatment:	5 for toxicity test; >2 dummy chambers for pore water ammonia (Day 0 and Day 28)
20. Diet:	Days 0-13, 20 mg TetraMin® per test chamber; Days 14-28, 40 mg TetraMin® per test chamber
21. Feeding schedule:	3 times per week (M-W-F) after water renewal.
22. Aeration and dissolved oxygen (DO):	aerate constantly with tickle flow of bubbles Daily limits: >3.6 mg/L (50% saturation) 28-d mean: >4.4 mg/L (60% saturation)
23. Overlying water:	Clean seawater, natural or reconstituted water; same source as used for culturing.

Table 11.1 (continued)

Parameter	Conditions
24. Overlying water quality and monitoring frequency:	Daily temperature in water bath or test or dummy chamber, daily min/max recommended; salinity, temperature, DO, and pH at test initiation and termination, and in one replicate per sediment treatment preceding water renewal during the test (three times per week); aeration rate daily in all containers; total ammonia on Days 0 and 28 in one replicate per treatment.
25. pH:	7.0 to 9.0 pH units
26. Pore water quality:	Total ammonia, salinity, temperature, and pH of pore water from surrogate containers on Days 0 and 28; recommended in bulk sediment prior to testing.
27. Test duration:	28 d
28. Test organism observations:	Observe condition and activity in each test chamber preceding water renewal (3 times per week).
29. Endpoints:	Survival, growth rate, and reproduction.
30. Test acceptability:	Minimum mean control survival of 80%, growth and reproduction measurable in all control replicates, and satisfaction of performance-based criteria outlined in Table 11.3.

Table 11.2 General Activity Schedule for Conducting a 28-d Sediment Toxicity Test with *Leptocheirus plumulosus*

Day	Activity
<u>Preparation</u>	
Pretest	Start or renew cultures approximately 6 to 8 weeks in advance of test initiation. Increase culture water temperature to about 25°C approximately 2 weeks in advance of test initiation.
Pretest	Determining pore water salinity of test sediment and acclimate *L. plumulosus* cultures to overlying water salinity to be used in testing.
Day-1	Layer sediment in test chambers, add overlying water. Measure pore water total ammonia in bulk sediment and begin purging procedures, if appropriate (Section 11.4.5). Measure tare weight of weigh boats for dry weights. Set up positive control reference-toxicity test chambers if appropriate.
<u>Initiation</u>	
Day 0	Measure pore water total ammonia, temperature, salinity, and pH from dummy jar. Measure salinity, temperature, DO, and pH in all test chambers. If water quality parameters are within test ranges, proceed with initiation; if not, correct problem and re-measure water quality. Obtain neonate test organisms, initiate test, and initiate positive control reference toxicant test if conducted. Only feed if a Monday, Wednesday, or Friday. Prepare 3 sets of 20 neonates for initial weight of growth rate endpoint; rinse in deionized water; dry overnight at 70°C, and weigh or measure length on Day 1 or later.
<u>Positive Control Reference-toxicity Test</u>	
Day 1 to 3	Measure and record water quality parameters in one replicate test chamber from each positive control treatment.
Day 4	Measure water quality parameters and record observations of amphipod activity in all positive control test chambers. Terminate the positive control references-toxicity control test if conducted.
<u>Maintenance of 28-d Test</u>	
Daily	Check aeration in all test chambers and test temperature (water bath, environmental chamber, or dummy chamber). If aeration is interrupted in a test chamber, measure and record DO prior to resumption of aeration. Check photoperiod controllers.
3 Times per Week (M-W-F)	Measure water quality in one replicate test chamber per sediment treatment. Record observations of amphipod activity and condition of sediment and water in all test chambers. Siphon off and replace 400 mL of water in all test chambers. Add food to all test chambers.
<u>Termination of 28-d Test</u>	
Day 28	Measure salinity, temperature, DO, and pH in all test chambers. Measure tare weight of weight boats for dry weight measurements. Terminate 28-d test: sieve adults and offspring from sediment, count surviving adults, prepare adults for drying, and dry to constant weight at 70°C. Count offspring, or preserve and stain offspring.
Day 29 or later	Measure dry weight or length of adults. If offspring were preserved, count them.

Table 11.3 Test Acceptability Requirements for a 28-d Sediment Toxicity Test with *Leptocheirus plumulosus*

A. It is recommended for conducting the 28-d test with *L. plumulosus* that the following performance criteria be met:

1. Neonate *L. plumulosus*, size-selected (retained between 0.25-mm and 0.6-mm screens) or age selected (<24h old), are used to initiate the test(s).
2. Average survival of amphipods in the negative control sediment must be greater than or equal to 80% at the end of the test, with no single replicate having 60% survival or less.
3. Measurable growth and reproduction should be observed in all replicates of the negative control treatment.
4. The time-weighted average of daily temperature readings must be within ±2°C of the desired temperature. The instantaneous temperature must always be within ±3°C of desired temperature.
5. The time-weighted average of daily salinity readings must be 5‰ ±2‰ or 20‰ ± 2‰. The instantaneous salinity readings must always be 5‰ ±3‰ or 20‰ ± 3‰.

B. Performance-based criteria for culturing *L. plumulosus* include the following:

1. Laboratories should perform periodic 96-h water-only reference-toxicity tests (at a minimum, one test every six months) to assess the sensitivity of culture organisms (Section 9.16).
2. Records should be kept on the frequency of restarting cultures.
3. Laboratories should record the pH and ammonia of the culture water at least quarterly. Dissolved oxygen and salinity should be measured weekly. Temperature should be recorded daily.
4. Laboratories should characterize and monitor background contamination and nutrient quality of food if problems are observed in culturing or testing organisms.

C. Additional requirements:

1. A negative-control sediment and appropriate solvent controls must be included in a test. The concentration of solvent used must not adversely affect test organisms.
2. All organisms in a test must be from the same source.
3. All test chambers should be identical and should contain the same amount of sediment and overlying water.
4. Natural physico-chemical characteristics of test sediment collected from the field should be within the tolerance limits of the test organisms.
5. Storage of sediments collected from the field should follow guidance outlined in Section 8.2.
6. Salinity, pH, and DO, in the overlying water, ammonia in pore water and test sediment grain size should be within test condition limits of the test species (Table 11.1), or else effects of the variables need to be considered during interpretation of test results.

11.3.3 Initial Measurements

11.3.3.1 On Day 0, water quality must be measured in all test chambers prior to adding amphipods to test chambers. If any water quality parameter is outside acceptable limits (Table 11.1), correct the problem in all replicate containers of that treatment, re-measure water quality parameters, and continue test initiation if water quality values are acceptable. Aberrant pH values might be caused by characteristics of certain sediments and therefore may be impractical to correct.

11.3.4 Acclimation

11.3.4.1 Test organisms should be cultured at a temperature near 25°C. Amphipod cultures held below 23°C need to be acclimated to test temperature of 25°C (±3°C) before test initiation. Ideally, test organisms should be cultured in the same water that will be used in testing.

11.3.4.2 Occasionally there is a need to perform evaluations at temperatures or salinity's different than those recommended in Table 11.1. Under these circumstances, it may be necessary to acclimate organisms to the desired test temperature or salinity to prevent thermal shock that could result when organisms are moved immediately from the culture temperature or salinity to the test temperature or salinity (ASTM, 2000a). Reproduction and growth rates in cultures may be greatly reduced at temperatures <20°C. However, reproduction and growth is not effected by salinity's ranging from 5‰ and 20‰ (DeWitt et al., 1997a). Acclimation can be achieved by exposing organisms to a gradual change in temperature or salinity; however, the rate of change should be relatively slow to prevent shock. A change in temperature or salinity not exceeding 3°C or 3‰ per 24-h period is strongly recommended (see Sections 10.2.3 and 10.3.2). Tests at temperatures other than 25°C need to be preceded by studies to determine expected performance under alternate conditions.

11.3.5 Addition of Amphipods

11.3.5.1 The test is initiated when amphipods are added to the test chambers. See Section 10.5 for procedures for obtaining neonates for testing. Amphipods should be randomly selected and placed in transfer containers (small dishes or eye cups) containing a small amount of test water. The number of amphipods in each dish should be verified by recounting before organisms are added to test chambers. To facilitate recounting, amphipods may be distributed to test chambers in batches of 5 or 10 instead of the full complement of 20. Because neonates are very small, extreme caution should be taken to ensure that each test chamber receives all 20 amphipods at test initiation. The distribution of amphipods to the test chambers needs to be done in a randomized fashion. Animals need to be added to test chambers as soon as possible following their collection to minimize handling stress and exposure to temperature changes. Three randomly selected sets of 20 neonates for initial weight determination needs to be set aside during initiation of the test.

11.3.5.2 To facilitate the initiation process, aeration should be stopped in test chambers immediately prior to adding the neonates. Sediment in test chambers should not be disrupted during the initiation procedure. Neonates from a transfer container should be poured into a test chamber. Any neonates remaining in transfer containers can be washed immediately into the test chamber using a gentle stream of water at appropriate temperature and salinity. Neonates trapped at the water's surface can be submerged by using a blunt probe or by gently dribbling a few drops of test or culture water onto the amphipod from above. A disk of 6-mil polyethylene, nylon, or Teflon® can be used on the water surface to minimize disruption of the sediment surface, if necessary. Rinse the disk after amphipods are added to ensure that none have stuck to the disk. The disk should be removed once the amphipods have been introduced. A separate disk should be used for each treatment to avoid cross contamination. Aeration is continued after amphipods are added to test chambers.

11.3.5.3 After the test organisms have been added, the test chambers should be examined for individuals that did not burrow into the sediment and might have been stressed or injured during the isolation, counting, or initiation processes.

Injured or stressed test organisms will not burrow into sediment and should be removed. Neonates that have not burrowed within 1 h should be replaced with test organisms from the same sieved population, unless they are repeatedly burrowing into the sediment and immediately emerging in an apparent avoidance response. In that case, the amphipods are not replaced. The number of amphipods that are replaced in each test chamber needs to be recorded.

11.3.6 Test Conditions

11.3.6.1 Test limits for the 28-d *L. plumulosus* test are provided in Table 11.1. Test sediments with characteristics that exceed these limits are subject to noncontaminant effects that needs to be considered during interpretation of test results.

11.3.6.2 Aeration. The overlying water in each test chamber needs to be aerated continuously after an initial settling period, except during introduction of the test organisms. Filtered, dry, clean air should be bubbled through a glass or plastic pipet via plastic tubing (about 3 bubbles/sec). The tip of the pipet should be suspended 2 cm to 3 cm above the surface of the sediment so that it does not disturb the sediment surface. The concentration of DO in the water overlying the sediment in the test chambers is maintained at or near saturation by gentle aeration. Ideally, air is bubbled through the water at a rate that maintains a high percentage of saturation (e.g., about 90%) but does not disturb the sediment surface. If air flow to one or more test chambers is interrupted (i.e., for more than 1 h), DO should be measured in those test chambers to determine whether DO concentrations have fallen below 4.4 mg/L. The 28-d mean should be >4.4 mg/L DO, and daily DO measurements should be >3.6 mg/L (50% saturation). Results may be unacceptable for test chambers in which aeration is interrupted or DO concentrations fall to below 50% of saturation.

11.3.6.3 Lighting. Laboratory lighting should be maintained on a 16 h light: 8 h dark photoperiod cycle throughout the test at an intensity of 500 to 1000 lux.

11.3.6.4 Feeding and Water Renewal. A TetraMin®-only diet is recommended for the 28-d sediment toxicity test with *L. plumulosus*. With this diet, 400 mL of overlying water is replaced three times per week (Monday-Wednesday-Friday), after which a TetraMin® slurry is delivered to each chamber in 1-mL aliquots. Water removal and replacement must be completed using procedures that minimize disturbance to sediment in the test chambers. Water can be removed by siphoning through a tube with fine-meshed screening over the intake to prevent uptake of amphipods. A pump can also be used to remove water. Water should not be poured from test chambers because this practice can resuspend and disturb the sediment. A separate turbulence-reducer should be used for each treatment when water is replaced to avoid cross contamination (see section 11.3.2). TetraMin® is fed at a rate of 20 mg per test chamber between Days 0-13 and 40 mg per test chamber between Days 14-28. To prepare the slurry, TetraMin® is finely ground with a food mill (blender, mortar and pestle, or a similar device) and sieved through a 0.25-mm screen. Test water is added to the appropriate amount of TetraMin®, and the slurry is mixed on a stir plate for 15 min. Appendix A provides a sample calculation for preparation of food rations. The slurry is prepared fresh for each use and needs to be mixed continuously during feeding to prevent the TetraMin® from settling.

11.3.6.4.1 Laboratory experimentation has shown that food ration can affect the response of test animals to sediment-associated contaminants. The food ration of TetraMin® recommended in this protocol was evaluated with two other food rations in an experiment in which test animals were exposed to sediments spiked with PCB29 at concentrations between 15 and 240 ppm (T. Bridges, USACE, personal communication). The feeding rates evaluated at each PCB29 concentration included 30 mg/60 mg (Days 0-13/Days 14-18), 20 mg/40 mg and 10 mg/20 mg per test chamber. Significant reductions in survival and growth were evident only in the highest PCB29 concentration for each of the food rations. Decreased reproduction was also evident at 240 ppm PCB29 at each food ration as well as at 120 ppm for the 20 mg/40 mg and 10 mg/20 mg rations (T. Bridges, USACE, personal communication). Given the generally lower

reproductive rates observed at the lowest food ration, the 20 mg/40 mg feeding rate is recommended for use in this protocol.

11.3.6.5 Water Temperature. The test temperature was selected to approximate summertime temperature experienced by *L. plumulosus* in the wild (Holland et al., 1988; McGee, 1998). The test temperature is 25°C with a daily maximum range of ±3°C and a 28-d weighted mean of 25°C ±2°C. Water used for renewal of test chambers needs to be adjusted to test temperature before use in renewals.

11.3.6.6 Salinity. The salinity of the water overlying the test sediment should be 5‰ ±3‰ (daily readings) when test sediment pore water salinity is 1‰ to 10‰; an overlying water salinity should be 20‰ ±3‰ when test sediment pore water salinity is >10‰. Selection of which overlaying water salinity should be based on the pore water salinity of the samples to be tested. If the suite of samples includes sediments with pore water salinity values spanning the range of both less-than and greater-than 10‰, use the appropriate overlying water salinity for each sample (i.e., 5‰ or 20‰), and include control-sediment treatments for both 5‰ and 20‰ overlying water salinity values. The 28-d mean salinity values should deviate no more than 2‰ from the recommended salinity (5‰ or 20‰). Pore water salinity of each test sediment should be measured prior to the initiation of a test. Sediment pore water can be measured in water overlying sediment in sample containers before homogenization of sediment. Alternatively, pore water salinity can be obtained by centrifugation (see Section 8.4.4.7).

11.3.7 Measurements and Observations

11.3.7.1 Temperature should be measured at least daily in a dummy chamber or from the water bath or environmental chamber. The temperature of the water bath or a test chamber should be continuously monitored with minimum and maximum temperature recorded daily. A dummy container identical to test containers is recommended for continuous temperature monitoring. The time-weighted average of daily temperature readings must be 25°C ±2°C. The instantaneous temperature must always be within ±3°C of the desired temperature.

11.3.7.2 Salinity, DO, temperature, and pH of the overlying water should be measured three times per week in at least one test chamber per treatment before renewal of water.

11.3.7.3 Total ammonia should be measured in overlying and pore water at test initiation (Day 0 or Day -1 for pore water) and at test termination (Day 28). Salinity, pH, and temperature should be measured with each ammonia measurement. Simultaneous measurements of ammonia, salinity, pH, and temperature in sediment pore water should be taken before test initiation. If test sediments are sieved (Section 4.3.2.3), pore water samples for ammonia should be collected before and after sieving. Pore water can be obtained by centrifugation or from overlying water in sample containers (prior to pretest homogenization). If ammonia levels exceed recommended limits (Table 11.1), then ammonia reduction procedures are advisable before test initiation. However, if ammonia is the chemical of concern in the test sediments, pore water ammonia concentrations should not be deliberately manipulated.

11.3.7.4 Each test chamber should be examined daily to ensure that airflow to the overlying water is acceptable. Daily checks for amphipods trapped at the water surface are recommended for the first three days of a test. Amphipods caught in the air-water interface should be gently pushed down into the water using a blunt glass probe or drops of dilution water. The number of amphipods swimming in the water column and trapped in the air-water interface should be noted and amphipods submerged before each water renewal. The number of apparently dead test organisms should be noted, but organisms should not be removed or otherwise disturbed during the test. Exuviae may be mistaken for dead amphipods; therefore, care should be taken in identifying animals as dead.

11.3.8 Ending a Test

11.3.8.1 The contents of each test chamber are sieved to isolate the test organisms. The mesh sizes for sieving the contents of the test chambers is 0.5 to 0.6 mm to isolate adults and 0.25 mm to isolate offspring. The 0.6-mm sieve should not be

stacked atop the 0.25-mm sieve for this process. Test water should be used for sieving. Material retained on each sieve should be washed into a sorting tray with clean test water. *L. plumulosus* are easily removed from the sediment by the sieving process.

11.3.8.2 Material that has been washed from the sieve into a sorting tray should be carefully examined for the presence of amphipods. A small portion of the material should be sorted through at a time, and amphipods should be removed as they are found. Amphipods and residual sediment retained on the 0.25-mm sieve should be rinsed briefly with freshwater to remove salts and washed into a labeled sample jar ($8 oz) using 70% alcohol (either ethyl or isopropyl). Use of a wide funnel supported by a ring stand facilitates this process. Because offspring are very small, great care is needed to ensure that all organisms are transferred from the screen to the sample jar. Add sufficient 70% alcohol to preserve the amphipods, and add about 3 mL of rose bengal solution (about 1 g/L) to stain the organisms. Offspring may be counted on test termination day, but waiting 2 to 3 d allows the amphipods to be more darkly stained.

11.3.8.3 Survival. Numbers of live and dead adult amphipods should be determined and recorded for each test chamber (see Figure 11.1). Missing adult organisms are assumed to have died, decomposed, and disintegrated during the test; they should be included in the number dead in calculations of the percentage survival for each replicate treatment. Amphipods that are inactive but not obviously dead are observed using a low-power dissecting microscope or a hand-held magnifying glass. Any organism that fails to exhibit movement (i.e., neuromuscular twitch of pleopods or antennae) upon gentle prodding with a probe should be considered dead. An independent count of survival in 10% of test chambers should be completed by a second observer. Based one the experience of one laboratory, the intralaboratory median CV for survival (sample size of 88 treatments) can be expected to be 11% (DeWitt et al. 1997b; see Section 13.5.1). Based on one study involving 10 laboratories, the interlaboratory CV for survival ranged from 4% to 19% (DeWitt el a. 1997b; see Section 13.5.2). It should be expected that intralaboratory CV for survival will decrease over time as a laboratory gains experience using this method. Similarly, the interlaboratory CV for survival should decrease from reported values here as more laboratories gain experience using this method.

11.3.8.4 Growth Rate. Growth rate of amphipods can be reported as daily change of average individual length or weight. However, measuring length is more laborious and therefore more expensive than measuring weight to determine growth rate, and does not result in an increase in sensitivity in *L. plumulosus* 28-d test (DeWitt et al., 1997a). Dry weight of amphipods can be determined as follows: (1) transferring the archived amphipods from a replicate out of the preservative into a crystallizing dish; (2) rinsing amphipods with deionized water; (3) transferring these rinsed amphipods to a preweighed aluminum pan; (4) drying these samples to constant weight at 60°C; and (5) weighing the pan and dried amphipods on a balance to the nearest 0.01 mg. Average dry weight of individual amphipods in each replicate is calculated from these data. Due to the small size of the amphipods, caution should be taken during weighing 20 dried amphipods after 28-d sediment exposure may weigh less than 25 mg). The average per-capita dry weight of adult amphipods for each replicate is the difference between the tared weight of the boat and the total weight of the boat plus dried amphipods, divided by the number of amphipods in the weigh boat. The growth rate endpoint (mg/d) is the difference between per capita adult and neonate dry weights, divided by 28 d. In other words, for each replicate, calculate: Growth Rate (mg/individual/day) = (mean adult dry weight - mean neonate dry weight)/28. Weigh pans need to be carefully handled using powder-less gloves and the balance should be calibrated with standard weights with each use. Use of small aluminum pans will help reduce variability in measurements of dry weight. Weigh boats can also be constructed from sheets of aluminum foil. Amphipod body length (±0.1 mm) can be measured from the base of the first antennae to the tip of the third uropod along the curve of the dorsal surface. The use of a digitizing system and microscope to measure length has been described

in Kemble et al. (1994) for *Hyalella azteca* and DeWitt et al. (1992a and 1997a) for *Leptocheirus plumulosus*. Based on the experience of one laboratory, the intralaboratory median CV for growth (sample size of 87 treatments) can be expected to be 3% (DeWitt et al. 1997b; see Section 13.5.1). Based on one study involving 10 laboratories, the interlaboratory CV for growth ranged from 14% to 38% (DeWitt el a. 1997b; see Section 13.5.2). It should be expected that intralaboratory CV for growth rate will decrease over time as a laboratory gains experience using this method. Similarly, the interlaboratory CV for growth rate should decrease from reported values here as more laboratories gain experience using this method.

