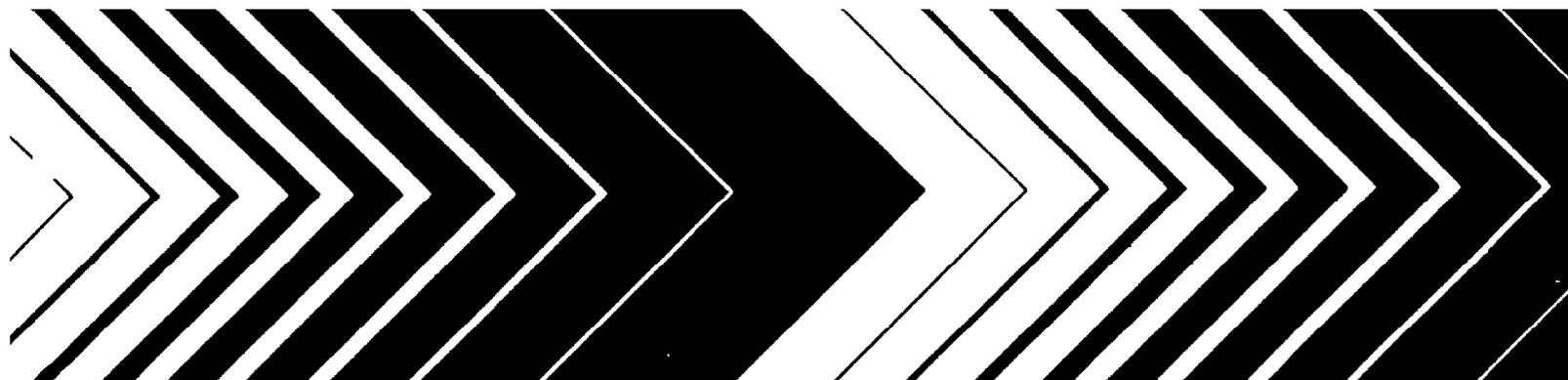




# Methods for Assessing the Toxicity of Sediment-associated Contaminants with Estuarine and Marine Amphipods



## Errata

Pages 80-82, Sections 11.4.5-11.4.5.3, Effects of Sediment-associated Ammonia

These sections describe a procedure that can be used to reduce ammonia concentrations in field-collected sediments prior to conducting laboratory toxicity tests. For dredged material testing under the Clean Water Act or the Marine Protection, Research, and Sanctuaries Act, the following alternative procedure should be used. This procedure was described in a December 21, 1993 guidance memorandum issued by the U.S. EPA Office of Wetlands, Oceans and Watersheds, U.S. EPA Office of Science and Technology, and U.S. Army Corps of Engineers Operations, Construction, and Readiness Division.

For dredged material testing the following procedure should be used if it is necessary to reduce interstitial water ammonia levels. Whenever chemical evidence of ammonia is present at toxicologically important levels, and ammonia is not a contaminant of concern, the laboratory analyst should reduce ammonia in the sediment interstitial water to species-specific no-effect concentrations (see table 11.4 on page 81). Ammonia levels in the interstitial water can be reduced by sufficiently aerating the sample and replacing two volumes of water per day. The analyst should measure interstitial ammonia each day until it reaches the appropriate species-specific no-effect concentration. After placing the test organism in the sediment, the analyst should ensure that ammonia concentrations remain within an acceptable range by conducting the toxicity test with continuous flow or volume replacement not to exceed two volumes per day.

EPA 600/R-94/025  
June 1994

**Methods for Assessing the Toxicity  
of Sediment-associated Contaminants with  
Estuarine and Marine Amphipods**

**Office of Research and Development  
U.S. Environmental Protection Agency  
Narragansett, Rhode Island 02882**

## **Disclaimer**

**This document has been reviewed in accordance with U.S. Environmental Protection Agency Policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.**

## 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 contaminants such as pesticides, herbicides, polychlorinated biphenyls (PCBs), polycyclic aromatic hydrocarbons, and metals such as lead, mercury, and arsenic. In-place contaminated sediment can result in depauperate benthic communities, while disposal of contaminated dredge material can potentially exert adverse effects on both pelagic and benthic systems. Historically, assessment of sediment quality has been limited to chemical characterizations. The United States Environmental Protection Agency (USEPA) is developing methodologies to calculate chemical-specific sediment quality criteria for use in the Agency's regulatory programs. However, quantifying contaminant concentrations alone cannot always provide enough information to adequately evaluate potential adverse effects that arise from interactions among chemicals, or that result from time-dependent availability of sediment-associated contaminants to aquatic organisms. Because relationships between concentrations of contaminants in sediment and bioavailability are not fully understood, determination of contaminated sediment effects on aquatic organisms may require the use of controlled toxicity and bioaccumulation tests.

As part of USEPA's Contaminated Sediment Management Strategy, all Agency programs have agreed to use the same 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 contaminated sediment. The sediment test methods in this manual will be used by USEPA to make decisions under a range of statutory authorities concerning such issues as: dredged material disposal, registration of pesticides and toxic substances, Superfund site assessment, and assessment and cleanup of hazardous waste treatment, storage, and disposal facilities. The use of uniform sediment testing procedures by USEPA programs is expected to increase data accuracy and precision, facilitate test replication, increase the comparative value of test results, and, ultimately, increase the efficiency of regulatory processes requiring sediment tests.

For additional guidance on the technical considerations in the manual, please contact Rick Swartz, USEPA, Newport, OR.

## Abstract

A laboratory method is described for determining the short-term toxicity of contaminated whole sediments using marine and estuarine amphipod crustaceans. Sediments may be collected from estuarine or marine environments or spiked with compounds in the laboratory. A test method is outlined that may be used with any of four amphipod species, including *Ampelisca abdita*, *Eohaustorius estuarius*, *Leptocheirus plumulosus*, and *Rhepoxynius abronius*. The toxicity test is conducted for 10 d in 1 L glass chambers containing 175 mL of sediment and 800 mL of overlying water. Overlying water is not renewed, and test organisms are not fed during the toxicity tests. Temperature and salinity of overlying water, and choice of negative control sediment, are species-specific. The choice of reference sediment may also be species-specific under certain applications. The endpoint in the toxicity test is survival, and reburial of surviving amphipods is an additional measurement that can be used as an endpoint. Procedures are described for use with sediments from oligohaline to fully marine environments.

## **Acknowledgements**

This document is a general purpose testing manual for estuarine and marine sediments. The approaches have also been described in Swartz et al. (1985), Scott and Redmond (1989), DeWitt et al. (1989), Schlekat et al. (1992), ASTM (1992) and ASTM (1994a). The manual incorporates general guidelines that reflect the consensus of the Freshwater Sediment Toxicity Assessment Committee and the U.S. Environmental Protection Agency (USEPA) Program Offices. Members of the Freshwater Sediment Toxicity Assessment Committee are G.T. Ankley, USEPA, Duluth, MN; D.A. Benoit, USEPA, Duluth, MN; G.A. Burton, Wright State University, Dayton, OH; F.J. Dwyer, National Biological Survey (NBS; formerly U.S. Fish and Wildlife Service), Columbia, MO; I.E. Greer, NBS, Columbia, MO; R.A. Hoke, SAIC, Hackensack, NJ; C.G. Ingersoll, NBS, Columbia, MO; P. Kosian, USEPA, Duluth, MN; P.F. Landrum, NOAA, Ann Arbor, MI; J.M. Lazorchak, USEPA, Cincinnati, OH; T.J. Norberg-King, USEPA, Duluth, MN; and P.V. Winger, NBS, Athens, GA.

The principal authors of this document are C.E. Schlekat and K.J. Scott, SAIC, Narragansett, RI and the document was prepared under contract No. 68-C0-004. Contributors to specific sections of this manual are listed alphabetically below.

### **1. Sections 1-9: General Guidelines**

G.T. Ankley, USEPA, Duluth, MN  
G.A. Burton, Wright State University, Dayton, OH  
F.J. Dwyer, NBS, Columbia, MO  
R.A. Hoke, SAIC, Hackensack, NJ  
C.G. Ingersoll, NBS, Columbia, MO  
T.J. Norberg-King, USEPA, Duluth, MN  
C.E. Schlekat, SAIC, Narragansett, RI  
K.J. Scott, SAIC, Narragansett, RI

### **2. Sections 10-11; Culture and Test Methods**

C.E. Schlekat, SAIC, Narragansett, RI  
K.J. Scott, SAIC, Narragansett, RI  
R. Swartz, USEPA, Newport, OR  
T. Dewitt, USEPA, Newport, OR

### **3. Section 12; Statistical Analysis**

J. Heltshe, SAIC, Narragansett, RI  
R.A. Hoke, SAIC, Hackensack, NJ  
H. Lee, USEPA, Newport, OR  
T.J. Norberg-King, USEPA, Duluth, MN  
C.E. Schlekat, SAIC, Narragansett, RI

#### **4. Section 13; Precision and Accuracy**

**G.A. Burton, Wright State University, Dayton, OH;  
T.J. Norberg-King, USEPA, Duluth, MN  
C.E. Schlekat, SAIC, Narragansett, RI  
K.J. Scott, SAIC, Narragansett, RI**

**Review comments from the following individuals are gratefully acknowledged: B.L. McGee, University of Maryland, Queenstown, MD and M.S. Redmond, Northwest Aquatic Science, Newport, OR; C. Philbrick Barr and P. Nolan, Region 1, Lexington, MA; D. Reed, Permits Division, OWEC, Washington, D.C.; P. Crocker, Technical Section and S. McKinney, Marine and Estuarine Section, Region 6, Dallas, TX; F. Schmidt, Monitoring Branch, OWOW, Washington, D.C.; T. Armitage, Standards and Applied Science Division, OST, Washington, D.C.; D. Klemm, EMSL and L. Gast, TAI, Newtown, OH; G. Hanson, OSW, Washington, D.C.; R. Swartz, S. Ferraro, ERL-N, Newport, OR; J. Arthur, R. Spehar, and C. Stephan, ERL-D, Duluth, MN.**

**Participation by the following laboratories in the round-robin testing is greatly appreciated: Aqua Survey, Incorporated, Flemington, NJ; Battelle Marine Science Laboratory, Sequim, WA; Environment Canada, Atlantic Regional Laboratory, Dartmouth, NS; EVS Environmental Consultants, North Vancouver, BC; Science Applications International Corporation, Environmental Testing Center, Narragansett, RI; USACE, Waterways Experiment Station, Vicksburg, MS; USEPA, Edison, NJ; USEPA, Gulf Breeze, FL; USEPA, Narragansett, RI; USEPA, Newport, OR.**

**USEPA's Office of Science and Technology provided support for the development of this manual.**

## Contents

Disclaimer	ii
Foreword	iii
Abstract	iv
Acknowledgements	v
Figures	xii
Tables	xiii
1. Introduction	1
1.1 Significance of Use	1
1.2 Program Applicability	3
1.3 Scope and Application	7
1.4 Performance-based Criteria	15
2. Summary of Method	16
2.1 Method Description and Experimental Design	16
2.2 Types of Tests	18
2.3 Test Endpoints	18
3. Definitions	19
4. Interferences	22
4.1 General Introduction	22
4.2 Non-Contaminant Factors	24
4.3 Changes in Bioavailability	25
4.4 Presence of Indigenous Organisms	26
5. Health, Safety, and Waste Management	27
5.1 General Precautions	27
5.2 Safety Equipment	27
5.3 General Laboratory and Field Operations	28
5.4 Disease Prevention	29
5.5 Safety Manuals	29
5.6 Pollution Prevention, Waste Management, and Sample Disposal	29
6. Facilities, Equipment, and Supplies	30
6.1 General	30
6.2 Facilities	30
6.3 Equipment and Supplies	31
7. Water, Reagents, and Standards	36
7.1 Water	36
7.2 Reagents	38
7.3 Standards	38

## Contents (continued)

8.	Sample Collection, Storage, Manipulation, and Characterization . . . . .	39
	8.1 Collection . . . . .	39
	8.2 Storage . . . . .	39
	8.3 Manipulation . . . . .	40
	8.4 Characterization . . . . .	43
9.	Quality Assurance and Quality Control . . . . .	45
	9.1 Introduction . . . . .	45
	9.2 Performance-based Criteria . . . . .	45
	9.3 Facilities, Equipment, and Test Chambers . . . . .	46
	9.4 Test Organisms . . . . .	46
	9.5 Water . . . . .	46
	9.6 Sample Collection and Storage . . . . .	46
	9.7 Test Conditions . . . . .	46
	9.8 Quality of Test Organisms . . . . .	47
	9.9 Quality of Food . . . . .	47
	9.10 Test Acceptability . . . . .	47
	9.11 Analytical Methods . . . . .	48
	9.12 Calibration and Standardization . . . . .	48
	9.13 Replication and Test Sensitivity . . . . .	48
	9.14 Demonstrating Acceptable Performance . . . . .	48
	9.15 Documenting Ongoing Laboratory Performance . . . . .	49
	9.16 Reference Toxicants . . . . .	49
	9.17 Record Keeping . . . . .	50
10.	Collection, Culture, and Maintaining of Test Organisms . . . . .	53
	10.1 Life History . . . . .	53
	10.2 Species Selection . . . . .	55
	10.3 Field Collection . . . . .	55
	10.4 Holding and Acclimation . . . . .	62
	10.5 Culture Procedure for <i>Leptocheirus plumulosus</i> . . . . .	64
11.	Test Method 100.4 <i>Ampelisca abdita</i> , <i>Eohaustorius estuarius</i> , <i>Leptocheirus plumulosus</i> , or <i>Rhepoxynius abronius</i> 10-d Survival Test for Sediments . . . . .	68
	11.1 Introduction . . . . .	68
	11.2 Recommended Test Method for Conducting a 10-d Sediment Toxicity Test with <i>Ampelisca abdita</i> , <i>Eohaustorius estuarius</i> , <i>Leptocheirus plumulosus</i> , or <i>Rhepoxynius abronius</i> . . . . .	68
	11.3 General Procedures . . . . .	72
	11.4 Interpretation of Results . . . . .	78

**Contents (continued)**

12.	Data Recording, Data Analysis and Calculations, and Reporting .....	83
12.1	Data Recording .....	83
12.2	Data Analysis .....	83
12.3	Data Interpretation .....	109
12.4	Reporting .....	110
13.	Precision and Accuracy .....	112
13.1	Determining Precision and Accuracy .....	112
13.2	Accuracy .....	113
13.3	Replication and Test Sensitivity .....	113
13.4	Demonstrating Acceptable Laboratory Performance .....	113
13.5	Precision of Sediment Toxicity Test Methods .....	115
	References .....	122
	Appendix	
A.	Sample data sheets .....	135

## Figures

Figure 12.1	Treatment response for a Type I and Type II error. . . . .	89
Figure 12.2	Power of the test vs percent reduction in treatment response relative to the control mean at various CV's (8 replicates, alpha = 0.05 (one-tailed)). . . . .	90
Figure 12.3	Power of the test vs percent reduction in treatment response relative to the control mean at various CV's (5 replicates, alpha = 0.05 (one-tailed)). . . . .	91
Figure 12.4	Power of the test vs percent reduction in treatment response relative to the control mean at various CV's (8 replicates, alpha = 0.10 (one-tailed)). . . . .	92
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 (alpha = 0.05 (One-tailed)). . . . .	94
Figure 12.6	Effect of alpha and beta on the number of replicates at various CV's (assuming combined alpha + beta = 0.25). . . . .	95
Figure 12.7	Decision tree for analysis survival and growth data subjected to hypothesis testing. . . . .	104
Figure 12.8	Decision tree for analysis of point estimate data. . . . .	106
Figure A.1	Field collection and laboratory holding data sheet . . . . .	136
Figure A.2	Data sheet for 96 h reference toxicant test. . . . .	137
Figure A.3	Data sheet for daily observations during the 10-d solid phase test. . . . .	138
Figure A.4	Data sheet for 10-d solid phase test. . . . .	139
Figure A.5	Data sheet for test summary. . . . .	140

## Tables

Table 1.1	Sediment quality assessment procedures . . . . .	4
Table 1.2	Statutory needs for sediment quality assessment . . . . .	6
Table 1.3	Rating of selection criteria for estuarine and marine sediment toxicity testing organisms . . . . .	11
Table 4.1	Advantages and disadvantages for use of sediment tests . . . . .	23
Table 6.1	Equipment and supplies for culturing and testing estuarine and marine amphipods. Supplies are for all species unless specified. . . . .	34
Table 9.1	Recommended test conditions for conducting reference-toxicity tests . . . . .	51
Table 10.1	Comparison of habitat characteristics and other life history parameters of four estuarine and marine amphipod species used in sediment toxicity tests . . . . .	57
Table 11.1	Test conditions for conducting a 10-d sediment toxicity test with <i>Ampelisca abdita</i> , <i>Eohaustorius estuarius</i> , <i>Leptocheirus plumulosus</i> , or <i>Rhepoxynius abronius</i> . . . . .	69
Table 11.2	General activity schedule for conducting a sediment toxicity test with <i>Ampelisca abdita</i> , <i>Eohaustorius estuarius</i> , <i>Leptocheirus plumulosus</i> , or <i>Rhepoxynius abronius</i> . . . . .	71
Table 11.3	Test acceptability requirements for a 10-d sediment toxicity test with <i>Ampelisca abdita</i> , <i>Eohaustorius estuarius</i> , <i>Leptocheirus plumulosus</i> , or <i>Rhepoxynius abronius</i> . . . . .	73
Table 11.4	Application limits for 10-d sediment toxicity tests with <i>Ampelisca abdita</i> , <i>Eohaustorius estuarius</i> , <i>Leptocheirus plumulosus</i> , or <i>Rhepoxynius abronius</i> . . . . .	81
Table 12.1	Suggested $\alpha$ levels to use for tests of assumptions . . . . .	87
Table 13.1	Inter-laboratory precision for survival of <i>Rhepoxynius abronius</i> in 10-d whole sediment toxicity tests using seven sediments . . . . .	117
Table 13.2	Inter-laboratory precision for survival of <i>Ampelisca abdita</i> in 10-d whole sediment toxicity tests using four sediments . . . . .	118
Table 13.3	Interlaboratory precision for survival of <i>Eohaustorius estuarius</i> in 10-d whole sediment toxicity tests using four sediments . . . . .	120

**Tables (continued)**

**Table 13.4** Inter-laboratory precision for survival of *Leptocheirus plumulosus* in 10-d whole sediment toxicity tests using four sediments . . . . . 121

## **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 USEPA Water Quality Criteria (WQC) are not exceeded, yet organisms in or near sediments are adversely affected (Chapman, 1989). The WQC were developed to protect organisms in the water column and were not intended to protect organisms in sediment. Concentrations of contaminants in sediment may 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; Di Toro et al., 1991). Although certain chemicals are highly sorbed to sediment, these compounds may still be available to the biota. Contaminated sediments may be directly toxic to aquatic life or can 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 contaminant 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 non-contaminant factors, including water quality fluctuations, physical parameters, and biotic interactions. In order to obtain a direct measure of sediment toxicity, laboratory tests have been developed in which surrogate organisms are exposed to sediments under controlled conditions. Sediment toxicity tests have evolved into effective tools providing direct, quantifiable evidence of biological consequences of sediment contamination that can only be inferred from chemical or benthic community analyses. The USEPA is developing a national inventory of contaminated sediment sites. This inventory will be used to develop a biennial report to Congress on sediment quality in the United States required under the Water Resources Development Act of 1992. The use of consistent sediment testing methods will provide high quality data needed for the national inventory and for regulatory programs to prevent, remediate, and manage contaminated sediment (Southerland et al., 1991).**

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 pathways of interactions among sediments and test organisms is not necessary in order to conduct the tests (Kemp and Swartz, 1988). 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 hazards of dredged material, (6) measure toxicity as part of product licensing or safety testing or chemical approval, (7) rank areas for clean up, and (8) set cleanup goals and estimate the effectiveness of remediation or management practices.

1.1.4 Results of toxicity tests on sediments spiked at different 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 (median lethal concentration), an EC50 (median effect concentration), an IC50 (inhibition concentration), or as a NOEC (no observed effect concentration) or LOEC (lowest observed effect concentration). However, spiked sediment may 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 affect responses.

1.1.5 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; Di Toro et al., 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 non-ionic organic compounds in sediment is often inversely correlated with the organic carbon 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 exposure concentration to an organism. Therefore, information on partitioning of chemicals between solid and liquid phases of sediment is useful for establishing effect concentrations (Di Toro et al., 1991).

1.1.6 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) whole sediment toxicity and sediment-spiking tests, (5) benthic community structure, and (6) Sediment Quality Triad and Range Effects median (see Chapman, 1989; USEPA, 1989a; USEPA, 1990a; USEPA, 1990b; and USEPA, 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., Apparent Effects Threshold; USEPA, 1992c). Numeric methods can be used to derive chemical-specific sediment quality criteria (SQC). Descriptive methods such as toxicity tests with field-collected sediment cannot be used alone to develop numerical SQC for individual chemicals. Although each approach can be used to make site-specific decisions, no one 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). 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; Burton, 1991).

## **1.2 Program Applicability**

1.2.1 The USEPA has authority under a variety of statutes to manage contaminated sediment. Until recently, the USEPA has not addressed sediment quality except in relation to disposal of material removed during navigational dredging (Table 1.2). Southerland et al. (1992) outlined four goals of a USEPA management strategy for contaminated sediments: (1) in-place sediment should be protected from contamination to ensure beneficial uses of surface waters, (2) protection of in-place sediment should be achieved through pollution prevention and source control, (3) in-place remediation should be limited to locations where natural recovery will not occur in an acceptable period of time, and (4) consistent methods should be used to trigger regulatory decisions.

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. The goal of the CWA is to restore and maintain physical, chemical, and biological integrity of the nation's waters (Southerland et al., 1992). Federal and state monitoring programs traditionally have focused on evaluating water column problems caused by point-source dischargers. During the next few years, the USEPA is developing a national inventory of contaminated sediment sites. This inventory will be used to develop a biennial report to Congress on sediment quality in the United States required under the Water Resources Development Act of 1992. The use of consistent sediment testing methods will provide high quality data needed for the national inventory and for regulatory program to prevent, remediate, and manage contaminated sediment (Southerland et al., 1992).