11.3.8.5 Reproduction. The offspring should be counted within 2 weeks of terminating the test. It may be possible to count the offspring the day the experiment is broken down. If not, preserve offspring in 70% alcohol (either ethyl or isopropyl). Transfer preserved, stained offspring to a fine screen (<0.25-mm mesh) and rinse with freshwater to remove alcohol and excess stain. Rinse the live or preserved neonates into a shallow dish and count them under magnification, such as a dissecting microscope. Record the number of offspring. For QA, 10% of the samples should be recounted by a second analyst. The reproduction endpoint is calculated as the number of offspring per living adult. Based on the experience of one laboratory, the intralaboratory median CV for reproduction (sample size of 88 treatments) can be expected to be 18% (DeWitt et al. 1997b; see Section 13.5.1). Based on one study involving 10 laboratories, the interlaboratory CV for survival ranged from 35% to 102% (DeWitt el a. 1997b; see Section 13.5.2). It should be expected that intralaboratory CV for reproduction will decrease over time as a laboratory gains experience using this method. Similarly, the interlaboratory CV for reproduction should decrease from reported values here as more laboratories gain experience using this method.

11.3.9 Control Performance Issues and Revisions to the Protocol

11.3.9.1 The *Leptocheirus plumulosus* 28-d sediment toxicity test, like all experimental systems, is subject to occasional failures. Because the *L. plumulosus* 28-d sediment toxicity test is more complex and of longer duration than any of the marine amphipod short-term sediment toxicity tests, there are more opportunities for problems to occur in this long-term test than in the short-term tests. Problems with the test are most readily detected by failure to meet test acceptability criteria in the control treatment (Tables 11.1 and 11.3), such as mortality <20% or failure of amphipods to grow or reproduce. Test failures usually can be attributed to a failure to maintain one or more test requirements described in Tables 11.1 and 11.3; however, tests sometimes fail inexplicably. Possible causes for unaccountable test failures have included overfeeding (e.g., leading to anoxia or increased production of hydrogen sulfide), poor health of test animals (i.e., culture failure), or accidental introduction of toxic materials into test chambers. Scientists from the USEPA and the USACE observe that the frequency of failure decreases as the laboratory and staff using the test gain more experience through conducting the test; however, neither agency has explicit data on the frequency of failure. Users of this test should be aware of this possibility and prepare for the possibility to rerun the test on occasion. Both agencies expect that the protocol for the *L. plumulosus* 28-d sediment toxicity test will be revised periodically, especially as new experimental data reveal test conditions that reduce the probability of possible test failure.

11.4 Interpretation of Results

11.4.1 This section describes information that is useful in helping to interpret the results of sediment toxicity tests with *L. plumulosus*. Section 12 provides additional information on analyses and reporting of toxicity test data.

11.4.2 Influence of Indigenous Organisms

11.4.2.1 Indigenous organisms may be present in field-collected sediment. The presence of organisms taxonomically similar to the test organism can make interpretation of treatment effects difficult. Predatory organisms can adversely affect test organism survival. For example, Redmond and Scott (1989) showed that

the polychaete *Nephtys incisa* will consume amphipods under test conditions. All control, reference, and test sediment should be press-sieved through 0.25-mm mesh to avoid these complications. If test sediment is not sieved, the number and species of indigenous organisms should be determined to better interpret results.

11.4.3 Effects of Sediment Grain Size

11.4.3.1 *L. plumulosus* tolerates a wide range of sediment types. There is generally little effect on survival, growth rate, or reproduction when coarse-grained (sand) or fine-grained (predominantly silt and clay) sediment is used. In some studies, *L. plumulosus* has exhibited >90% survival in clean sediment ranging from nearly 100% sand to nearly 100% silt + clay (SAIC, 1993a; 1993b; Schlekat et al., 1992; J. Kavanaugh, University of West Florida, Gulf Breeze, FL, personal communication). However, adverse effects can occur in sediment with very high levels of clay or sand. Laboratory studies have shown significant reduction in survival when clay content exceeded 84%, and survival, growth and reproduction were significantly reduced in 100% sand (Emery et al., 1997). Results have been equivocal from controlled tests with mixed grained sediments (between 10% and 90% silt/clay). Emery et al. (1997) found an increase in growth as sediment coarseness increased up to 75% sand. DeWitt et al. (1997a) reported enhanced growth in finer-grained sediment as compared with more coarse-grained material, but the difference in growth was not considered to be biologically significant (DeWitt et al., 1997a). Therefore, *L. plumulosus* should be tested with sediment with silt/clay content between 5% and 85% (Table 11.1). If sediment characteristics exceed these bounds, an appropriate clean control/reference sediment should be incorporated into the test to separate effects of sediment-associated contaminants from effects of particle size.

11.4.4 Effects of Pore Water Salinity

11.4.4.1 The range of salinity in which a given species can survive when the overlying water salinity is matched to that of the pore water salinity is the salinity tolerance range. The potential for a toxic response caused by salinity alone exists if a species is exposed to conditions outside of its range of tolerance. For estuarine sediment, it is important to know the pore water salinity of each sediment before testing is started and to use overlying water of an appropriate salinity. *L. plumulosus* is not recommended for testing with truly freshwater sediments (0‰ pore water salinity) or with sediments having pore water salinity >35‰ until further testing is completed to confirm acceptable response in organisms (DeWitt et al., 1997a). This methods manual recommends use of standard salinity of overlying water for testing (i.e., 5‰ or 20‰; Table 11.1).

11.4.4.2 *L. plumulosus*, a euryhaline species, can survive and thrive in a wide range of salinity conditions. The salinity tolerance and application range for this amphipod is 1‰ to 35‰ (DeWitt et al., 1989; DeWitt et al., 1992a; SAIC, 1993b; Schlekat et al., 1992; DeWitt et al., 1997a). Although there is some evidence of salinity-related stress for *L. plumulosus* at salinity extremes, the breadth of salinity tolerance exhibited by this species is most likely sufficient for application to the majority of sediments that might be encountered in an estuarine system (i.e., interstitial salinity from 1‰ to >30‰).

11.4.4.3 This method recommends testing with an overlying water salinity of either 5‰ or 20‰; the choice of overlying water salinity is dependant on the pore water salinity of test sediment.

11.4.4.4 Although matching overlying and pore water salinity values in test containers might be appropriate for some study designs, this practice is logistically complicated and normally impractical to accomplish. Acclimation of amphipod cultures to the appropriate salinity is required. Moreover, if sediment samples to be tested have different pore water salinity values, extreme care needs to be exercised to ensure that renewals are completed with water of the appropriate salinity.

11.4.5 Effects of Sediment-associated Ammonia

11.4.5.1 Field-collected sediment may contain concentrations of pore water ammonia that are toxic to amphipods. The water-only NOEC for *L. plumulosus* is 60 mg/L (USEPA, 1994d). If ammonia concentrations are above this value at

test initiation, mortality may be due in part to effects of ammonia. Depending on test application, it might be desirable to lower the ammonia concentration by manipulating the test system prior to introduction of test organisms if measured ammonia in the pore water or overlying water is greater than the NOEC. However, if ammonia is the chemical of concern in the test sediments, pore water ammonia concentrations should not be deliberately manipulated. If sediment toxicity tests are conducted to evaluate the acceptability of dredged material for disposal, the manipulations could be performed. Section 12.3.6 discusses methods for conducting TIEs to determine whether ammonia is contributing to the toxicity of sediment samples. Manipulations involve flushing the test system by renewing a specified amount of overlying water until ammonia concentrations are reduced. The effects of dilution of ammonia on pore water concentration is not known. Due to this uncertainty, one option could be to monitor pore water concentrations.

11.4.5.2 If ammonia is of concern to the regulatory application associated with the sediment toxicity test, overlying water should be sampled approximately 1 cm above the sediment surface prior to introduction of test organisms on Day 0. Pore water ammonia should be measured when sediment samples are prepared for testing. If both the pore water and overlying water ammonia concentrations are ≤60 mg/L, then the test may proceed normally. If the ammonia concentration is >60 mg/L in a given sample, then ammonia level can be reduced by aerating the sample to saturation and replacing 2 volumes of overlying water per day. Purging pore water ammonia (up to 60 mg/L) from test sediments prior to starting the toxicity test, and employing the routine replacement of overlying water in each test chamber every other day (M-W-F) did result in a consistently reduced pore water ammonia concentration throughout the 28 days from approximately 60 mg/L to approximately 1 mg/L (DeWitt et al., 1997a). Similar results were obtained by other researchers (Moore et al. 1997; Moore et al. 1995). The analyst should measure the pore water ammonia concentration each day until it is ≤60 mg/L. The pore water ammonia threshold for the chronic sediment toxicity test was based on 28-d exposures of the amphipods to sediments with experimentally-elevated pore water ammonia (up to 60mg/L), employing the specified purging technique prior to starting the toxicity test exposure, and employing the routine replacement of overlying water (M-W-F) (DeWitt et al., 1997a). No lethal or sublethal toxicity was observed in this experiment at any one of the tested pore water ammonia concentrations, which is most likely caused by loss of ammonia from the test system due to diffusion of pore water ammonia from the sediments to the overlying water and the replacement of the overlying water three times per week. Because dummy test containers are required for pore water measurements, a minimum of two dummy containers are required (one for Day 0 and one for Day 28). Additional dummy containers should be prepared if pore water ammonia levels are high enough to require several successive days for pore water ammonia reduction. When ammonia concentrations are reduced to ≤60 mg/L, testing should be initiated by adding test organisms.

11.4.6 Future Research

11.4.6.1 Research to find methods that reduce the variability of the growth rate and reproduction endpoints could lead to improvements in the statistical power of the *L. plumulosus* chronic toxicity test. A second “round-robin” study, using only laboratories with considerable experience running this toxicity test, could provide improved estimates of the interlaboratory accuracy and precision of each endpoint. Additional research is needed to evaluate the relative toxicological sensitivity of the lethal and sublethal endpoints to a wide variety of chemicals spiked in sediment and to mixtures of chemicals in sediments from pollution gradients in the field. Additional research is needed to evaluate the ability of the test’s lethal and sublethal endpoints to estimate the responses of populations and communities of benthic invertebrates to contaminated sediments. Research is also needed to link the toxicity test’s endpoints to a field-validated population model of *L. plumulosus* that would then generate estimates of population-level responses of the amphipod to test sediments and thereby provide additional ecologically relevant interpretive guidance for the toxicity test.

Section 12
Data Recording, Data Analysis and Calculations, and Reporting

12.1 Data Recording

12.1.1 Quality assurance project plans with data quality objectives and SOPs should be developed before starting a test. Procedures should be developed by each laboratory to record, verify, and archive data (USEPA, 1999).

12.1.2 A file should be maintained for each sediment test or group of tests on closely related samples (Section 9). This file should contain a record of the sample chain-of-custody; a copy of the sample log sheet; the original bench sheets for the test organism responses during the sediment test(s); chemical analysis data on the sample(s); control data sheets for reference toxicants; detailed records of the test organisms used in the test(s), such as source, age, date of receipt, and other pertinent information relating to their history and health; information on the calibration of equipment and instruments; test conditions used; and results of reference toxicity tests. Original data sheets should be signed and dated by the laboratory personnel performing the tests. A record of the electronic files of data should also be included in the file.

12.1.3 Example data sheets are in Appendix A.

12.2 Data Analysis

12.2.1 Statistical methods are used to make inferences about populations, based on samples from those populations. In most sediment toxicity tests, test organisms are exposed to contaminated sediment to estimate the response of the population of laboratory organisms. The organism response to these sediments is usually compared with the response to a control or reference sediment. In any toxicity test, summary statistics, such as means and standard errors for response variables (e.g., survival), should be provided for each treatment (e.g., pore water concentration, sediment).

12.2.1.1 Types of Data

12.2.1.1.1 Two types of data and three endpoints (survival, growth rate, and reproduction) will be obtained from the 28-d *L. plumulosus* chronic test. Survival is a dichotomous or categorical type of data. Growth rate and reproduction are representative of continuous data.

12.2.1.2 Sediment Testing Scenarios

12.2.1.2.1 Sediment tests are conducted to determine whether contaminants in sediment are harmful to benthic organisms. Sediment tests are commonly used in studies designed to 1) evaluate dredged material, 2) assess site contamination in the environment (e.g., to rank areas for cleanup), and 3) determine effects of specific contaminants, or combinations of contaminants, through the use of sediment spiking techniques. Each of these broad study designs has specific statistical design and analytical considerations, which are detailed below.

12.2.1.2.2 Dredged Material Disposal Suitability In these studies, each site is compared individually to a reference sediment. The statistical procedures appropriate for these studies are generally pairwise comparisons. Additional information on toxicity testing of dredged material and analysis of data from dredged material disposal suitability evaluations is available in USEPA-USACE (1991; 1998).

12.2.1.2.3 Site Assessment of Field Contamination. Surveys of sediment toxicity often are included in more comprehensive analyses of biological, chemical, geological, and hydrographic data. Statistical correlation can be improved and costs can be reduced if subsamples

are taken simultaneously for sediment toxicity or bioaccumulation tests, chemical analyses, and benthic community structure determinations. There are several statistical approaches to field assessments, each with a specific purpose. If the objective is to compare the response or residue level at all sites individually with a control or reference sediment, then the pairwise comparison approach described below is appropriate. If the objective is to compare all sites in the study area, then a multiple comparison procedure that employs an experiment-wise error rate is appropriate. If the objective is to compare among groups of sites, then orthogonal contrasts are a useful data analysis technique.

12.2.1.2.4 Sediment-Spiking Experiments. Sediment spiked with known concentrations of contaminants can be used to establish cause-and-effect relationships between chemicals and biological responses. Results of toxicity tests with test materials spiked into sediment at different concentrations can be reported in terms of an LC50, EC50, IC50, NOEC, or LOEC. The statistical approach outlined above for spiked-sediment toxicity tests also applies to the analysis of data from sediment dilution experiments or water-only reference-toxicity tests.

12.2.2 Experimental Design

12.2.2.1 The guidance outlined below on the analysis of sediment toxicity test data is adapted from a variety of sources including USEPA (1991a; 1994a; 1994b; 1994c), and USEPA-USACE (1998). The objectives of a sediment toxicity test are to quantify contaminant effects on test organisms exposed to natural or spiked sediment or dredged materials and to determine whether these effects are statistically different from those occurring in a control or reference sediment. Each experiment consists of at least two treatments: the control and one or more test treatment(s). The test treatment(s) consist(s) of the contaminated or potentially contaminated sediment(s). A control sediment is always required to ensure that no contamination is introduced during the experimental set-up and that test organisms are healthy. A control sediment is used to judge the acceptability of the test (Table 11.3). Designs other than those for sediment-spiking experiments also require a reference sediment that represents an environmental condition or potential treatment effect of interest. The reference sediment is defined as a relatively uncontaminated sediment and is used as the standard with which all test sediments are compared. Testing a reference sediment provides a site-specific basis for evaluating toxicity of the test sediments. Comparisons of test sediments to multiple reference or control sediments representative of the physical characteristics of the test sediment (i.e., grain size, organic carbon) may be useful in these evaluations (Section 2.1.2).

12.2.2.2 Experimental Unit

12.2.2.2.1 During toxicity testing, each test chamber to which a single application of treatment is applied is an experimental unit. The important concept is that the treatment (sediment) is applied to each experimental unit as a discrete unit. Experimental units should be independent and should not differ systematically.

12.2.2.3 Replication

12.2.2.3.1 Replication is the assignment of a treatment to more than one experimental unit. The variation among replicates is a measure of the within-treatment variation and provides an estimate of within-treatment error for assessing the significance of differences between treatments.

12.2.2.4 Minimum Detectable Difference (MDD)

12.2.2.4.1 As the minimum difference between treatments which the test is required or designed to detect decreases, the number of replicates required to meet a given significance level and power increases. Because no consensus currently exists on what constitutes a biologically acceptable minimum detectable difference (MDD), the appropriate statistical minimum significant difference should be a data quality objective (DQO) established by the individual user (e.g., program considerations) based on their data requirements, the logistics and economics of test design, and the ultimate use of the sediment toxicity test results.

12.2.2.5 Minimum Number of Replicates

12.2.2.5.1 Five replicates are recommended for 28-d marine and estuarine sediment toxicity testing for each control or experimental treatment. However, it is always prudent to include as many replicates in the test design as are economically and logistically possible. Both the 10-d and 28-d sediment toxicity testing methods recommend the use of 20 organisms per replicate for marine testing (USEPA, 1994a). An increase in the number of organisms per replicate in all treatments, including the control, is allowable only if (1) test performance criteria for the recommended number of replicates are achieved, and (2) it can be demonstrated that no change occurs in contaminant availability as a result of increased organism loading. See Table 11.1 for a description of the number of replicates and test organisms/replicate recommended for long-term testing of *L. plumulosus*.

12.2.6.6 Randomization

12.2.2.6.1 Randomization is the unbiased assignment of treatments within a test system and to the exposure chambers ensuring that no treatment is favored and that observations are independent. It is also important to (1) randomly select the organisms (but not the number of organisms) for assignment to the reference and test treatments (e.g., a bias in the results might occur if all the largest animals are placed in the same treatment), (2) randomize the allocation of sediment (e.g., not take all the sediment in the top of a jar for the control and the bottom for spiking), and (3) randomize the location of exposure units.

12.2.2.7 Pseudoreplication

12.2.2.7.1 The appropriate assignment of treatments to the replicate exposure chambers is critical to the avoidance of a common error in design and analysis termed "pseudoreplication" (Hurlbert, 1984). Pseudoreplication occurs when inferential statistics are used to test for treatment effects even though the treatments are not replicated or the replicates are not statistically independent (Hurlbert, 1984). The simplest form of pseudoreplication is the treatment of subsamples of the experimental unit as true replicates. For example, two aquaria are prepared, one with reference sediment, the other with test sediment, and 10 organisms are placed in each aquarium. Even if each organism is analyzed individually, the 10 organisms only replicate the biological response and do not replicate the treatment (i.e., sediment type). In this case, the experimental unit is the 10 organisms and each organism is a subsample. A less obvious form of pseudoreplication is the potential systematic error due to the physical segregation of exposure chambers by treatment. For example, if all the reference exposure chambers are placed in one area of a room and all the test exposure chambers are in another, spatial effects (e.g., different lighting, temperature) could bias the results for one set of treatments. Random physical intermixing of the exposure chambers or randomization of treatment location might be necessary to avoid this type of pseudoreplication. Pseudoreplication can be avoided or reduced by properly identifying the experimental unit, providing replicate experimental units for each treatment, and applying the treatments to each experimental unit in a manner that includes random physical intermixing (interspersion) and independence. However, avoiding pseudoreplication may be difficult or impossible given resource constraints.

12.2.2.8 Optimum Design of Experiments

12.2.2.8.1 An optimum design is one which obtains the most precise answer for the least effort. It maximizes or minimizes one of many optimality criteria, which are formal, mathematical expressions of certain properties of the model that are fit to the data. Optimum design of experiments using specific approaches described in Atkinson and Donev (1992) has not been formally applied to sediment testing; however, it might be desirable to use the approaches in experiments. The choice of optimality criterion depends on the objective of the test, and composite criteria can be used when a test has more than one goal. A design is optimum only for a specific model, so it is necessary to know beforehand which models might be used (Atkinson and Donev, 1992).

12.2.2.9 Compositing Samples

12.2.2.9.1 Decisions regarding compositing of samples depend on the objective of the test. Compositing is used primarily in bioaccumulation experiments when the biomass of an individual organism is insufficient for chemical analysis. Compositing consists of combining samples (e.g., organisms, sediment) and chemically analyzing the mixture rather than the individual samples. The chemical analysis of the mixture provides an estimate of the average concentration of the individual samples making up the composite. Compositing also may be used when the cost of analysis is high. Each organism or sediment sample added to the composite should be of equal size (i.e., wet weight) and the composite should be completely homogenized before taking a sample for chemical analysis. If compositing is performed in this manner, the value obtained from the analysis of the composite is the same as the average obtained from analyzing each individual sample (within any sampling and analytical errors). If true replicate composites (not subsample composites) are made, the variance of the replicates will be less than the variance of the individual samples, providing a more precise estimate of the mean value. This increases the power of a test between means of composites over a test between means of individuals or samples for a given number of samples analyzed. If compositing reduces the actual number of replicates, however, the power of the test will be reduced. If composites are made of individuals or samples varying in size, the value of the composite and the mean of the individual organisms or sediment samples are no longer equivalent. The variance of the replicate composites will increase, decreasing the power of any test between means. In extreme cases, the variance of the composites can exceed the population variance (Tetra Tech, 1986). Therefore, it is important to keep the individuals or sediment samples comprising the composite equivalent in size. If sample sizes vary, consult the tables in Schaeffer and Janardan (1978) to determine whether replicate composite variances will be higher than individual sample variances, which would make compositing inappropriate.

12.2.3 Hypothesis Testing and Power

12.2.3.1 The purpose of the 28-d *L. plumulosus* chronic toxicity test is to determine whether the biological response to a treatment sample differs from the response to a control or reference sample. Figure 12.1 presents the possible outcomes and decisions that can be reached in a statistical test of such a hypothesis. The nullypothesis is that no difference exists among the mean reference and treatment responses. The alternative hypothesis of greatest interest in sediment tests is that the treatments are toxic relative to the reference sediment.

12.2.3.2 Statistical tests of hypotheses can be designed to control for the chances of making incorrect decisions. In Figure 12.1, alpha (") represents the probability of making a Type I statistical error. A Type I statistical error in this testing situation results from the false conclusion that the treated sample is toxic or contains chemical residues not found in the control or reference sample. Beta ($) represents the probability of making a Type II statistical error, or the likelihood that one erroneously concludes there are no differences among the mean responses in the treatment and control or reference samples. Traditionally, acceptable values have ranged from 0.1 to 0.01, with 0.05 or 5% used most commonly. This choice should depend upon the consequences of making a Type I error. Historically, having chosen ", environmental

		True State of Nature	
	DECISION	TR = Control	TR > Control
Experimental Results	TR = Control	Correct 1 - α	Type II Error β
	TR > Control	Type I Error α	Correct 1 - β (Power)

NOTE: Treatment response (TR); alpha (α) represents the probability of making a Type I statistical error (false positive); beta (β) represents the probability of making a Type II statistical error (false negative).