1.2.3 The Office of Water (OW), the Office of Pesticide Programs (OPP), the Office of Pollution Prevention and Toxic Substances (OPPT), the Office of Solid Waste (OSW), and the Office of Emergency and Remedial Response (OERR) are all committed to the

**Table 1.1 Sediment quality assessment procedures**

Method	Type			Approach
	Numeric	Descriptive	Combination	
Equilibrium Partitioning	*			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	*			Safe sediment concentrations of specific chemicals are established by determining the sediment chemical concentration that results in acceptable tissue residues.
Interstitial Water Toxicity	*	*	*	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		*		Environmental degradation is measured by evaluating alterations in benthic community structure.

Note: Modified from USEPA (1992c).

**Table 1.1 Sediment quality assessment procedures (continued)**

Method	Type			Approach
	Numeric	Descriptive	Combination	
<b>Whole sediment Toxicity and Sediment Spiking</b>	*	*	*	Test organisms are exposed to sediments 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.
<b>Sediment Quality Triad</b>	*	*	*	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.
<b>Apparent Effects Threshold (AET)</b>	*	*	*	The sediment concentration of a contaminant above which statistically significant biological effects (e.g., sediment toxicity) are always expected. AET values are empirically derived from paired field data for sediment chemistry and a range of biological effects indicators.

Table 1.2 Statutory needs for sediment quality assessment

Law <sup>1</sup>	Area of Need
CERCLA	<ul style="list-style-type: none"> <li>Assess need for remedial action with contaminated sediments; assess degree of clean-up required, disposition of sediments</li> </ul>
CWA	<ul style="list-style-type: none"> <li>NPDES permitting, especially under Best Available Technology (BAT) in water-quality-limited water</li> <li>Section 403(c) criteria for ocean discharges; mandatory additional requirements to protect marine environment</li> <li>Section 301(g) waivers for publicly owned treatment works (POTWS) discharging to marine waters</li> <li>Section 404 permits for dredge and fill activities (administered by the Corps of Engineers)</li> </ul>
FIFRA	<ul style="list-style-type: none"> <li>Review uses of new and existing chemicals</li> <li>Pesticide labeling and registration</li> </ul>
MPRSA	<ul style="list-style-type: none"> <li>Permits for ocean dumping</li> </ul>
NEPA	<ul style="list-style-type: none"> <li>Preparation of environmental impact statements for projects with surface water discharges</li> </ul>
TSCA	<ul style="list-style-type: none"> <li>Section 5: Pre-manufacture notice reviews for new chemicals</li> <li>Sections 4, 5 and 6: Reviews for existing chemicals</li> </ul>
RCRA	<ul style="list-style-type: none"> <li>Assess suitability (and permit) on-land disposal or beneficial use of contaminated sediments considered "hazardous"</li> </ul>

<sup>1</sup> 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

Note: Modified from Dickson et al. (1984) and Southerland et al. (1992).

principle of consistent tiered testing outlined in the Agency-wide Contaminated Sediment Strategy (Southerland et al., 1992). Agency-wide consistent testing is desirable because all USEPA programs will use similar methods to evaluate whether a sediment poses an ecological or human health risk, and comparable data would be produced. It will also provide the basis for uniform cross-program decision-making within the USEPA. Each program will, however retain the flexibility of deciding whether identified risks would trigger regulatory actions.

1.2.4 Tiered testing should include a hierarchy of tests with the tests in each successive tier becoming progressively more rigorous, complex, and costly (Southerland et al. 1992). Guidance needs to be developed to explain how information within each tier would trigger regulatory action. The guidance could be program specific, describing decisions based on a weight of evidence approach, a pass-fail approach, or comparison to a reference site depending on statutory and regulatory requirements. There are now two approaches for tiered testing used by USEPA: (1) the Office of Water-U.S. Army Corps of Engineers dredged material testing framework and (2) the OPP ecological risk assessment tiered testing framework. Tier 1 of the dredged material testing framework consists of a review of existing chemical and biological data or an inventory of nearby sources. In Tier 2, chemical data are compared to water and sediment quality criteria. Tier 3 evaluations consist of acute toxicity and bioaccumulation testing, and a comparison of the results to a reference area. Tier 4 studies consist of site-specific field studies. The OPP testing framework consists of acute toxicity testing in Tier 1, followed by chronic (early life stage) toxicity testing in Tier 2 and further chronic toxicity testing (full life cycle) in Tier 3. 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 are toxicity tests, bioaccumulation tests, chemical criteria, 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 testing estuarine and marine amphipod crustaceans in the laboratory to evaluate the toxicity of contaminants associated with whole sediments. Sediments may be collected from the field or spiked with compounds in the laboratory. A toxicity method is outlined for four species of estuarine and marine sediment-burrowing amphipods found within United States coastal waters. The species are *Ampelisca abdita*, a marine species that inhabits marine and mesohaline portions of the Atlantic coast, the Gulf of Mexico, and San Francisco Bay; *Eohaustorius estuarius*, a Pacific coast estuarine species; *Leptocheirus plumulosus*, an Atlantic coast estuarine species; and *Rhepoxynius abronius*, a Pacific coast marine species. Generally, the method described may be applied to all four species, although acclimation procedures and some test conditions (i.e., temperature and salinity) will be species-specific (Sections 10 and 11). The toxicity test is conducted for 10 d in 1 L glass chambers containing 175 mL of sediment and 800 mL of overlying seawater. Exposure is static (i.e., water is not renewed), and the animals are not fed over the 10 d exposure period. The endpoint in the toxicity test is survival, and reburial of surviving amphipods is an additional

measurement that can be used as an endpoint. Procedures are described for use with sediments with pore water salinity ranging from >0 ‰ to fully marine.

1.3.2 Additional research and methods development are now in progress to: (1) develop standard chronic sediment toxicity tests (e.g., 28-d exposures with *Leptocheirus plumulosus*), (2) develop standard sediment bioaccumulation tests (i.e., 28-d exposures with the bivalve *Macoma nasuta* and the polychaete *Nereis virens*) (Lee et al., 1989), (3) refine sediment spiking procedures, (4) refine sediment dilution procedures, (5) refine sediment Toxicity Identification Evaluation (TIE) procedures, and (6) produce additional data on confirmation of responses in laboratory tests with natural populations of benthic organisms. This information will be described in future editions of the manual.

1.3.3 This methods manual serves as a companion to the freshwater sediment testing method manual (USEPA, 1994a).

1.3.4 Procedures described in this manual are based on the following documents: Swartz et al. (1985), DeWitt et al. (1989), Scott and Redmond (1989), Schlekot et al. (1992), ASTM (1992), and Environment Canada (1992). This USEPA manual outlines specific test methods for evaluating the toxicity of sediments with *A. abdita*, *E. estuarius*, *L. plumulosus*, and *R. abronius*. While 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 the effect of shipping on organism sensitivity, additional performance criteria for organism health, and confirmation of responses in laboratory tests with natural benthos populations.

1.3.5 General procedures described in this manual might be useful for conducting tests with other estuarine or marine organisms (e.g., *Corophium* spp., *Grandidierella japonica*, *Lepidactylus dytiscus*, *Streblospio benedicti*), although modifications may be necessary. Altering the procedures described in this manual may alter bioavailability and produce results that are not directly comparable with results of acceptable 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 methods different from those described in this manual, additional tests are required to determine comparability of results.

1.3.6 Methods have been described for culturing and testing indigenous species that may be as sensitive or more sensitive than the species recommended in this manual. However, the USEPA allows the use of indigenous species only where State regulations require their use or prohibit importation of the recommended species. Where state regulations prohibit the importation or use of the recommended test species, permission should be requested from the appropriate regulatory agency before their using indigenous species.

1.3.7 Where States have developed culturing and testing methods for indigenous species other than those recommended in the manual, data comparing the sensitivity of the substitute species and one or more of the recommended species 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.

### 1.3.8 Selection of Test Organisms

1.3.8.1 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 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 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, 1993a). Methods utilizing selected organisms should also be (11) peer reviewed (e.g., journal articles, ASTM guides) and (12) confirmed with responses with natural populations of benthic organisms.

1.3.8.2 Of these criteria (Table 1.3), a database demonstrating relative sensitivity to contaminants, contact with sediment, ease of culture in the laboratory or availability for field-collection, ease of handling in the laboratory, tolerance to varying sediment physico-chemical characteristics, and confirmation with responses with natural benthic populations were the primary criteria used for selecting *A. abdita*, *E. estuarius*, *L. plumulosus*, and *R. abronius* for the current edition of the manual. The species chosen for this method are intimately associated with sediment, due to their tube-dwelling or free-burrowing, and sediment ingesting nature. 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; Schlekot et al., 1992; ASTM, 1992). The selection of test species for this manual followed the consensus of experts in the field of sediment toxicology who participated in a workshop entitled "Testing Issues for Freshwater and Marine Sediments". The workshop was sponsored by USEPA Office of Water, Office of Science and Technology, and Office of Research and Development, and was held in Washington, D.C. from 16-18 September 1992 (USEPA, 1992a). Of the candidate species discussed at the workshop, *A. abdita*, *E. estuarius*, *L. plumulosus*, and *R. abronius* best fulfilled the selection criteria, and presented the availability of a combination of one estuarine and one marine species each for both the Atlantic (the estuarine *L. plumulosus* and the marine *A. abdita*) and Pacific (the estuarine *E. estuarius* and the marine *R. abronius*) coasts. *Ampelisca abdita* is also native to portions of the Gulf of Mexico. The existence of established ASTM methods for each species (ASTM, 1992) supported their selection. Many other organisms that might be appropriate for sediment testing do not now meet these selection criteria because little emphasis has been placed on developing standardized testing procedures for benthic organisms. For example, a fifth species for which an ASTM method is

available. *Grandidierella japonica* (ASTM, 1992), was not selected because workshop participants felt that the use of this species was not sufficiently broad to warrant standardization of the method. Environment Canada (1992) has recommended the use of the following amphipod species for sediment toxicity testing: *Amphiporeia virginiana*, *Corophium volutator*, *Eohaustorius washingtonianus*, *Foxiphalus xiximeus*, and *Leptocheirus pinguis*. A database similar to those available for *A. abdita*, *E. estuarius*, *L. plumulosus*, and *R. abronius* must be developed in order for these and other organisms to be included in future editions of this manual.

1.3.8.3 An important consideration in the selection of specific species for test method development is the existence of information concerning relative sensitivity of the organisms both to single chemicals and complex mixtures. Several studies have evaluated the sensitivities of *A. abdita*, *E. estuarius*, *L. plumulosus*, or *R. abronius*, either relative to one another, or to other commonly tested estuarine or marine species. For example, the sensitivity of marine amphipods was compared to other species that were used in generating saltwater Water Quality Criteria. Seven amphipod genera, including *Ampelisca abdita* and *Rhepoxynius abronius*, were among the test species used to generate saltwater Water Quality Criteria for 12 chemicals. Acute amphipod toxicity data from 4-d water-only tests for each of the 12 chemicals was compared to 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

**Table 1.3 Rating of selection criteria for estuarine and marine sediment toxicity testing organisms**

Criterion	<i>Ampelisca abdita</i>	<i>Eohaustorius estuarius</i>	<i>Leptocheirus plumulosus</i>	<i>Rhepoxynius abronius</i>
Relative sensitivity toxicity database	+	+	+	+
Round-robin studies conducted	+	+	+	+
Contact with sediment	+	+	+	+
Laboratory culture	+/-	-	+	-
Maintain in laboratory	+	+	+	+
Taxonomic identification	+	+	+	+
Ecological importance	+	+	+	+
Geographical distribution <sup>1</sup>	ATL, PAC, GLF	PAC	ATL	PAC
Sediment physico-chemical tolerance	+	+	+	+
Field-validated	+	-	-	+
Peer-reviewed	+	+	+	+
Endpoints monitored	Survival	Survival, reburial	Survival	Survival, reburial

<sup>1</sup> ATL = Atlantic Coast, PAC = Pacific Coast, GLF = Gulf of Mexico

Note: A "+" or "-" indicates a positive or negative attribute, respectively.

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.

1.3.8.4 Williams et al. (1986) compared the sensitivity of the *R. abronius* 10-d whole sediment test, the oyster embryo (*Crassostrea gigas*) 48-h abnormality test, and the bacterium (*Photobacterium phosphoreum*) 1-h luminescence inhibition test (i.e., the Microtox<sup>®</sup> test) to sediments collected from 46 contaminated sites in Commencement Bay, WA. *Rhepoxynius abronius* were exposed to whole sediment, while the oyster and bacterium tests were conducted with sediment elutriates and extracts, respectively. Microtox<sup>®</sup> was the most sensitive test, with 63% of the sites eliciting significant inhibition of luminescence. Significant mortality of *R. abronius* was observed in 40% of test sediments, and oyster abnormality occurred in 35% of sediment elutriates. Complete concordance (i.e., sediments that were either toxic or not-toxic in all three tests) was observed in 41% of the sediments. Possible sources for the lack of concordance at other sites include interspecific differences in sensitivity among test organisms, heterogeneity in contaminant types associated with test sediments, and differences in routes of exposure inherent in each bioassay. These results highlight the importance of using multiple assays when performing sediment assessments.

1.3.8.5 Several studies have compared the sensitivity of combinations of the four amphipods to sediment contaminants. For example, there are several comparisons between *A. abdita* and *R. abronius*, between *E. estuarius* and *R. abronius*, and between *A. abdita* and *L. plumulosus*. There are fewer examples of direct comparisons between *E. estuarius* and *L. plumulosus*, and no examples comparing *L. plumulosus* and *R. abronius*. There is some overlap in relative sensitivity from comparison to comparison within each species combination, which appears to indicate that all four species are within the same range of relative sensitivity to contaminated sediments.

1.3.8.5.1 Word et al. (1989) compared the sensitivity of *A. abdita* and *R. abronius* to contaminated sediments in a series of experiments. The experiments followed protocols developed specifically for each species; thus, *A. abdita* was tested at 20°C, whereas *R. abronius* was tested at 15°C. Experiments were designed to compare the sensitivity of the protocols rather than to provide a comparison of the response of the organism. Sediments collected from Oakland Harbor, CA, were used for the comparisons. Twenty-six sediments were tested in one comparison, while 5 were tested in the other. Analysis of results using Kruskal Wallance rank sum test for both experiments demonstrated that *R. abronius* exhibited greater sensitivity to the sediments than *A. abdita*. Long and Buchman (1989) also compared the sensitivity of *A. abdita* and *R. abronius* to sediments from Oakland Harbor, CA. They also determined that *A. abdita* showed less sensitivity than *R. abronius*, but they also showed that *A. abdita* was less sensitive to sediment grain size factors than *R. abronius*.

1.3.8.5.2 DeWitt et al. (1989) compared the sensitivity of *E. estuarius* and *R. abronius* to sediment spiked with flouranthene and field-collected sediment from industrial waterways in Puget Sound, WA, in 10-d tests, and to aqueous cadmium (CdCl<sub>2</sub>) in a 4-d water-only test. The sensitivity of *E. estuarius* was from two (to flouranthene-spiked sediment) to seven (to one Puget Sound, WA, sediment) times less sensitive than *R. abronius* in sediment tests, and ten times less sensitive to CdCl<sub>2</sub> in the water-only test. These results are supported by the findings of Pastorak and Becker (1990) who found the acute sensitivity of *E. estuarius* and *R. abronius* to be generally comparable to each other, and both were more sensitive than *Neanthes* (survival and biomass endpoints), *Panope* (survival), and *Dendraster* (survival).

1.3.8.5.3 *Leptocheirus plumulosus* was as sensitive as the freshwater amphipod *Hyaella 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 sensitivity of *L. plumulosus* with three other amphipod species, two molluscs, and one polychaete to highly contaminated sediment collected from Baltimore Harbor, MD, that was serially diluted with clean sediment. *Leptocheirus plumulosus* was more sensitive than the amphipods *Hyaella azteca* and *Lepidactylus dytiscus* and exhibited equal sensitivity with *E. estuarius*. Comparisons using dilutions of sediment collected from Black Rock Harbor, CT, show that *A. abdita* shows greater sensitivity than *L. plumulosus* when the latter is tested at 20°C (SAIC, 1993a). However, *L. plumulosus* is more sensitive at 25°C, the temperature at which chronic test methods with this species are being developed (DeWitt, 1992a), than *A. abdita* at 20°C (SAIC, 1993a).

1.3.8.6 Limited comparative data is available for concurrent water-only exposures of all four species in single-chemical tests. Studies that do exist generally show that no one species is consistently the most sensitive.

1.3.8.6.1 The relative sensitivity of the four amphipod species to ammonia was determined in ten-d water only toxicity tests in order to aid interpretation of results of tests on sediments where 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 this manual. Sensitivity to total ammonia increased with increasing pH for all four species. The rank sensitivity was *R. abronius* = *A. abdita* > *E. estuarius* > *L. plumulosus*.

1.3.8.6.2 Cadmium chloride has been a common reference toxicant for all four species in 4-d exposures. DeWitt et al. (1992a) reports the rank sensitivity as *R. abronius* > *A. abdita* > *L. plumulosus* > *E. estuarius* at a common temperature and salinity of 15°C and 28 ‰. A series of 4-d exposures to cadmium that were conducted at species-specific temperatures and salinities showed the following rank sensitivity: *A. abdita* = *L. plumulosus* ≈ *R. abronius* > *E. estuarius* (SAIC, 1993a; SAIC, 1993b; and SAIC, 1993c).

**1.3.8.6.3 Relative species sensitivity frequently varies among contaminants:** consequently, a battery of tests including organisms representing different trophic levels may be needed to assess sediment quality (Craig, 1984; Williams et al., 1986a; Long et al., 1990; Ingersoll et al., 1990; Burton and Ingersoll, 1994). For example, Reish (1988) reported the relative toxicity of six metals (arsenic, cadmium, chromium, copper, mercury, and zinc) to crustaceans, polychaetes, pelecypods, and fishes and concluded that no one species or group of test organisms was the most sensitive to all of the metals.

**1.3.8.7 The sensitivity of an organism is related to route of exposure and biochemical response to contaminants.** Sediment-dwelling organisms can receive exposure via 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 and detritus may receive most of their body burden directly from materials attached to sediment or from actual sediment ingestion. In some 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 may also accumulate contaminants by direct adsorption to the body wall or by absorption through the integument (Knezovich et al., 1987).

**1.3.8.8 Despite the potential complexities in estimating the dose that an animal receives from sediment, the toxicity and bioaccumulation of many contaminants in sediment such as Kepone<sup>®</sup>, flouranthene, organochlorines, and metals have been correlated with either the concentration of these chemicals in interstitial water or in the case of non-ionic organic chemicals, concentrations in sediment on an organic carbon normalized basis (Di Toro et al., 1990; Di Toro et al., 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.

**1.3.8.9 The use of *A. abdita*, *E. estuarius*, and *R. abronius* in laboratory toxicity studies has been field validated with natural populations of benthic organisms (Swartz et al., 1994 for *E. estuarius*; Swartz et al., 1982 for *R. abronius*).** While no laboratory information is available, 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; Reinhartz, 1981).

**1.3.8.9.1 Data from USEPA Office of Research and Development's Environmental Monitoring and Assessment program were examined to evaluate the relationship between survival of *Ampelisca abdita* in sediment toxicity tests and the presence of amphipods, particularly ampeliscids, in field samples.** Over 200 sediment samples from two years of sampling in the Virginian Province (Cape Cod, MA, to Cape Henry, VA)

were available for comparing synchronous measurements of *A. abdita* survival in toxicity tests to benthic community enumeration. Although species of this genus were among the more frequently occurring taxa in these samples, ampeliscids were totally absent from stations that exhibited *A. abdita* test survival <60% of that in control samples. Additionally, ampeliscids were found in very low densities at stations with amphipod test survival between 60 and 80% (J. Scott, SAIC, Narragansett, RI, personal communication). These data indicate that tests with this species are predictive of contaminant effects on sensitive species under natural conditions.

1.3.8.9.2 Swartz et al.(1982) compared sensitivity of *R. abronius* to sediment collected from sites in Commencement Bay, WA, to benthic community structure at each site. Mortality of *R. abronius* was negatively correlated with amphipod density, and phoxocephalid amphipods were ubiquitously absent from the most contaminated areas. Schlekat et al., (1994) reported general good agreement between sediment tests with *H. azteca* and benthic community responses in the Anacostia River, Washington, DC.

1.3.8.9.3 Sediment toxicity to amphipods in 10-d toxicity tests, field contamination, and field abundance of benthic amphipods were examined along a sediment contamination gradient of DDT (Swartz et al., 1994). Survival of *E. estuarius* and *R. abronius* in laboratory toxicity tests was positively correlated to abundance of amphipods in the field and along with the survival of *H. azteca*, was negatively correlated to DDT concentrations. The threshold for 10-d sediment toxicity in laboratory studies was about 300 µg DDT (+metabolites)/g organic carbon. The threshold for abundance of amphipods in the field was about 100 µg DDT (+metabolites)/g organic carbon. Therefore, correlations between toxicity, contamination, and biology indicate that acute sediment toxicity tests can provide reliable evidence of biologically adverse sediment contamination in the field.

## 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 which 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 having healthy organisms of known quality. A performance-based criteria approach was selected as the preferred method through which individual laboratories should evaluate culture methods or the quality of field-collected organisms rather than by control-based criteria. This method was chosen to allow each laboratory to optimize culture methods, determine the quality field-collected organisms, and minimize effects of test organism health on the reliability and comparability of test results. See Table 11.3 for a listing of performance criteria used to assess the quality of cultured (i.e., *L. plumulosus*) and field-collected amphipods, and to determine the acceptability of 10-d sediment toxicity tests.

## Section 2 Summary of Method

### 2.1 Method Description and Experimental Design

#### 2.1.1 Method Description

2.1.1.1 This manual describes a laboratory method for determining the short-term toxicity of contaminated whole sediments using marine and estuarine amphipod crustaceans. Test sediments may be collected from estuarine or marine environments or spiked with compounds in the laboratory. A single test method is outlined that may be used with any of four amphipod species, including *Ampelisca abdita*, *Eohaustorius estuarius*, *Leptocheirus plumulosus*, and *Rhepoxynius abronius*. The toxicity test is conducted for 10 d in 1 L glass chambers containing 175 mL of sediment and 800 mL of overlying water. Overlying water is not renewed, and test organisms are not fed during the toxicity tests. Temperature and salinity of overlying water, and choice of control sediment (i.e., negative control), are species-specific. The choice of reference sediment may be species-specific under certain applications. The endpoint in the toxicity test is survival, and reburial of surviving amphipods is an additional measurement that can be used as an endpoint. Procedures are described for use with sediments from oligohaline to fully marine environments.

2.1.2 **Experimental Design.** The following section is a general summary of experimental design. See Section 12 for additional detail.

#### 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 a sediment that is essentially free of contaminants and is used routinely to assess the acceptability of a test and 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 non-point sources (Lee et al., 1994). 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 collected near an area of concern 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.

2.1.2.1.2 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 must 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 including grain size and organic carbon can be evaluated if the limits are exceeded in a test sediment. See Section 10.1 for tolerance limits of each species for physico-chemical characteristics. If the physico-chemical characteristic(s) of a test sediment exceed the tolerance limits of the test organism, it

may be desirable to include a control sediment that encompasses those characteristics, or to choose a test organism with tolerance limits that are not exceeded by the physico-chemical characteristics in question. The effects of some sediment characteristics on the results of sediment tests may be able to be addressed with regression equations (DeWitt et al., 1988; Ankley et al., 1994).

**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 sediments, the number of treatments and replicates, and water quality characteristics. For instance, the purpose of the study might be to determine a specific endpoint such as an LC50 and may include a control sediment, a positive control, and several concentrations of sediment spiked with a chemical. A useful 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., ANOVA), 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 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 characteristics of each of these grabs would be required for 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, 1993b).

**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. Sediment 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 pre-dredging 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 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 exchange between test chambers are kept to a minimum. As the number of test chambers per treatment increases, the number of degrees of freedom increases, and, therefore, the width of the confidence interval on a point estimate, such as an LC50, decreases, and the power of a significance test increases (Section 12). Because of factors that might affect results within test chambers and results of a test, 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 non-biased.

## **2.2 Types of Tests**

2.2.1 A toxicity method is outlined for four species of estuarine and marine amphipod, including *Ampelisca abdita*, *Eohaustorius estuarius*, *Leptocheirus plumulosus*, and *Rhepoxynius abronius* (Section 11). The manual describes procedures for testing sediments from oligohaline to fully marine environments.

## **2.3 Test Endpoints**

2.3.1 The primary endpoint measured in the toxicity test is survival. Reburial of surviving amphipods in control sediment is an additional measurement that can be used as an endpoint. Behavior of test organisms should be qualitatively observed daily in all tests (e.g., avoidance of sediment).

## **Section 3 Definitions**

### **3.1 Terms**

The following terms were defined in Lee (1980), NRC (1989), USEPA (1989b), USEPA-USCOE (1991), USEPA-USCOE (1994), Lee et al. (1994), ASTM (1994a), or ASTM (1993b).

#### **3.1.1 Technical Terms**

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

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

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

**3.1.1.4 *Control sediment.*** A 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 does not reflect any substantial input from local or non-point sources. Comparing test sediments to control sediments is a measure of the toxicity of a test sediment beyond inevitable background contamination.

**3.1.1.5 *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 dredge material evaluations.

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

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

**3.1.1.8 *Reference-toxicity test.*** A test conducted in conjunction with sediment tests to determine possible changes in condition of the test organisms. Deviations outside an established normal range indicate a change in the condition of the test organism population. Reference-toxicity tests are most often performed in the absence of sediment.

3.1.1.9 *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.

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

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

3.1.1.12 *No observable 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 significant different from the controls).

3.1.1.13 *Lowest observable Effect Concentration (LOEC)*. The lowest concentration of a toxicant to which organisms are exposed in a test which causes an adverse effect on the test organisms (i.e., where the value for the observed response is statistically significant different from the controls).

3.1.1.14 *Lethal concentration (LC)*. The toxicant concentration that would cause death in a given percent 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.

3.1.1.15 *Effect concentration (EC)*. The toxicant concentration that would cause an effect in a given percent 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 death in 50% of the test population.

3.1.1.16 *Inhibition concentration (IC)*. The toxicant concentration that would cause a given percent reduction in a non-quantal 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.

### 3.1.2 Grammatical Terms

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

3.1.2.2 "Must" is used to express an absolute requirement, that is, to state that a test ought to be designed 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.

3.1.2.3 "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.

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

3.1.2.5 "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 that can potentially affect test organism survival aside from those related to sediment-associated contaminants. These interferences can potentially confound interpretation of test results in two ways: (1) toxicity is observed in the test when contamination is not present, or there is more toxicity than expected; and (2) no toxicity is observed when contaminants are present at elevated concentrations, or there is less toxicity than expected.

**4.1.2** There are three categories of interfering factors: those characteristics of sediments affecting survival independent of chemical concentration (i.e., non-contaminant factors); changes in chemical bioavailability as a function of sediment manipulation or storage; and 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.

**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 may alter 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, mixing, storage, and testing is extremely difficult and may complicate the interpretation of effects. Direct comparisons of organisms exposed 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) and aging (Word et al., 1987; Landrum, 1989; Landrum and Faust, 1992) of spiked sediment can affect responses of organisms.

**4.1.3.1** Laboratory sediment 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 Toxicity Identification Evaluations (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). 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).

**4.1.4** Methods which measure sublethal effects are either not available or have not been routinely used 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.,  $\leq 10$  d; USEPA-

Table 4.1 Advantages and disadvantages for use of sediment tests

---

**Advantages**

- Measure bioavailable fraction of contaminant(s).
- Provide a direct measure of benthic effects, assuming no field adaptation or amelioration of effects.
- Limited special equipment is required.
- Methods are rapid and inexpensive.
- Legal and scientific precedence exists for use; ASTM standard guides are available.
- 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.

---

Note: Modified from Swartz (1989).

USCOE, 1977; USEPA-USCOE, 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. 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 (Scott, 1989). Additional methods development of chronic sediment testing procedures and culturing of infaunal organisms with a variety of feeding habits including suspension and deposit feeders is needed.

4.1.5 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. 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 two to four weeks. Legal and scientific precedents exist for use of toxicity and bioaccumulation tests in regulatory decision-making (e.g., USEPA, 1986a). 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 Non-Contaminant Factors

4.2.1 Results of sediment tests can be used to predict effects that may occur with aquatic organisms in the field as a result of exposure under comparable conditions. Yet motile organisms might avoid exposure in the field. 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). Fluorescent light does not contain UV light, but natural sunlight does. Lighting can therefore affect toxicological responses and is an important experimental variable for photoactivated chemicals. However, lighting typically used to conduct laboratory tests does not include the appropriate spectrum of ultraviolet radiation to photoactivate compounds (Oris and Giesy, 1985) and thus laboratory tests may not account for toxicity expressed by this mode of action.

4.2.2 There are a number of non-contaminant factors that may influence amphipod survival in these tests. The most important and variable factors include sediment particle size, pore water salinity, and pore water ammonia. The physico-chemical properties of each test sediment must be within the tolerance limits of the test organism. Tolerance limits of the four amphipod species described in this manual for the factors listed above are well defined and are presented in Section 11.4 and summarized in Table 11.4. If a particular sediment characteristic exceeds the tolerance of the proposed test species, another, more appropriate species may be used or control samples (treatments) for exceeded factors may be included in the test design. The effects of sediment characteristics can also be extrapolated with regression equations (DeWitt et al., 1988; Ankley et al., 1994) that estimate the proportion of toxicity that may be due to the non-contaminant factor alone.

### **4.3 Changes in 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. Although the tests are not designed to simulate natural conditions, there is concern that contaminant availability in laboratory toxicity test is different from what it is representative of in-place sediments in the field.**

**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. These manipulation processes are generally thought to increase availability of organic compounds because of disruption of the equilibrium with organic carbon in the pore water per particle system. Similarly, oxidation of anaerobic sediments increases the availability of certain metals (Di Toro et al., 1990). Because the availability of contaminants may be a function of the degree of manipulation, this manual recommends that handling, storage, and preparation of the sediment for actual testings be as consistent as possible. Although very disruptive of natural sediment physical features, all test sediments should be press-sieved sometime before testing and re-homogenized immediately before introduction to the test chambers if warranted (See Section 8.3.1). Press-sieving is performed primarily to remove predatory organisms, large debris, or organisms taxonomically similar to the test species. Certain USEPA program offices may recommend that sediments should not be press-sieved. Also, it may not be necessary to press-sieve sediments if previous experience has demonstrated the absence of potential interferences, including predatory or competitive organisms or large debris, or if large debris or predators can be removed with forceps or other suitable tools. The presence of an abundance of amphipods that are taxonomically similar to the test species should prompt press-sieving. This is particularly true if endemic Ampeliscidae are present and *A. abdita* is the test species because it may be difficult to remove all of the resident amphipods from their tubes. If sediments must be sieved, it may be desirable to perform select analyses (e.g., pore-water metals or DOC, AVS, TOC) on samples before and after sieving to document the influence of sieving on sediment chemistry.**

**4.3.3 Testing sediments at temperatures different from that in the field might affect contaminant solubility, partitioning coefficients, or other physical and chemical characteristics. Interaction between sediment and overlying water and the ratio of sediment to overlying water may influence bioavailability (Stemmer et al., 1990b).**

**4.3.4 Depletion of aqueous and sediment-sorbed contaminants resulting from uptake by an organism or test chamber may also influence 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 may limit uptake rates. Within minutes to hours, a major portion of the total chemical may be inaccessible to the organisms because of depletion of available residues. The desorption of a particular compound from sediment may range from easily reversible (labile; within minutes) to irreversible (non-labile; within days or months; Karickhoff and Morris, 1985). Inter-particle diffusion or advection and the quality and quantity of sediment organic carbon can also affect sorption kinetics.**

**4.3.5** 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 non-ionic 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).

**4.3.6** Salinity of the overlying water is an additional factor that can affect the bioavailability of metals. Some metals (e.g., cadmium) are more bioavailable at lower salinities. 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 suite of species provided in this manual allow these tests to be conducted over the range of pore water salinities routinely encountered in field-collected sediments from North American estuarine and marine environments.

#### **4.4 Presence of Indigenous Organisms**

**4.4.1** Indigenous organisms may be present in field-collected sediments. An abundance in the sediment sample of the test organism, or organisms taxonomically similar to the test organism, may make interpretation of treatment effects difficult. The presence of predatory organisms can also adversely affect test organism survival. For example, Redmond and Scott (1989) showed that the polychaete *Nephtys incisa* will consume *Ampelisca abdita* under toxicity test conditions. Previous investigators have inhibited the biological activity of sediment with sieving, heat, mercuric chloride, antibiotics, or gamma irradiation (Day et al., 1992). Although further research is needed to determine effects on contaminant bioavailability from treating sediment to remove or destroy indigenous organisms, estuarine and marine sediments must be press-sieved before the start of a sediment toxicity test if the presence of predatory organisms is suspected (See Section 8.3.1.1).

## **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 which may involve hazardous materials, operations, and equipment, and 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 the manual to encompass all safety requirements necessary to conduct sediment tests.

5.1.3 Collection and use of sediments 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 sediments under a ventilated hood or 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 which 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 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 safety equipment, such as rubber aprons, laboratory coats, respirators, gloves, safety glasses, face shields, hard hats, and safety shoes as appropriate. The degree of protection should vary according to the level contamination associated with the test sediments. Generally, a higher degree of coverage should be adopted in all aspects of testing sediments that may harbor hazardous levels of

compounds. Coverage for testing control or moderately contaminated sediment does not have to be as stringent.

## **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 fountains.

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 the procedures described in this manual. Routinely encountered chemicals include acids and organic solvents. Special handling and precautionary guidance in Material Safety Data Sheets should be followed for reagents and other chemicals purchased from supply houses. All containers should be adequately labeled to indicate their contents.

5.3.2 Work with some sediments may 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 fumes 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 in 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 results.

## **5.4 Disease Prevention**

**5.4.1** Personnel handling samples which are known or suspected to contain human wastes should be given the opportunity to be immunized against hepatitis B, tetanus, typhoid fever, and polio.

## **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 hods and bench operations. Also, compliance is require 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 your responsibility to know and comply with the pertinent regulations applicable in the State in which you are operating. Refer to the Bureau of National Affairs, Inc. (1986) for the citations of the Federal requirements.

## 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 "non-toxicant" test with each potential test species, in which all test chambers contain a control sediment (sometimes called the negative control), and clean overlying water for each amphipod species to be tested. Survival of the test organism will demonstrate whether facilities, water, control sediment, and handling techniques are adequate to achieve acceptable species-specific control survival. Evaluations may also be conducted of the magnitude of the within- and between-chamber variance in a test.

### 6.2 Facilities

6.2.1 The facility should include separate areas for culturing and 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 in which sediment tests are conducted, stock solutions or where 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. Enclosure of the test systems 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 is adequate (about 500 to 1000 lux using wide-spectrum fluorescent lights; e.g., cool-white or day-light) for culturing and testing. Lux is the unit selected for reporting luminance in this manual. Multiply units of lux by 0.093 to convert to units of foot candles. Multiply units of lux by  $6.91 \times 10^{-3}$  to convert to units of  $\mu\text{E}/\text{m}^2/\text{s}^1$  (assuming an average wavelength of 550 nm ( $\mu\text{mol}^{-2} \text{s}^{-1} = \text{W m} \times \lambda(\text{nm}) \times 8.36 \times 10^{-3}$ )) (ASTM, 1994b). Luminance should be measured at the surface of the water. Illumination should be uniform and must be continuous throughout holding, acclimation, and the test period. Continuous overhead lighting has been previously demonstrated to inhibit nocturnal emergence of amphipods, thus maximizing exposure to test sediments. A 16:8 light:dark photoperiod should be used for culturing *L. plumulosus* (Section 10.4.8) and for holding and acclimating *A. abdita* in the laboratory before testing (Section 10.4.5). This photoperiod can be achieved using automatic timers.

6.2.3 During 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. Air used for aeration must be free of oil and fumes. Filters to remove oil, water, and bacteria are desirable. Oil-free air pumps should be used where possible. Particulates 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<sup>®</sup>, C-1 filter), or equivalent. 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 loosely-fitting doors.

### **6.3 Equipment and Supplies**

6.3.1 Equipment and supplies that contact stock solutions, sediments 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, high-density polyethylene, polycarbonate and fluorocarbon plastics should be used whenever possible to minimize leaching, dissolution, and sorption. High-density plastic containers are recommended for holding, acclimation, and culture chambers. These materials should be washed in detergent, acid rinsed, and soaked in flowing water for a week or more before use. Copper, brass, lead, galvanized metal, and natural rubber should 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 (5 to 25°C).

6.3.3.2 Water purification system capable of producing at least 1 mega-ohm water (USEPA, 1993a).

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 which are at the upper and lower ends of the range of the weight values 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 non-toxic 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 -- PROPIPET<sup>®</sup>, or equivalent.

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 Glass or electronic thermometers -- for measuring water temperature.

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

6.3.3.13 Dissolved oxygen, pH/selective ion, and salinity meters for routine physical and chemical measurements. Unless a test is being conducted to specifically measure the effect of one of these measurements, a portable field-grade instrument is acceptable. A temperature compensated salinity refractometer is useful for measuring salinity of water overlying field collected sediment.

6.3.3.14 Equipment for measuring ammonia (i.e., ammonia-specific probe) is also necessary.

6.3.3.15 See Table 6.1 for a list of additional equipment and supplies.

#### **6.3.4 Test Chambers**

6.3.4.1 The test chambers to be used in sediment toxicity tests are 1 liter glass containers (beakers or wide-mouthed jars) with an internal diameter of 10 cm. Each test chamber should have a cover. Acceptable test chamber covers include watch-glasses, plastic lids, and 9 cm diameter glass culture dishes. It may be necessary to drill a hole in each cover to allow for the insertion of a pipette for aeration.

#### **6.3.5 Cleaning**

6.3.5.1 All non-disposable sample containers, test chambers, tanks, and other equipment that has 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 (use a fume hood or canopy).
6. Rinse three times with deionized water.

**6.3.5.2 All test chambers and equipment should be thoroughly rinsed or soaked with the toxicity test diluent water immediately before use in a test.**

**6.3.5.3 Many organic solvents 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, 1988), but the solution might attack silicone adhesive and leave chromium residues on glass. A alternative to use of dichromate-sulfuric acid could be to heat glassware for 8 h at 450°C.**

Table 6.1                      Equipment and supplies for culturing and testing estuarine and marine amphipods. Supplies are for all species unless specified.

---

**A. Biological Supplies**

Brood stock of test organisms  
TetraMin® (LP)  
dried wheat leaves (LP)  
dried alfalfa leaves (LP)  
Neo-Novum® (LP)  
Algae (e.g., *Pseudoisochrysis paradoxa* and *Phaeodactylum tricornutum* [optional]) (LP)

**B. Glassware**

Culture chambers (30 cm × 45 cm × 15 cm plastic wash bin)  
Test chambers (1 L glass jar or beaker)  
Glass bowls  
Wide-bore pipets (4 to 6 mm ID)  
Glass disposable pipets  
Graduated cylinders (assorted sizes, 10 mL to 4 L)

**C. Instruments and Equipment**

Dissecting microscope  
Stainless-steel sieves (e.g., U.S. Standard No. 25, 30, 35, 40, 50 mesh)  
Photoperiod timers  
Light meter  
Temperature controllers  
Thermometer  
Continuous recording thermometer  
Photoperiod timer  
Dissolved oxygen meter  
pH meter  
Selective ion meter  
Ammonia electrode (or ammonia kit)  
Salinity meter/temperature compensating salinity refractometer  
Drying oven  
Desiccator  
Balance (0.01 mg sensitivity)  
Refrigerator  
Freezer

---

Note: LP = *Leptocheirus plumulosus*

Table 6.1                      Equipment and supplies for culturing and testing estuarine and marine amphipods. Supplies are for all species unless specified (continued)

---

Light box  
Hemocytometer  
Mortar and pestle or blender (LP)

**D. Miscellaneous**

Ventilation system for test chambers  
Air supply and air stones (oil free and regulated)  
Glass hole-cutting bits  
Glass glue  
Aluminum weighing pans  
Fluorescent light bulbs  
Deionized water  
Air line tubing  
White plastic dish pan  
Water squirt bottles  
Shallow pans (plastic (light-colored), glass, stainless steel)  
Sieve cups (mesh size  $\leq 0.5$  mm)  
Dissecting probes

**E. Chemicals**

Detergent (non-phosphate)  
Acetone (reagent grade)  
Hexane (reagent grade)  
Hydrochloric acid (reagent grade)  
Reagents for preparing synthetic seawater (reagent grade  $\text{CaCl}_2 \cdot 2 \text{H}_2\text{O}$ , KBr, KCl,  $\text{MgCl}_2 \cdot 6 \text{H}_2\text{O}$ ,  $\text{Na}_2\text{B}_4\text{O}_7 \cdot 10 \text{H}_2\text{O}$ , NaCl, Na  $\text{HCO}_3$ ,  $\text{Na}_2\text{SO}_4$ ,  $\text{SrCl}_2 \cdot 6 \text{H}_2\text{O}$ )  
Formalin  
Ethanol  
Rose bengal  
Cadmium chloride  
Sodium dodecyl sulfate  
Copper sulfate

---

Note: LP = *Leptocheirus plumulosus*.

## **Section 7**

### **Water, Reagents, and Standards**

#### **7.1 Water**

##### **7.1.1 Requirements**

**7.1.1.1** Sea water used to test and culture organisms should be uniform in quality. Acceptable sea water should allow satisfactory survival, growth, or reproduction of the test organisms. Test organisms should not show signs of disease or apparent stress (e.g., discoloration, unusual behavior). If problems are observed in the culturing or testing of organisms, it is desirable to evaluate the characteristics of the water. See USEPA (1993a) and ASTM (1994a) for a recommended list of chemical analyses of the water supply.

##### **7.1.2 Source**

**7.1.2.1** The source of natural water will depend to some extent on the objective of the test and the test organism that is being used. All natural waters should be obtained from an uncontaminated surface-water source upstream from or beyond the influence of known discharges. Water should be collected at slack high tide, or within one h after high tide. Suitable surface 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. Full strength sea water should be obtained from areas where the salinity does not fall below 28 ‰. For estuarine tests, water having a salinity as near as possible to the desired test salinity should be collected from an uncontaminated area. Alternatively, it may be desirable to dilute full strength sea water with an appropriate fresh water source. Sources of fresh water (i.e., 0 ‰) for dilution include deionized water, uncontaminated well or spring water, or an uncontaminated surface-water source. Municipal-water supplies may be variable and may contain unacceptably high concentrations of materials such as copper, lead, zinc, fluoride, chlorine, or chloramines. Chlorinated water should not be used to dilute water utilized 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 for diluting sea water to the desired salinity since dechlorination is often incomplete (ASTM, 1994a; USEPA, 1993a).

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

**7.1.2.3** Water that might be contaminated with facultative pathogens may be passed through a properly maintained ultraviolet sterilizer equipped with an intensity meter and flow controls or passed through a filter with a pore size of 0.45  $\mu\text{m}$  or less.

**7.1.2.4** Natural sea water might need aeration using air stones, surface aerators, or column aerators. Adequate aeration will stabilize pH, bring concentrations of dissolved oxygen and other gases into equilibrium with air, and minimize oxygen demand and concentrations of volatiles. The concentration of dissolved oxygen in source water should be between 90 to 100% saturation to help ensure that dissolved oxygen concentrations are acceptable in test chambers. Natural sea water used for holding/acclimating, culturing, and testing amphipods should be filtered ( $\leq 5 \mu\text{m}$ ) shortly before use to remove suspended particles and organisms.

**7.1.2.5** Water that is prepared from natural sea water should be stored in clean, covered containers at 4°C and used within 2 d.