Figure 12.1 Treatment response for a Type I and Type II error

researchers have ignored $ and the associated power of the test (1-$).

12.2.3.3 Fairweather (1991) presented a review of the need for, and the practical implications of, conducting power analysis in environmental monitoring studies. This review also includes a comprehensive bibliography of recent publications on the need for, and use of, power analyses in environmental study design and data analysis. The consequences of a Type II statistical error in environmental studies should never be ignored and may, in fact, be the most important criteria to consider in experimental designs and data analyses that include statistical hypothesis testing. According to Fairweather (1991), the commitment of time, energy and people to a false positive (a Type I error) will only continue until the mistake is discovered. In contrast, the cost of a false negative (a Type II error) will have both short- and long-term costs (e.g., ensuring environmental degradation and the cost of its rectification).

12.2.3.4 The critical components of the experimental design associated with the test of hypotheses outlined above are (1) the required MDD between the treatment and reference responses, (2) the variance among treatment and reference replicate experimental units, (3) the number of replicate units for the treatment and reference samples, (4) the number of animals exposed within a replicate exposure chamber, and (5) the selected probabilities of Type I (") and Type II ($) errors.

12.2.3.5 Sample size or number of replicates might be fixed because of cost or space considerations, or might be varied to achieve *a priori* probabilities of " and $. The MDD should be established ahead of time based upon biological and program considerations. The investigator has little control of the variance among replicate exposure chambers. However, this variance component can be minimized by selecting test organisms that are as biologically similar as possible and maintaining test conditions within prescribed QC limits.

12.2.3.6 The MDD can be expressed as the absolute or relative (i.e., percentage) change from the mean reference response. In this technical manual, MDD is expressed as the absolute change from the mean reference response (Section 13). To test the equality of the reference and a treatment response, a two-sample t-test with its associated assumptions is an appropriate parametric analysis. If the desired MDD, the number of replicates per treatment, the number of organisms per replicate, and an estimate of typical among replicate variability (CV) are available, it is possible to use a graphical approach as in Figure 12.2 to determine how likely it is that a 20% reduction will be detected in the treatment response relative to the reference response. The CV is defined as 100% x (standard deviation divided by the mean). In a test design with eight replicates per treatment and with an " level of 0.05, high power (i.e., >0.8) to detect a 20% reduction from the reference mean occurs only if the CV is 15% or less (Figure 12.2). The choice of these variables also affects the power of the test. If five replicates are used per treatment (Figure 12.3), the CV needs to be 10% or lower to detect a 20% reduction in response relative to the reference mean with a power of 90%.

12.2.3.7 Relaxing the " level of a statistical test increases the power of the test. Figure 12.4 duplicates Figure 12.3 except that " is 0.10 instead of 0.05. Selection of the appropriate " level of a test is a function of the costs associated with making Type I and II statistical errors. Evaluation of Figure 12.3 illustrates that with a CV of 15% and an " level of 0.05, there is approximately 60% probability (power) of detecting a 20% reduction in the mean treatment response relative to the reference mean. However, if " is set at 0.10 (Figure 12.4) and the CV remains at 15%, then there is approximately 80% probability (power) of detecting a 20% reduction relative to the reference mean. The latter example would be preferable if an environmentally conservative analysis and interpretation of the data is desirable.

12.2.3.8 Increasing the number of replicates per treatment will increase the power to detect a 20% reduction in treatment response relative to the reference mean (Figure 12.5). Note, however, that for less than eight replicates per treatment, it is difficult to have high power (i.e., >0.80) unless the CV is less than 15%. If space or cost limit the number of replicates to fewer than eight per

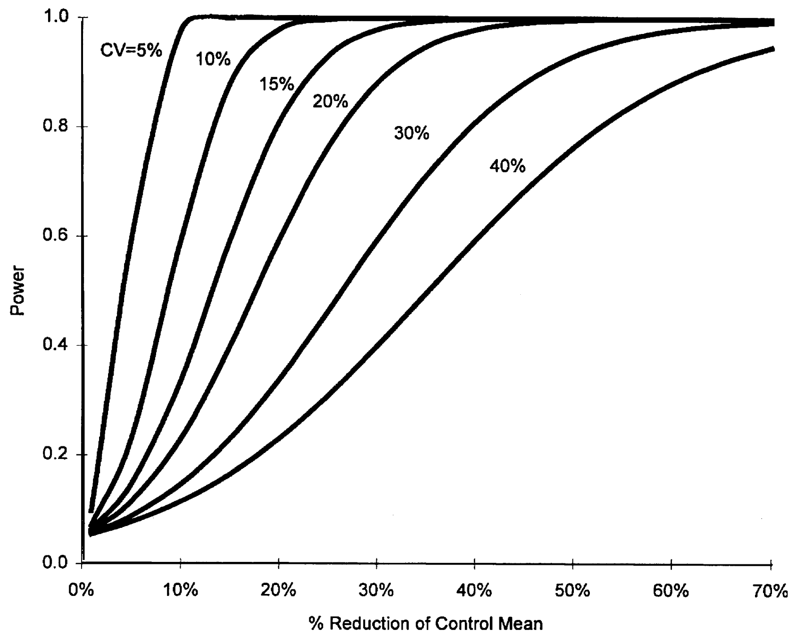


Figure 12.2 Power of the test vs. percentage reduction in treatment response relative to the control mean at various CVs (eight replicates, " = 0.05 [one-tailed]

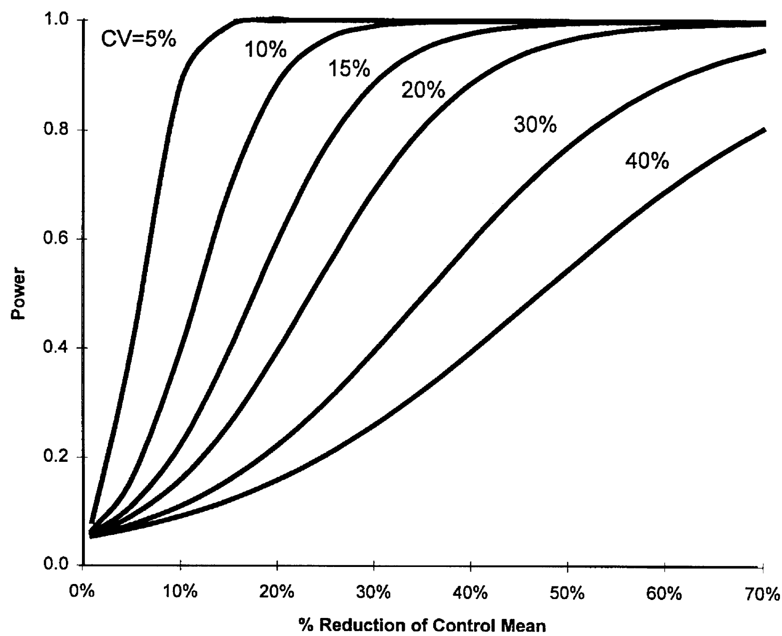


Figure 12.3 Power of the test vs. percentage reduction in treatment response relative to the control mean at various CVs (five replicates, " = 0.05 [one-tailed]

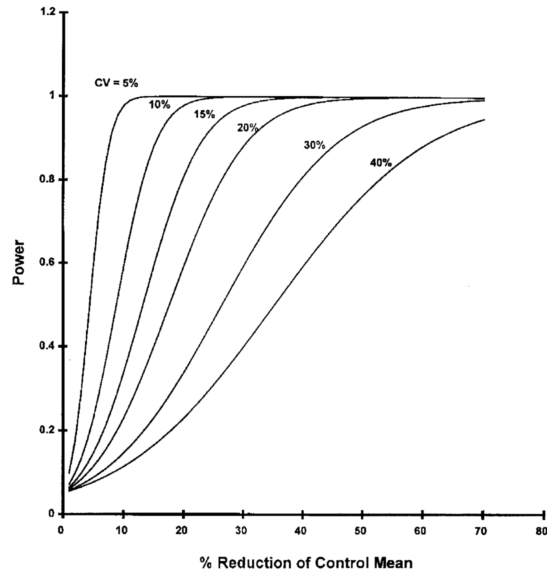


Figure 12.4 Power of the test vs. percentage reduction in treatment response relative to the control mean at various CVs (5 replicates, " = 0.10 [one-tailed])

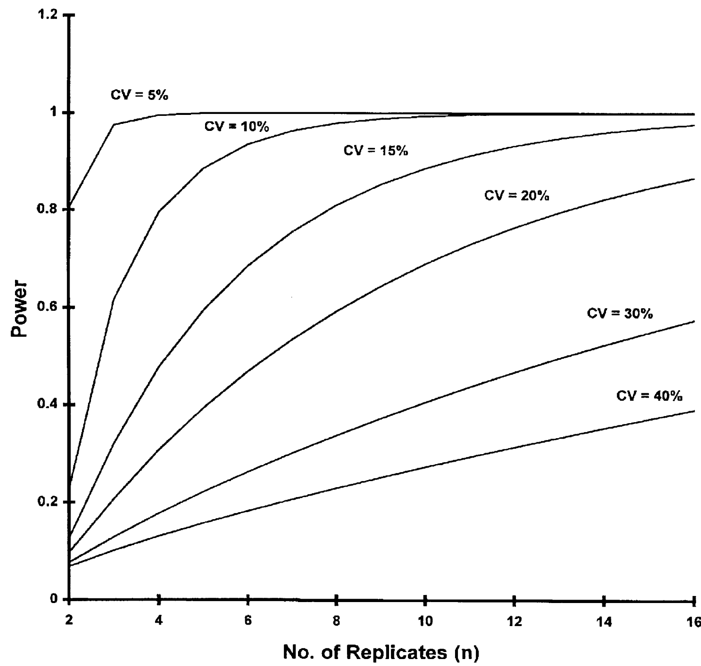


Figure 12.5 Effect of CV and number of replicates on the power to detect a 20% decrease in treatment response relative to the control mean (" = 0.05 [one-tailed])

treatment, then it may be necessary to find ways to reduce the among-replicate variability. Options that are available to increase the power of the test include selecting more uniform organisms to reduce biological variability and/or increasing the α level of the test. For CVs in the range of 30% to 40%, even eight replicates per treatment is inadequate to detect small reductions (<20%) in response relative to the control mean.

12.2.3.9 The effect of the choice of α and β on number of replicates for various CVs, assuming the combined total probability of Type I and Type II statistical errors is fixed at 0.25, is illustrated in Figure 12.6. An α of 0.10, therefore, establishes a β of 0.15. In Figure 12.6, if α = β = 0.125, the number of replicates required to detect a difference of 20% relative to the reference is at a minimum. As α or β decrease, the number of replicates required to detect the same 20% difference relative to the reference increases. However, the curves are relatively flat over the range of 0.05 to 0.20, and their shape will change dramatically if the combined total of α + β is changed. Limiting the total of α + β to 0.10 greatly increases the number of replicates necessary to detect a preselected percentage reduction in mean treatment response relative to the control mean.

12.2.4 Comparing Means

12.2.4.1 Figure 12.7 outlines a decision tree for analysis of survival, growth rate, and reproduction data subjected to hypothesis testing. In the tests described herein, samples or observations refer to replicates of treatments. Sample size n is the number of replicates (i.e., exposure chambers) in an individual treatment, not the number of organisms in an exposure chamber. Overall sample size N is the combined total number of replicates in all treatments. The statistical methods discussed in this section are described in general statistics texts such as Steel and Torrie (1980), Sokal and Rohlf (1981), Dixon and Massey (1983), Zar (1984), and Snedecor and Cochran (1989). It is recommended that users of this manual have at least one of these texts and associated statistical tables on hand. A nonparametric statistics text such as Conover (1980) might also be helpful.

12.2.4.2 Mean

12.2.4.2.1 The sample mean is the average value, or $\Sigma x_{i/n}$ where

n = number of observations (replicates)
x_i = ith observation
Σx_i = every x summed = $x_1 + x_2 + x_3 + \ldots + x_n$

12.2.4.3 Standard Deviation

12.2.4.3.1 The sample standard deviation (S) is a measure of the variation of the data around the mean and is equivalent to $\sqrt{s^2}$. The sample variance, s^2, is given by the following "machine" or "calculation" formula:

$$s^2 = \frac{\sum_{i=1} x_i^2 - (\bar{x})^2}{n-1}$$

12.2.4.4 Standard Error of the Mean

12.2.4.4.1 The standard error of the mean (SE or $s/\sqrt{n}$) estimates variation among sample means rather than among individual values. The SE is an estimate of the standard deviation among means that would be obtained from several samples of n observations each. Most of the statistical tests in this manual compare means with other means (e.g., dredged sediment mean with reference mean) or with a fixed standard (e.g., Food and Drug Administration [FDA] action level; ASTM, 2000d). Therefore, the "natural" or "random" variation of sample means (estimated by SE), rather than the variation among individual observations (estimated by S), is required for the tests.

12.2.4.5 Tests of Assumptions

12.2.4.5.1 In general, parametric statistical analyses such as *t* tests and analysis of variance are appropriate only if (1) there are independent, replicate experimental units for each treatment, (2) the observations within each treatment follow a normal distribution, and (3) variances for both treatments are equal or similar. The first

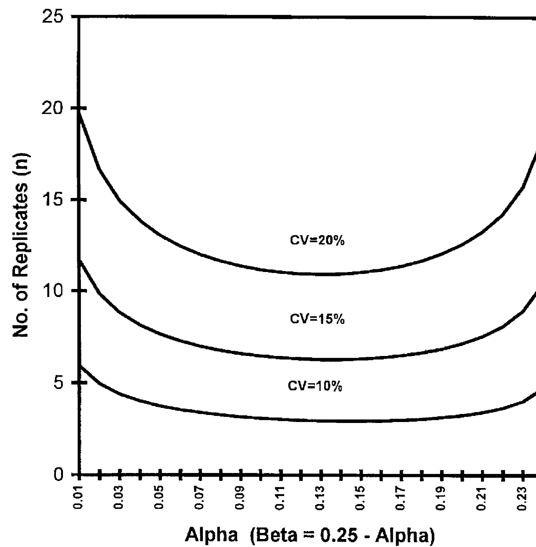


Figure 12.6 Effect of alpha and beta on the number of replicates at various CVs (assuming combined " + $ = 0.25)

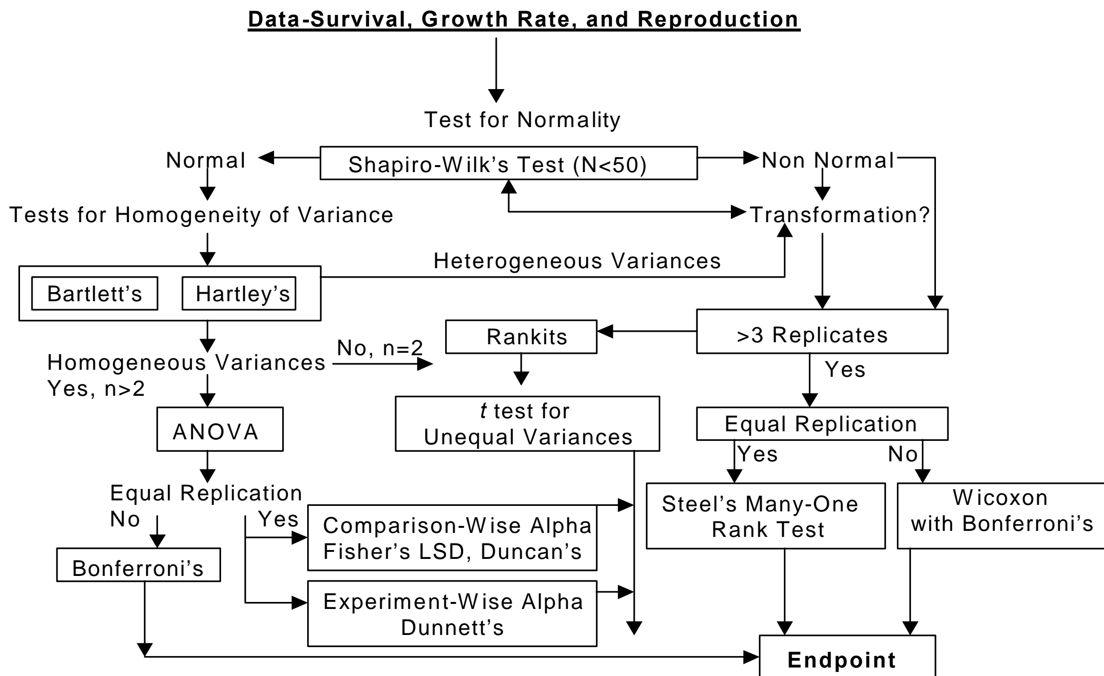


Figure 12.7 Decision tree for analysis of survival, growth rate, and reproduction data subjected to hypothesis testing

assumption is an essential component of experimental design. The second and third assumptions can be tested using the data obtained from the experiment. Therefore, before conducting statistical analyses, tests for normality and equality of variances should be performed.

12.2.4.5.2 Outliers. Extreme values and systematic departures from a normal distribution (e.g., a log-normal distribution) are the most common causes of departures from normality or equality of variances. An outlier is an inconsistent or questionable data point that appears unrepresentative of the general trend exhibited by the majority of the data. Outliers may be detected by tabulation of the data, by plotting, or by analysis of residuals. An explanation should be sought for any questionable data points. Sometimes an investigator knows from past experience that occasional wild observations occur, though the process is otherwise stable. Except in such cases, statisticians warn against automatic rejection rules based on tests of significance, particularly if there appear to be several outliers. The apparent outliers may reflect distributions of the observations that are skewed or have long tails and are better handled by methods being developed for nonnormal distributions. (Snedecor and Cochran, 1989). If there is no explanation, the analysis should be performed both with- and without- the outlier, and the results of both analyses should be reported. An appropriate transformation, such as the arcsine square root transformation, will normalize many distributions (USEPA, 1985). Problems with outliers can usually be solved only by using nonparametric tests, but careful laboratory practices can reduce the frequency of outliers.

12.2.4.5.3 Tests for Normality. The most commonly used test for normality for small sample sizes (n<50) is the Shapiro-Wilk's Test. This test determines whether residuals are normally distributed. Residuals are the differences between individual observations and the treatment mean. Residuals, rather than raw observations, are tested because subtracting the treatment mean removes any differences among treatments. This scales the observations so that the mean of residuals for each treatment and over all treatments is zero. The Shapiro-Wilk's Test provides a test statistic W, which is compared to values of W expected from a normal distribution. W will generally vary between 0.3 and 1.0, with lower values indicating greater departure from number of replicates (n) and design. A balanced design means that all treatments have an equal number (n) of replicate exposure chambers. A design is considered normality. Because normality is desired, one looks for a high value of W with an associated probability greater than the pre-specified alpha level.

12.2.4.5.3.1 Table 12.1 provides alpha levels to determine whether departures from normality are significant. Normality should be rejected when the probability associated with W (or other normality test statistic) is less than " for the appropriate total unbalanced when the treatment with the largest number of replicates (n_{max}) has at least twice as many replicates as the treatment with the fewest replicates (n_{min}). Note that higher " levels are used when the number of replicates is small or when the design is unbalanced, because these are the cases in which departures from normality have the greatest effects on *t* tests and other parametric comparisons. If data fail the test for normality, even after transformation, nonparametric tests should be used for additional analyses (Section 12.2.7.17 and Figure 12.7).

12.2.4.5.3.2 Tables of quantiles of W can be found in Shapiro and Wilk (1965), Gill (1978), Conover (1980), and other statistical texts. These references also provide methods of calculating W, although the calculations can be tedious. For that reason, commonly available computer programs or statistical packages are preferred for the calculation of W.

12.2.4.5.4 Tests for Homogeneity of Variances. There are a number of tests for equality of variances. Some of these tests are sensitive to departures from normality, which is why a test for normality should be performed first. Bartlett's test or other tests such as Levene's test or Cochran's test (Winer, 1971; Snedecor and Cochran, 1989) all have similar power for small, equal sample sizes (n=5) (Conover et al., 1981). The data must be normally distributed for Bartlett's test. Many

Table 12.1 Suggested alpha Levels to Use for Tests of Assumptions

Test	Number of Observations[1]	" When Design Is	
		Balanced	Unbalanced[2]
Normality	N = 2 to 9	0.10	0.25
	N = 10 to 19	0.05	0.10
	N = 20 or more	0.01	0.05
Equality of Variances	N = 2 to 9	0.10	0.25
	N = 10 or more	0.05	0.10

1 N = total number of observations (replicates) in all treatments combined; n = number of observations (replicates) in an individual treatment.

2 n_{max} • $2n_{min}$.

software packages for *t* tests and ANOVA provide at least one of the tests.

12.2.4.5.4.1 If no tests for equality of variances are included in the available statistical software, Hartley's F_{max} can easily be calculated:

F_{max} = (larger of s_1^2, s_2^2) / (smaller of s_1^2, s_2^2)

When F_{max} is large, the hypothesis of equal variances is more likely to be rejected. F_{max} is a two-tailed test, because it does not matter which variance is expected to be larger. Some statistical texts provide critical values of F_{max} (Winer, 1971; Gill, 1978; Rohlf and Sokal, 1981).

12.2.4.5.4.2 Levels of " for tests of equality of variances are provided in Table 12.1. These levels depend upon the number of replicates in a treatment (n) and allotment of replicates among treatments. Relatively high " values (i.e., $0.10) are recommended, because power of the above tests for equality of variances is rather low (about 0.3) when n is small. Equality of variances is rejected if the probability associated with the test statistic is less than the appropriate ".

12.2.4.6 Transformations of the Data

12.2.4.6.1 When the assumptions of normality or homogeneity of variance are not met, transformations of the data may remedy the problem, so that the data can be analyzed by parametric procedures, rather than by a nonparametric technique. The first step in these analyses is to transform the responses, expressed as the proportion surviving, by the arcsine-square root transformation. The arcsine-square root transformation is commonly used on proportionality data to stabilize the variance and satisfy the normality requirement. If the data do not meet the assumption of normality and there are four or more replicates per group, then the nonparametric test, Wilcoxon Rank Sum Test, can be used to analyze the data. If the data meet the assumption of normality, Bartlett's Test or Hartley's F test for equality of variances is used to test the homogeneity of variance assumption. Failure of the homogeneity of variance assumption leads to the use of a modified *t* test and the degrees of freedom for the test are adjusted.

12.2.4.6.2 The arc sine-square root transformation consists of determining the angle (in radians) represented by a sine value. In this

transformation, the proportion surviving is taken as the sine value, the square root of the sine value is calculated, and the angle (in radians) for the square root of the sine value is determined. When the proportion surviving is 0 or 1, a special modification of the transformation should be used (Bartlett, 1937). An example of the arcsine-square root transformation and modification are provided below.

1. Calculate the response proportion (RP) for each replicate within a group, where
   RP = (number of surviving organisms)/(number exposed)
2. Transform each RP to arcsine, as follows:
   a. For RPs greater than zero or less than one:

      Angle (in radians) = arc sine $\sqrt{(RP)}$

   b. Modification of the arc sine when RP = 0.

      Angle (in radians) = arc sine $\sqrt{\frac{1}{4n}}$

where n = number animals/treatment

replicate.

   c. Modification of the arc sine when RP = 1.0.

   Angle = 1.5708 radians – (radians for RP=0)

12.2.4.7 Two Sample Comparisons (n=2)

12.2.4.7.1 The true population mean (:) and standard deviation (F) are known only after sampling the entire population. In most cases, samples are taken randomly from the population, and the S calculated from those samples is only an estimate of F. Student's *t*-values account for this uncertainty. The degrees of freedom for the test, which are defined as the sample size minus one (n-1), should be used to obtain the correct *t*-value. Student *t*-values decrease with increasing sample size because larger samples provide a more precise estimate of : and F.

12.2.4.7.2 When using a *t* table, it is crucial to determine whether the table is based on one-tailed probabilities or two-tailed probabilities. In formulating a statistical hypothesis, the alternative hypothesis can be one-sided (one-tailed test) or two-sided (two-tailed test). The null hypothesis (H_o) is always that the two values being analyzed are equal. A one-sided alternative hypothesis (H_a) is that there is a specified relationship between the two values (e.g., one value is greater than the other) versus a two-sided alternative hypothesis (H_a), which is that the two values are simply different (i.e., either larger or smaller). A one-tailed test is used when there is an *a priori* reason to test for a specific relationship between two means, such as the alternative hypothesis that the treatment mortality or tissue residue is greater than the reference mortality or tissue residue. In contrast, the two-tailed test is used when the direction of the difference is not important or cannot be assumed before testing.

12.2.4.7.3 Because control organism mortality and sediment contaminant concentrations are presumed lower than reference or treatment sediment values, conducting one-tailed tests is recommended in most cases. For the same number of replicates, one-tailed tests are more likely to detect statistically significant differences between treatments (e.g., have a greater power) than two-tailed tests. This is a critical consideration when dealing with a small number of replicates (such as five/treatment). The other alternative for increasing statistical power is to increase the number of replicates, which increases the cost of the test.

12.2.4.7.4 There are cases when a one-tailed test is inappropriate. When no *a priori* assumption can be made as to how the values vary in relationship to one another, a two-tailed test should be used. An example of an alternative two-sided hypothesis is that the reference sediment TOC content is different (greater or lesser) from the reference sediment TOC.

12.2.4.7.5 The *t* value for a one-tailed probability can be found in a two-tailed table by looking up *t* under the column for twice the desired one-tailed probability. For example, the one-tailed *t* value for " = 0.05 and df = 20 is 1.725, and is found in a two-tailed table using the column for " = 0.10.

12.2.4.7.6 The usual statistical test for comparing two independent samples is the two-sample *t* test (Snedecor and Cochran, 1989). The *t* statistic for

testing the equality of means $\bar{x}_1$ and $\bar{x}_2$ from two independent samples with n_1 and n_2 replicates and unequal variances is where s_1^2 and s_2^2 are the

$$t = \left(\bar{x}_1 - \bar{x}_2\right) / \sqrt{s_1^2 / n_1 + s_2^2 / n_2}$$

sample variances of the two groups. Although the equation assumes that the variances of the two groups are unequal, it is equally useful for situations in which the variances of the two groups are equal. This statistic is compared with the student's *t* distribution with degrees of freedom given by Satterthwaite's (1946) approximation:

$$df = \frac{\left(s_1^2 / n_1 + s_2^2 / n_2\right)^2}{\left(s_1^2 / n1\right)^2 / \left(n_1 - 1\right) + \left(s_2^2 / n_2\right)^2 / \left(n_2 - 1\right)}$$

This formula can result in fractional degrees of freedom, in which case one should round the degrees of freedom down to the nearest integer in order to use a *t* table. Using this approach, the degrees of freedom for this test will be less than the degrees of freedom for a *t* test assuming equal variances. If there are unequal numbers of replicates in the treatments, the *t* test with Bonferroni's adjustment can be used for data analysis (USEPA, 1994b; 1994c). When variances are equal, an *F* test for equality is unnecessary.

12.2.4.8 Nonparametric Tests

12.2.4.8.1 Tests such as the *t* test, which analyze the original or transformed data, and which rely on the properties of the normal distribution, are referred to as parametric tests. Nonparametric tests, which do not require normally distributed data, analyze the ranks of data and generally compare medians rather than means. The median of a sample is the middle or fiftieth percentile observation when the data are ranked from smallest to largest. In many cases, nonparametric tests can be performed simply by converting the data to ranks or normalized ranks (rankits) and conducting the usual parametric test procedures on the ranks or rankits. Rankits are simply the z-scores expected for the rank in a normal distribution. Thus, using rankits imposes a normal distribution over all the data, although not necessarily within each treatment. Rankits can be obtained by ranking the data, then converting the ranks to rankits using the following formula:

$$rankit = z_{\left[\left(rank - 0.375\right) / \left(N + 0.25\right)\right]}$$

where z is the normal deviate and N is the total number of observations. Alternatively, rankits may be obtained from standard statistical tables such as Sokal and Rohlf (1981).

12.2.4.8.2 Nonparametric tests are useful because of their generality, but have less statistical power than corresponding parametric tests when the parametric test assumptions are met. If parametric tests are not appropriate for comparisons because the normality assumption is not met, data should be converted to normalized ranks (rankits).

12.2.4.8.3 If normalized ranks are calculated, the ranks should be converted to rankits using the formula above. In comparisons involving only two treatments (n=2), there is no need to test assumptions on the rankits or ranks; simply proceed with a one-tailed *t* test for unequal variances using the rankits or ranks.

12.2.4.9 Analysis of Variance (n=2)

12.2.4.9.1 Some experiments are set up to compare more than one treatment with a control, whereas others may also be interested in comparing the treatments with one another. The basic design of these experiments is the same as for experiments evaluating pairwise comparisons. After the applicable comparisons are determined, the data must be tested for normality to determine whether parametric statistics are appropriate and whether the variances of the treatments are equal. If normality of the data and equal variances are established, then ANOVA may be performed to address the hypothesis that all the treatments including the control are equal. If normality or equality of variance are not established, then transformations of the data might be appropriate, or nonparametric statistics can be used to test for equal means. Tests for normality of the data should be performed on the treatment residuals. A residual is defined as the observed value minus the treatment mean, that is, $r_{ik} = o_{ik}$ -(k^{th} treatment mean). Pooling residuals provides an adequate sample size to test the data for normality.

12.2.4.9.2 The variances of the treatments should also be tested for equality. Currently there is no easy way to test for equality of the treatment means using analysis of variance if the variances are not equal. In a toxicity test with several treatments, one treatment might have 100% mortality in all of its replicates, or the control treatment may have 100% survival in all of its replicates. These responses result in 0 variance for a treatment that results in a rejection of equality of variance in these cases. No transformation will change this outcome. In this case, the replicate responses for the treatment with 0 variance should be removed before testing for equality of variances. Only those treatments that do not have 0 replicate variance should be used in the ANOVA to get an estimate of the within treatment variance. After a variance estimate is obtained, the means of the treatments with 0 variance can be tested against the other treatment means using the appropriate mean comparison. Equality of variances among the treatments can be evaluated with the Hartley F_{max} test or Bartlett's test. The option of using nonparametric statistics on the entire set of data is also an alternative.

12.2.4.9.3 If the data are not normally distributed or the variances among treatments are not homogeneous, even after data transformation, nonparametric analyses are appropriate. If there are four or more replicates per treatment and the number of replicates per treatment is equal, the data can be analyzed with Steel's Many-One Rank test. Unequal replication among treatments requires data analysis with the Wilcoxon Rank Sum test with Bonferroni's adjustment. Steel's Many-One Rank test is a nonparametric test for comparing treatments with a control. This test is an alternative to the Dunnett's test and can be applied to data when the normality assumption has not been met. Steel's test requires equal variances across treatments and the control, but is thought to be fairly insensitive to deviations from this condition (USEPA, 1991a). Wilcoxon's Rank Sum test is a nonparameteric test to be used as an alternative to the Steel's test when the number of replicates are not the same within each treatment. A Bonferroni's adjustment of the pairwise error rate for comparison of each treatment versus the control is used to set an upper bound of alpha on the overall error rate. This is in contrast to the Steel's test with a fixed overall error rate for alpha. Thus, Steel's tests is a more powerful test (USEPA, 1991a).

12.2.4.9.4 Different mean comparison tests are used depending on whether an " percent comparison-wise error rate or an " percent experiment-wise error rate is desired. The choice of a comparison-wise or experiment-wise error rate depends on whether a decision is based on a pairwise comparison (comparison-wise) or from a set of comparisons (experiment-wise). For example, a comparison-wise error rate would be used for deciding which stations along a gradient were acceptable or not acceptable relative to a control or reference sediment. Each individual comparison is performed independently at a smaller " than that used in an experiment-wise comparison, such that the probability of making a Type I error in the entire series of comparisons is not greater than the chosen experiment-wise " level of the test. This results in a more conservative test when comparing any particular sample to the control or reference. However, if several samples were taken from the same area and the decision to accept or reject the area were based on all comparisons with a reference, then an experiment-wise error rate should be used. When an experiment-wise error rate is used, the power to detect real differences between any two means decreases as a function of the number of treatment means being compared with the reference treatment.

12.2.4.9.5 The recommended procedure for pairwise comparisons that have a comparison-wise " error rate and equal replication is to do an ANOVA followed by a one-sided Fisher's Least Significant Difference (LSD) test (Steel and Torrie, 1980). A Duncan's mean comparison test should give results similar to the LSD. If the treatments do not contain equal numbers of replicates, the appropriate analysis is the *t test* with Bonferroni's adjustment. For comparisons that maintain an experiment-wise " error rate, Dunnett's test is recommended for comparisons with the control.

12.2.4.9.6 Dunnett's test has an overall error rate of " , which accounts for the multiple comparisons with the control. Dunnett's procedure uses a pooled estimate of the variance, which is equal to the error value calculated in an ANOVA.

12.2.4.9.7 To perform the individual comparisons, calculate the *t* statistic for each treatment and control combination, as follows:

$$t_i = \frac{\left(\overline{Y}_1 - \overline{Y}_i\right)}{S_w\sqrt{\left(1/n_1\right) + \left(1/n_i\right)}}$$

where $\overline{Y}_i$ = mean for each treatment

$\overline{Y}_i$ = mean for the control

S_w = square root of the within mean square

n_1 = number of replicates in the control

n_i = number of replicates for treatment "i"

To quantify the sensitivity of the Dunnett's test, the minimum significant difference (MSD = MDD) may be calculated with the following formula:

$$MSD = d\, S_w\sqrt{\left(1/n_1\right) + \left(1/n\right)}$$

where d = critical value for the Dunnett's Procedure

S_w = square root of the within mean square

n = number of replicates per treatment, assuming an equal number of replicates at all treatment concentrations

n_1 = number of replicates in the control

12.2.5 Methods for Calculating LC50s, EC50s, and ICps

12.2.5.1 Figure 12.8 outlines a decision tree for analysis of point estimate data. USEPA manuals (USEPA, 1991a; 1994b; 1994c) discuss in detail the mechanics of calculating LC50 (or EC50) or values using the most current methods. The most commonly used methods are the Graphical, Probit, trimmed Spearman-Karber, and the Linear Interpolation Methods. Methods for evaluating point estimate data using logistic regression are outlined in Snedecor and Chochran (1989). In general, results from these methods should yield similar estimates. Each method is outlined below, and recommendations are presented for the use of each method.

12.2.5.2 Data for at least five test concentrations and the control should be available to calculate an LC50, although each method can be used with fewer concentrations. Survival in the lowest concentration must be at least 50%, and an LC50 should not be calculated unless at least 50% of the organisms die in at least one of the serial dilutions. When less than 50% mortality occurs in the highest test concentration, the LC50 is expressed as greater than the highest test concentration.

12.2.5.3 Due to the intensive nature of the calculations for the estimated LC50 and associated 95% confidence interval using most of the following methods, it is recommended that the data be analyzed with the aid of computer software. Computer programs to estimate the LC50 or ICp values and associated 95% confidence intervals using the methods discussed below (except for the Graphical Method) were developed by USEPA and can be obtained by sending a diskette with a written request to USEPA, National Exposure Research Laboratory, 26 W. Martin Luther King Drive, Cincinnati, OH 45268 or calling 513/569-7076.

12.2.5.4 The Graphical Method

12.2.5.4.1 This procedure estimates an LC50 (or EC50) by linearly interpolating between points of a plot of observed percentage mortality versus the base 10 logarithm ($\log_{10}$) of treatment concentration. The only requirement for its use is that treatment mortalities bracket 50%.

12.2.5.4.2 For an analysis using the Graphical Method, the data should first be smoothed and adjusted for mortality in the control replicates. The procedure for smoothing and adjusting the data is detailed in the following steps: Let p_0, p_1, ..., p_k denote the observed proportion mortalities for the control and the k treatments. The first step is to smooth the p_i if they do not satisfy p_o-p_i-...-p_k. The smoothing process replaces any adjacent p_is that

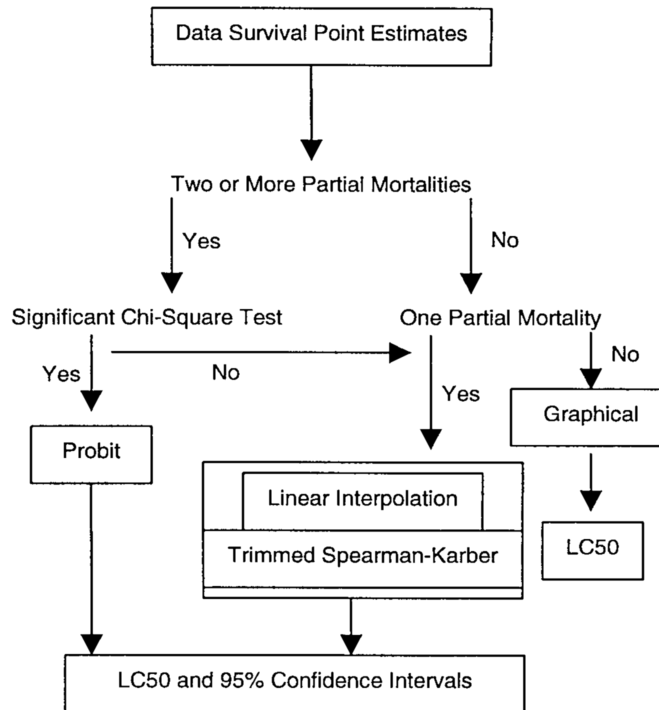


Figure 12.8 Decision tree for analysis of point estimate data

do not conform to p_o-p_i-...-p_k with their average. For example, if p_i is less than p_{i-1} then:

$$p_{i-1}^s = p_i^s = (p_i + p_{i-1})/2$$

where p_i^s = the smoothed observed proportion mortality for concentration i.

Adjust the smoothed observed proportion mortality in each treatment for mortality in the control group using Abbott's formula (Finney, 1971). The adjustment takes the form:

$$p_i^a = (p_i^s - p_o^s)/(1 - p_o^s)$$

where p_o^s = the smoothed observed proportion mortality for the control

p_i^s = the smoothed observed proportion mortality for concentration i.

12.2.5.5 The Probit Method

12.2.5.5.1 This method is a parametric statistical procedure for estimating the LC50 (or EC50) and the associated 95% confidence interval (Finney, 1971). The analysis consists of transforming the observed proportion mortalities with a Probit transformation, and transforming the treatment concentrations to $\log_{10}$. Given the assumption of normality for the $\log_{10}$ of the tolerances, the relationship between the transformed variables mentioned above is about linear. This relationship allows estimation of linear regression parameters, using an iterative approach. A Probit is the same as a z-score: for example, the Probit corresponding to 70% mortality is $z_{.70}$ or = 0.52.

The LC50 is calculated from the regression and is the concentration associated with 50% mortality or z=0. To obtain a reasonably precise estimate of the LC50 with the Probit Method, the observed proportion mortalities must bracket 0.5 and the $\log_{10}$ of the tolerance should be normally distributed. To calculate the LC50 estimate and associated 95% confidence interval, two or more of the observed proportion mortalities must be between zero and one. The original percentage of mortalities should be corrected for control mortality using Abbott's formula (Section 12.2.5.4.2; Finney, 1971) before the Probit transformation is applied to the data.

12.2.5.5.2 A goodness-of-fit procedure with the Chi-square statistic is used to determine whether data fit the Probit model. If many data sets are to be compared with one another, the Probit Method is not recommended because it may not be appropriate for many of the data sets. This method is also only appropriate for percent mortality data sets and should not be used for estimating endpoints that are a function of the control response, such as inhibition of growth or reproduction. Most computer programs that generate Probit estimates also generate confidence interval estimates for the LC50. These confidence interval estimates on the LC50 might not be correct if replicate mortalities are pooled to obtain a mean treatment response (USEPA-USACE, 1998). This can be avoided by entering the Probit-transformed replicate responses and doing a least-squares regression on the transformed data.

12.2.5.6 The Trimmed Spearman-Karber Method

12.2.5.6.1 The trimmed Spearman-Karber Method is a modification of the Spearman-Karber, nonparametric statistical procedure for estimating the LC50 and the associated 95% confidence interval (Hamilton et al., 1977). This procedure estimates the trimmed mean of the distribution of the $\log_{10}$ of the tolerance. If the log tolerance distribution is symmetric, this estimate of the trimmed mean is equivalent to an estimate of the median of the log tolerance distribution. Use of the trimmed Spearman-Karber Method is only appropriate for lethality data sets and when the requirements for the Probit Method are not met (USEPA, 1994b; 1994c).

12.2.5.6.2 To calculate the LC50 estimate with the trimmed Spearman-Karber Method, the smoothed, adjusted, observed proportion mortalities must bracket 0.5. To calculate a confidence interval for the LC50 estimate, one or more of the smoothed, adjusted, observed proportion mortalities must be between zero and one.

12.2.5.6.3 Smooth the observed proportion mortalities as described for the Probit Method. Adjust the smoothed observed proportion mortality in each concentration for mortality in the control group using Abbott's formula (see Probit Method; Section 12.2.5.5). Calculate the amount of trim to use in the estimation of the LC50 as follows:

$$Trim = max\left(p_1^a, 1 - p_k^a\right)$$

where

p_1^a = the smoothed, adjusted proportion mortality for the lowest treatment concentration, exclusive of the control

p_k^a = the smoothed, adjusted proportion mortality for the highest treatment concentration

k = the number of treatment concentrations, exclusive of the control.

12.2.5.7 The Linear Interpolation Method

12.2.5.7.1 The Linear Interpolation Method calculates a toxicant concentration that causes a given percent reduction (e.g., 25%, 50%, etc.) in the endpoint of interest and is reported as an ICp value (where p = the percent effect). The procedure was designed for general applicability in the analysis of data from chronic toxicity tests and for the generation of an endpoint from a continuous model that allows a traditional quantitative assessment of the precision of the endpoint, such as confidence limits for the endpoint of a single test or a mean and coefficient of variation for the endpoints of multiple tests.

12.2.5.7.2 As described in USEPA (1994b; 1994c), the Linear Interpolation Method of calculating an ICp assumes that the responses (1) are monotonically nonincreasing, where the mean response for each higher concentration is less than or equal to the mean response for the previous concentration, (2) follow a piecewise linear response function, and (3) are from a random, independent, and representative sample of test data. If the data are not monotonically nonincreasing, they are adjusted by smoothing (averaging). In cases where the responses at the low toxicant concentrations are much higher than those in the controls, the smoothing process may result in a large upward adjustment in the control mean. In the Linear Interpolation Method, the smoothed response means are used to obtain the ICp estimate reported for the test. No assumption is made about the distribution of the data except that the data within a group being resampled are independent and identically distributed.