### **7.1.3 Reconstituted/Synthetic Seawater**

**7.1.3.1** Although reconstituted water is acceptable, natural seawater is preferable, especially for tests involving chemicals whose bioavailability is affected by seawater chemistry. Reconstituted water is prepared by adding specified amounts of reagent-grade chemicals to high-purity distilled or deionized water (ASTM, 1988; USEPA, 1993a). Acceptable high-purity water can be prepared using deionization, distillation, or reverse-osmosis units (Section 6.3.3.2; USEPA, 1993a). Test water can also be prepared by diluting natural water with deionized water (Kemble et al., 1993).

**7.1.3.2** Deionized water should be obtained from a system capable of producing at least 1 mega-ohm water. If large quantities of high quality deionized water are needed, it may be advisable to supply the laboratory grade water deionizer with preconditioned water from a mixed-bed water treatment system.

**7.1.3.3** Reconstituted sea water is prepared by adding specified amounts of a suitable salt reagent to high-purity distilled or deionized water (ASTM, 1988; USEPA, 1991a). Suitable salt reagents can be reagent grade chemicals, commercial sea salts, such as Forty Fathoms<sup>®</sup>, Instant Ocean<sup>®</sup>, or HW Marinemix<sup>®</sup>. Pre-formulated brine (e.g., 60 to 90 ‰), prepared with dry ocean salts or heat-concentrated natural sea water, can also be used.

**7.1.3.4** A synthetic sea formulation called GP2 is prepared with reagent grade chemicals that can be diluted with a suitable high-quality water to the desired salinity (USEPA, 1994c).

**7.1.3.5** To obtain the desired holding or acclimation salinity, sea salts or brine can be added to a suitable freshwater or distilled water, or the laboratory's sea water supply may be diluted with a suitable freshwater or distilled water.

**7.1.3.6** The suitability and consistency of a particular salt formulation for use in holding and acclimation should be verified by laboratory tests because some formulations can produce unwanted toxic effects or sequester contaminants (Environment Canada, 1992). Salinity and pH should be measured on each batch of reconstituted water.

**7.1.3.7** Salinity, pH, and dissolved oxygen should be measured on each batch of reconstituted water. The reconstituted water should be aerated before use to adjust pH and dissolved oxygen to the acceptable ranges (e.g., Section 7.1.3.4.1). Reconstituted sea water should be filtered ( $\leq 5 \mu\text{m}$ ) shortly before use to remove suspended particles and should be used within 24 h of filtration. USEPA (1993a) recommends using a batch of reconstituted water within a two week period.

## **7.2 Reagents**

**7.2.1** 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 on formulation 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 sediments that might contain unknown quantities of toxic contaminants (Section 5).

**8.1.2** A benthic grab (i.e., PONAR, Smith-MacIntyre, Van Veen) or core sampler are preferred sediment samplers because disturbance of sediment samples with these devices is minimized relative to dredge samplers. Although selective sub-sampling, compositing, and homogenization of sediment samples are necessary for most routine applications addressed by this manual, collection and handling in the field should involve as little disruption as possible. Disruption of sediment samples will cause the loss of sediment integrity, and may cause changes in chemical speciation and chemical equilibrium (ASTM, 1994a). Sediments are spatially and temporally variable (Stemmer et al., 1990a). Replicate samples should be collected to determine variance in sediment characteristics. Sediments should be collected to a depth appropriate for the study objectives. For example, samples collected for evaluations of dredged material should include all sediment to project depth. Surveys of the toxicity of surficial sediment are often based on cores of the upper 2-cm sediment depth.

**8.1.3** Exposure to direct sunlight during collection should be minimized, especially if the sediment contains photolytic compounds. Removal of sediment from the sampling device and subsequent allocation to storage containers or homogenization should be accomplished using spoons, trowls, etc. made of, or coated in, inert materials (e.g., Teflon®, kynar). Sediment samples should be cooled to 4°C in the field before return to the laboratory or shipment (ASTM, 1994a). Dry ice can be used to cool samples in the field; however, sediments should never be frozen. 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 see ASTM (1994a).

#### **8.2 Storage**

**8.2.1** Manipulation or storage can alter bioavailability of contaminants in sediment (Burton and Ingersoll, 1994); however, the alterations that occur may not substantially affect toxicity. Storage of sediment samples for several months at 4°C did not result in significant changes in chemistry or toxicity (T. Dillon and H. Tatem, USCOE, Vicksburg, MS, personal communication; G.T. Ankley and D. Foe, USEPA, Duluth, MN, unpublished data); however, others have demonstrated changes in spiked sediment within days to weeks (e.g., Burton, 1991; Stemmer et al., 1990a). Sediments primarily

contaminated with non-ionic, nonvolatile organic compounds will probably change little during storage because of their relative resistance to biodegradation and sorption to solids. However, metals and metalloids may be affected by changing redox, oxidation, or microbial metabolism (such as with arsenic, selenium, mercury, lead, and tin; all of which are methylated by a number of bacteria and fungi). Metal contaminated sediments may need to be tested relatively soon after collection with as little manipulation as possible (Burton and Ingersoll, 1994).

8.2.2 Given that the contaminants of concern and the influencing sediment characteristics are not always known *a priori*, it is desirable to hold sediments in the dark at 4°C and start tests soon after collection from the field. Recommended sediment holding time ranges from less than two (ASTM, 1994a) to less than eight weeks (USEPA-USCOE, 1994). If whole sediment tests are started after two weeks of collection, it may be desirable to conduct additional characterizations of sediment to evaluate possible effects of storage on sediment. For example, concentrations of contaminants of concern could be measured in pore water within two weeks from sediment collection and at the start of the sediment test (Kemble et al., 1993). Ingersoll et al. (1993) recommend conducting a toxicity test with pore water within two weeks from sediment collection and at the start of the sediment test. Freezing and longer storage might further change sediment properties such as grain size or contaminant partitioning and should be avoided (ASTM, 1994a; Schuytema et al., 1989; K.E. Day, Environment Canada, Burlington, Ontario, personal communication). Sediment should be stored with no air over the sealed samples (no headspace) at 4°C before the start of a test (Shuba et al., 1978; ASTM, 1994a). Sediment may be stored in containers constructed of suitable materials as outlined in Section 6. It is desirable to avoid contact with metals, including stainless steel and brass sieving screens, and some plastics.

### 8.3 Manipulation

#### 8.3.1 Homogenization

8.3.1.1 Sediment 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. If warranted, sediment samples should be press-sieved through a 1 or 2 mm mesh stainless steel screen to remove indigenous organisms. Press-sieving is performed primarily to remove predatory organisms, large debris, or organisms taxonomically similar to the test species. Certain USEPA program offices may recommend that sediments should not be press-sieved. Also, it may not be necessary to press-sieve sediments if previous experience has demonstrated the absence of potential interferences, including predatory or competitive organisms or large debris, or if large debris or predators can be removed with forceps or other suitable tools. The presence of an abundance of amphipods that are taxonomically similar to the test species should prompt press-sieving. This is particularly true if endemic Ampeliscidae are present and *A. abdita* is the test species because it may be difficult to remove all of the resident amphipods from their tubes. If sediments must be sieved, it may be desirable to perform select analyses (e.g., pore-water metals or DOC, AVS, TOC) on samples before and after sieving 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 using stirring or a rolling mill, feed mixer, or other suitable apparatus (see ASTM, 1994a). It is preferable to homogenize sediments by gentle hand mixing. Although potentially disruptive, large numbers of sediments may demand the use of a mechanical aid. Mechanical homogenization of sediment can be accomplished using a modified 30-cm bench-top drill press (Dayton Model 3Z993) or a variable-speed hand-held drill outfitted with a stainless steel auger (diameter 7.6 cm, overall length 38 cm, auger bit length 25.4 cm; Augers Unlimited, Exton, PA; Kemble et al., 1994). These procedures could also be used to mix test sediment with a control sediment in dilution experiments.

### **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 responses. 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. Consistent spiking procedures should be followed in order to make interlaboratory comparisons. It is recommended that spiked sediment be aged at least one month before starting a test; however equilibration for some chemicals may not be achieved for long periods of time.

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 (lowest observable effect concentration: 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 on formulation 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) estimated toxicity to the test organism and to humans, (4) 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 (5) recommended handling and disposal procedures.

8.3.2.2.1 Organic compounds have been added in the dry form or coated on the inside walls of the mixing container (Ditsworth et al., 1990). Metals are generally added in an aqueous solution (ASTM, 1994a; Carlson et al., 1991; Di Toro et al., 1990). If an organic solvent is used, the solvent in the sediment should be at a concentration that does not affect the test organisms. Concentrations of the chemical in the pore water and in whole sediment should be monitored at the beginning and end of a test.

**8.3.2.3** Use of a solvent other than water should be avoided if possible. Addition of organic solvents may dramatically influence the concentration of dissolved organic carbon in pore water (G.T. Ankley, USEPA, Duluth, MN, personal communication). If an organic solvent must be used, both a solvent-control and a negative-control sediment must be included in a test. The solvent in the sediment should be at a concentration that does not affect the test organism. 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, 1988). The same concentration of solvent should be used in all treatments. If an organic solvent is used as a carrier, it may be possible to perform successive washes of sediment to remove most of the solvent while leaving the compound of study (Harkey et al., 1994).

**8.3.2.4** If the concentration of solvent is not the same in all test solutions that contain test material, 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.

**8.3.2.4.1** 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, 1992).

### **8.3.2.5 Test Concentration(s) for Laboratory Spiked Sediments**

**8.3.2.5.1** If a test is intended to generate an LC50, the selected test concentrations should bracket the predicted LC50. The prediction might be based on the results of a test on the same or a similar test material with the same or a similar test organism. The LC50 of a particular compound may vary depending on physical and chemical sediment characteristics. If a useful prediction is not available, it is 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).

**8.3.2.5.2** Bulk-sediment chemical concentrations might be normalized to factors other than dry weight. For example, concentrations of non-polar 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).

**8.3.2.5.3** In some situations it might be necessary to only 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.

**8.3.2.6** Addition of test material(s) to sediment may be accomplished using various methods, such as a: (1) rolling mill (preferred), (2) feed mixer, or (3) hand mixing (ASTM, 1994a). 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, 1994a). 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 is advisable to determine the degree of mixing homogeneity (Ditsworth et al., 1990). Moreover, results from sediment-spiking studies should be compared with the response of test organisms to chemical concentrations in natural sediments (Lamberson and Swartz, 1988).

## **8.4 Characterization**

**8.4.1** All sediments should be characterized and at least the following determined: salinity, pH, and ammonia of the pore water, organic carbon content (total organic carbon, TOC), particle size distribution (percent sand, silt, clay), and percent water content (ASTM, 1994a; Plumb, 1981). Salinity of sediment pore water should be measured on the supernatant of an aliquot of the sediment using a refractometer or conductivity meter. See Section 8.4.4.7 for methods to isolate pore water.

**8.4.2** Other analyses on sediments might include: biological oxygen demand, chemical oxygen demand, cation exchange capacity, Eh, total inorganic carbon, total volatile solids, acid volatile sulfides, metals, synthetic organic compounds, oil and grease, petroleum hydrocarbons, as well as interstitial water analyses for various physico-chemical parameters.

**8.4.3** Macrobenthos may be quantified by subsampling the field-collected sediment. If direct comparisons are to be made, subsamples for toxicity testing should be collected from the same sample for analysis of sediment physical and chemical characterizations. Qualitative descriptions of the sediment may 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 Methodology**

**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: that is, sediment, water, 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 sediments 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 might be added to these extra chambers depending on the objective of the study.**

**8.4.4.5 Measurement of test material(s) concentration in water can be accomplished by pipeting water samples from about 1 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 test material(s) concentration 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 A variety of procedures have been used to isolate interstitial water including centrifugation, filtration, pressure, or by using an interstitial water sampler; however, centrifugation without filtration is the recommended procedure (Ankley and Schubauer-Berigan, 1994). Filtration may reduce concentrations of materials in interstitial water (Schults et al., 1992). Care should be taken to ensure that contaminants do not transform, degrade, or volatilize during isolation or storage of the interstitial water sample.**

## **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 (4) provision of adequate, qualified technical staff and suitable space and equipment to assure reliable data. Additional guidance for QA can be obtained in USEPA (1989c).

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) 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 carried out within the scope of the overall QA program. For more detailed discussion of quality assurance, and general guidance on good laboratory practices related to testing see FDA (1978), USEPA (1979a), USEPA (1980a), USEPA (1980b), USEPA (1993a), USEPA (1994b), USEPA (1994c), DeWoskin (1984), and Taylor (1987).

#### **9.2 Performance-based Criteria**

9.2.1 USEPA Environmental Monitoring Management Council (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 which permits the use of appropriate 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 Therefore, a performance-based criteria approach was selected as the preferred method through which individual laboratories should evaluate culture methods or the quality of field-collected organisms rather than by control-based criteria. 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. See Table 11.3 for a listing of performance

criteria for culturing *L. plumulosus*, determining the quality of field-collected organisms, and evaluating the outcome of sediment tests.

### **9.3 Facilities, Equipment, and Test Chambers**

9.3.1 Separate areas for test organism culturing and testing areas must be provided to avoid loss of cultures due to 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 which comes in contact with the sediment or overlying water.

9.3.3 Before a sediment test is conducted in a new facility, a "non-contaminant" 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 (e.g., <15%) in cultures, during holding, and in test controls. 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. See Section 7 for guidance on selection and preparation of high quality test water.

### **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 the each test. Temperatures should be maintained within the limits specified for each test. Dissolved oxygen, salinity, ammonia, and pH should be checked as prescribed in Section 11.3.

## **9.8 Quality of Test Organisms**

**9.8.1** If test organisms are obtained from culture, monthly reference-toxicity tests should be conducted on all test organisms using procedures outlined in Section 9.16. If reference-toxicity tests are not conducted monthly, the lot of organisms used to start a sediment test must be evaluated using a reference toxicant.

**9.8.2** The quality of test organisms obtained from an outside source, regardless of whether they are from culture or collected from the field, must be verified by conducting a reference-toxicity test concurrently with the sediment test. 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. If the supplier has not conducted five reference toxicity tests with the test organism, it is the responsibility of the testing laboratory to conduct these five reference toxicity tests before starting a sediment test (Section 9.14.1).

**9.8.3** The supplier should also certify the species identification of the test organisms, and provide the taxonomic references, or name(s) of the taxonomic expert(s) consulted.

## **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 (see Section 10.5.8). Additionally, survival in sediment tests conducted with *A. abdita* and *L. plumulosus* may be affected by the nutritional suitability of food provided during holding and acclimation.

## **9.10 Test Acceptability**

**9.10.1** For the test results to be acceptable, survival at 10 d must equal or exceed 90% for all four amphipod species in the control sediment. See Table 11.3 for additional requirements for acceptability of the tests.

**9.10.2** 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 11.3). The acceptability of a test will depend on the experience and professional judgment of the laboratory analyst and the 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 (USEPA, 1979a; USEPA, 1979b; USEPA, 1993d).

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, dissolved oxygen, temperature, and salinity should be calibrated before use each day 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, 1979b). Calibration data should be recorded in a permanent log.

9.12.2 A known-quality water should be included in the analyses of each batch of water samples (e.g., water hardness, alkalinity, conductivity).

## **9.13 Replication and Test Sensitivity**

9.13.1 The sensitivity of sediment tests will depend in part on the number of replicates per 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 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 It is the responsibility of a laboratory to demonstrate its ability to obtain consistent, precise results with reference toxicants before it performs sediment tests (see Section 9.16). 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 should 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 as a point of reference for future testing. A reference toxicant concentration series (0.5 or higher) should be selected that will consistently provide partial mortalities at two or more concentrations of the test chemical (Section 12).

9.14.2 Before conducting tests with contaminated sediment, the laboratory should demonstrate its ability to conduct tests by conducting five exposures in control sediment

as outlined in Table 11.1. It is recommended that these five exposures with control sediment be conducted concurrently with the five reference toxicity tests described in Section 9.14.1.

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

## **9.15 Documenting Ongoing Laboratory Performance**

**9.15.1** Satisfactory laboratory performance on a continuing basis is demonstrated by conducting monthly reference-toxicity tests with each test organism. For a given test organism, 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 13).

**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 datum from a given test falls outside the "expected" range (e.g.,  $\pm 2$  SD), the sensitivity of the organisms and the credibility of the test results are 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 automatically 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  $\pm 2$  SD, by definition, will be exceeded 5% of the time, regardless of how well a laboratory performs. For this reason, good laboratories which develop very narrow control limits may be penalized if a test result which 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** Reference-toxicity tests should be conducted in conjunction with sediment tests to determine possible changes in condition of a test organism (Lee, 1980). Water-only reference-toxicity tests on cultured organisms should be conducted monthly, and should be performed on each batch of field-collected organisms used for testing. Deviations

outside an established normal range may indicate a change in the condition of the test organism population. Results of reference-toxicity tests also enable interlaboratory comparisons of test organism sensitivity.

9.16.2 Reference toxicants such as cadmium (available as cadmium chloride ( $\text{CdCl}_2$ )), copper (available as copper sulfate ( $\text{CuSO}_4$ )), and sodium dodecyl sulfide (SDS) are suitable for use. No one reference toxicant can be used to measure the condition of test organisms in 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.

9.16.3 Test conditions for conducting reference-toxicity tests with *A. abdita*, *E. estuarius*, *L. plumulosus*, and *R. abronius* are outlined in Table 9.1.

## 9.17 Record Keeping

9.17.1 Proper record keeping is important. 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; test conditions used; and results of reference toxicant 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**

<b>Parameter</b>	<b>Conditions</b>
1. Test type:	Water-only test
2. Dilution series:	Control and at least 5 test concentrations (0.5 dilution factor)
3. Toxicant:	Cd, Cu, Sodium dodecyl sulfate (SDS)
4. Temperature:	15°C for <i>E. estuarius</i> and <i>R. abronius</i> 20°C for <i>A. abdita</i> 25°C for <i>L. plumulosus</i>
5. Salinity:	28 ‰ for <i>A. abdita</i> and <i>R. abronius</i> 20 ‰ for <i>E. estuarius</i> and <i>L. plumulosus</i>
6. Light quality:	Chambers should be kept in dark or covered with opaque material
7. Photoperiod:	24 h D
8. Renewal of water:	None
9. Age and size of test organisms:	<i>A. abdita</i> : 3 - 5 mm (no mature males or females) <i>E. estuarius</i> : 3 - 5 mm <i>L. plumulosus</i> : 2 - 4 mm (no mature males or females) <i>R. abronius</i> : 3 - 5 mm
10. Test chamber:	1 L glass beaker or jar
11. Volume of water:	800 mL (minimum)
12. Number of organisms/chamber:	n = 20 if 1 per replicate; n = 10 (minimum) if >1 per replicate

**Table 9.1 Recommended test conditions for conducting reference-toxicity tests (continued)**

<b>Parameter</b>	<b>Conditions</b>
<b>13. Number of replicate chambers/treatment:</b>	<b>1 minimum; 2 recommended</b>
<b>14. Aeration:</b>	<b>Recommended; but not necessary if &gt;90% dissolved oxygen saturation can be achieved without aeration</b>
<b>15. Dilution water:</b>	<b>Culture water, surface water, site water, or reconstituted water.</b>
<b>16. Water quality:</b>	<b>Salinity, pH, and dissolved oxygen at beginning and end of test. Temperature daily</b>
<b>17. Test duration:</b>	<b>96 h</b>
<b>18. Endpoint:</b>	<b>Survival (LC50); Reburial (EC50) optional for <i>E. estuarius</i> and <i>R. abronius</i></b>
<b>19. Test acceptability:</b>	<b>90% control survival</b>

## Section 10 Collection, Culture, and Maintaining of Test Organisms

### 10.1 Life History

10.1.1 *Ampelisca abdita*: *A. abdita* is a tube-building amphipod in the family Ampeliscidae. It occurs on the Atlantic coast from central Maine to central Florida, although it is also found in the eastern portion of the Gulf of Mexico (Bousfield, 1973). On the Pacific coast, it is present in San Francisco Bay, CA (Nichols et al., 1985; Hopkins, 1986). They are small (adult length 4 to 8 mm), laterally compressed amphipods. Healthy animals are opalescent pink and will remain tightly curled, whereas unhealthy animals tend to be translucent white, and may uncurl (T. Thompson, SAIC, Bothell, WA, personal communication). Often dominant members of the benthic community, *A. abdita* forms thick mats of tubes with amphipod densities up to 110,000/m<sup>2</sup>, and are often a dominant food source for bottom-feeding fish (Richards, 1963). The tubes are narrow and approximately 2 to 3 cm in length. A filter feeder, *A. abdita* feeds on both particles in suspension and those from surficial sediment surrounding the tube. *Ampelisca abdita* is euryhaline, and has been reported in waters that range in salinity from fully marine to 10 ‰ (Hyland, 1981). Laboratory tests have shown the salinity application range of *A. abdita* in sediments is from 0 to 34 ‰ when the salinity of overlying water is  $\geq 28$  ‰ (Weisberg et al., 1992). This species generally inhabits sediments from fine sand to mud and silt without shell fragments, although it can also be found in relatively coarser sediments with a sizeable fine component. It is often abundant in sediments with a high organic content. Analysis of historical data shows little effect of sediment grain size on survival of *A. abdita* during 10 d sediment toxicity tests (Long and Buchman, 1989; Weisberg et al., 1992). There is evidence that sediments with >95% sand may elicit excessive mortality (J. Scott, SAIC, Narragansett, RI, personal communication). *Ampelisca abdita* have been collected at water temperatures ranging from -2 to 27°C (J. Scott, SAIC, Narragansett, RI, and M.S. Redmond, Northwest Aquatic Sciences, Newport, OR, unpublished data). Reproduction patterns of *A. abdita* vary geographically. In the colder waters of its range, *A. abdita* produces two generations per year, an over-wintering population that broods in the spring, and a second that breeds in mid- to late-summer (Mills, 1963). In warmer waters south of Cape Hatteras, NC, breeding might be continuous throughout the year (Nelson, 1980). Juveniles are released after approximately two weeks in the brood pouch. Juveniles take approximately 40 to 80 d for newly released juveniles to become breeding adults under laboratory conditions at 20°C (Scott and Redmond, 1989).