12.2.8.7.3 The Linear Interpolation Method assumes a linear response from one concentration to the next. Thus, the IC is estimated by linear interpolation between two concentrations whose responses bracket the response of interest, the (p) percent reduction from the control.

12.2.5.7.4 If the assumption of monotonicity of test results is met, the observed response means $(\bar{Y}_1)$ should stay the same or decrease as the toxicant concentration increases. If the means do not decrease monotonically, the responses are "smoothed" by averaging (pooling) adjacent means. Observed means at each concentration are considered in order of increasing concentration, starting with the control mean $(\bar{Y}_1)$. If the mean observed response at the lowest toxicant concentration $(\bar{Y}_2)$ is equal to or smaller than the control mean $(\bar{Y}_1)$, it is used as the response. If it is larger than the control mean, it is averaged with the control, and this average is used for both the control response (M_1) and the lowest toxicant concentration response (M_2). This mean is then compared with the mean observed response for the next higher toxicant concentration $(\bar{Y}_3)$. Again, if the mean observed response for the next higher toxicant concentration is smaller than the mean of the control and the lowest toxicant concentration, it is used as the response. If it is higher than the mean of the first two, it is averaged with the first two, and the resulting mean is used as the response for the control and two lowest concentrations of toxicant. This process is continued for data from the remaining toxicant concentrations. Unusual patterns in the deviations from monotonicity might require an additional step of smoothing. Where Y_I decrease monotonically, the $(\bar{Y}_1)$ become M_i without smoothing.

12.2.5.7.5 To obtain the ICp estimate, determine the concentrations C_J and C_{J+1} that bracket the response M_1 (1 - p/100), where M_1 is the smoothed control mean response and p is the percent reduction in response relative to the control response. These calculations can easily be done by hand or with a computer program as described below. The linear interpolation estimate is calculated as follows:

$$ICp = C_J + \left[M_1(1 - p/100) - M_J\right]\frac{(C_{J+1} - C_J)}{(M_{J+1} - M_J)}$$

where
- C_j = tested concentration whose observed mean response is greater than M_1(1 - p/100)
- C_{J+1} = tested concentration whose observed mean response is less than M_1(1 - p/100)
- M_1 = smoothed mean response for the control
- M_J = smoothed mean response for concentration J
- M_{J+1} = smoothed mean response for concentration J + 1
- p = percent reduction in response relative to the control response
- ICp = estimated concentration at which there is a percent reduction from the smoothed mean control response

12.2.5.7.6 Standard statistical methods for calculating confidence intervals are not applicable

for the ICp. The bootstrap method, as proposed by Efron (1982), is used to obtain the 95% confidence interval for the true mean. In the boostrap method, the test data Y_{ji} are randomly resampled with replacement to produce a new set of data Y_{ji}* that is statistically equivalent to the original data, but which produces a new and slightly different estimate of the ICp (ICp*). This process is repeated at least 80 times (Marcus and Holtzman, 1988), resulting in multiple "data" sets, each with an associated ICp* estimate. The distribution of the ICp* estimates derived from the sets of resampled data approximates the sampling distribution of the ICp estimate. The standard error of the ICp is estimated by the standard deviation of the individual ICp* estimates. Empirical confidence intervals are derived from the quantiles of the ICp* empirical distribution. For example, if the test data are resampled a minimum of 80 times, the empirical 2.5% and the 97.5% confidence limits are about the second smallest and second largest ICp* estimates (Marcus and Holtzman, 1988). The width of the confidence intervals calculated by the bootstrap method is related to the variability of the data. When confidence intervals are wide, the reliability of the IC estimate is in question. However, narrow intervals do not necessarily indicate that the estimate is highly reliable, because of undetected violations of assumptions and the fact that the confidence limits based on the empirical quantiles of a bootstrap distribution of 80 samples may be unstable.

12.3 Data Interpretation

12.3.1 Sediments spiked with known concentrations of contaminants can be used to establish cause -and -effect relationships between chemicals and biological responses. Results of toxicity tests with test materials spiked into sediments at different concentrations may be reported in terms of an LC50, an EC50, an IC50, or as an NOEC or LOEC (Section 3). Consistent spiking procedures should be followed in order to make interlaboratory comparisons (Section 8.3). The data interpretation of USEPA program specific regulatory decisions will be developed by the respective USEPA program office.

12.3.2 Evaluating effect concentrations for chemicals in sediment requires knowledge of factors controlling the bioavailability. Similar concentrations of a chemical in units of mass of chemical per mass of sediment dry weight often exhibit a range in toxicity in different sediment (Di Toro et al., 1991; USEPA, 1992c). Effect concentrations of chemicals in sediment have been correlated to interstitial water concentrations, and effect concentrations in interstitial water are often similar to effect concentrations in water-only exposures. The bioavailability of nonionic organic compounds are often inversely correlated with the organic carbon concentration of the sediment. Whatever the route of exposure, the correlations of effect concentrations to interstitial water concentrations indicate that predicted or measured concentrations in interstitial water can be useful for quantifying the exposure concentration to an organism. Therefore, information on partitioning of chemicals between solid and liquid phases of sediment can be useful for establishing effect concentrations.

12.3.3 Toxic units can be used to help interpret the response of organisms to multiple chemicals in sediment. A toxic unit is the concentration of a chemical divided by an effect concentration. For example, a toxic unit of exposure can be calculated by dividing the measured concentration of a chemical in pore water by the water-only LC50 for the same chemical (Ankley et al., 1991). Toxicity expressed as toxic units can be summed, and this may provide information on the toxicity of chemical mixtures (Ankley et al., 1991).

12.3.4 Field surveys can be designed to provide either a qualitative reconnaissance of the distribution of sediment contamination or a quantitative statistical comparison of contamination among sites (Burton and Ingersoll, 1994). Surveys of sediment toxicity are usually part of more comprehensive analyses of biological, chemical, geological, and hydrographic data. Statistical correlation can be improved and costs reduced if subsamples are taken simultaneously for sediment toxicity or bioaccumulation tests, chemical analyses, and benthic community structure.

12.3.5 Descriptive methods, such as toxicity tests with field-collected sediment, should not be used alone to evaluate sediment contamination. An integration of several methods using the weight of evidence is needed to assess the effects of contaminants associated with sediment (Long and Morgan, 1990; Ingersoll et al., 1996; 1997; Macdonald et al., 1996). Hazard evaluations integrating data from laboratory exposures, chemical analyses, and benthic community assessments provide strong complementary evidence of the degree of pollution-induced degradation in aquatic communities (Burton, 1991; Canfield et al., 1994; 1996; 1998; Chapman et al., 1992; 1997).

12.3.6 TIE procedures can be used to provide insights as to specific contaminants responsible for toxicity in sediment (USEPA, 1991b; Ankley and Thomas, 1992). For example, the toxicity of contaminants such as metals, ammonia, hydrogen sulfide, and non ionic organic compounds can be identified using TIE procedures.

12.4 Reporting

12.4.1 The record of the results of an acceptable sediment test should include the following information either directly or by referencing available documents:

12.4.1.1 Name of test and investigator(s), name and location of laboratory, and dates of start and end of test.

12.4.1.2 Source of control, reference, or test sediment, and method for collection, handling, shipping, storage, and disposal of sediment.

12.4.1.3 Source of test material, lot number if applicable, composition (identities and concentrations of major ingredients and impurities if known), known chemical and physical properties, and the identity and concentration(s) of any solvent used.

12.4.1.4 Source and characteristics of overlying water, description of any pretreatment, and results of any demonstration of the ability of an organism to survive or grow in the water.

12.4.1.5 Source, history, and age of test organisms; source, history, and age of brood stock, culture procedures; and source and date of collection of test organisms, scientific name, name of person who identified the organisms and the taxonomic key used, age or life stage, means and ranges of weight or length, observed diseases or unusual appearance, treatments used, and holding procedures.

12.4.1.6 Source and composition of food; concentrations of test material and other contaminants; procedure used to prepare food; and feeding methods, frequency and ration.

12.4.1.7 Description of the experimental design and test chambers, the depth and volume of sediment and overlying water in the chambers, lighting, number of test chambers and number of test organisms/treatment, date and time test starts and ends, temperature measurements, DO (: g/L) and any aeration used before starting a test and during the conduct of a test.

12.4.1.8 Methods used for physical and chemical characterization of sediment.

12.4.1.9 Definition(s) of the effects used to calculate LC50 or EC50s, biological endpoints for tests, and a summary of general observations of other effects.

12.4.1.10 A table of the biological data for each test chamber for each treatment, including the control(s), in sufficient detail to allow independent statistical analysis.

12.4.1.11 Methods used for statistical analyses of data.

12.4.1.12 Summary of general observations on other effects or symptoms.

12.4.1.13 Anything unusual about the test, any deviation from these procedures, and any other relevant information.

12.4.2 Published reports should contain enough information to clearly identify the methodology used and the quality of the results.

Section 13

Precision and Accuracy

13.1 Determining Precision and Accuracy

13.1.1 Precision is a term that describes the degree to which data generated from replicate measurements differ and reflects the closeness of agreement between replicates. Accuracy is the difference between the value of the measured data and the true value and is the closeness of agreement between an observed value and an accepted reference value. Quantitative determination of precision and accuracy in sediment testing of aquatic organisms is difficult or may be impossible in some cases, as compared to analytical (chemical) determinations. This is due in part to the many unknown variables that affect organism response. Determining the accuracy of a sediment test using field samples is not possible since the true values are not known. Because there is no acceptable reference material suitable for determining the accuracy of sediment tests, accuracy of the test methods has not been determined (Section 13.2).

13.1.2 Sediment tests exhibit variability due to several factors (Section 9). Test variability can be described in terms of two types of precision, either single laboratory precision (intralaboratory or repeatability; Section 13.5.1) or multilaboratory (interlaboratory or reproducibility; Section 13.5.2) precision. Intralaboratory precision reflects the ability of trained laboratory personnel to obtain consistent results repeatedly when performing the same test on the same organism using the same toxicant. Interlaboratory precision (also referred to as round-robin or ring tests) is a measure of reproducibility of a method when tests are conducted by a number of laboratories using that method and the same organism and samples. Generally, intralaboratory results are less variable than interlaboratory results (USEPA, 1991; 1991d; 1994b; 1994c; Hall et al., 1989; Grothe and Kimerle, 1985).

13.1.3 A measure of precision can be calculated using the mean and relative standard deviation (percent coefficient of variation, or CV% = standard deviation/mean x 100)of the calculated endpoints from the replicated endpoints of a test. However, precision reported as the CV should not be the only approach used for evaluating precision of tests and should not be used for the NOEC levels derived from statistical analyses of hypothesis testing. The CVs can be very high when testing extremely toxic samples. For example, if there are multiple replicates with no survival and one with low survival, the CV might exceed 100%, yet the range of response is actually quite consistent. Therefore, additional estimates of precision should be used, such as range of responses and minimum detectable difference (MDD) compared with control survival or growth rate. Several factors can affect the precision of the test, including test organism age, condition and sensitivity; handling and feeding of the test organisms; overlying water quality; and the experience of the investigators in conducting tests. For these reasons, it is recommended that trained laboratory personnel conduct the tests in accordance with the procedures outlined in Section 9. Quality assurance practices should include the following: (1) single laboratory precision determinations that are used to evaluate the ability of the laboratory personnel to obtain precise results using reference toxicants for test organisms and (2) preparation of control charts (Section 13.4) for each reference toxicant and test organism. The single laboratory precision determinations should be made before conducting a sediment test and should be periodically performed as long as whole-sediment tests are being conducted at the laboratory.

13.1.4 Intralaboratory precision data are routinely calculated for test organisms using water-only 96-h exposures to a reference toxicant, such as cadmium chloride ($CdCl_2$). Intralaboratory precision data should be tracked using a control chart. Each laboratory's reference-toxicity data will reflect conditions unique to that facility, including dilution water, culturing, and other variables (Section 9). However, each laboratory's reference-toxicity CVs should reflect good repeatability.

13.1.5 One interlaboratory precision (round-robin) test has been completed on the 28-d chronic test with *Leptocheirus plumulosus* (DeWitt et al., 1997b). Ten laboratories participated in the round-robin study, which used a dilution series of highly contaminated Black Rock Harbor sediment from a Superfund site in Connecticut mixed with uncontaminated, diluent sediment from Sequim Bay, Washington. The results of this round-robin study are described in Section 13.5.

13.2 Accuracy

13.2.1 The relative accuracy of toxicity tests cannot be determined because there is no acceptable reference material. The relative accuracy of the reference-toxicity tests can only be evaluated by comparing test responses to control charts.

13.3 Replication and Test Sensitivity

13.3.1 The sensitivity of sediment tests will depend in part on the number of replicates per concentration, the selected probability levels (" and $) selected, and the type of statistical analysis. For a specific level of variability, the sensitivity of the test will increase as the number of replicates is increased. The minimum recommended number of replicates varies with the objectives of the test and the statistical method used for analysis of the data (Section 12).

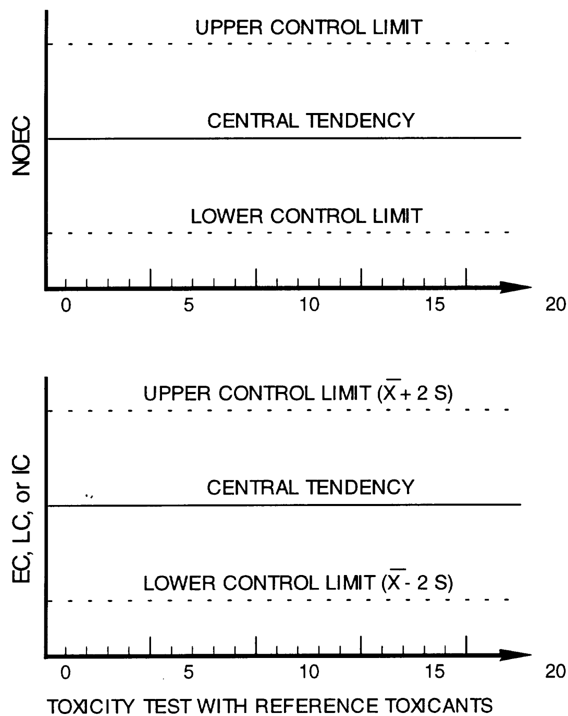
13.4 Demonstrating Acceptable Laboratory Performance

13.4.1 Intralaboratory precision, expressed as a CV, can be determined by performing five or more tests with different batches of test organisms, using the same reference toxicant, at the same concentrations, with the same test conditions (e.g., the same test duration, type of water, age of test organisms, feeding), and same data analysis methods. A reference-toxicity concentration series (dilution factor of 0.5 or higher) should be selected that will provide partial mortalities at two or more concentrations of the test chemical (Section 9.14,Table 9.1). See Section 9.16 for additional detail on reference-toxicity testing.

13.4.2 Test animals should only be obtained from culture. It is likely to be impractical to obtain test-sized neonates directly from a supplier because of their sensitivity to physical disturbances and their rapid growth. Instead, test laboratories will likely want to establish their own cultures of *L. plumulosus* from which to harvest neonates.

13.4.3 Before conducting tests with potentially contaminated sediment, it is strongly recommended that the laboratory conduct the tests with control sediment(s) alone. Results of these preliminary studies should be used to determine if use of the control sediment and other test conditions (i.e., water quality) result in acceptable performance in the tests as outlined in Tables 11.1 and 11.3.

13.4.4 A control chart should be prepared for each combination of reference toxicant and test organism. Each control chart should include the most current data. Endpoints from five tests are adequate for establishing the control charts. In this technique, a running plot is maintained for the values (X_i) from successive tests with a given reference toxicant (Figure 13.1), and the endpoints (LC50, NOEC, ICp) are examined to determine whether they are within prescribed limits. Control charts as described in USEPA (1991a) and USEPA (1993b) are used to evaluate the cumulative trend of results from a series of samples. The mean and upper and lower control limits (±2 SD) are recalculated with each successive test result. After 2 years of data collection, or a minimum of 20 data points, the control (cusum) chart should be maintained using only the 20 most recent data points.



$$\bar{x} = \frac{\sum_{i=1} x_i}{n}$$

$$s = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n-1}}$$

where x_1 = Successive toxicity values of toxicity tests

N = Number of tests

$\bar{x}$ = Mean toxicity value

S = Standard deviation

Figure 13.1 Control charts: (A) hypothesis testing results; and (B) point estimates (LC, EC, or IC)

13.4.5 The outliers, which are values falling outside the upper and lower control limits, and trends of increasing or decreasing sensitivity, are readily identified using control charts. With an " of 0.05, one in 20 tests would be expected to fall outside of the control limits by chance alone. During a 30-d period, if two reference-toxicity tests out of a total previous 20 fall outside the control limits, the sediment toxicity tests conducted during the time in which the second reference-toxicity test failed are suspect and should be considered as provisional and subject to careful review.

13.4.5.1 A sediment test may be acceptable if specified conditions of a reference-toxicity test fall outside the expected ranges (Section 9). Specifically, a sediment test should not necessarily be judged unacceptable if the LC50 for a given reference-toxicity test falls outside the expected range or if mortality in the control of the reference-toxicity test exceeds 10% (Table 9.1). All the performance criteria outlined in Table 11.3 needs to be considered when determining the acceptability of a sediment test. The acceptability of the sediment test would depend on the experience and judgement of the investigator and the regulatory authority.

13.4.6 If the value from a given test with the reference toxicant falls more than two standard deviation (SD) outside the expected range, the sensitivity of the organisms and the overall credibility of the test system may be suspect (USEPA, 1991a). In this case, the test procedure should be examined for defects and should be repeated with a different batch of test organisms.

13.4.7 Performance should improve with experience, and the control limits for point estimates should gradually narrow. However, control limits of ±2 SD, by definition, will be exceeded 5% of the time, regardless of how well a laboratory performs. Highly proficient laboratories that develop a very narrow control limit may be unfairly penalized if a test that falls just outside the control limits is rejected *de facto*. For this reason, the width of the control limits should be considered in determining whether or not an outlier is to be rejected. This determination may be made by the regulatory authority evaluating the data.

13.4.8 The recommended reference-toxicity test consists of a control and five or more concentrations in which the endpoint is an

estimate of the toxicant concentration that is lethal to 50% of the test organisms in the time period prescribed by the test. The LC50 is determined by an appropriate procedure, such as the trimmed Spearman-Karber Method, Probit Method, Graphical Method, or the Linear Interpolation Method (Section 12).

13.4.9 The point estimation analysis methods recommended in this manual have been chosen primarily because they are well-tested, well-documented, and are applicable to most types of test data. Many other methods were considered in the selection process, and it is recognized that the methods selected are not the only possible methods of analysis for toxicity data.

13.5 Precision of the 28-d Chronic Sediment Toxicity Test Method

13.5.1 Intralaboratory Performance

13.5.1.1 Studies described in DeWitt et al. (1997b) provide additional data to characterize intralaboratory precision with the 28-day long-term toxicity test with *L. plumulosus*. This data set provides an estimate of intralaboratory precision from a single laboratory from a total of 88 treatments (Table 13.1). To be consistent with standard statistical procedures, these data were transformed to reduce the heterogeneity of within class variance. Percent survival was transformed to the arcsine-square root of the value; growth rate was transformed to the natural logarithm of the value; and reproduction (offspring per survivor) was transformed to the arcsine -square root of the value. A CV was calculated on the transformed data for each treatment within an experiment. The observed distribution obtained from the resulting sample of CVs from all experiments was then characterized. This distribution of CVs then provides an appropriate range on which to base sample size calculations for future experiments. The median CVs were 11% for survival, 3% for growth rate, and 18% for reproduction (Table 13.1). The range between the first and third quartiles provides a useful nonparametric interval bounding the distribution. This range was 8% to 14% for survival, 2% to 6% for growth rate, and 13% to 36% for reproduction (Table 13.1).

13.5.1.2 These values are similar to CVs for intralaboratory precision calculated for survival from 10-d tests with control sediment using *Hyalella azteca* and *Chironomus tentans* (7.2% and 5.7%, respectively; USEPA, 2000).

13.5.2 Interlaboratory Precision

13.5.2.1 Interlaboratory precision for *L. plumulosus* in the 28-d whole sediment toxicity test using the methods described in this manual (Table 11.1) was evaluated by round-robin testing (DeWitt et al., 1997b). Ten laboratories, including federal and state government laboratories, contract laboratories, and academic laboratories with demonstrated experience in chronic toxicity testing using *L. plumulosus*, participated in round-robin toxicity testing (DeWitt et al., 1997b).

Table 13.1 Intralaboratory Precision Distribution of the Coefficient of Variation for Each Test Endpoint (DeWitt et al. 1997a)

End Point	Sample Size	Mean	Median	Minimum	Maximum	1st Quartile	3rd Quartile
% Survival (Arcsine transformed)	88	14%	11%	0%	173%	8%	14%
Growth rate (log transformed)	87	4%	3%	0%	16%	2%	6%
Reproduction (square root transformed)	88	31%	18%	0%	141%	13%	36%

The experimental design required each laboratory to conduct the 28-d chronic test using a dilution series of Black Rock Harbor sediment (BRH; a Superfund site in Connecticut) mixed with clean, diluent sediment from Sequim Bay, Washington. Each sediment treatment was prepared in a single batch that was subsampled and shipped to testing laboratories. A total of four concentrations of BRH sediment and one negative control sediment were tested. Across all treatments, total organic carbon averaged 2.6% dry weight, total solids averaged 33%, and grain size averaged 15% sand, 42% silt, and 43% clay. In general, cadmium, chromium, copper, lead, nickel, and zinc, as well as total PAHs, increased along the dilution series gradient. Table 13.2 summarizes the concentration ranges for the inorganic contaminants.