10.1.2 *Eohaustorius estuarius*: *Eohaustorius estuarius* is a free-burrowing amphipod in the family Haustoriidae. It is found on protected and semi-protected beaches from the lower intertidal to shallow subtidal waters exclusively on the Pacific coast from British Columbia south to central California (Environment Canada, 1992). They are stout (adult size range 3 to >5 mm) cup- or bell-shaped, dorsally compressed amphipods that are grayish-brown or yellowish-brown in color (Environment Canada, 1992; ASTM, 1992). *Eohaustorius estuarius* are thought to be deposit feeders. It is an estuarine species and has been reported in areas where pore water salinity ranges from 1 to 25 ‰

(Environment Canada, 1992; R. Caldwell, Northwest Aquatic Sciences, Newport, OR, personal communication). Laboratory studies have shown a salinity application range in control sediments for *E. estuarius* from 0 to 34 ‰. *Eohaustorius estuarius* inhabits clean, medium-fine sand with some organic content. The species has exhibited acceptable (i.e., >90%) survival when exposed to clean sediments with a wide range of grain sizes, with generally little effect on survival whether coarse-grained or fine-grained (i.e., predominantly silt and clay) clean sediments are used (Environment Canada, 1992). However, some correlation between survival and grain size exists (DeWitt et al., 1989). *Eohaustorius estuarius* has been collected from water temperatures from 0 to 23°C (ASTM, 1992; R. Caldwell, personal communication). *Eohaustorius estuarius* apparently has an annual life cycle (Environment Canada, 1992; DeWitt et al., 1989). Gravid females are abundant in intertidal sediments from February through July (DeWitt et al., 1989; ASTM, 1992). However, reproduction might occur year-round because juveniles are found throughout most of the year (DeWitt et al., 1989).

**10.1.3 *Leptocheirus plumulosus*:** *Leptocheirus plumulosus* is a tube-building member of the family Aoridae. It is found infaunally in subtidal portions of Atlantic Coast brackish estuaries from Cape Cod, MA, to northern Florida (Bousfield, 1973; DeWitt et al., 1992a). It is common in protected embayments but has been collected in channels of estuarine rivers up to depths of 10 m (Holland et al., 1988; Schlekot et al., 1992), and has been reported to occur in depths up to 13 m (Shoemaker, 1932). They are relatively large amphipods (adult length up to 1.3 cm) with cylindrically shaped bodies that are brownish-grey in color. A distinguishing feature is a series of dark bands or stripes that cross the dorsal surface of the pereopods and pleons. In Chesapeake Bay, densities of *L. plumulosus* can reach 28,987/m<sup>2</sup> and 24,133/m<sup>2</sup> in sandy and muddy sediments, respectively (Holland et al., 1988). It feeds on particles that are in suspension and on the sediment surface (DeWitt et al., 1992a). *Leptocheirus 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; Schlekot et al., 1992; McGee et al., 1994). Laboratory studies have demonstrated a salinity application range in control sediments of 0 to 32 ‰ (Schlekot et al., 1992; SAIC, 1993b). It is most often found in fine-grained sediment with a high proportion of particulate organic material, although it has been collected in fine sand with some organic content (Jordan and Sutton, 1984; Holland et al., 1988; Marsh and Tenore, 1990; DeWitt et al., 1992a; Schlekot et al., 1992; McGee et al., 1994). Analysis of historical data for *L. plumulosus* reveals no effect of sediment grain size on survival in control sediment. Populations of *L. plumulosus* may be seasonally ephemeral, with major population growth occurring in spring and large population declines occurring in summer due to actions of predatory fish (Hines et al., 1986) or absence of essential micronutrients (Marsh and Tenore, 1990).

**10.1.4 *Rhepoxynius abronius*** is a free-burrowing amphipod in the family Phoxocephalidae. It occurs on the Pacific Coast from Puget Sound, WA, to central California in lower intertidal and nearshore subtidal zones to depths of 274 m offshore (ASTM, 1992; Environment Canada, 1992; Lamberson and Swartz, 1988; Kemp et al., 1985; Barnard and Barnard, 1982). Densities in the field are reported to range from 150 to 2200/m<sup>2</sup> (Lamberson and Swartz, 1988; Swartz et al., 1985). It is a medium-sized

(adult length from 3 to >5 mm) amphipod with a stout, somewhat rounded body shape. Color may range from salmon pink to yellowish, grayish-brown to white with a pinkish-brown hue (Environment Canada, 1992). *Rhepoxynius abronius* is a meiofaunal predator, but it also ingests sedimentary organic material (Oakden, 1984). In the field, *R. abronius* is found where pore water salinity is no lower than 20 ‰ (Environment Canada, 1992). Laboratory tests have indicated that salinities below 25 ‰ may be toxic to *R. abronius* (SAIC, 1993b; Swartz et al., 1985). *Rhepoxynius abronius* should therefore not be chosen as the test species when the sediment pore water is <25 ‰ (Swartz et al., 1985). *Rhepoxynius abronius* naturally inhabits clean, fine, sandy sediments (ASTM, 1992). A number of studies have shown some reduction in survival when this species is held in very fine-grained (predominantly silt and clay) sediment (DeWitt et al., 1988; Long et al., 1990; McLeay et al., 1991; SAIC, 1993a; SAIC, 1993b). Normally collected at temperatures ranging from 8 to 16°C, *R. abronius* has survived at temperatures ranging from 0 to 20°C under laboratory conditions (ASTM, 1992). Reproduction of *R. abronius* is annual, with peak production occurring from late winter through spring (Kemp et al., 1985).

## 10.2 Species Selection

10.2.1 All four species have been routinely used to test sediments with a range of grain size characteristics and pore water salinities. Selection of one or more of the four species for a particular test/investigation should take into consideration the geographic location of the testing facility and study area, the pore water salinity regime of the study area, and the grain size characteristics of the sediment being tested. The species that is used must exhibit tolerance to the physicochemical properties of every sediment included in a particular study. Pore water ammonia concentrations may also enter into selection of one species over others because the four species exhibit differential sensitivity to aqueous ammonia. Most often it will not be necessary to discriminate among the four species, and the decision to test one species above the rest may be driven by practical or logistical concerns. For example, a testing facility may choose to primarily test one species with a suitable local population in order to prevent potential complications associated with shipping. However, sediments may be encountered with characteristics that are outside of the tolerance range of one or more of the species. For example, grain size limitations for *A. abdita* and *R. abronius* are <10% and >90% fines, respectively. If these species are exposed to sediments that exhibit textural characteristics outside of these extremes, any mortality that is observed could be due to effects of grain size independent of contaminants associated with the sediment. Ambiguity in interpretation may be avoided by careful consideration of the test species given the sediment to be tested. Comparative information is available for the four species on sediment grain size sensitivity (Section 11.4.3), salinity application ranges (Section 11.4.4), and sensitivity to aqueous ammonia (Section 11.4.5).

## 10.3 Field Collection

10.3.1 Field collection is presently the most common method for obtaining estuarine and marine amphipods for sediment testing. All four species are commonly collected, shipped, and held in the laboratory. Commercial vendors are available for all four

species. The availability of the appropriate size class for each species may vary seasonally. The collection site chosen should be one for which the presence of abundant organisms of the correct size and age has been demonstrated previously, and identification of the species has been confirmed taxonomically (e.g., Bousfield, 1973; Barnard and Barnard, 1982). Collection areas should be relatively free of contamination. All individuals in a test must be from the same source, because different populations may exhibit different sensitivities to contaminants. The four species are found in distinctly different habitats (Table 10.1).

### 10.3.2 Species-Specific Habitat Characteristics

10.3.2.1 *Ampelisca abdita* is found mainly in protected areas from the low intertidal zone to depths of 60 m. On the Atlantic Coast, *A. abdita* ranges from central Maine to south-central Florida and the eastern Gulf of Mexico. It can also be found on the Pacific Coast in San Francisco Bay. *Ampelisca abdita* is euryhaline, and has been reported in waters that range from fully marine to 10 ‰. This species generally inhabits sediments from fine sand to mud and silt without shell fragments, although it can also be found in relatively coarser sediments with a sizeable fine component. This species is often abundant in sediments with a high organic content. Aggregations of *A. abdita* are indicated by an abundance of tubes on the sediment surface, location of which can be facilitated by looking through a glass-bottom bucket. Although populations may be seasonally ephemeral, *A. abdita* is routinely collected year-round for toxicity testing from subestuaries of Narragansett Bay, RI. It is also routinely collected in San Francisco Bay, CA.

10.3.2.2 *Eohaustorius estuarius* is found on protected and semi-protected beaches from mid-water level to shallow subtidal, within the upper 10 cm of sediment along the Pacific coast from British Columbia south to at least central California (ASTM 1992; Environment Canada, 1992). *Eohaustorius estuarius* can be found on open coasts in beds of freshwater streams flowing into the ocean, and in sand banks in estuaries, above the level of other regional eohaustorids (*E. sawyeri* and *E. washingtonianus*) (Environment Canada, 1992). It is an estuarine species, and has been reported in areas where pore water salinity ranges from 1 to 25 ‰ (Environment Canada, 1992; R. Caldwell, Northwest Aquatic Sciences, Newport, OR, personal communication). *Eohaustorius estuarius* inhabits clean, medium-fine sand with some organic content. It is routinely collected for toxicity tests from Beaver Creek near Newport, OR, and on the west coast of Vancouver Island, BC, Canada.

10.3.2.3 *Leptocheirus plumulosus* is found in subtidal portions of Atlantic Coast brackish estuaries from Cape Cod, MA, to northern Florida. It is common in protected embayments, but has been collected in channels of estuarine rivers up to depths of 10 m. *Leptocheirus plumulosus* is an estuarine species and has been reported in areas where salinity at the sediment-water interface ranges from 1 to 15 ‰ (DeWitt et al., 1992a; Schlekot et al., 1992; McGee et al., 1994). It is most often found in fine-grained sediment with a high proportion of particulate organic material, although it has been collected in fine silty sand with some organic content. Populations of *L. plumulosus* may be seasonally ephemeral, with major population growth occurring in spring, and

**Table 10.1 Comparison of habitat characteristics and other life history parameters of four estuarine and marine amphipod species used in sediment toxicity tests**

<b>Characteristic</b>	<b><i>Ampelisca abdita</i></b>	<b><i>Eohaustorius estuarius</i></b>	<b><i>Leptocheirus plumulosus</i></b>	<b><i>Rhepoxynius abronius</i></b>
<b>Substrate Relation</b>	<b>Tube dwelling, closed<sup>1</sup></b>	<b>Free burrowing<sup>2</sup></b>	<b>Tube dwelling, open<sup>1</sup></b>	<b>Free burrowing<sup>3</sup></b>
<b>Zoogeography</b>	<b>Atlantic-Gulf<sup>1</sup>- San Francisco Bay<sup>4,5</sup></b>	<b>Pacific<sup>2,6</sup></b>	<b>Atlantic<sup>1</sup></b>	<b>Pacific<sup>3</sup></b>
<b>Habitat</b>	<b>Poly-upper mesohaline<sup>1</sup></b>	<b>Oligo-mesohaline<sup>2,6</sup></b>	<b>Oligo- mesohaline<sup>1</sup></b>	<b>Polyhaline<sup>3,7</sup></b>
<b>Life Cycle</b>	<b>40 to 80 days<sup>8</sup></b>	<b>Annual<sup>2</sup></b>	<b>30 to 40 days<sup>9</sup></b>	<b>Annual<sup>10</sup></b>
<b>Availability</b>	<b>Field- Potential culture<sup>1</sup></b>	<b>Field<sup>2</sup></b>	<b>Field-Culture<sup>9,11,12</sup></b>	<b>Field<sup>7</sup></b>
<b>Ecological Importance</b>	<b>High</b>	<b>High</b>	<b>High</b>	<b>High<sup>9</sup></b>

- <sup>1</sup> Bousfield, 1973  
<sup>2</sup> DeWitt et al., 1989  
<sup>3</sup> Barnard and Barnard, 1982  
<sup>4</sup> Nichols et al., 1985  
<sup>5</sup> Hopkins, 1986  
<sup>6</sup> Environment Canada, 1992  
<sup>7</sup> Swartz et al., 1985  
<sup>8</sup> Scott and Redmond, 1989  
<sup>9</sup> DeWitt et al., 1992a  
<sup>10</sup> Kemp et al., 1985  
<sup>11</sup> Schlekot et al., 1992  
<sup>12</sup> McGee et al., 1993

large population declines occurring in summer due to actions of predatory fish (Hines et al., 1986) or absence of essential micronutrients (Marsh and Tenore, 1990). *Leptocheirus plumulosus* has been collected for toxicity tests from several areas in the Maryland portion of Chesapeake Bay, including the Magothy, Chester, Corsica, and Wye rivers. Organisms have been collected for testing year-round from the Magothy River sub-estuary of Chesapeake Bay (C. Schlekot, SAIC, Narragansett, RI, and B. McGee, University of Maryland, Queenstown, MD, unpublished data/personal communication).

10.3.2.4 *Rhepoxynius abronius* occurs in lower intertidal and nearshore subtidal zones on the Pacific Coast from Puget Sound, WA, to central California. Primary habitats of *R. abronius* include nearshore subtidal zones on the Pacific Ocean coastline, and sub- and intertidal zones within polyhaline portions of estuaries in the Pacific Northwest. It is found where the pore water salinity is no lower than 20 ‰. *Rhepoxynius abronius* naturally inhabits clean, fine sand. It has been collected for use in toxicity tests from Lower Yaquina Bay, OR (Swartz et al., 1985), and West Beach, Whidbey Island, WA (Ramsdell et al., 1989; Word et al., 1989).

### 10.3.3 Collection Methods

10.3.3.1 Collection methods are species-dependent. *Ampelisca abdita* and *L. plumulosus* are subtidal, and can be collected with a small dredge or grab (e.g., PONAR, Smith-McIntyre, or Van Veen), or by skimming the sediment surface with a long-handled, fine-mesh net. *Eohaustorius estuarius* and *R. abronius* occur both intertidally and subtidally. Subtidal populations can be collected as above, and intertidal populations can be collected using a shovel. At least one-third more amphipods should be collected than are required for the test.

10.3.3.2 All apparatus used for collecting, sieving, and transporting amphipods and control-site sediment should be clean and made of non-toxic material. They should be marked "live only" and must never be used for working with formalin or any other toxic materials and should be stored separately from the aforementioned. The containers and other collection apparatus 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 source before use.

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

10.3.3.4 Amphipods can be isolated from collection site sediment by gentle sieving. The mesh size of the sieve will depend on the species collected. Sieves with 0.5 mm mesh should be used for sediment containing *A. abdita* and *L. plumulosus*. Larger *A. abdita*, which should not be used in the test, should be excluded by sieving first with a 1.0 mm screen. When sieving *A. abdita*, only about half of the amphipods will be extracted from

their tubes. The tube mat should be placed undisturbed for 20 to 30 min coax the remaining animals out. Sieves with 1.0 mm mesh should be used for *E. estuarius* and *R. abronius*.

10.3.3.5 Collection-site water should be used to sieve sediment in the field. A 2-cm thick layer of sieved collection site sediment should be placed in transport containers, and this sediment covered 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.

10.3.3.6 The salinity and temperature of surface and bottom sea water at the collection site should be measured and recorded. An adequate portion of collection site sediment should be returned with the amphipods to serve as both laboratory holding sediment and for use as control sediment in the toxicity test.

10.3.3.7 During transport to the laboratory, amphipods should be kept in sieved collection-site sediment at or below the collection site temperature. Containers of amphipods and sediment should be transported to the laboratory in coolers with ice-packs, and the water in the containers of amphipods should be aerated if transport time exceeds 1 h.

10.3.3.8 An alternate collection method for *A. abdita* involves transporting intact field-collected tubes to the laboratory for isolation of amphipods. This method is advantageous because separation of *A. abdita* from its tubes may be time-consuming when attempted in the field, a practice which may be impractical in cold winter months. Amphipod tubes are collected as described in 10.3.1. and placed on a 0.5 mm sieve. The sieve should be shaken vigorously to remove most of the sediment, leaving the intact tubes. The tubes should be placed into a covered bucket that contains a sufficient quantity of collection site water to cover the collected material, and transported to the laboratory as described in Section 10.2.3.7. In the laboratory, the tubes should be removed from the collection buckets and placed on a sieve series consisting of a 2 mm mesh sieve over a 0.5 mm mesh sieve. Amphipods should be forced from their tubes by spraying collection-temperature sea water on the material present on the 2 mm sieve. When all the tube material has been sprayed, the 0.5 mm sieve should be shaken vigorously to separate amphipods from any material that is present. The 0.5 mm sieve should then be completely submersed, at which point the amphipods will float on the water surface. The amphipods should then be skimmed from the surface with a small aquarium net and transported to a container with sea water at the appropriate temperature. The shaking process should be continued until only a few amphipods remain in the sieve.

#### 10.3.4 Life Stage and Size

10.3.4.1 The life stage for amphipods used in sediment toxicity tests will depend on the species tested. For *A. abdita* and *L. plumulosus*, sub-adult individuals should always be selected for testing. The life cycle of these species is relatively short, so the likelihood of senescence and any effects that could be associated with reproductive

development/maturation are minimized if young individuals are selected. *Eohaustorius estuarius* and *R. abronius* are annual species with longer life spans than *A. abdita* and *L. plumulosus*. Mature individuals can be used providing they are within the recommended size range.

**10.3.4.2** The size range of test animals should be kept to a minimum regardless of the chosen species. For all species, mature female amphipods, which are distinguishable by the presence of embryos in the brood pouch or oviduct, should not be selected for testing. Additionally, mature male *A. abdita* and *L. plumulosus* should not be used. Recommended size ranges for the four species are as follows:

**10.3.4.3** *Ampelisca abdita*: 3 to 5 mm; or those amphipods retained on a 0.71 mm sieve after passing through a 1.0 mm sieve. Adult male animals must not be tested; they are active swimmers and die shortly after mating.

**10.3.4.4** *Eohaustorius estuarius*: 3 to 5 mm; or those amphipods retained on a 1.0 mm sieve. Large individuals (i.e.,  $\geq 5$  mm) should not be tested because they might be senescent.

**10.3.4.5** *Leptocheirus plumulosus*: 2 to 4 mm; or those amphipods retained on a 0.5 mm sieve after passing through a 0.71 mm sieve (P. Adolphson, Old Dominion University, Norfolk, VA, personal communication).

**10.3.4.6** *Rhepoxynius abronius*: 3 to 5 mm; or those amphipods retained on a 1.0 mm sieve. Large individuals (i.e.,  $\geq 5$  mm) should not be tested because they might be senescent.

### **10.3.5 Shipping Methods**

**10.3.5.1** All four species have been routinely shipped from the collection site to the laboratory for sediment toxicity testing. Currently, shipping from the collection site is necessary for many testing laboratories because culture methods are not available for all four species. It is critical that standard, demonstrated shipping methods are utilized to ensure that consistently healthy animals are used in successive toxicity tests. Additionally, the amphipods that are received by a laboratory must meet the shipping acceptance criteria recommended for each species. Shipping methods and acceptance criteria will vary depending on the species used.

**10.3.5.2** *Ampelisca abdita*: Collected amphipods should be shipped within 24 h of collection. Acceptable methods are available for shipping *A. abdita* in sediment and in water. For shipping in sediment, small plastic "sandwich" containers (approximately 500 mL) with sealable lids should be used. The containers are filled three-quarters full with a minimum depth of 2 cm of sieved fine-grain collection-site sediment and then to the top with well-aerated seawater. No more than 200 amphipods should be added to each container. Amphipods should be allowed to burrow into the sediment and build tubes before the containers are sealed. Containers should be sealed with lids under water to eliminate any air pockets. For shipping in water-only, sealable plastic bags

(approximately 1 L) should be used. Amphipods in their tubes should be placed in bags and a sufficient amount of collection site water should be added to keep the tubes moist. The bag should be filled with pure oxygen before sealing, and then placed into a second bag. Bags should be placed in a container that has a layer of material (i.e., styrofoam or newspaper) sufficiently thick to prevent excessive movement over a layer of ice-packs. The shipping container should be marked to prevent it from being inverted.

**10.3.5.3 *Eohaustorius estuarius* and *Rhepoxynius abronius*:** Shipping methods for these organisms are essentially the same. Small plastic "sandwich" containers (approximately 500 mL) with sealable lids should be used. The containers are filled three-quarters full with sieved collection site sediment (fine sand) and then with a 1 cm layer of collection site sea water. Not more than 100 amphipods should then be added and allowed to burrow. After the animals have burrowed, the overlying water should be poured off, but the sediment should be moist. The containers are then sealed and ready for shipment.

**10.3.5.4 *Leptocheirus plumulosus*:** Several methods have successfully been used to ship *L. plumulosus*, including the *A. abdita* sediment/overlying water method (10.2.5.2) and the *E. estuarius*/*R. abronius* wet-sediment method (10.2.5.3) as described above. Additionally, *L. plumulosus* have been successfully shipped in "sandwich" containers, cubitainers, and thick plastic bags containing only well-aerated collection-site sea water.

### **10.3.6 Performance Criteria for Shipped Amphipods**

**10.3.6.1** The process of ensuring the availability of healthy amphipods on the day that the test is set up begins when the animals arrive in the laboratory from the supplier. Although the ultimate performance criterion for amphipods utilized in sediment toxicity tests is achievement of >90% survival in control sediment, it would be desirable to assess the quality and acceptability each batch of shipped amphipods using the criteria that follow. For all four species, biological criteria should include an exhibition of active swimming behavior upon placement in water, full digestive tracts, and an acceptable color. *Ampelisca abdita* should be opalescent pink, *E. estuarius* should be grayish- or yellowish white, *L. plumulosus* should be brownish or orangish-gray, and *R. abronius* should be salmon pink, grayish- or yellowish-brown, or white with a pinkish-brown hue. Mortality among the shipped animals should not exceed 5%. No sexually mature animals should be included in shipments of *A. abdita* or *L. plumulosus*. The shipping containers should arrive intact, and the temperature of water or sediment in shipping containers should be between 4 and 10°C. Information on physical parameters of the collection site, 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 during the acclimation period, and for use as control sediment in the test. It may be desirable for the testing facility to stipulate these criteria to the supplier when the animals are ordered. If these criteria are not met, the animals may have experienced stress during shipment, and >90% survival in control sediment may not be achieved.