13.5.2.2 Approximately 4 months before the start of the round-robin study, laboratories not currently maintaining cultures of *L. plumulosus* were supplied with amphipods, sediment, food, and culturing methods by the Battelle Marine Sciences Laboratory (MSL). Each laboratory maintained cultures following the culturing method detailed in DeWitt et al. (1997a). Each laboratory used its own source of clean seawater.

Table 13.2 Ranges of the BRH Sediment Dilution Series Chemical Concentrations (mg/kg dry wt; from DeWitt et al., 1997b)

	Low (BRH treatment)	High (BRH treatment)
Cadmium	4.09 (0.0%)	13.5 (15.1%)
Chromium	104 (0.0%)	767 (15.1%)
Copper	104 (0.0%)	1503 (15.1%)
Lead	31.1 (0.0%)	209 (15.1%)
Nickel	91.2 (0.0%)	150 (15.1%)
Zinc	189 (0.0%)	736 (15.1%)
Total PAHs	9.85 (1.4%)	17.5 (15.1%)

13.5.2.3 Of the ten laboratories participating in the round-robin, only five laboratories had $80% survival in the negative control sediment, and thereby met this performance criterion for test acceptability (Top of Table 13.3). Analysis of the data resulting from the round-robin included only these five laboratories. Mean survival in the negative control sediment was 93.6%, the CV was 4.2%, and the range was from 89% to 98% (Table 13.3). The CVs across laboratories from the five treatments ranged from 3.1% to 12.8%, with a mean of 8.4%, and increased with dose. None of the laboratories produced less than 70% survival, even in the highest concentration of BRH sediment. Further, none of the laboratories produced a monotonic dose response for survival. This suggests that the test did not contain a wide enough series of dilutions to adequately measure the response of survival. For those laboratories that showed a statistically significant decrease in survival in the highest concentration of BRH (n=4), an average of 16% change in survival was produced between the control and the highest concentration of BRH sediment.

13.5.2.4 For the five laboratories that met the performance criterion, interlaboratory precision for this study was characterized by the maximum and minimum CV for each endpoint. The minimum interlaboratory CV averaged about 4% for survival, 14% for growth rate, and 35% for reproduction (Table 13.4). Maximum interlaboratory CV averaged 19% for survival, 38% for growth rate, and 102% for reproduction. The interlaboratory MDD for survival ranged from 8% to 31%, and the intralaboratory MDD for survival ranged from 10% to 26%. The interlaboratory MDD for growth rate ranged from 0.011 to 0.017 mg/ind/d, and the intralaboratory MDD for growth rate ranged from 0.009 to 0.024 mg/ind/d. The interlaboratory MDD for reproduction ranged from 0.33 to 2.86 offspring per survivor, and the intralaboratory MDD for reproduction ranged from 0.92 to 2.73 offspring per survivor. These MDD's should be interpreted cautiously, because they are derived from one study consisting of a small number of comparisons. Although the technical staff for laboratories participating in the round-robin had extensive sediment toxicity testing experience, many had limited testing experience specifically with *L. plumulosus*. Therefore, these values for

Table 13.3 Results of Round-robin Interlaboratory Precision of Endpoint Sensitivity for *L. plumulosus* in a 28-d Long-term Toxicity Test using Black Rock Harbor Sediments (DeWitt et al., 1997b)

A) Results for Laboratories that met Control Performance Criteria

Lab	0.0%	1.4%	4.6%	8.3%	15.1%	
			Mean Percent Survival (%CV)			MDD %
4	89 (11.5)	92 (3.0)	82 (17.6)	76 (16.4)	73 (13.4)	16
6	96 (6.8)	93 (2.9)	97 (4.6)	95 (7.4)	96 (5.7)	8
7	90 (6.8)	88 (9.5)	84 (12.9)	92 (6.2)	82 (11.9)	13
8	95 (6.4)	92 (6.2)	72 (42.4)	74 (42.0)	70 (18.2)	31
9	98 (2.8)	96 (2.3)	84 (15.4)	91 (10.6)	86 (14.5)	14
Mean	93.6	92.2	83.8	85.6	81.4	
%CV	4.2	3.1	10.6	11.5	12.8	
MDD %	10	7	26	24	16	
			Mean Growth Rate mg/d(%CV)			MDD mg/ind/d
6	0.059 (9.8)	0.054 (6.0)	0.046 (19.0)	0.039 (11.7)	0.020 (24.1)	0.009
7	0.084 (4.4)	0.075 (4.9)	0.063 (8.5)	0.053 (7.2)	0.035 (28.0)	0.009
8	0.045 (18.3)	0.031 (12.7)	0.036 (25.1)	0.024 (27.5)	0.014 (14.1)	0.010
9	0.089 (8.7)	0.078 (13.4)	0.065 (12.7)	0.060 (12.0)	0.045 (11.6)	0.012
Mean	0.063	0.057	0.049	0.039	0.025	
%CV	35.8	35.7	29.8	45.1	59.4	
MDD	0.014	0.014	0.017	0.012	0.011	
			Mean Offspring per Survivor (%CV)			MDD # offspring
4	0.27 (141)	2.26 (72.3)	0.6 5 (149)	0.35 (56.5)	0.33 (81.2)	1.33
6	4.37 (41.0)	2.96 (53.8)	2.58 (27.5)	1.70 (43.4)	0.18 (76.6)	1.77
7	5.22 (55.7)	3.99 (40.5)	3.61 (42.5)	2.21 (75.4)	0.48 (65.6)	2.73
8	1.66 (65.8)	1.10 (54.2)	1.52 (29.8)	0.25 (91.5)	0.10 (108)	0.92
9	7.09 (30.8)	5.43 (21.9)	3.48 (29.8)	1.65 (60.7)	0.19 (99.0)	1.96
Mean	3.72	3.15	2.37	1.23	0.25	
%CV	73.8	52.5	53.8	71.2	59.5	
MDD	2.86	2.10	1.53	1.42	0.33	

Concentration of Black Rock Harbor Sediment (column headings 0.0%–15.1%)

B) Results for Laboratories that did not meet the Control Performance Criteria

Lab	0.0%	1.4%	4.6%	8.3%	15.1%
			Mean Percent Survival (%CV)		
1	53 (31.7)	74 (13.0)	65 (38.5)	58 (18.9)	39 (64.4)
2	0 (—)	10 (—)	27 137.1)	15 (—)	0 (—)
3	72 (34.6)	85 (17.1)	74 (15.4)	61 (21.2)	55 (24.9)
5	60 (56.5)	88 (18.7)	66 (29.5)	84 (24.7)	76 (11.8)
10	69 (29.6)	59 (49.9)	58 (44.2)	37 (70.0)	25 (58.3)
			Mean Growth Rate mg/ind/d (%CV)		
1	0.024 (81.7)	0.032 (37.7)	0.012 (74.9)	0.012 (67.9)	0.008 (71.2)
2	0 (—)	0.027 (—)	0.028 (49.0)	0.017 (—)	0 (—)
3	0.050 (50.2)	0.067 (21.0)	0.055(33.3)	0.034(52.4)	0.025 (32.0)
5	0.058 (16.0)	0.062 (31.7)	0.037 (67.6)	0.036 (43.0)	0.024 (12.0)
10	0.006 (54.5)	0.014 (139)	0.007 (47.1)	0.003 (54.0)	0.003 (80.2)
			Mean Offspring per Survivor (%CV)		
1	0.7 (45.2)	1.7 (57.0)	0.4 (206)	0.1 (163)	0 (—)
2	0 (—)	1.3 (—)	1.2 (18.0)	0.6 (—)	0 (—)
3	4.8 (42.5)	3.7 (51.5)	3.4 (34.9)	0.4 (92.4)	0 (138)
5	3.1 (80.8)	2.3 (25.5)	1.1 (136)	0.8 (113)	0.6 (117)
10	0.1 (131)	1.4 (111)	0.5 (98.3)	0.8 (157)	0.3 (144)

Concentration of Black Rock Harbor Sediment (column headings 0.0%–15.1%)

Table 13.4 Summary of Interlaboratory Precision at Five Laboratories for the 28-Day *Leptocheirus plumulosus* Chronic Test Using Five Dilutions of Black Rock Harbor Sediment (DeWitt et al. 1997b)

	Lab-4	Lab-6	Lab-7	Lab-8	Lab-9
Survival					
Min CV (%)	3	3	6	6	2
Max CV (%)	18	7	13	42	15
Growth rate					
Min. CV (%)	36	6	4	13	9
Max CV (%)	96	24	28	27	13
Offspring per Survivor					
Min CV (%)	56	27	40	30	22
Max CV (%)	149	77	75	108	99

interlaboratory precision may be higher than would be expected from laboratories with routine experience testing with this species.

13.5.2.5 A cost-power analysis was conducted on round-robin data to determine the number of replicates required per treatment for the 28-d whole-sediment standard testing using *L. plumulosus* (DeWitt et al., 1997b). This analysis involved evaluating both the improvement in statistical power of the test to detect a difference between treatment means and the additional expense of adding more replicates. For this analysis, the cost of a replicate was assumed to be proportionate to the time required to conduct all of the tasks associated with one treatment. If cost was not a concern, 14 replicates would be optimal and would provide 80% power for detecting a 30% difference in reproduction at a CV of approximately 36%. This number of replicates is impractical because of costs and logistics. The cost-power analysis for the *L. plumulosus* chronic test indicated that six replicates per treatment gives the greatest statistical power at the most efficient cost. However, this conclusion was based on the assumption that every 1% increase in improved detection equals a 1% increase in cost. The decision to specify 5 replicates per treatment in this manual was based primarily on an effort to keep the cost of performing this test to a minimum. Based on the median CVs for growth rate, survival, and reproduction calculated from a large data set (3%, 11%, and 18%, respectively; see Section 13.5.1.4), five replicates will provide high power ($0.80) to detect a 20% decrease in survival and growth rate endpoints relative to the control (Figure 12.5). For the reproduction endpoint, the power to detect a 20% decrease will be closer to 0.40 using five replicates and 0.50 using six replicates. With power fixed at 80% and at a CV of 20%, the median CV demonstrated for reproduction with five replicates would be suitable to detect approximately 18% reduction in reproduction and with six replicates approximately 16% reduction. Thus, there is relatively little gained by increasing the number of replicates from five to six. Nevertheless, if reproduction is the assessment endpoint of most concern, then incorporation of more than five replicates should be considered. Because space and cost considerations make use of five replicates desirable, this method would benefit from additional research to find ways to reduce the among-replicate variability for the reproduction endpoint.

13.5.2.6 The mean growth rates across the laboratories for each dose decreased with increasing concentration of BRH sediment (Table 13.4). Thus, the growth rate was a more sensitive measure to the concentration of BRH survival. The CVs across the laboratories from the five treatments ranged from 29.8% to 59.4% , with a mean of 41.2%, and were on average five times greater for growth rate than for survival (Table 13.4). Of the five laboratories that met the performance criterion for control survival, three laboratories produced a monotonic dose response to growth rate. The percentage of change in the growth rate between control and the highest concentration of BRH sediment was on average 58% for these three laboratories.

13.5.2.7 The mean reproduction across laboratories for each dose decreased with increasing concentration of BRH sediment. Thus, the measure of reproduction was a more sensitive to the concentration of BRH than was survival; however, the CVs across laboratories are on average eight times greater for reproduction than for survival. The CVs for the five treatments ranged from 52.5% to 73.8%, with a mean of 62.2%. Of the five laboratories that met the performance criteria for control survival, three laboratories produced a monotonic dose response in reproduction. The percentage of change in reproduction (offspring/survivor) between the control and the highest concentration of BRH sediment was on average 95% for these three laboratories.

13.5.2.8 USEPA (2000) included a review of a series of round-robin studies from which interlaboratory precision was analyzed. CVs for survival in 10-d whole-sediment tests with *H. azteca* ranged from 6% to 114% in three test sediments. Similar tests with *C. tentans* produced CVs of 8% to 181% in three test sediments. In 28-d whole-sediment tests with *H. azteca*, CVs from five test sediments ranged from 7% to 28% for survival , from 52% to 78% for growth (dry weight), and from 66% to 193% for reproduction.

13.5.2.9 The *Leptocheirus* round-robin study exhibited similar or better intra- and interlaboratory precision than many chemical analyses and toxicity test methods (USEPA, 1991a; 1991d; 1998). The cause(s) of the high failure rate among laboratories participating in the round-robin study is not known. Several of the laboratories had not conducted this toxicity test previously, and inexperience with the procedures may have contributed to some of the test failures. Some of the laboratories suggested that uneaten food might have accumulated during early days of the experiment, which might have led to lethal low-dissolved oxygen stress to the young amphipods (DeWitt et al 1997b). Because of this potential problem, additional experiments were conducted (Section 11.3.6.4.1) to find the minimum food ration that would minimize the build-up of excess food, minimize mortality, produce significant growth rate and reproduction endpoints of the 28-d *L. plumulosus* sediment toxicity test. The diet recommended in this manual (Section 11.3.6.4) is based on the results of that experiment.

Section 14
References

Adams, W.U., R.A. Kimerle, and R.G. Mosher. 1985. An approach for assessing the environmental safety of chemicals sorbed to sediments. In *Aquatic Toxicology and Hazard Assessment*: 7th Volume, ASTM STP 854, R.D. Cardwell, R. Purdy, and R.C. Bahner, eds., pp. 429-453.

American Chemical Society. Waste Management Manual for Laboratory Personnel, Department of Government Relations and Science Policy, Washington, D.C.

Ankley, G.T. and M. K. Schubauer-Berigan. 1994. Comparison of techniques for the isolation of pore water for sediment toxicity testing. *Arch. Environ. Contam. Toxicol*. 27:507-512.

Ankley, G. and N. Thomas. 1992. Interstitial water toxicity identification evaluation approach. In *Sediment Classification Methods Compendium*, pp. 5-1 to 5-14. EPA-823/R-92-006, Washington, D.C.

Ankley, G.T., D.A. Benoit, J.C. Balough, T.B. Reynoldson, K.E. Day, and R.A. Hoke. 1993. Development and evaluation of test methods for benthic invertebrates and sediments: Effects of flow rate and feeding on water quality and exposure conditions. *Arch. Environ. Contam. Toxicol*. 25:12-19.

Ankley, G.T., D.A. Benoit, J.C. Balough, T.B. Reynoldson, K.E. Day, and R.A. Hoke. 1994a. Evaluation of potential confounding factors in sediment toxicity tests with three freshwater benthic invertebrates. *Environ. Toxicol. Chem*. 13: 627-635.

Ankley, G.T., S.A. Collyard, P.D. Monson, and P.A. Kosian. 1994b. Influence of ultraviolet light on the toxicity of sediments contaminated with polycyclic aromatic hydrocarbons. *Environ. Toxicol. Chem*. 11:1791-1796.

Ankley, G.T., P.M. Cook, A.R. Carlson, D.J. Call, J.A. Swenson, H.F. Corcoran, and R.A. Hoke. 1992a. Bioaccumulation of PCBs from sediments by oligochaetes and fishes: Comparison of laboratory and field studies. *Can. J. Fish. Aquat. Sci*. 49:2080-2085.

Ankley, G.T., M.K. Schubauer-Berigan, and J.R. Dierkes. 1991. Predicting the toxicity of bulk sediments to aquatic orgnisms using aqueous test fractions: Pore water versus elutriate. *Environ. Toxicol. Chem*. 10: 1359-1366.

Ankley, G.T., M.K. Schubauer-Berigan, and R.A. Hoke. 1992b. Use of toxicity identification evaluation techniques to identify dredged material disposal options: A proposed approach. *Environ. Management*. 16:1-6.

ASTM. 2000a. Standard guide for conducting 10-day static sediment toxicity tests with marine and estuarine amphipods. E1367-99 *Annual Book of ASTM Standards* Vol. 13.04, West Conshohocken, PA.

ASTM. 2000b. Standard test methods for measuring the toxicity of sediment-associated contaminants with freshwater invertebrates. E1706-95b. In *Annual Book of ASTM Standards*, Vol. 11.05, West Conshohocken, PA.

ASTM. 2000c. Standard guide for collection, storage, characterization, and manipulation of sediments for toxicological testing. E1391-94. In *Annual Book of ASTM Standards*, Vol. 11.05, West Conshohocken, PA.

ASTM. 2000d. Standard guide for the determination of bioaccumulation of sediment-associated contaminants by benthic invertebrates. E1688-00. In *Annual Book of ASTM Standards*, Vol. 11.05, West Conshohocken, PA.

ASTM. 2000e. Standard guide for designing biological tests with sediments. E1525-94a. In *Annual Book of ASTM Standards*, Vol. 11.05, West Conshohocken, PA.

ASTM. 2000f. Standard guide for conducting acute toxicity tests with fishes, macroinvertebrates, and amphibians. E729-96 In *Annual Book of Standards*, Vol. 11.05, West Conshohocken, PA.

ASTM. 2000g. Standard guide for conducting early life-stage toxicity test with fishes. E1241-98 In *Annual Book of Standards*, Vol. 11.05, West Conshohocken, PA.

ASTM. 2000h. Standard terminology relating to biological effects and environmental fate. E 943-00a. In *Annual Book of ASTM Standards*, Vol. 11.05, West Conshohocken, PA.

ASTM. 2000i. Standard practice for conducting bioconcentration tests with fishes and saltwater bivalve mollusks. E1022-94, In *Annual Book of Standards* Vol. 11.05, West Conshohocken, PA.

Atkinson, A.C. and Donev, A.N. 1992. *Optimum Experimental Designs.* Oxford: Clarendon Press. 328 p.

Bagarinao, T. 1992. Sulfide as an environemental factor and toxicant: tolerance and adaptations in aquatic organisms. *Aquatic Tox*. 24: 21-62.

Bartlett, M.S. 1937. Some examples of statistical methods of research in agriculture and applied biology. *J. Royal Statist. Soc. Suppl.* 4:137-183.

Becker, D.S., and T.C. Ginn. 1995. Effects of storage time on toxicity of sediments from Puget Sound, Washington. *Environ. Toxicol. Chem.* 14:829-836.

Besser, J.M., C.G. Ingersoll, E. Leonard, and D.R. Mount. 1998. Effect of zeolite on toxicity of ammonia and metals to freshwater invertebrates and implications for sediment toxicity identification evaluation (TIE) procedures. *Environ. Toxicol. Chem.*: In review.

Boese B.L., Lee II, H., Specht, D.T., Randall, R.C., and Winsor, M.H. 1990. Comparison of aqueous and solid-phase uptake for hexachlorobenzene in the tellinid clam *Macoma nasuta* A mass balance approach. *Environ. Toxicol. Chem.* 9:221-231.

Bousfield, E.L. 1973. *Shallow Water Gammeridean Amphipoda of New England*. Ithaca, NY: Cornell University Press, 312 pp.

Bridges, T.S., J.D. Farra, and B.M. Duke. 1997. The influence of food ration on sediment toxicity in *Neanthes arenaceodentata* (Annelida:Polychaeta). *Environ. Toxicol. Chem*. 16:1659-1665.

Bufflap, S.T., and Allen, H.E. 1995a. Sediment pore water collection methods for tace metal analysis: A review. *Wat Res.* 29:165-177.

Bufflap, S.T. and Allen, H.E. 1995b. Comparison of pore water sampling techniques for trace metals. *Wat Res.* 29:2015-2054.

Bureau of National Affairs, Inc. 1986. U.S. Environmental Protection Agency General Regulation for Hazardous Waste Management. Washington, D.C.

Burton, G.A. 1991. Assessment of freshwater sediment toxicity. *Environ. Toxicol. Chem*. 10: 1585-1627.

Burton, G.A., Jr. 1992. Sediment collection and processing factors affecting realism. In *Sediment Toxicity Assessment*. ed. G.A. Burton, Jr., 37-66. Lewis Publ. Boca Raton, FL.

Burton, G.A., Jr. and C.G. Ingersoll. 1994. Evaluating the toxicity of sediments. In *The Assessment of Contaminated Great Lakes sediment.* US Environmental Protection Agency Report, Region V, Chicago, IL. EPA-905-B94-002.

Burton, G.A., M.K. Nelson, and C.G. Ingersoll. 1992. Freshwater benthic toxicity tests. In *Sediment Toxicity Assessment*, G.A. Burton, ed.. Lewis Publishers, Chelsea, MI. pp. 213-240.

Canfield, T.J., E.L. Brunson, F.J. Dwyer, C.G. Ingersoll, and N.E. Kemble. 1998. Assessing sediments from the upper Mississippi river navigational pools using benthic invertebrate community evaluation and the sediment quality triad. *Arch. Environ. Contam. Toxicol.* 35:202-212

Canfield, T.J., F.J. Dwyer, J.F. Fairchild, P.S. Haverland, C.G Ingersoll, N.E. Kemble, D.R. Mount, T.W. LaPoint, G.A. Burton, and M.C. Swift. 1996. Assessing contamination in Great Lakes sediments using benthic invertebrate communities and the sediment quality triad approach. *J. Great Lakes Res*. 22:565-583.

Canfield, T.J., N.E. Kemble, W.G. Brumbaugh, F.J. Dwyer, C.G. Ingersoll, and J.F. Fairchild. 1994. Use of benthic invertebrate community structure and the sediment quality triad to evaluate metal-contaminated sediment in the Upper Clark Fork River, Montana. *Environ. Toxicol. Chem*. 13:1999-2012.

Carlson, A.R., G.L. Phipps, V.R. Mattson, P.A. Kosian, and A.M. Cotter. 1991. The role of acid-volatile sulfide in determining cadmium bioavailability and toxicity in freshwater sediments. *Environ. Toxicol. Chem*. 10:1309-1319.