## 10.4 Holding and Acclimation

10.4.1 Density. Amphipods should be held and acclimated (if necessary) in containers (4 to 8 L volume) that contain a 2 to 4 cm layer of collection site sediment that has been sieved through a 0.5 mm mesh screen. Approximately 350 amphipods should be added to each 8 L container. Amphipod density should not exceed 1 amphipod/cm<sup>3</sup>.

10.4.2 Duration. Depending on temperature and salinity at the collection site, amphipods may have to be acclimated to standard test conditions. If necessary, changes in temperature or salinity to bring amphipods from the collection site conditions to the test conditions should be made gradually. Once test conditions are achieved, amphipods should be maintained at these conditions for at least two d before testing to allow for acclimation. Amphipods held for more than ten d should not be used for testing because they may not satisfy performance control criteria. Temperature and salinity should be measured at least daily during the period when amphipods are being adjusted to the conditions of the test water. Thereafter, temperature, salinity, pH, and dissolved oxygen should be measured in the holding containers at least at the start and end of the acclimation period, and preferably daily.

10.4.3 Temperature. Overlying water temperature must not be changed by more than 3°C per day during acclimation to the test temperature. Once the test temperature is reached, amphipods must be maintained at that temperature for a minimum of 2 d. A water bath, an incubator, or temperature-regulated room can be used for temperature acclimation.

10.4.4 Salinity. It is unlikely that either *A. abdita* or *R. abronius* will require salinity acclimation because the collection site salinity for these two species will likely be within 3 ‰ of the test salinity of 28 ‰. Salinity of water used for temperature acclimation for these species, if necessary, should be the test salinity, or 28 ‰. The target test salinity for *E. estuarius* and *L. plumulosus* is 20 ‰, and it is likely that the collection site salinity will be considerably lower than this for both species. Upon arrival in the laboratory, the water used to hold *E. estuarius* and *L. plumulosus* should be adjusted to 20 ‰ by adjusting the salinity in the holding container at a rate that must not exceed 5 ‰ per 24 h. The amphipods should be maintained at 20 ‰ for 2 d before testing.

10.4.5 Lighting. Lighting must be constant and continuous throughout the holding and acclimation period for all species except *A. abdita*. *Ampelisca abdita* requires a 16:8 L:D photoperiod to promote feeding. Fluorescent lights should be used, and they should provide from 500 to 1000 lux at the surface of the sediment in holding containers. Some chemicals are photoactivated by ultraviolet (UV) light (Ankley et al., 1994). Fluorescent light does not contain UV light, but natural sunlight does. Lighting can therefore affect toxicological responses and is an important experimental variable photoactivated chemicals.

## 10.4.6 Water

10.4.6.1 Provided that it is acceptable to the test organisms, either an uncontaminated supply of natural sea water or reconstituted sea water can be used for holding and acclimation (Section 7). At a minimum, healthy amphipods must exhibit acceptable survival in holding water, and must not exhibit signs of stress, such as unusual behavior or changes in appearance.

10.4.6.2 If natural sea water is used, it should be obtained from an uncontaminated area known to support a healthy, reproducing population of the test species or comparatively sensitive species. Reconstituted sea water is prepared by adding commercially available sea salts to water from a suitable source, in quantities sufficient to provide the desired salinity. Pre-formulated brine (e.g., 60 to 90 ‰) prepared with dry ocean salts or heat-concentrated natural sea water can also be used. To obtain the desired holding or acclimation salinity, sea salts or brine can be added to a suitable fresh water, natural estuarine water, or the laboratory's sea water supply. The suitability and consistency of a particular salt formulation for use in holding and acclimation should be verified by laboratory tests because some formulations can produce unwanted toxic effects or sequester contaminants (Environment Canada, 1992). Reconstituted water should be intensively aerated for two weeks before use (Environment Canada, 1992). Suitable sources of water used for preparing reconstituted sea water include deionized water or distilled water, or an uncontaminated natural surface water or ground water. Chlorinated water must never be used because residual chlorine and chlorine-produced oxidants are highly toxic to many aquatic animals. Dechlorinated municipal drinking water should be used only as a last resort because dechlorination is often incomplete.

10.4.6.3 Assessments of the quality of the water used for holding and acclimation and for preparing reconstituted sea water should be performed as frequently as required to document acceptability. Analyses of variables including salinity, temperature, suspended solids, pH, dissolved oxygen, total dissolved gasses, ammonia, nitrite, pesticides, and metals are recommended. Sea water used for holding and acclimating amphipods should be filtered ( $\leq 5 \mu\text{m}$ ) shortly before use to remove suspended particles and organisms. Holding/acclimation water prepared from natural sea water should be used within 2 d of filtration/sterilization whereas reconstituted sea water should be used within 24 h of filtration/sterilization.

10.4.7 Feeding. *Ampelisca abdita* and *L. plumulosus* require supplemental feeding during holding/acclimation. *Ampelisca abdita* should be fed daily, whereas *L. plumulosus* should be fed every other day. Both species should be supplied with an algal ration consisting of *Pseudoisochrysis paradoxa* or *Phaeodactylum tricornutum* that is provided in conjunction with sea water renewal. See Stein (1973) for procedures to culture algae. After 75% of the overlying water has been removed, each holding container should be renewed with sea water at the appropriate salinity that contains algae at a concentration of at least  $1 \times 10^6$  cells/mL. *Leptocheirus plumulosus* should also be provided with dry food ration, consisting at a minimum of finely powdered TetraMin®. It may be desirable to grind the dry food in a blender. Each container should receive approximately 0.5 g

dry food/350 amphipods. *Eohaustorius estuarius* and *R. abronius* will utilize organic material in the holding sediment as food and do not require supplemental feeding.

**10.4.8 Acceptability of animals.** Amphipods counted into the holding/acclimation chambers should be active and appear healthy. Any individuals that fail to burrow or fail to make tubes (i.e., *A. abdita*) in holding sediment or that appear unhealthy during the holding/acclimation period should be discarded. Apparently dead individuals should also be discarded. If greater than 5% of the amphipods emerge or appear unhealthy during the 48 h preceding the test, the entire group should be discarded and not used in the test. Additionally, the group should be discarded if more than 10% of the amphipods die or become inactive during the holding period before testing.

## **10.5 Culture Procedure for *Leptocheirus plumulosus***

**10.5.1** The culturing method below is based on a method described in DeWitt et al. (1992a). USEPA sponsored a workshop was held from 12 to 13 January 1994 in Newport, OR, on sediment testing (both acute and chronic) and culturing of *Leptocheirus plumulosus*. Participants arrived at a consensus on recommending that laboratories with little or no experience follow what has worked for other laboratories first. A consensus among workshop participants was reached on a recommended diet for cultures, and this will be stated. These procedures should not be considered definitive because many issues contributing to optimal culture productivity have yet to be addressed. A periodic-renewal culture system is used. It consists of culture bins that contain aerated sea water over a thin (~1 cm) layer of clean, fine-grained sediment in which the amphipods burrow. Culturing areas must be separate from testing areas to avoid loss of cultures because of cross-contamination.

**10.5.2 Starting a Culture.** Amphipods for starting a laboratory culture of *L. plumulosus* should be obtained from a source with a verified culture (Table 10.1). Alternatively, *L. plumulosus* can be obtained from field populations. The taxonomy of the animals must be confirmed before they are introduced into existing laboratory populations. In addition, the ability of the wild population of sexually reproducing organisms to cross-breed with existing laboratory populations must be determined. Sensitivity of the wild population to select contaminants should also be documented. The temperature and salinity of the shipped water containing amphipods should be gradually adjusted to 20°C and 20 ‰, respectively, at rates not exceeding 3°C per d and 5 ‰ per d. Feeding and regular maintenance should begin once the acclimation period is over. Cultures should be started with approximately 300 mixed-age animals, of which only 100 should be reproductively active adults (i.e., individuals >3 mm in length).

**10.5.3 Culture Chambers.** Culture chambers should be amenable to easy maintenance. Plastic dishpans (30 cm × 40 cm × 15 cm) have been used successfully by several laboratories (DeWitt et al., 1992a), and are recommended. Aeration, provided through an aquarium air stone, should be vigorous and constant.

**10.5.4 Culture Sediment.** Cultures should be established with a thin layer (~1 cm) of sediment that is spread on the bottom of the culture chamber. The sediment that is

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 another area that harbors sediment within the physico-chemical tolerance of *L. plumulosus* that are listed in Table 4.1. Culture sediment should be press sieved through a <0.5 mm screen before using to avoid the presence of indigenous organisms in culture.

**10.5.5 Culture Water.** Culture water should be from the same source used for holding and acclimating test organisms and for conducting toxicity tests. See Section 7.1.2 for acceptable sources of seawater. Cultures of *L. plumulosus* are maintained at a salinity of 20 ‰. To obtain this salinity, sea water should be diluted with deionized water. Seawater used for culturing amphipods should be filtered ( $\leq 5 \mu\text{m}$ ) shortly before use to remove suspended particles and organisms. Water that may be contaminated with pathogens should be treated shortly before use by filtration ( $\leq 0.45 \mu\text{m}$ ), either alone or in combination with ultraviolet sterilization. Culture water prepared from natural sea water should be used within 2 d of filtration/sterilization whereas reconstituted sea water should be used within 24 h of filtration/sterilization. Culture water should be renewed three times per week in conjunction with feeding.

**10.5.6 Temperature and Photoperiod.** Cultures should be maintained at 20°C. The reproductive rate of *L. plumulosus* increases at temperatures greater than 20°C, necessitating more frequent culture thinning. Temperatures below 20°C may not foster sufficiently prolific reproductive rates. Fluorescent lights should be on a 16L:8D photoperiod at a light intensity of 500 to 1000 lux.

### **10.5.7 Food and Feeding**

**10.5.7.1** The following section is based on recommendations made at the USEPA-sponsored "*Leptocheirus plumulosus* Workshop" held in Newport, OR, from 12 to 13 January 1994. The recommendations follow the consensus of experts who culture *L. plumulosus* for use in sediment toxicity testing. It was concluded that laboratories unfamiliar with this species should utilize the specific diet recommended below. Modifications to the diet could then proceed by laboratories in order to optimize culture practices as long as the modifications satisfied the performance criteria.

**10.5.7.2** Culture chambers should be provided with food in conjunction with water renewal. Three times a week, approximately 60% of culture water should be siphoned from each culture chamber and replaced with the same volume of renewal water. The renewal water should consist of seawater, cultured phytoplankton, and deionized water combined to a salinity of 20 ‰ and an algal density of approximately  $10^6$  cells  $\text{mL}^{-1}$ . The proportions will vary depending upon the salinity of the seawater and the density of the cultured phytoplankton. The algae used should be the chrysophyte *Pseudoisochrysis paradoxa* and the diatom *Phaeodactylum tricornutum* mixed 1:1 v/v. Other algal species can be used if it can be demonstrated that they foster amphipod growth and reproductive rates equal to those of the aforementioned algal species.

10.5.7.3 The cultures should also be provided with dry food just after the water change is complete. The recommended dry food consists of a mixture of TetraMin<sup>®</sup>, dried alfalfa, dried wheat leaves, and Neo-Novum<sup>®</sup>. TetraMin<sup>®</sup> is a fish food flake, and is available at most aquarium-supply stores. Dried alfalfa and dried wheat leaves are available at health food stores. Neo-Novum<sup>®</sup> is a maturation feed for use in shrimp mariculture, and is available from Argent Chemical Laboratories, Redmond, WA. The mixture should be prepared at the following proportions: 48.5% TetraMin<sup>®</sup> + 24% dried alfalfa + 24% dried wheat leaves + 4.5% Neo-Novum<sup>®</sup>. The dry food mixture should be made in advance in 400 g batches. The mixture should be thoroughly homogenized into a fine powder. A blender will provide best results; alternatively, a mortar-and-pestle can be used.

10.5.7.4 The amount of dry food added will depend on the density of each culture bin. Newly started culture bins (i.e., those with 300 mixed-age animals) should receive 0.25 to 0.3 g of the dry food mixture per feeding for the first two weeks. Thereafter, 0.5 g of the dry food mixture should be provided per feeding. Less food is provided during the first two weeks as excess food may result in microbial build-up on the sediment surface. The appropriate amount of food should be measured out and sprinkled on the water surface at the center of the culture bin, where it will disperse and settle evenly.

### **10.5.8 Culture Maintenance**

10.5.8.1 Observations and Measurements. Cultures should be observed daily to ensure that aeration is adequate in all culture chambers. Inspection for the presence of oligochaete and polychaete worms and copepods should be conducted weekly. The presence of excessive densities of these or other competing or predacious organisms should prompt renewal of culture sediment after separating amphipods from other organisms. Cultures should be inspected for the presence of microbial build-up on the sediment surface in conjunction with water changes and feeding. This build-up appears as a white or gray growth that may originate near uneaten food. Presence of microbial build-up may indicate that the amount dry food is in excess required by the amphipods. Addition of the dry food mixture to culture chambers with surficial microbial build-up should temporarily cease until the build-up is no longer present.

10.5.8.2 Healthy cultures are characterized by an abundance of burrow-openings on the sediment surface. Although amphipods may leave their burrows to search for food or mates, they will ordinarily remain in their burrows under the daylight portion of the photoperiod. Amphipod density may therefore only be estimated by examining the number of burrow openings. An abundance of animals (e.g., >15 per culture bin) on the sediment surface could indicate inadequate sediment quality, low dissolved oxygen concentrations, or overcrowding. A culture chamber with an abundance of amphipods or unhealthy individuals on the sediment surface should be examined closely, and the dissolved oxygen concentration should be measured. If the dissolved oxygen concentration is below 60% saturation the culture chamber should be sieved, and the population and culture sediment examined. If the population is too dense (i.e., >1500 adults per culture chamber), it should be thinned as described below. If the sediment becomes an unacceptable habitat, i.e., if it is black and sulfurous below the sediment

surface, or contains an excess of competitive or predacious organisms, the healthy surviving amphipods should be placed in a new culture chamber with newly prepared culture sediment.

10.5.8.3 Water temperature should be measured daily and dissolved oxygen should be measured weekly. Cultures should be continuously aerated. Salinity should be measured after water renewal, or 3 times per week. Ammonia and pH should be measured with each new batch of sediment.

10.5.8.4 Culture Density. *Leptocheirus plumulosus* can be prolific, and care must be taken to ensure that culture chambers do not get overcrowded. Amphipods in overcrowded culture chambers may be stressed because of food and space limitations, causing the fecundity of females to drop below 5 eggs per female. Culture density must not exceed 1.5 amphipods cm<sup>2</sup> (i.e., 1500 animals per culture chamber) and should ideally be maintained at approximately 0.4 amphipods cm<sup>2</sup> (i.e., 400 adults per culture chamber). To avoid overcrowding, cultures should be thinned every two months by sieving through a 1 mm mesh screen, allowing young amphipods to pass through the screen and remain in the sediment. Approximately 100 to 150 adult amphipods should be selected from the sieved population and returned to the culture tub. The remainder should be used to start new cultures or discarded.

#### 10.5.9 Obtaining Amphipods for Starting a Test

10.5.9.1 *Leptocheirus plumulosus* used in tests should be started with pre-reproductive animals that are 2 to 4 mm in length. To obtain animals in this size range, sediment from culture chambers containing mixed-size amphipods should be poured over a sieve series that consists of the following sequence of mesh sizes: 0.71 mm, 0.50 mm, and 0.25 mm. Animals retained on the 0.50 mm mesh screen should be washed into a shallow glass pan. The smaller animals from this group should be selected for toxicity testing. Gravid females should be avoided.

10.5.9.2 Alternatively, test animals within a narrow size range are obtained by isolating the smallest amphipods which are allowed to grow until they reach a testable size. To obtain the smallest amphipods, first transfer sediment from culture chambers containing mixed-size amphipods over a sieve series that consists of the following sequence of mesh sizes: 1.0 mm, 0.5 mm, and 0.25 mm. Animals retained on the 0.25 mm mesh screen should be small juveniles that are 1.1 to 2.0 mm in length. They will take approximately two weeks to reach testable size after isolation. The amphipods retained on the 0.25 mm screen should be washed into a culture chamber that is set up as a normal culture chamber, i.e., containing a thin (~1 cm) sediment layer and maintained under culture conditions for the two week interim period. By the end of the two week grow-out period, the animals should be of testable size (i.e., 2 to 4 mm), and be within a narrow size and age range.

## Section 11 Test Method 100.4

### *Ampelisca abdita*, *Eohaustorius estuarius*, *Leptocheirus plumulosus*, or *Rhepoxynius abronius* 10-d Survival Test for Sediments

#### 11.1 Introduction

11.1.1 *Ampelisca abdita*, *Eohaustorius estuarius*, *Leptocheirus plumulosus*, and *Rhepoxynius abronius* have been used extensively to test the toxicity of estuarine and marine sediments. The choice of these amphipod species as test organisms is based on sensitivity to sediment-associated contaminants, availability and ease of collection, tolerance of environmental conditions (e.g., temperature, salinity, grain-size), ecological importance, and ease of handling in the laboratory. Additionally, the species chosen for this method are intimately associated with sediment by nature of their burrowing or tube-dwelling and feeding habits. Field validation studies have shown that amphipods are absent or have reduced abundances at sites where toxicity 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. Test guidance for *A. abdita*, *E. estuarius*, *L. plumulosus*, and *R. abronius* has previously been developed (ASTM, 1992). The four species chosen are representative of both estuarine and marine habitats and sediments that span the spectrum of particle sizes from fine-to coarse-grained sediment. Thus, either alone or in combination, they may be used to measure toxicity of any commonly encountered estuarine or marine sediment.

11.1.2 Specific test methods for conducting the 10-d sediment toxicity test for the amphipods *Ampelisca abdita*, *Eohaustorius estuarius*, *Leptocheirus plumulosus*, and *Rhepoxynius abronius* are described in Section 11.2. Test method 100.4 was developed based on Swartz et al. (1985); DeWitt et al. (1989); Scott and Redmond (1989); Schlekot et al. (1992); ASTM (1992); and Environment Canada (1992). Results of tests using procedures different from the procedures described in Section 11.2 may not be comparable and these different procedures may alter bioavailability. Comparison of results obtained using modified versions of these procedures might provide useful information concerning new concepts and procedures for conducting sediment tests with estuarine or marine organisms. If tests are conducted with procedures different from the procedures described in the manual, additional tests are required to determine comparability of results (Section 1.3).

#### 11.2 Recommended Test Method for Conducting a 10-d Sediment Toxicity Test with *Ampelisca abdita*, *Eohaustorius estuarius*, *Leptocheirus plumulosus*, or *Rhepoxynius abronius*.

11.2.1 Recommended conditions for conducting a 10-d sediment toxicity test with *A. abdita*, *E. estuarius*, *L. plumulosus*, and *R. abronius* 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, number of test

**Table 11.1** Test conditions for conducting a 10-d sediment toxicity test with *Ampelisca abdita*, *Eohaustorius estuarius*, *Leptocheirus plumulosus*, or *Rhepoxynius abronius*

Parameter	Conditions
1. Test type:	Whole sediment toxicity test, static.
2. Temperature:	15°C: <i>E. estuarius</i> and <i>R. abronius</i> 20°C: <i>A. abdita</i> 25°C: <i>L. plumulosus</i>
3. Salinity:	20 ‰: <i>E. estuarius</i> and <i>L. plumulosus</i> 28 ‰: <i>A. abdita</i> and <i>R. abronius</i>
4. Light quality:	Wide-spectrum fluorescent lights
5. Illuminance:	500 - 1000 lux
6. Photoperiod:	24L:0D
7. Test chamber:	1-L glass beaker or jar with ~10 cm I.D.
8. Sediment volume:	175 mL (2 cm)
9. Overlying water volume:	800 mL
10. Renewal of overlying water:	None
11. Size and life stage of amphipods:	<i>A. abdita</i> : 3 - 5 mm (no mature males or females) <i>E. estuarius</i> : 3 - 5 mm <i>L. plumulosus</i> : 2 - 4 mm (no mature males or females) <i>R. abronius</i> : 3 - 5 mm
12. Number of organisms/chamber:	20 per test chamber
13. Number of replicate chambers/treatment:	Depends on objectives of test. At a minimum, four replicates must be used.
14. Feeding:	None

**Table 11.1** Test conditions for conducting a 10-d sediment toxicity test with *Ampelisca abdita*, *Eohaustorius estuarius*, *Leptocheirus plumulosus*, or *Rhepoxynius abronius* (continued)

Parameter	Conditions
15. Aeration:	Water in each test chamber should be aerated overnight before start of test, and throughout the test; aeration at rate that maintains $\geq 90\%$ saturation of dissolved oxygen concentration
16. Overlying water:	Clean sea water, natural or reconstituted water
17. Overlying water quality:	Temperature daily. pH, ammonia, salinity, and DO of overlying water at least at test start and end. Salinity, ammonia, and pH of pore water
18. Test duration:	10 d
19. Endpoints:	Survival (reburial optional for <i>E. estuarius</i> , <i>L. plumulosus</i> , and <i>R. abronius</i> )
20. Test acceptability:	Minimum mean control survival of 90% and satisfaction of performance-based criteria specifications outlined in Table 11.3.

**Table 11.2** General activity schedule for conducting a sediment toxicity test with *Ampelisca abdita*, *Eohaustorius estuarius*, *Leptocheirus plumulosus*, or *Rhepoxynius abronius*.

Day	Activity
-10 to -3	Collect or receive amphipods from supplier and place into collection site sediment. Alternatively, separate 2 - 4 mm <i>L. plumulosus</i> from cultures.
-9 to -2	Acclimate and observe amphipods to species-specific test conditions. Feed <i>A. abdita</i> and <i>L. plumulosus</i> . Monitor water quality (e.g., temperature, salinity, and dissolved oxygen).
-1	Observe amphipods, monitor water quality. Add sediment to each test chamber, place chambers into exposure system, and start aeration.
0	Measure pore water total ammonia, salinity, and pH. Measure temperature of overlying water in test chambers. Transfer 20 amphipods into each test chamber. Archive 20 test organisms for length determination.
1	Measure temperature. Observe behavior of test organisms and ensure that each test chamber is receiving air. Measure dissolved oxygen in test chambers to which aeration has been cut-off.
2	Measure total water quality (pH, temperature, dissolved oxygen, salinity, total ammonia of overlying water). Observe behavior of test organisms and ensure that each test chamber is receiving air.
3 to 7 and 9	Same as Day 1.
8	Same as Day 2.
10	Measure temperature. End the test by collecting the amphipods with a sieve.

chambers/treatment, and water quality characteristics should be based on the purpose of the test and the methods of data analysis (Section 12). The number of replicates and concentrations tested depends in part on the significance level selected and the type of statistical analysis. When variability remains constant, the sensitivity of a test increases as the number of replicates increase.

11.2.2 The recommended 10-d sediment toxicity test with *A. abdita*, *E. estuarius*, *L. plumulosus*, and *R. abronius* must be conducted at the species-specific temperature and salinity with a 24 h light photoperiod at a illuminance of about 500 to 1000 lux (Table 11.1). Test chambers are 1 L glass chambers containing 175 mL of sediment and 800 mL of overlying seawater. Twenty amphipods are added to each test chamber at the start of a test. The size range of the amphipods will depend on species that is being tested (see Section 10.3.4 for allowable size range for each species). The number of replicates/treatment depends on the objective of the test. Five replicates are recommended for routine testing (see Section 12). Exposure is static (i.e., water is not renewed), and the animals are not fed over the 10 d exposure period. Overlying water can be culture water, surface water, site water, or reconstituted water. For site-specific evaluations, the characteristics of the overlying water should be as similar as possible to the site where sediment is collected. For all other applications, the characteristics of the overlying water for each species should be chosen according to Table 11.1. Requirements for test acceptability are summarized in Table 11.3.

### 11.3 General Procedures

11.3.1 Introduction of Sediment. On the day before the addition of amphipods (Day -1), each test sediment (either field collected or laboratory spiked) should be homogenized by stirring in the sediment storage container or by using a rolling mill, feed mixer, or other suitable apparatus. Control and reference sediments are included. Sediment should be visually inspected to judge the extent of homogeneity. Excess water on the surface of the sediment can indicate separation of solid and liquid components. If a quantitative measure of homogeneity is required, replicate subsamples should be taken from the sediment batch and analyze for TOC, chemical concentrations, and particle size. Spiked sediments should not be homogenized before introduction into test chambers because the equilibrium between the spiked contaminant and the sediment partitioning factors may be disrupted.

11.3.1.1 A 175-mL aliquot of thoroughly homogenized sediment is added to each test chamber. It is important that an identical volume be added to each replicate test chamber; at a minimum the volume added should equate to a depth of 2 cm in the test chamber. The sediment added to the test chamber should be settled either by tapping the side of the test chamber against the side of the hand or by smoothing the sediment surface with a nylon, fluorocarbon, or polyethylene spatula. Highly contaminated sediment should be added to test chambers in a certified laboratory fume hood.

11.3.2 Addition of Overlying Water. To minimize disruption of sediment as test seawater is added, a turbulence reducer should be used. The turbulence reducer may be

**Table 11.3 Test acceptability requirements for a 10-d sediment toxicity test with *Ampelisca abdita*, *Eohaustorius estuarius*, *Leptocheirus plumulosus*, or *Rhepoxynius abronius*.**

---

- A.** It is recommended for conducting a 10-d test with *A. abdita*, *E. estuarius*, *L. plumulosus*, or *R. abronius* that the following performance criteria are met:
- 1.** Size, life stage, and reproductive stage of amphipods must be within the prescribed species-specific ranges at the end of the test (Section 10.3.4).
  - 2.** Average survival of amphipods in the control sediment must be greater than or equal to 90% at the end of the test.
  - 3.** Salinity, pH, and ammonia in the overlying water and sediment grain size are within tolerance limits of test species.
- B.** Performance-based criteria for culturing *L. plumulosus* include:
- 1.** Laboratories should perform monthly 96-h water-only reference-toxicity tests to assess the sensitivity of culture organisms. If reference-toxicity tests are not conducted monthly, the lot of organisms used to start a sediment test must be evaluated using a reference toxicant (Section 9.16).
  - 2.** Records should be kept on the frequency of restarting cultures.
  - 3.** Laboratories should record the pH and ammonia of the cultures 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.** Performance-based criteria for field-collected amphipods include:
- 1.** Laboratories should perform reference-toxicant tests on each batch of field-collected amphipods received used in a sediment test (Section 9.16).
  - 2.** Acclimation rates to test salinity and temperature should not exceed 3 ‰ and 5 ‰ per 24 h.
-

Table 11.3 Test acceptability requirements for a 10-d sediment toxicity test with *Ampelisca abdita*, *Eohaustorius estuarius*, *Leptocheirus plumulosus*, or *Rhepoxynius abronius* (continued)

---

3. Amphipods received from commercial suppliers must exhibit active swimming behavior upon placement in water, have full digestive tracts, and display an acceptable color.

D. Additional requirements:

1. All organisms in a test must be from the same source.
  2. It is desirable to start tests as soon as possible after collection of sediment from the field (see Section 8.2 for additional detail).
  3. All test chambers (and compartments) should be identical and should contain the same amount of sediment and overlying water.
  4. Negative-control sediment must be included in a test.
  5. The time-weighted average of daily temperature readings must be within  $\pm 1^{\circ}\text{C}$  of the desired temperature. The instantaneous temperature must always be within  $\pm 3^{\circ}\text{C}$  of the desired temperature.
  6. Natural physico-chemical characteristics of test sediment collected from the field should be within the tolerance limits of the test organisms.
-

either a disk cut from polyethylene, nylon, or Teflon® sheeting (4 to 6 mil), or a glass petri dish attached (open face up) to a glass pipette. If a disk is used as the turbulence reducer, it should fit the inside diameter of the test chamber and have attached a length of nylon monofilament (or nontoxic equivalent) line. The turbulence reducer is positioned just above the sediment surface and raised as sea water is added to the 750-mL mark on the side of the test chamber. The turbulence reducer is removed and rinsed with test sea water between replicates of a treatment. A separate turbulence reducer is used for each treatment. The test chambers should be covered, placed in a temperature controlled water bath (or other acceptable equivalent) and gently aerated. A test begins when the organisms are added to the test chambers (Day 0).

**11.3.3.1 Addition of Amphipods.** On the following day (Day 0), amphipods are added to the test chambers. Approximately one-third more amphipods than are needed for the test should be sieved from the culture or control sediment in the holding container(s), and transferred to a sorting tray. The additional animals allow for the selection of healthy, active individuals. The sieve size for isolating amphipods from the culture or control sediment will depend upon the selected species. *Ampelisca abdita* and *L. plumulosus* should be isolated using a 0.5 mm sieve, whereas *E. estuarius* and *R. abronius* should be isolated using a 1.0 mm sieve. Sieving should be conducted with sea water of the same temperature and salinity as the holding and test water. Once isolated, active amphipods should be randomly selected using a transfer pipette or other suitable tool (not forceps), and distributed among dishes or cups containing approximately 150 mL of test sea water until each container has twenty amphipods. The number of amphipods in each dish should be verified by recounting before adding 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. The distribution of amphipods to the test chambers must be executed in a randomized fashion.

**11.3.3.2 Amphipods should be added to test chambers without disruption of the sediment by placing a 6-mil polyethylene, nylon, or Teflon® disk on the water surface and gently pouring the water and amphipods from the sorting container over the disk into the test chamber. The disk should be removed once the amphipods have been introduced. Alternatively, amphipods from the sorting container can be poured into a sieve cup (mesh size  $\leq 0.5$  mm) and gently washed into the test chamber with test sea water. Any amphipods remaining in the sorting container should be gently washed into the test chamber using test sea water. The water level should be brought up to the 950-mL mark, the test chamber covered, and aeration continued.**

**11.3.3.3 After the addition of the animals, the test chambers should be examined for animals that may have been injured or stressed during the isolation, counting, or addition processes. Injured or stressed animals will not burrow into sediments, and should be removed. The period of time allowed for healthy amphipods to bury into test sediments will depend upon the species used. *Eohaustorius estuarius*, *L. plumulosus*, and *R. abronius* should be allowed 5 to 10 min to bury into the test sediment. *Ampelisca abdita*, which may take longer to build tubes, should be allowed 1 h. Amphipods that have not burrowed within the prescribed time should be replaced with animals 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 removed must be recorded.

#### 11.3.4 Test Conditions

11.3.4.1 Aeration. The overlying sea water in each test chamber must be aerated continuously after the water is added (i.e., Days -1 through 10) except during introduction of the test organisms. Compressed air, previously filtered and free of oil, should be bubbled through a glass or plastic pipette and attached plastic tubing. The tip of the pipette should be suspended 2 to 3 cm above the surface of the sediment layer so as to not disturb the sediment surface. The concentration of dissolved oxygen (DO) in the water overlying the sediment in the test chambers is maintained at or near saturation by gently aerating the water. Air is bubbled through the test chamber at a rate that maintains a  $\geq 90\%$  DO concentration, but does not cause turbulence or disturb the sediment surface. If air flow to one or more test chambers is interrupted for more than one h, DO must be measured in those test chambers to determine whether DO concentrations have fallen below 60% of saturation. Results may be unacceptable for test chambers in which aeration was interrupted and DO concentrations fell to below 60% saturation.

11.3.4.2 Lighting. Lights must be left on continuously at an intensity of 500 to 1000 lux during the 10 d exposure period. The constant light increases the tendency of the organisms to remain buried in the sediment, and thus to remain exposed to the test material.

11.3.4.3 Feeding. The four species of amphipods used in this method must not be fed during the 10-d exposure period.

11.3.4.4 Water Temperature. The test temperature will depend on the species that is tested. Test temperatures were selected to be near the summertime thermal maximum that each species would be expected to encounter in the environment. *Eohaustorius estuarius* and *R. abronius*, the Pacific Coast amphipods, must be tested at 15°C. *Ampelisca abdita* must be tested at 20°C and *L. plumulosus* at 25°C.

11.3.4.5 Salinity. The salinity of the water overlying the test sediment will vary depending on the selected test species. For routine testing, *A. abdita* and *R. abronius* should be tested at an overlying water salinity of 28 ‰, whereas *E. estuarius* and *L. plumulosus* should be tested at 20 ‰. Pore water salinity of each test sediment must be measured prior to the initiation of a test. Sediment pore water should be obtained by centrifugation. Alternatively, salinity can be measured before homogenization in the water that comes to the surface in the sample container as the sediment settles. The pore water salinity of the test sediment must be within the salinity application range of the chosen amphipod species (Table 10.2). *Rhepoxynius abronius* cannot be tested when sediment pore water salinities are  $< 25$  ‰. Another species must be used for such sediments. *Ampelisca abdita*, *E. estuarius*, and *L. plumulosus* can be tested over the entire pore water salinity range (i.e., 0 to 34 ‰) when the recommended species-specific

overlying salinity is used. Depending on the objectives of the study, *E. estuarius* and *L. plumulosus* may be tested with overlying water at salinities ranging from 1 to 32 ‰.

### **11.3.5 Measurements and Observations**

**11.3.5.1** Temperature should be measured at least daily in at least one test chamber from each treatment. The temperature of the water bath or the exposure chamber should be continuously monitored. The time-weighted average of daily temperature readings must be within  $\pm 1^{\circ}\text{C}$  of the desired temperature. The instantaneous temperature must always be within  $\pm 3^{\circ}\text{C}$  of the desired temperature.

**11.3.5.2** Salinity, dissolved oxygen, and pH of the overlying water should be measured daily in at least one test chamber per treatment, and at a minimum, they must be measured in every test chamber at the beginning and the end of a test.

**11.3.5.3** Ammonia must be measured in overlying water towards the beginning (e.g., day 2) and towards the end of the test (e.g., day 8). Measurement of overlying water pH and temperature should accompany each ammonia measurement. Simultaneous measurements of ammonia, pH, and temperature in sediment pore water should be measured at the beginning of the test. Pore water should be extracted after the sediment has been press-sieved and homogenized. Samples of pore water should be obtained by centrifugation.

**11.3.5.4** Each test chamber must be examined at least daily during the 10 d test period to ensure that airflow to the overlying sea water is acceptable. The number of amphipods swimming in the water column and trapped in the air-water interface should be noted. Amphipods caught in the air-water interface must be gently pushed down into the water using a glass rod or pipette. The number of apparently dead animals should be noted.

**11.3.6** Ending a Test. Laboratories should demonstrate the ability of their personnel to recover an average of at least 90% of the organisms from control sediment. For example, test organisms could be added to control sediment and recovery could be determined after 1 h (Tomasovic et al., 1994).

**11.3.6.1** The contents of the test chambers must be sieved to isolate the test animals. The mesh size for sieving the contents of the test chambers must be no larger than 0.5 mm. Test water should be used for sieving. Material retained on the sieve should be washed into a sorting tray with clean test sea water. *Ampelisca abdita* are tube-builders, and it will be necessary to make an effort to ensure that no tubes remain on the sieve. The sieve should be slapped forcefully against the surface of the water to ensure that all of the amphipods and tubes are dislodged from the screen. *Eohaustorius estuarius*, *L. plumulosus*, and *R. abronius* are easily removed from the sediment by the sieving process.

**11.3.6.2** Material that has been washed from the sieve into the sorting tray should be carefully examined for the presence of amphipods. A small portion of the material

should be sorted through at a time, removing amphipods as they are found. Material from tests conducted with *A. abdita* will include tubes built by the amphipods during the test. The tubes must be carefully examined and teased apart under a dissecting microscope or magnifying glass because *A. abdita* will often remain in the tubes even after vigorous sieving. Numbers of live, missing, and dead amphipods should be determined and recorded for each test chamber. Missing animals are assumed to have died and decomposed during the test and disintegrated; they should be included in the number dead in calculations of the percent survival for each replicate treatment. Amphipods that are inactive but not obviously dead must be observed using a low-power dissecting microscope or a hand-held magnifying glass. Any animal that fails to exhibit movement (i.e. neuromuscular twitch of pleopods or antennae) upon gentle prodding with a probe should be considered dead.

**11.3.7 Test Data.** Survival is the primary endpoint recorded at the end of the 10-d sediment toxicity test with *A. abdita*, *E. estuarius*, *L. plumulosus*, and *R. abronius*. The ability of surviving amphipods to rebury in clean control sediment can be used to calculate *effective mortality*, that is, the sum of dead animals plus those survivors that fail to rebury. This endpoint has been used for *E. estuarius*, *L. plumulosus*, and *R. abronius*. If it is desired to determine reburial, surviving amphipods should be transferred to containers holding a 2-cm layer of 0.5 mm sieved control sediment and an overlying layer ( $\geq 2$  cm) of test sea water. Salinity of the test sea water for reburial should be the same as that measured in the test chamber. The number of surviving amphipods unable to rebury in control sediment after 1 h is recorded for each test chamber and is used to calculate effective mortality.

## **11.4 Interpretation of Results**

**11.4.1** Section 12 describes general information for interpretation of test results. The following sections describe species-specific information that is useful in helping to interpret the results of sediment toxicity tests with *A. abdita*, *E. estuarius*, *L. plumulosus*, and *R. abronius*.

**11.4.2 Influence of Indigenous Organisms.** Indigenous organisms may be present in field-collected sediments. An abundance in the sediment sample of the test organism, or organisms taxonomically similar to the test organism, may make interpretation of treatment effects difficult. The presence of predatory organisms can also adversely affect test organism survival. For example, Redmond and Scott (1989) showed that the polychaete *Nephtys incisa* will consume *Ampelisca abdita* under toxicity test conditions.

**11.4.3 Effect of Sediment Grain Size.** All four species show tolerance to most sediment types, with generally little effect on survival whether coarse-grained or fine-grained (i.e., predominantly silt and clay) clean sediments are used. However, adverse effects due to the grain-size distribution of test sediment may occur when sediments that are either extremely sandy or fine depending on the species of amphipod used. In order to separate effects of sediment-associated contaminants from effects of particle size, an appropriate clean control/reference sediment should be incorporated into the test when test sediments are within the range of concern for each species. Alternatively, another

species that is tolerant of the sediment extreme in question might be tested in conjunction with the chosen species. Ranges of concern are outlined below.

11.4.3.1 *Ampelisca abdita*. Survival of *Ampelisca abdita* in sediment that is  $\geq 95\%$  sand may elicit excess mortality, but this has not been quantified (J. Scott, SAIC, Narragansett, RI, personal communication). Toxicity tests conducted with *A. abdita* on sediments that are  $\geq 95\%$  sand should be conducted with a clean control sediment characteristic of that test sediment.

11.4.3.2 *Leptocheirus plumulosus*. *Leptocheirus plumulosus* has exhibited  $\geq 90\%$  survival in clean sediments ranging from ~ 100% sand to ~ 100% silt + clay (SAIC, 1993a; SAIC, 1993b; Schlekot et al., 1992; J. Kavanaugh, University of West Florida, Gulf Breeze, FL, personal communication).

11.4.3.3 *Eohaustorius estuarius*. *Eohaustorius estuarius* has exhibited acceptable (100%) survival when exposed to clean sediments ranging from 0.6 to 100% sand (USEPA ERL-Narragansett, Pacific Ecosystems Branch, Newport, OR, unpublished data). However, *E. estuarius* naturally inhabits sandy sediments, and some correlation between survival and grain size has been reported by DeWitt et al. (1989) and SAIC (1993a; 1993b), with increased mortality associated with increased proportions of fine-grained sediment. Therefore, it may be desirable to include clean control sediments with a range of particle sizes characteristic of those of the test sediment(s) in toxicity tests conducted with *E. estuarius*.

11.4.3.4 *Rhepoxynius abronius*. *Rhepoxynius abronius* has been used to test sediments with a wide range of sediment grain sizes. However, *R. abronius* naturally inhabits clean, fine, sandy sediments, and a number of studies have shown some reduction in survival when this species is held in very fine-grained (predominantly silt and clay) sediment (DeWitt et al., 1988; Long et al., 1990; McLeay et al., 1991; SAIC, 1993a; SAIC, 1993b). Therefore, when test sediments are predominantly silts or clays, the experimental design include a silt-clay control sediment with a range of particle sizes characteristic of the test sediment(s). Alternatively, when the particle size of test sediments are known, regression techniques can be used to evaluate potential effects of fines on *R. abronius* survival (see DeWitt et al., 1988).

11.4.4 Effects of Pore Water Salinity. The four amphipod species exhibit variability in their salinity tolerance ranges. There are two options available for laboratory sediment testing regarding the choice of overlying water salinity for a given sediment. The options are to either use the standard salinity for each test species, or to match the salinity to that of the pore water. The range of pore water salinities in which a given species can survive for ten days when using the species-specific overlying water salinity is the salinity application range. The range of salinity in which a given species can survive for ten days when the overlying water salinity is matched to that of the pore water salinity is the salinity tolerance range. In either scenario, the potential for a toxic response due to salinity alone exists if a species is exposed to conditions outside of its range of tolerance. For estuarine sediments, it is very important to know the pore water

salinity of each sediment before testing is started, to choose a species that will not be affected by the pore water salinity, and to use overlying water of an appropriate salinity.

11.4.4.1 Salinity tolerance ranges for each species are as follows: *Ampelisca abdita*: 20 to 32 ‰ (SAIC, 1993b); *Eohaustorius estuarius*: 2 to 34 ‰; *Leptocheirus plumulosus*: 1.5 to 32 ‰; *Rhepoxynius abronius*: 25 to 32 ‰. While there is some evidence of salinity-related stress for *E. estuarius* and *L. plumulosus* at salinity extremes, the breadth of salinity tolerance exhibited by these species (DeWitt et al., 1989; Schlekot et al., 1992; SAIC, 1993b) is most likely sufficient for application to the majority of sediments that may be encountered in an estuarine system (i.e., interstitial salinity between 2 and 28 ‰). If it is desirable to have matching overlying and pore water salinity from areas where pore water salinities are 0 to 2 ‰, an organism that has been demonstrated to tolerate this salinity range should be used, either instead of or in addition. The amphipod *Hyaella azteca* is one such species. Likewise, sediments collected from areas of high salinity (i.e., >32 ‰ for *L. plumulosus*) should probably utilize *A. abdita*, *E. estuarius*, or *R. abronius*.

11.4.4.2 Salinity application ranges for each species are as follows: *Ampelisca abdita* with overlying water salinity of 28 to 32 ‰; 0 to 34 ‰ (Weisberg et al., 1992; SAIC, 1993b); *Eohaustorius estuarius* with overlying water salinity of 20 ‰; ≤2 to 34 ‰ (DeWitt et al., 1989; SAIC, 1993b); *Leptocheirus plumulosus* with overlying salinity of 20 ‰; ≤1.5 to 32 ‰ (Schlekot et al., 1992; SAIC, 1993b) and *Rhepoxynius abronius* with overlying water salinity of 28 to 32 ‰; 25 to 34 ‰ (Swartz et al., 1985; Lamberson and Swartz, 1988).

11.4.5 Effects of Sediment-associated Ammonia. Field-collected sediments may contain concentrations of ammonia that are toxic to amphipods. Water column no effect concentrations for the four amphipod species are presented in Table 11.4. If ammonia concentrations are above these values, mortality occurring after 10 d may be due in part to effects of ammonia. Depending on test application, it may be desirable to lower the ammonia concentration by manipulating the test system prior to introduction of test organisms if measured ammonia in the overlying water is greater than the species-specific no effect concentration. If sediment toxicity tests are conducted to evaluate the acceptability of dredge material for disposal, the manipulations must be performed. Manipulations involve flushing the test system by renewing a specified amount of overlying water for up to two consecutive 24 h periods.

11.4.5.1 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 animals on Day 0. If overlying water ammonia concentration are less than or equal to the species-specific no effect concentration listed in Table 11.4, then the test may proceed normally. If overlying water ammonia concentration is greater than the species-specific no effect concentration listed in Table 11.4, then the test system must be flushed for 24 h at a rate of 6 volume replacements/24 h.