Carr, R.S. and D.C. Chapman. 1995. Comparison of methods for conducting marine and estuarine sediment porewater toxicity tests, extraction, storage, and handling techniques. *Arch. Environ. Contam. Toxicol*. 28: 69-77.

Chapman, P.M. 1989. Current approaches to developing sediment quality criteria. *Environ. Toxicol. Chem.* 8:589-599.

Chapman, P.M., B. Anderson, S. Carr, V. Engle, R. Green, J. Hameedi, M. Harmon, P. Haverland, J. Hyland, C. Ingersoll, E. Long, J. Rodgers, M. Salazar, P.K. Sibley, P.J. Smith, R.C. Swartz, B. Thompson, and H. Windom. 1997. General guidelines for using the sediment quality triad. *Mar. Pollut. Bull*. 34:368-372.

Chapman, P.M., E.A. Power, and G. A. Burton, Jr. 1992. Integrative assessments in aquatic ecosystems. In *Sediment Toxicity Assessment*. G.A. Burton, Jr., ed. Lewis Publishers, Ann Arbor, MI. pp. 313-340.

Ciarelli, S., W.A..P.M.A. Vonck, N.M. van Straalen, and J. Stronkhorst. 1998. Ecotoxicology assessment of contaminated dredged material with the marine amhpipod *Corophium volutator*. *Arch Environ. Contam*. Toxicol. 34:350-356.

CWA (Clean Water Act). Citation Pending

Conover, W.J. 1980. *Practical Nonparametric Statistics*. Second edition. John Wiley and Sons, New York, NY, 493 p.

Conover, W.J., M.E. Johnson, and M.M. Johnson. 1981. A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics* 23: 351-361.

Craig, G.R. 1984. Bioassessment of sediments: Review of previous methods and recommendations for future test protocols. IEC Beak consultants, Ltd. Mississauga, Ontario.

Davenport, R. and A. Spacie. 1991. Acute phototoxicity of harbor and tributary sediments from lower Lake Michigan. *J. Great Lakes Res*. 17:51-56.

Day, K.E., R.S. Kirby, and T.B. Reynoldson. 1992. The effects of sediment manipulations on chronic sediment bioassays with three species of benthic invertebrates. Abstract. Presented at the 13th Annual Meeting, Society of Environmental Toxicology and Chemistry, Cincinnati, OH, Nov. 8-12.

DeFoe, D.L. and G.T. Ankley. 1998. Influence of storage time on toxicity of freshwater sediments to benthic macroinvertebrates. *Environ. Pollut*. 99:123-131.

DeWitt, T.H., G.R. Ditsworth, and R.C. Swartz. 1988. Effects of natural sediment features on survival of the Phoxocephalid amphipod, *Rhepoxynius abronius*. *Marine Envir. Res.,* 25:99-124.

DeWitt, T.H., M.S. Redmond, J.E. Sewall, and R.C. Swartz. 1992a. Development of a chronic sediment toxicity test for marine benthic amphipods. U.S. Environmental Protection Agency. CBP/TRS/89/93.

DeWitt, T.H., B.L. McGee, R. Alden, H. Phelps, and A. Pinkney. 1992b. A multi-species, multi-lab comparison of the toxicity of a Chesapeake Bay sediment. Presented at the Thirteenth Annual Meeting of the Society for Environmental Toxicology and Chemistry, Cincinnati, OH, November 8-12.

DeWitt, T.H., L.A. Niewolny, S.L. Nieukirk, B.D. Gruendell, W.W. Gardiner, and A.B. Borde. 1997a. Support for development of a standard chronic sediment toxicity protocol with the estuarine amphipod, *Leptocheirus plumulosus*. Prepared for the U.S. Environmental Protection Agency, Office of Water. PNWD-2369. Battelle Pacific Northwest Division, Richland, WA.

DeWitt, T.H., M.R. Pinza, L.A. Niewolny, V.I. Cullinan, and B.D. Gruendell. 1997b. Development and evaluation of standard marine/estuarine chronic sediment toxicity test method using *Leptocheirus plumulosus*. Prepared for the U.S. Environmental Protection Agency, Office of Science and Technology. PNNL-11768. Pacific Northwest National Laboratory, Richland, WA.

DeWitt, T.H, M.R. Pinza, and L.A. Niewonly. 1997c. Phase I (Dilution Series Study) and Phase II (Field-Collected Sediment Study) of the PTI Chronic *Leptocheirus plumulosus* Project. Prepared for PTI Environmental, Bellevue, Washington. PNWD-2384. Battelle Pacific Northwest Division, Richland. WA..

DeWitt, T.H., R.C. Swartz, and J.O. Lamberson. 1989. Measuring the acute toxicity of estuarine sediments. *Environ. Toxicol. Chem*. 8:1035-1048.

DeWoskin, R.S. 1984. Good laboratory practice regulations: A comparison. Research Triangle Institute, Research Triangle Park, NC. 63 p.

Dickson, K.L., A.W. Maki, and W.A. Brungs (eds). 1987. Fate and effects of sediment-bound chemicals in aquatic systems: proceedings of the sixth Pellston Environmental Workshop, 1984. Pergamon Press, New York, NY.

Dillon, T.M. and A.B. Gibson. 1986. Bioassessment methodologies for the regulatory testing of freshwater dredged material. Miscellaneous Paper EL-86-6, U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS.

Dillon, T.M., D.W. Moore, and A.S. Jaruis. 1994. The effects of storage temperature and time on sediment toxicity. *Arch. Environ. Contam. Toxicol*. 27:51-53.

Di Toro, D.M., J.D. Mahony, D.J. Hansen, K.J. Scott, M.B. Hicks, S.M. Mays, and M.S. Redmond. 1990. Toxicity of cadmium in sediments: The role of acid volatile sulfide. *Environ. Toxicol. Chem*. 9:1489-1504.

Di Toro, D.M., C.S Zarba, D.J. Hansen, W.J. Berry, R.C. Swartz, C.E. Cowan, S.P. Pavlou, H.E. Allen, N. A. Thomas, and P.R. Paquin. 1991. Technical basis for establishing sediment quality criteria for nonionic organic chemicals using equilibrium partitioning. *Environ. Toxicol. Chem*. 10:1541-1583.

Ditsworth, G.R., D.W. Schults, and J.K.P. Jones. 1990. Preparation of benthic substrates for sediment toxicity testing. *Environ. Toxicol. Chem*. 9:1523-1529.

Dixon, W.J. and F.J. Massey, Jr. 1983. *Introduction to Statistical Analysis*. Fourth edition. McGraw-Hill Book Company, New York, NY. p. 678.

Duan, Y., S.I. Guttman, and J.T.Oris. 1997. Genetic differentiation among laboratory populations of *Hyalella azteca* : Implications for toxicology. *Environ. Toxicol. Chem*. 16:691-695.

Efron, B. 1982. The Jackknife, the Bootstrap, and other resampling plans. CBMS 38, Soc. Industr. Appl. Math., Philadelphia, PA.

Emery, V.L., Jr., and D.W. Moore. 1996. Preliminary protocol for conducting 28-day chronic sublethal bioassays using the estuarine amphipod *Leptocheirus plumulosus* (Shoemaker). Technical Note EEDP-01-36. U.S. Army Corps of Engineers, Vicksbury, MS.

Emery, V.L., Jr., D.W. Moore, B.R. Gray, B.M. Duke, A.B. Gibson, R.B. Wright, and J.D. Farrar. 1997. Development of a chronic sublethal sediment biodassay using the estuarine amphipod *Leptocheirus plumulosus* (Shoemaker). *Environ. Toxic. Chem.* 16:1912-1920.

Environment Canada. 1992. Biological test method: Acute test for sediment toxicity using marine or estuarine amphipods. Environment Canada, Conservation and Protection, Ottawa, Ontario. 83 p., EPS 1/RM/26.

Enviroment Canada. 1994. Guidance Document on Measurement of Toxicity Test Precision Using Control Sediments Spiked with a Reference Toxicant. Report EPS 1/RM/30.

Fairweather, P.G. 1991. Statistical power and design requirements for environmental monitoring. *Aust. J. Mar. Freshwater Res.* 42:555-567.

Finney, D.J. 1971. *Probit Analysis*. Third edition, Cambridge, University Press, London. 333 p., 1971.

FDA. 1978. Good laboratory practices for non-clinical laboratory studies. Part 58. Fed. Reg. 43(247):60013-60020 (December 22, 1978).

Gill, J.L. 1978. *Design and Analysis of Experiments in the Animal and Medical Sciences*. Vol. 3. Appendices. The Iowa State University Press, Ames, IA. 173 p.

Gray, B.R., V.L. Emery, D.L. Brandon, R.B. Wright, B.M. Duke, J.D. Farrar, and D.W. Moore. 1998. Selection and optimal measures of growth and reproduction for the sublethal *Leptocheirus plumulosus* sediment bioassay. *Environ. Toxicol. and Chemistry*. 17: 2288-2297.

Green, A.S. and G.T. Chandler. 1996. Life-table evaluation of sediment-associated chlorpyrifos chronic toxicity to the benthic copepod, *Amphiascus tenuiremis. Arch. Environ. Contam. Toxicol.* 31:77-83.

Green, R.H. 1979. *Sampling Design and Statistical Methods for Environmental Biologists*. Wiley Interscience. New York, NY. 257 p.

Grothe, D.R. and Kimerle, R.A. 1985. Inter-and Intra-laboratory variability in *Daphnia magna* effluent toxicity test results. *Environ. Toxicol. Chem.* 4:189-192.

Hall, W.S., Patroczka, J.B., Mirenda, R.J.,Porter, B.A., and Miller, E. 1989. Acute toxicity of industrial surfactants to *Mysidopisis bahia. Arch. Environ. Contam. Toxicol.* 18:765-772.

Hamilton, M.A., R.C. Russo, and R.V. Thurston. 1977. Trimmed Spearman-Karber method for estimating median lethal concentrations in toxicity bioassays. *Environ. Sci. Technol*. 11:714-719. Correction 12:417 (1978).

Harkey, G.A., S. Kane Driscoll, and P. Landrum. 1997. Effect of feeding in 30-day bioaccumulation assays using *Hyalella azteca* in fluoranthene-dosed sediment. *Environ. Toxicol. Chem*. 16:762-769.

Harkey, G.A., P.F. Landrum, and S.J. Klaine. 1994. Preliminary studies on the effect of feeding during whole sediment bioassays using *Chironomus riparius* larvae. *Chemosphere,* 28:597-606.

Hines, A.H., P.J. Haddon, J.J. Miklas, L.A. Wiechert, and A.M. Haddon. 1986. Estuarine invertebrates and fish: Sampling design and constraints for long-term measurement of population dynamics. In *New Approaches to Monitoring Aquatic Ecosystems*. STP 940, T.P. Boyle, ed., American Society for Testing and Materials.

Holland, A.F., A.T. Shaughnessy, L.C. Scott, V.A. Dickens, J.A. Ranasinghe, and J.K. Summers. 1988. Progress report: Long-term benthic monitoring and assessment program for the Maryland portion of Chesapeake Bay (July 1986-October 1987). PPRP-LTB/EST-88-1. VERSAR, Columbia, MD.

Horwitz, A.J., Elrick, K.A., and Colberg, M.R. 1992. The effect of membrance filtration artifacts on dissolved trace element concentrations. *Wat Res.* 26:753-763.

Hurlbert, S.H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecol. Mono*. 54: 187-211.

Ingersoll, C.G. 1991. Sediment toxicity and bioaccumulation testing. *Standardization News* 19: 28-33.

Ingersoll, C.G. 1995. Sediment toxicity tests. In, *Fundamentals of Aquatic Toxicology* , Second edition, G. Rand, ed. Taylor and Francis, Washington, D.C. pp. 231-255.

Ingersoll, C.G., F.J. Dwyer, S.A. Burch, M.K. Nelson, D.R. Buckler, and J.B. Hunt. 1992. The use of freshwater and saltwater animals to distinguish between toxic effects of salinity and contaminants in irrigation drainwater. *Environ Toxicol Chem*. 11:503-511.

Ingersoll, C.G. and M.K. Nelson. 1990. Testing sediment toxicity with *Hyalella azteca* (Amphipoda) and *Chironomus riparius* (Diptera). In *Aquatic Toxicology and Risk Assessment,* 13th volume. ASTM STP 1096 W.G. Landis and W.H. van der Schalie, eds.). Philadelphia, PA. pp. 93-109, 1990.

Ingersoll, C.G., T. Dillon, R.G. Biddinger (eds.). 1997. Methodological uncertainty in sediment ecological risk assessment. In *Ecological Risk Assessments of Contaminated Sediment*. SETAC Press, Pensacola, FL. 389 p.

Ingersoll, C.G., E.L. Brunson, F.J. Dwyer, D.K. Hardesty, and N.E. Kemble. 1998. Use of sublethal endpoints in sediment toxicity tests with the amphipod *Hyalella azteca*. *Environ. Toxicol. Chem.* 17:1508-1523.

Ingersoll, C.G., D.R. Buckler, E.A. Crecelius, and T.W. LaPoint. 1993. U.S. Fish and Wildlife Service and Battelle final report for the USEPA GLNPO assessment and remediation of contaminated sediment (ARCS) project: Biological assessment of contaminated Great Lakes sediment. EPA-905-R93-006, Chicago, IL.

Ingersoll C.G., F.J. Dwyer, and T.W. May. 1990. Toxicity of inorganic and organic selenium to *Daphnia magna* (Cladocera) and *Chironomus riparius* (Diptera). *Environ. Toxicol. Chem*. 9:1171-1181.

Ingersoll, C.G., P.S. Haverland, E.L. Brunson, T.J. Canfield, F.J. Dwyer, C.E. Henke, and N.E. Kemble. 1996. Calculation and evaluation of sediment effect concentrations for the amphipod *Hyalella azteca* and the midge *Chironomus riparius*. *J. Great Lakes Res*. 22:602-623.

Jordan, R.A. and C.E. Sutton. 1984. Oligohaline benthic invertebrate communities at two Chesapeake Bay power plants. *Estuaries* 7:192-212.

Karickhoff, S.W. and K.R. Morris. 1985. Sorption dynamics of hydrophobic pollutants in sediment suspensions. *Environ. Toxicol. Chem*. 4:469-479.

Kemble, N.E., W.G. Brumbaugh, E.L. Brunson, F.J. Dwyer, C.G. Ingersoll, D.P. Monda, and D.F. Woodward. 1994. Toxicity of metal-contaminated sediments from the upper Clark Fork RIver, MT, to aquatic invertebrates in laboratory exposures. *Environ. Toxicol. Chem*. 13:1985-1997.

Kemp, P.F. and R.C. Swartz. 1988. Acute toxicity of interstitial and particle-bound cadmium to a marine infaunal amphipod. *Mar. Environ. Res*. 26:135-153.

Knezovich, J.P., F.L. Harrison, and R.G. Wilhelm. 1987. The bioavailability of sediment-sorbed organic chemicals: A review. *Water Air Soil Pollut.* 32:233-245.

Knezovich, J.P., D.J. Steichen, J.A. Jelinski, and S.L. Anderson. 1996. Sulfide tolerance of four marine species used to evaluate sediment and pore-water toxicity. *Bull. Environ. Contam. Toxicol.* 57: 450-457.

Kohn, N.P., J.Q. Word, D.K. Niyogi, L.T. Ross, T. Dillon, and D.W. Moore. 1994. Acute toxicity of ammonia to four species of marine amphipod. *Mar. Environ. Res.* 38: 1-15.

Lamberson, J.O. and R.C. Swartz. 1988. Use of bioassays in determining the toxicity of sediment to benthic organisms. In *Toxic Contaminants and Ecosystem Health: A Great Lakes Focus*., M.S. Evans, ed. John Wiley and Sons,New York, NY. pp. 257-279.

Lamberson, J.O. and R.C. Swartz. 1992. Spiked-sediment toxicity test approach. In *Sediment Classification Methods Compendium*. EPA-823-R-92-006, Washington, D.C. pp. 4-1 to 4-10.

Landrum, P.F. 1989. Bioavailability and toxicokinetics of polycyclic aromatic hydrocarbons sorbed to sediments for the amphipod, *Pontoporeia hoyi*. *Environ. Sci. Tech*. 23:588-595.

Landrum, P.F. and W.R. Faust. 1992. Variation in the bioavailability of polycyclic aromatic hydrocarbons sorbed to sediments for the amphipod *Pontoporeia hoyi.* *Environ. Toxicol. Chem.* 11:1197-1208.

Lee, D.R. 1980. Reference toxicants in quality control of aquatic bioassays. In, *Aquatic invertebrate bioassays*. ASTM STP 715, A.L. Buikema and J. Cairns, Jr., eds, Philadelphia, PA. pp. 188-199.

Lee, II, H., B.L. Boese, and P.F. Landrum. 1994. Guide for determination of the bioaccumulation of sediment-associated contaminants by benthic invertebrates (8/94 draft in ASTM review).

Lee, II, H., B.L. Boese, J. Pelletier, M. Winsor, D.T. Specht, and R.C. Randall. 1989. Guidance manual: Bedded sediment bioaccumulation tests. USEPA Report No. 600/x-89/302, ERL-Narragansett. RI. 232 p.

Levin, L., H. Caswell, T. Bridges, C. Dibacco, D. Cabrera, and G. Plaia. 1996. Demographic responses of estuarine polychaetes to pollutants:life table response experiments. *Ecological Applications*. 6:1295-1313.

Long, E.R. and L.G. Morgan. 1990. The potential for biological effects of sediment-sorbed contaminants in the NOAA National Status and Trends program. NOAA Technical Memorandum NOS OMA 52, Seattle, WA. 175 pp + appendices.

Long, E.R., M.F. Buchman, S.M. Bay, R.J. Breteler, R.S. Carr, P.M. Chapman, J.E. Hose, A.L. Lissner, J. Scott, and D.A. Wolfe. 1990. Comparative evaluation of five sediment toxicity tests with sediments from San Francisco Bay and Tomales Bay, CA. *Environ Toxicol. Chem.* 9:1193-1214.

MacDonald, D.D., R.S. Carr, F.D. Calder, E.R. Long, and C.G. Ingersoll. 1996. Development and evaluation of sediment quality guidelines for Florida coastal waters. *Ecotoxicology* 5:253-278.

Marcus, A.H. and A.P. Holtzman. 1988. A robust statistical method for estimating effects concentrations in short-term fathead minnow toxicity tests. Manuscript submitted to the Criteria and Standards Division, U. S. Environmental Protection Agency, by Battelle Washington Environmental Program Office, Washington, D.C., June 1988, under EPA Contract No. 69-03-3534. 39 p.

Marsh, A.G. and K.R. Tenore. 1990. The role of nutrition in regulating the population dynamics of opportunistic, surface deposit feeders in a mesohaline community. *Limnol. Oceanogr.* 35(3):710-724.

McGee, B.L., C.E. Schlekat, D.M Boward, and T.L. Wade. 1995. Sediment contamination and biological effects in a Chesapeake Bay marina. *Ecotoxicology*. 4:39-59.

McGee, B.L., C.E. Schlekat, and E. Reinharz. 1993. Assessing sublethal levels of sediment contamination with the estuarine amphipod, *Leptocheirus plumulosus*. *Environ Toxicol. Chem.* 12:577-588.

McGee, B.L. 1998. Population dynamics of the estuarine amphipod *Leptocheirus plumulosus*: Implications for sediment toxicity tests. Ph.D. dissertation, University of Maryland, College Park, MD.

McGee, B.L., D.A Wright and D.J. Fisher. 1998. Biotic factors modifying the acute toxicity of aqueous cadmium to the estuarine amphipod *Leptocheirus plumulosus*. *Archives of Environmental Contamination and Toxicology* 35:34-40.

McGee, B.L. and D.J. Fisher. 1999. Field validation of the chronic sediment bioassay with the estuarine amphipod *Leptocheirus plumulosus*. Final Report. Prepared for the U.S. Environmental Protection Agency, Office of Science and Technology by the University of Maryland, Wye Research and Education Center, Queenstown, MD.

McGee, B.L., D.J. Fisher, L.T. Yonkos, G.P. Ziegler, and S. Turley. 1999. Assessment of sediment contamination, acute toxicity and population viability of the estuarine amphipod Leptocheirus plumulosus in Baltimore Harbor. *Environ. Toxicol. Chem*. 18:2151-2160.

McNulty, E.W., F.J. Dwyer, M.R. Ellersieck, I.E. Greer, C.G. Ingersoll, and C.F. Rabeni. 1999. Evaluation of the ability of reference-toxicity tests to identify stress in laboratory populations of the amphipod *Hyatella azteca*. *Environ. Toxicol. Chem*. 18:544-548.

Meador, J.P. and C.A. Rice. 2001. Impaired growth in the polychaete Armandia brevis exposed to tributyltin in sediment. *Marine Environmental Research.* 51:113-129.

Moore, D.M., T. M. Dillon, J.Q. Word, and J.A. Ward. 1994. Quality assurance/quality control (QA/QC) guidance for laboratory dredged material bioassays; Results of a QA/QC workshop held May 26-27, 1993, in Seattle, Washington. Miscellaneous Paper D-94-3, U.S. Army Engineer Waterways Experiment Station, Vicksbury, MS.

Moore, D.M. and V.L. Emery, Jr. 1996. Preliminary protocol for conducting 28-day chronic sublethal sediment bioassays using the estuarine amphipod *Leptocheirus plumulosus* (Shoemaker). Environmental Effects of Dredging Technical Notes. EEDP-01-36. March 1996. U.S. Army Engineer Watersays Experiment Station, Vicksburg, MS.

Moore, D.W., T.S. Bridges, B.R. Gray, and B.M. Duke. 1997. Risk of ammonia toxicity during sediment bioassays with the estuarine amphipod *Leptocheirus plumulosus. Environ. Toxicol. Chem.* 16:1020-1027.

Moore, D.W., T.M. Dillon, and E.W. Gamble. 1996. Long-term storage of sediments: Implications for sediment toxicity testing. *Environ. Pollut.* 89:147-155.

National Research Council (NRC). 1989. *Contaminated Marine Sediments--Assessment and Remediation*, NRC, National Academy Press, Washington, D.C.

Niewolny, L.A., T.H. DeWitt, and B.D. Gruendell. 1997. ALCOA Point Comfort Sediment Quality Triad Project: *Leptocheirus plumulosus* Chronic Survival Sediment Toxicity Test. Prepared for Parametrix, Inc. Kirkland, Washington. PNWD-2381. Battelle Pacific Northwest Division, Richland. WA..

Oris, J.T. and J.P. Giesy. 1985. The photoenhanced toxicity of anthracene to juvenile sunfish (*Lepomis* spp.). *Aquat. Toxicol.* 6:133-146.

Pfitezenmeyer, H. 1975. Benthos. In *A Biological Study of Baltimore Harbor*, T. Koo, ed., Contribution No. 621. Center for Environmental and Estuarine Studies, University of Maryland, Solomons, MD.

Plumb, R.H. 1981. Procedures for handling and chemical analyses of sediment and water samples. U.S. Environmental Protection Agency/Corps of Engineers Technical Committee on Criteria for Dredged and Fill Material. Contract EPA-4805572010, EPA/CE-81-1. 478p.

Redmond, M.S. and K.J. Scott. 1989. Amphipod predation by the infaunal polychaete, *Nephtys incisa*. *Estuaries*. 12(3):205-207.

Reinharz, E. 1981. Animal sediment relationships: A case study of the Patapsco River. Open File No.6. Maryland Geological Survey, Baltimore, MD.

Reish, D.J. 1988. The use of toxicity testing in marine environmental research, Chapter 10. In *Marine Organisms as Indicators.* D.F. Soule and G.S. Kleppel eds. Springer-Verlag, New York, NY. pp. 213-245.

Rohlf, F.J. and R.R. Sokal. 1981. *Statistical Tables*. W.H. Freeman and Company, New York, NY.

SAIC. 1993a. Draft data report on comparative sensitivity of four amphipod species to sediment-associated contaminants. Submitted to USEPA. EPA contract No. 68-CO-0044. Prepared by Science Applications International Corporation, Narragansett, RI. 15 pp.

SAIC. 1993b. Draft data report on salinity tolerance and application for four amphipod species. Submitted to USEPA. EPA contract No. 68-CO-0044. Prepared by Science Applications International Corporation, Narragansett, RI. 22 pp.

SAIC. 1993c. Draft data report on comparative sensitivity of four amphipod species to aqueous ammonia. Submitted to USEPA. EPA contract No. 68-CO-0044. Prepared by Science Applications International Corporation, Narragansett, RI. 30 pp.

Sanders, H.L., P.C. Mangelsdorf, Jr., and G.R. Hampson. 1965. Salinity and faunal distribution in the Pocasset River, Massachusetts. Limnology and Oceanography Supplement to Vol 10:R216-R229.

Sarda, N., and G.A. Burton 1995. Ammonia variation in sediments: Spatial, temporal, and method-related effects. *Environ. Toxicol. Chem.* 14:1499-1506.

Sasson-Brickson, G. and G.A. Burton. 1991. In-Situ and laboratory sediment toxicity testing with *Ceriodaphnia dulsia. Environ. Toxicol. Chem.* 10:201-208.

Satterthwaite, F.W. 1946. An approximate distribution of estimates of variance components. *Biom. Bull.* 2:110-114.

Schaeffer, D.J. and Janardan, K.G. 1978. Theoretical comparision of grab and composite sampling programs. *Biom. J.* 20:215-227.

Schlekat, C.E., B.L. McGee, and E. Reinharz. 1992. Testing sediment toxicity in Chesapeake Bay with the amphipod *Leptocheirus plumulosus*: An evaluation. *Environ Toxicol. Chem.* 11:225-236.

Schlekat C.E, K.J. Scott, R.C. Swartz, B. Albrecht, L. Antrim, K. Doe, S. Douglas, J.A. Ferretti, D.J. Hansen, D.W. Moore, C. Mueller, A. Tang. 1995. Interlaboratory comparison of a 10-day sediment toxicity test method using *Ampelisca abdita, Eohaustorius estuarius* and *Leptocheirus-plumulosus*. Environmental Toxicology and Chemistry 14:2163-2174.

Schlekat, C.E. 1998. Assimilation of metals by sediment-ingesting invertebrates: Effects of natural sediment qualities and intrinsic metal characteristics. Ph.D. Thesis. University of South Carolina, Columbia, SC.

Schults, D.W., S.P. Ferraro, L.M. Smith, F.A. Roberts, and C.K. Poindexter. 1992. A comparison of methods for collecting interstitial water for trace organic compounds and metal analyses. *Wat. Res.* 26:989-995.

Schuytema, G.S., A.V. Nebeker, W.L. Griffis, and C.E. Miller. 1989. Effects of freezing on sediments contaminated with DDT and endrin. *Environ. Toxicol. Chem.* 8:883-891.

Scott, J., C. Mueller, W. Starkel, and R. Mayer. 1996. Chronic response of *Leptocheirus plumulosus* in 28-day exposures to contaminated estuarine sediments. Poster presented at the Society of Environmental Toxicology and Chemistry 17th Annual Meeting, November 17-21, 1996, Washington, D.C.

Scott, K.J. 1989. Effects of contaminated sediments on marine benthic biota and communities. In *Contaminated Marine Sediments Assessment and Remediation*. National Research Council. National Academy Press, Washington, D.C. pp. 132-154.

Scott, K.J. and M.S. Redmond. 1989. The effects of a contaminated dredged material on laboratory populations of the tubicolous amphipod, *Ampelisca abdita*. In *Aquatic Toxicology and Hazard Assessment*: 12th Volume, ASTM STP 1027, U.M. Cowgill and L.R. Williams, eds. American Society for Testing and Materials, Philadelphia, PA. pp. 289-303.

Scott. K.J. 1998. Citation pending.

Shapiro, S.S. and M.B. Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52:591-611.

Shoemaker, C.R. 1932. A new amphipod of the genus *Leptocheirus* from Chesapeake Bay. *Wash. Acad. Sci.* 22: 548-551.

Shuba, P.J., H.E. Tatem, and J.H. Carroll. 1978. Biological assessment methods to predict the impact of open-water disposal of dredged material. U.S. Army Corps of Engineers Technical Report D-78-5Q, Washington, D.C.

Sims, J.G. and D.W. Moore. 1995. Risk of pore water hydrogen sulfide toxicity in dredged material bioassays. Miscellaneous paper, D-95-4, U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS.

Sijm, R.T.H., M. Haller, and S.M. Schrap. 1997. Influence of storage on sediment characteristics and of drying sediment on sorption coefficients of organic contaminants. *Bull. Environ. Contam. Toxicol*. 58:961-968.

Snrchek, J.C. and Zeeman, M. 1993. Assessing risks to ecological systems from chemicals. Chapter 3 In: *Handbook for Environmental Risk Assessment and Management*, ed. P.Colow, pp. 24-90.

Snedecor, G.W. and G.C. Cochran. 1989. *Statistical Methods.* Eighth edition. The Iowa State University Press, Ames, IA. 507 p.

Sokal, R.R. and F.J. Rohlf. 1981. Biometry, second edition. W.H. Freeman and Company, New York, NY.

Southerland, E., M. Kravitz, and T. Wall. 1992. Management framework for contaminated sediments (The USEPA Sediment Management Strategy). In *Sediment Toxicity Assessment*, G.A. Burton, ed. Lewis Publishers, Ann Arbor, MI. pp. 341-370.

Steel, R.G.D. and J.H. Torrie. 1980. *Principles and Procedures of Statistics*. McGraw Hill Book Co, New York, NY.

Stemmer, B.L., G.A. Burton Jr., and G. Sasson-Brickson. 1990a. Effect of sediment spatial variance and collection method on cladoceran toxicity and indigenous microbial activity determinations. *Environ. Toxicol. Chem.* 9:1035-1044.

Stemmer, B. L., G.A. Burton, Jr., and S. Leibfritz-Frederick. 1990b. Effect of sediment test variables on selenium toxicity to *Daphnia magna*. Environ. Toxicol. Chem. 9:1035-1044.

Stephan, C.E., D.I. Mount, D.J. Hansen, J.H. Gentile, G.A. Chapman, and W.A. Brungs. 1985. Guidelines for deriving numerical national water quality criteria for the protection of aquatic organisms and their uses. PB85-227049, National Technical Information Service, Springfield, VA.

Swartz, R.C. 1989. Marine sediment toxicity tests. In *Contaminanted Marine Sediments-Assessment and Remediation.* National Research Council, National Academy Press, Washington, DC. pp. 115-129.

Swartz, R.C., W.A. DeBen, J.K.P. Jones, J.O. Lamberson, and F.A. Code. 1985. Phoxocephalid amphipod bioassay for marine sediment toxicity. In *Aquatic Toxicology and Hazard Assessment*: 7th Volume, ASTM STP 854, R.D. Cardwell, R. Purdy, and R.C. Bahner, eds., American Society for Testing and Materials,Philadelphia, PA. pp. 284-307.

Swartz, R.C., P.F. Kemp, D.W. Schultz, and J.O. Lamberson. 1988. Effects of mixtures of sediment contaminants on the marine infaunal amphipod, *Rhepoxynius abronius*. *Environ. Toxicol. Chem*. 7:1013-1020.

Taylor, J.K. 1987. *Quality Assurance of Chemical Measurements*. Lewis Publ., Inc., Chelsea, MI.

Tetra Tech, Inc. 1986. ODES statistical power analysis. Draft report, prepared for the Office of Marine and Estuarine Protection, USEPA Contract No. 68-01-6938, TC-3953-03. Bellevue, WA.

Tomasovic, M. et al. 1994. Recovery of *Hyalella azteca* from sediment. *Environ. Toxicol. Chem*. 13:1311-1314.

USEPA. 1979a. Handbook for analytical quality assurance in water and wastewater laboratories. EPA-600/4-79-019, Cincinnati, OH.

USEPA. 1979b. Methods for chemical analysis of water and wastes. EPA-600/4-79-020, Cincinnati, OH.

USEPA. 1980a. Proposed good laboratory practice guidelines for toxicity testing. Paragraph 163.60-6. Fed. Reg. 45:26377-26382 (April 18, 1980).

USEPA. 1980b. Physical, chemical, persistence, and ecological effects testing; Good laboratory practice standards (proposed rule). 40 CFR 772, Fed. Reg. 45:77353-77365 (November 21, 1980).

USEPA. 1985. Methods for measuring the acute toxicity of effluents to freshwater and marine organisms. Third edition. EPA-600/4-85/013, Cincinnati, OH.

USEPA. 1986. Occupational health and safety manual. Office of Administration, EPA, Washington, D.C.

USEPA. 1989a. Evaluation of the apparent effects threshold (AET) approach for assessing sediment quality. Report of the sediment criteria subcommittee. SAB-EETFC-89-027, Washington, D.C.

USEPA. 1989b. Toxic Substance Control Act (TSCA); Good laboratory practice standards, Final Rule, Federal Register 54: 34034-34050. August 17, 1989.

USEPA. 1990a. Evaluation of the equilibrium partitioning (EqP) approach for assessing sediment quality. Report of the sediment criteria subcommittee of the ecological processes and effects committee. EPA-SAB-EPEC-90-006, Washington, D.C.

USEPA. 1990b. Evaluation of the sediment classification methods compendium. Report of the sediment criteria subcommittee of the ecological processes and effects committee. EPA-SAB-EPEC-90-018, Washington, D.C.

USEPA. 1990c. Managing contaminated sediments. EPA decision-making process. EPA/506/6-90/002.

SEPA. 1991a. Methods for measuring the acute toxicity of effluents to freshwater and marine organisms. Fourth edition. EPA 600/4-90/027F, Cincinnati, OH.

USEPA. 1991b. Sediment toxicity identification evaluation: Phase I (Characterization), Phase II (Identification) and Phase III (Confirmation). Modifications of effluent procedures. EPA 600/6-91/007, Duluth, MN.

USEPA. 1991c. Technical support document for water quality-based toxic control. EPA-505/2-90/001, Washington, D.C.

USEPA. 1992a. Proceedings from workshop on tiered testing issues for freshwater and marine sediments. Washington, D.C., September 16-18, 1992.

USEPA. 1992b. An SAB report: Review of sediment criteria development methodology for non-ionic organic contaminants. Report of the sediment criteria subcommittee of the ecological processes and effects committee. EPA-SAB-EPEC-93-002, Washington, D.C.

USEPA. 1992c. Sediment classification methods compendium. EPA-823-R-92-006, Washington, D.C.

USEPA. 1994a. Procedures for conducting sediment toxicity and bioaccumulation tests using freshwater invertebrates. First Edition. EPA-600/R-94/024, Duluth, MN.

USEPA. 1994b. Short-term methods for estimating the chronic toxicity of effluents and surface waters to freshwater organisms. Third edition. EPA-600/4-91/002, Cincinnati, OH.

USEPA. 1994c. Short-term methods for estimating the chronic toxicity of effluents and surface waters to marine and estuarine organisms. Second edition. EPA-600/4-91/003, Cincinnati, OH.

USEPA. 1994d. Methods for assessing the toxicity of sediment-associated contaminants with estuarine and marine amphipods. EPA/600/R-94/025, Office of Research and Development, Washington, D.C.

USEPA. 1995. QA/QC guidance for sampling and analysis of sediments, water, and tissues for dredged material evaluations—chemical evaluations. EPA 823-B-95-001. Office of Water, Washington, D.C.

USEPA. 1996. Marine toxicity identification evaluation (TIE). Phase I Guidance document. EPA/600/R-96/054. Burgess, R.M., K.T. Ho, G.E. Morrison, G. Chapman, D.L. Denton, eds. U.S. Environmental Protection Agency, National Health and Environmental Effects Research Laboratory, Narragansett, RI.

USEPA. 1997a. The Incidence and Severity of Sediment Contamination in Surface Waters of the United States, Volume 1: National Sediment Quality Survey. EPA 823-R-97-006. Office of Science and Technology, Washington, D.C.

USEPA. 1997b. The Incidence and Severity of Sediment Contamination in Surface Waters of the United States, Volume 2: Data summaries for areas of probable concern. EPA 823-R-97-007. Office of Science and Technology, Washington, D.C.

USEPA. 1997c. The Incidence and Severity of Sediment Contamination in Surface Waters of the United States, Volume 3: National sediment contaminant point source inventory. EPA 823-R-97-008. Office of Science and Technology, Washington, D.C.

USEPA. 1998. Contaminated sediment management strategy. EPA 823-R-98-001. Office of Water, Washington, D.C.

USEPA. 1999. EPA requirements for quality assurance project plans for environmental data operations. Draft interim final. EPA QA/R-5. Office of Research and Development, Washington, D.C.

USEPA. 2000. Methods for measuring the toxicity and bioaccumulation of sediment-associated contaminants with freshwater invertebrates, second edition. EPA600/R-99/064, Duluth, MN.

USEPA. 2001. Methods for collection, storage, and manipulation of sediments for Chemical and Toxicological Analyses (January, 2001 draft).

USEPA-USACE. 1977. Ecological evaluation of proposed discharge of dredged material into ocean waters. Technical commitee on criteria for dredged and fill material. Environmental Effects Laboratory. U.S. Army Engineer Waterways Experiment station. Vicksburg, MS.

USEPA-USACE. 1991. Evaluation of dredged material proposed for ocean disposal (Testing manual). EPA-68-C8-0105. USEPA, Office of Marine and Estuarine Protection, Washington, D.C.

USEPA-USACE. 1998. Evaluation of dredged material proposed for discharge in waters of the U.S. testing manual. EPA-823-B-98-004 Washington, D.C.

Van Rees, K.C.J., E.A. Sudlicky, P. Suresh, C. Rao, and K.R. Reddy. 1991. Evaluation of laboratory techniques for measuring diffusion coefficients in sediments. *Environ. Sci. Technol*. 25:1605-1611.

Walters, D.B. and C.W. Jameson. 1984. *Health and Safety for Toxicity Testing*. Butterworth Publications, Woburn, MA.

Wiederholm, T., A.M. Wiederholm, and M. Goran. 1987. Fresh-water oligochaetes. *Water Air Soil Pollut*. 36:131-154.

Winger, P.V. and Lasier, P.J. 1993. Vacuum-extracted pore water toxicity testing. In *Environmental Toxicology and Risk Assessment,* 2nd volume, eds. J.W. Gorsuch, F.J. Dwyer, C.G. Ingersoll, T.W. LaPoint, 640-662. ASTM STP 1173, Philadelphia, PA.

Winger, P.V., Lasier, P.J., and Jackson, B.P. 1998. The influence of extration procedure on ion concentration in sediment pore water. *Arch Environ. Contam. Toxicol.* 35:8-13.

Williams, D.A. 1971. A test for differences between treatment means when several dose levels are compared with a zero dose control. *Biometrics*. 27: 103-117

Williams, L.G., P.M. Chapman, and T.C. Ginn. 1986. A comparative evaluation of marine sediment toxicity using bacterial luminescence, oyster embryo, and amphipod bioassays. *Marine Environmental Research*, 19: 225-249.

Williams, L.R. 1993. Methods for the EPA's regulatory program. *Environmental Testing and Analysis.* 2:36.

Winer, B.J. 1971. *Statistical Principles in Experimental Design*. Second edition. McGraw-Hill Book Company, New York, NY. 907 p.

Word, J.Q., J.A. Ward, L.M. Franklin, V.I. Cullinan, and S.L. Kiesser. 1987. Evaluation of the equilibrium partitioning theory for estimating the toxicity of the nonpolar organic compound DDT to the sediment dwelling amphipod, *Rhepoxynius abronius*. Battelle Marine Science Research Laboratory Report, Task 1, WA56, Sequim, WA. 60 pp.

Word, J., W. Gardiner, and M. Pinza. 1997. Confounding factors related to marine sediment ecotoxicological testing. PNWD-SA-4435. 1997 Short Course, Society of Environmental Toxicology and Chemistry.

Zar, J.H. 1984. *Biostatistical Analysis*. Second edition. Prentice-Hall, Inc., Englewood Cliffs, NJ. 717 p.

Appendix A

Example Data Sheets For Use with the
28-d Chronic *L. plumulosus* Test

Water Quality Measurements

Project Name: ____________________ Project No: ____________________

Duration: ____________________ Test Day: ____________________

Test type: ____________________ Date: ____________________

Species: ____________________ Page ______ of ______

Position	Treatment		Rep.	Temperature (24-26°C)*	pH (7.0 - 9.0)*	Dissolved Oxygen (>4.0 mg/l)	Salinity (18-22 ppt)

Recorder:________

* Test acceptability limits; take corrective action if values are outside limits.

Reveiewed by: __________________ Date: _____________

Daily Observations

Project Name: ____________________
Duration: ____________________
Test type: ____________________
Species: ____________________

Project No: ____________
Test Day: ____________
Date: ____________
Page ____ of ______

Position Number	Number on Sediment	Number Floating	Comments

Recorder: ____________

Reviewed by ________________ Date ___________

Overlying Water Renewal

Project Name: ____________________ Project No: ____________________

Duration: ____________________ Test Day: ____________________

Test type: ____________________ Date: ____________________

Species: ____________________ Page ___ of ____________________

RENEWAL: **Monday, Wednesday, and Friday**
With the designated small peristaltic pump and correct hose for the type of container, remove 400 mL overlying water from each jar and then replace it with 400 mL of 20‰ seawater at test temperature

Date	M,W, or F	Test Day	Water Renewed	Time	Initials	Animals Fed	Time	Initials

Reviewed by: ___________________ Date: _________

Animal Feeding

Project Name:	Project No:
Duration:	Test Day:
Test type:	Date:
Species:	Page ____ of ____

FEEDING: **Monday, Wednesday, and Friday**
Test Days 0-13 = 20 mg ground Tetramin/test chamber
Test Days 14-28 = 40 mg ground Tetramin/test chamber
Example calculation if you have 60 test chambers- prepare 1200 mg of tetramin in 60 mLs of 20 ppt seawater for day 0 - 13 and 2400 mg of tetramin in 60 mLs of 20 ppt seawater for days 14 - 28

Date	Test Day	Time	Amt. of Food per Container	Initials

Reviewed by: ________________ Date:______________

Termination Form

Project Name: ____________________ Project No: ____________________

Duration: ____________________ Test Day: ____________________

Test type: ____________________ Date: ____________________

Species: ____________________ Page ________ of ________

Position	Replicate	Treatment	# Live	# Dead	# Missing	Comments

Recorder: ____________

Reviewed by____________ Date ______________

Neonate Counting Form

Project Name: ______________________ Project No: ____________

Duration: ______________________ Test Day: ____________

Test type: ______________________ Date: ____________

Species: ______________________ Page ____ of ______

Position Number	Treatment	Replicate	Count 1	Initials	Count 2	Initials

Comments: ____

Reviewed by ________________ Date __________

Neonate Counting Form

Project Name: ______________________________ Project No: ______________

Duration: ______________________________ Test Day: ______________

Test type: ______________________________ Date: ______________

Species: ______________________________ Page ______ of ______

Position Number	Treatment	Replicate	Count 1	Initials	Count 2	Initials

Reviewed by __________________ Date ___________