### **Errata**

#### **Pages 80-82, Sections 11.4.5-11.4.5.3, Effects of Sediment-associated Ammonia**

These sections describe a procedure that can be used to reduce ammonia concentrations in field-collected sediments prior to conducting laboratory toxicity tests. For dredged material testing under the Clean Water Act or the Marine Protection, Research, and Sanctuaries Act, the following alternative procedure should be used. This procedure was described in a December 21, 1993 guidance memorandum issued by the U.S. EPA Office of Wetlands, Oceans and Watersheds, U.S. EPA Office of Science and Technology, and U.S. Army Corps of Engineers Operations, Construction, and Readiness Division.

For dredged material testing the following procedure should be used if it is necessary to reduce interstitial water ammonia levels. Whenever chemical evidence of ammonia is present at toxicologically important levels, and ammonia is not a contaminant of concern, the laboratory analyst should reduce ammonia in the sediment interstitial water to species-specific no-effect concentrations (see table 11.4 on page 81). Ammonia levels in the interstitial water can be reduced by sufficiently aerating the sample and replacing two volumes of water per day. The analyst should measure interstitial ammonia each day until it reaches the appropriate species-specific no-effect concentration. After placing the test organism in the sediment, the analyst should ensure that ammonia concentrations remain within an acceptable range by conducting the toxicity test with continuous flow or volume replacement not to exceed two volumes per day.

**Table 11.4 Application limits for 10-d sediment toxicity tests with *Ampelisca abdita*, *Eohaustorius estuarius*, *Leptocheirus plumulosus*, or *Rhepoxynius abronius***

<b>Parameter</b>	<b><i>Ampelisca abdita</i></b>	<b><i>Eohaustorius estuarius</i></b>	<b><i>Leptocheirus plumulosus</i></b>	<b><i>Rhepoxynius abronius</i></b>
<b>Temperature (°C)</b>	<b>20</b>	<b>15</b>	<b>25</b>	<b>15</b>
<b>Overlying Salinity (‰)</b>	<b>&gt; 10</b>	<b>0 - 34</b>	<b>1.5 - 32</b>	<b>&gt; 25</b>
<b>Grain Size (% silt/clay)</b>	<b>&gt; 10</b>	<b>full range</b>	<b>full range</b>	<b>&lt; 90</b>
<b>Ammonia (total mg/l, pH 7.7)</b>	<b>&lt; 30</b>	<b>&lt; 60</b>	<b>&lt; 60</b>	<b>&lt; 30</b>
<b>Ammonia (UI<sup>1</sup> mg/l, pH 7.7)</b>	<b>&lt; 0.4</b>	<b>&lt; 0.8</b>	<b>&lt; 0.8</b>	<b>&lt; 0.4</b>
<b>Sulfides</b>	<b>NA</b>	<b>NA</b>	<b>NA</b>	<b>NA</b>

<sup>1</sup> UI = unionized ammonia

**11.4.5.2** After 24 h, the overlying water ammonia concentration must be measured again. If it is less than or equal to the species-specific no effect concentration, testing should be initiated by adding animals. The system must be flushed at a rate of 6 volume replacements/24 h over the course of the test. Overlying water ammonia should be measured again on Day 10 of the test.

**11.4.5.3** If after the initial 24 h flushing period (i.e., that described in 11.4.5.1) the overlying water ammonia concentration is still greater than the species-specific no effect concentration, the system must be flushed for again 24 h at a rate of 6 volumereplacements/24 h. After the second flushing, ammonia concentrations in the overlying water should be measured again, and if concentrations are less than or equal to the species-specific no effect concentration listed in Table 11.4, then the test may proceed as described in Section 11.4.5.2. If overlying water ammonia concentrations still exceed the species-specific no effect concentration, it must be concluded that ammonia cannot be reduced to no effect concentrations without concern for flushing other contaminants from the sediment. At this point, the test should still be conducted as described in Section 11.4.5.2 After 24 h, the overlying water ammonia concentration must be measured again. If it is less than or equal to the species-specific no effect concentration, testing should be initiated by adding animals. The system must be flushed at a rate of 6 volume replacements/24 h over the course of the test. Overlying water ammonia should be measured again on Day 10 of the test.

**11.4.5.3** If after the initial 24 h flushing period (i.e., that described in 11.4.5.1) the overlying water ammonia concentration is still greater than the species-specific no effect concentration, the system must be flushed for again 24 h at a rate of 6 volume

## **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 standard operating procedures should be developed before starting a test. Procedures should be developed by each laboratory to verify and archive data.

**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 species, 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 toxicant tests. Original data sheets should be signed and dated by the laboratory personnel performing the tests.

**12.1.3** Example data sheets are included 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 contaminated sediments is usually compared with the response to a control or reference sediment. In any toxicity, 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.* Two types of data can be obtained from sediment toxicity tests. The most common endpoint in toxicity testing is mortality, which is a dichotomous or categorical type of data.

**12.2.1.2** *Sediment Testing Scenarios.* Sediment tests are conducted to determine whether contaminants in sediment are harmful to or are bioaccumulated in benthic organisms. Sediment tests are commonly used in studies designed to: (1) evaluate hazards of dredged material, (2) assess site contamination in the environment (e.g., to rank areas for clean-up), 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.

**12.2.1.2.1 Dredged Material Hazard Evaluation.** In these studies, *n* sites are 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 hazard evaluations is available in USEPA-USCOE (1994).

**12.2.1.2.2 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 may be reduced if subsamples are taken simultaneously for sediment toxicity, 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 to a control sediment, then the pairwise comparison approach described below is appropriate. If the objective is to compare among 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.

**12.2.1.2.3 Sediment-Spiking Experiments.** 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, 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 toxicant tests.

**12.2.2** The guidance outlined below on the analysis of sediment toxicity test data is adapted from a variety of sources including Lee et al. (1994), USEPA (1993a), USEPA (1993b), USEPA (1993c), and USEPA-USCOE (1994). The objectives of a sediment toxicity test is to quantify contaminant effects on test organisms exposed to natural or spiked sediments 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. Some designs will also require a reference sediment that represents an environmental condition or potential treatment effect of interest.

**12.2.2.1 Experimental Unit.** 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.

**12.2.2.2 Replication.** 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 observed differences between treatments.

**12.2.2.3 *Minimum Detectable Difference (MDD).*** 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 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.

**12.2.2.4 *Minimum number of replicates.*** Four replicates per treatment or control are the absolute minimum number of replicates for a sediment toxicity test. However, USEPA recommends five replicates for marine testing or eight replicates for freshwater testing (USEPA, 1994a) for each control or experimental treatment. It is always prudent to include as many replicates in the test design as are economically and logistically possible. USEPA sediment toxicity testing methods recommend the use of 20 organisms per replicate for marine testing or 10 organisms per replicate for freshwater 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 due to the increased organism loading.

**12.2.2.5 *Randomization.*** 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 control and test treatments (e.g., a bias in the results may 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.

**12.2.2.6 *Pseudoreplication.*** 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 control 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 control 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 may 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 completely may be difficult or impossible given resource constraints.

**12.2.3** The purpose of a toxicity test is to determine if the biological response to a treatment sample differs from the response to a control sample. Table 12.1 presents the possible outcomes and decisions that can be reached in a statistical test of such a hypothesis. The null hypothesis is that no difference exists among the mean control and treatment responses. The alternative hypothesis of greatest interest in sediment tests is that the treatments are toxic, or contain concentrations of bioaccumulable compounds, relative to the control or reference sediment.

**12.2.3.1** Statistical tests of hypotheses can be designed to control for the chances of making incorrect decisions. In Table 12.1, alpha ( $\alpha$ ) 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 ( $\beta$ ) 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, control or reference samples. Traditionally, acceptable values for  $\alpha$  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  $\alpha$ , environmental researchers have ignored  $\beta$  and the associated power of the test ( $1-\beta$ ).

**12.2.3.2** Fairweather (1991) presents 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 which include statistical hypothesis testing. To paraphrase 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., ensuing environmental degradation and the eventual cost of its rectification)."

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

**Table 12.1 Suggested  $\alpha$  levels to use for tests of assumptions**

Test	Number of Observations <sup>1</sup>	$\alpha$ When Design Is	
		Balanced	Unbalanced <sup>2</sup>
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

<sup>1</sup> N = total number of observations (replicates) in all treatments combined; n = number of observations (replicates) in an individual treatment

<sup>2</sup>  $n_{\max} \geq 2n_{\min}$

12.2.3.4 Sample size or number of replicates may be fixed due to cost or space considerations, or may be varied to achieve *a priori* probabilities of  $\alpha$  and  $\beta$ . 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 quality control (QC) limits.

12.2.3.5 The MDD is expressed as a percentage change from the mean control response. To test the equality of the control and treatment responses, a two-sample t-test with its associated assumptions is the 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, such as the coefficient of variation (CV) from a control sample, are available, it is possible to use a graphical approach as in Figure 12.1 to determine how likely it is that a 20% reduction will be detected in the treatment response relative to the control response. The CV is defined as 100% x (standard deviation divided by the mean). In a test design with 8 replicates per treatment and with an  $\alpha$  level of 0.05, high power (i.e., >0.80) to detect a 20% reduction from the control mean occurs only if the CV is 15% or less (Figure 12.1). The choice of these variables also affects the power of the test. If 5 replicates are used per treatment (Figure 12.2), the CV needs to be 10% or lower to detect a 20% reduction in response relative to the control mean with a power of 90%.

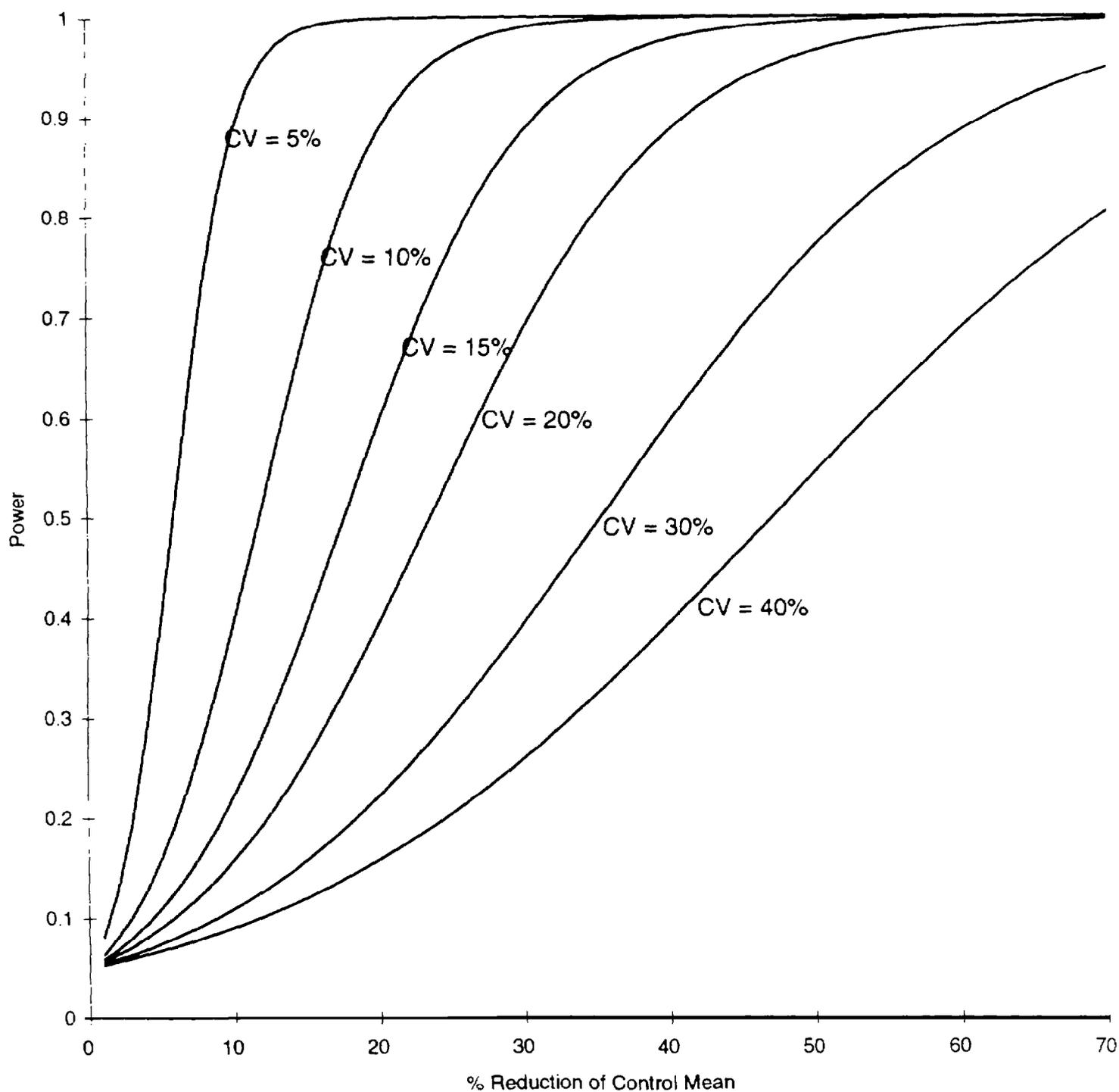
12.2.3.6 Relaxing the  $\alpha$  level of a statistical test increases the power of the test. Figure 12.3 duplicates Figure 12.1 except that  $\alpha$  is 0.10 instead of 0.05. Selection of the appropriate  $\alpha$  level of a test is a function of the costs associated with making Type I and II statistical errors. Evaluation of Figure 12.1 illustrates that with a CV of 15% and an  $\alpha$  level of 0.05, there is an 80% probability (power) of detecting a 20% reduction in the mean treatment response relative to the control mean. However, if  $\alpha$  is set at 0.10 (Figure 12.3) and the CV remains at 15%, then there is a 90% probability (power) of detecting a 20% reduction relative to the control mean. The latter example would be preferable if an environmentally conservative analysis and interpretation of the data is desirable.

12.2.3.7 Increasing the number of replicates per treatment will increase the power to detect a 20% reduction in treatment response relative to the control mean (Figure 12.4). Note, however, that for less than 8 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 8 per treatment, then it may be necessary to find ways to reduce the among replicate variability and consequently the CV. Options that are available include selecting more uniform organisms to reduce biological variability or increasing the  $\alpha$  level of the test. For CVs in the range of 30% to 40%, even eight replicates per treatment is inadequate to detect small reductions ( $\leq 20\%$ ) in response relative to the control mean.

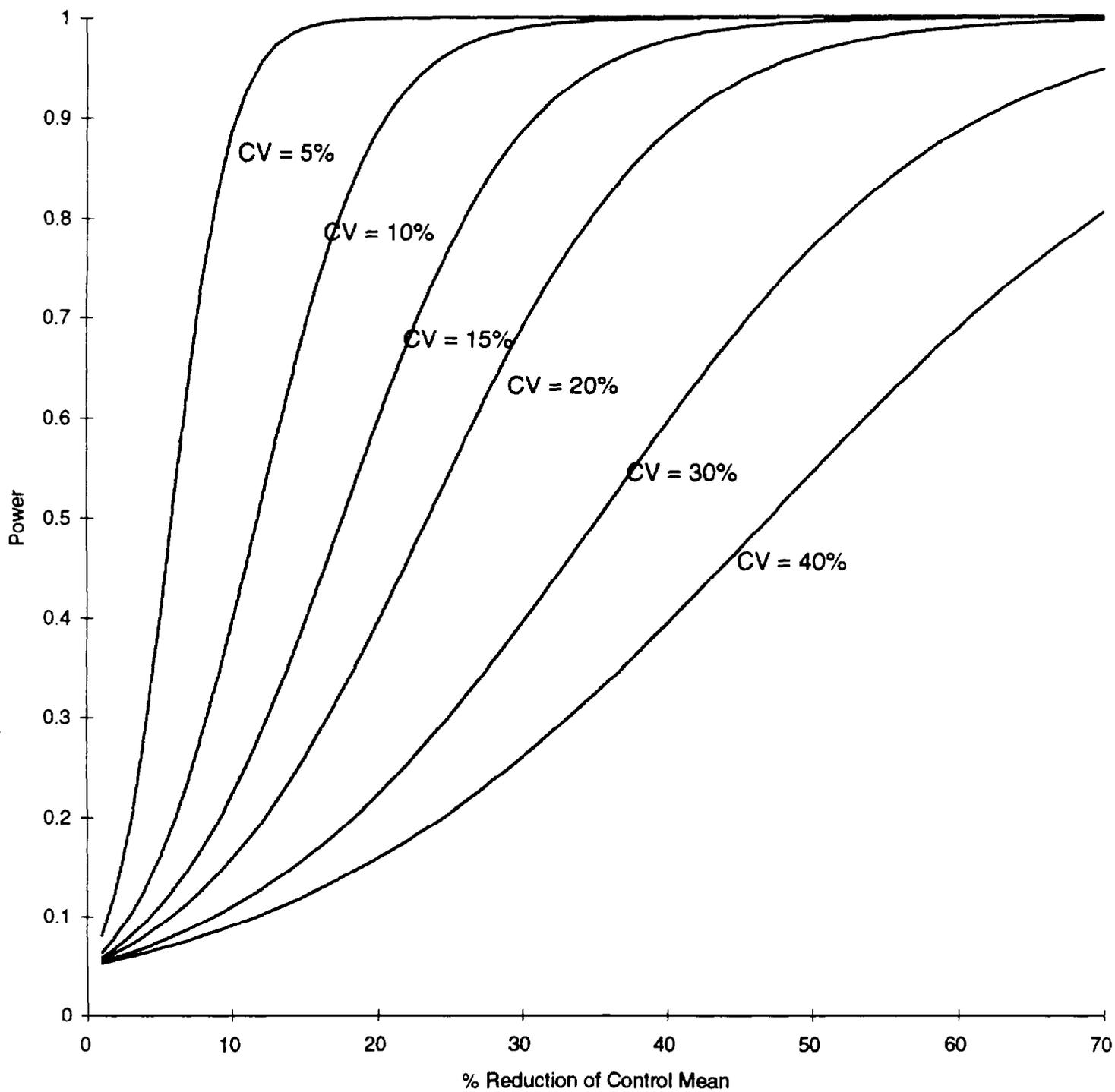
	<b>TR = Control</b>	<b>TR &gt; Control</b>
<b>DECISION</b>		
<b>TR = Control</b>	<p><b>Correct</b> <math>1 - \alpha</math></p>	<p><b>Type II Error</b> <math>\beta</math></p>
<b>TR &gt; Control</b>	<p><b>Type I Error</b> <math>\alpha</math></p>	<p><b>Correct</b> <math>1 - \beta</math> <b>(Power)</b></p>

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

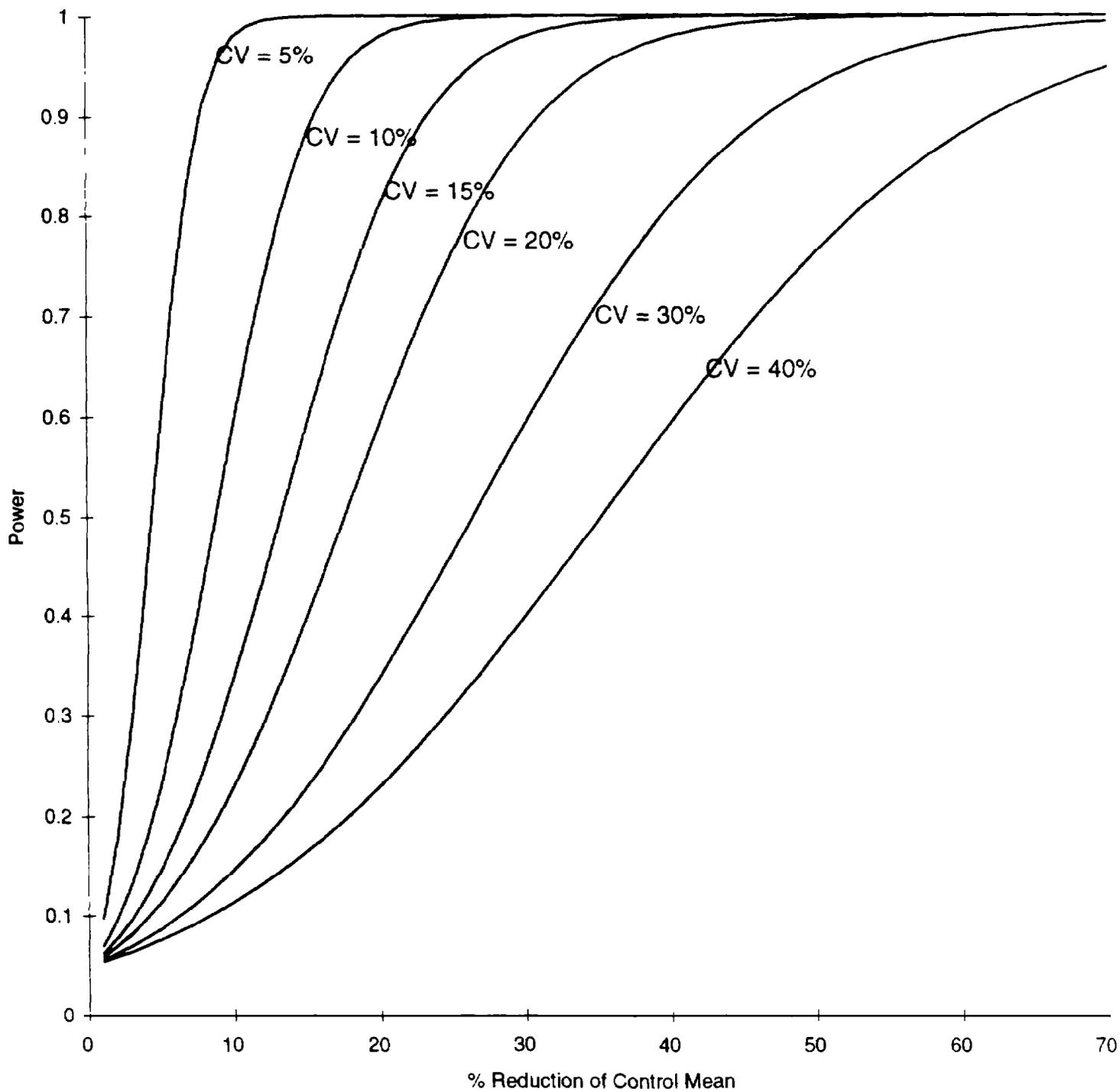
**Figure 12.1** Treatment response for a Type I and Type II error.



**Figure 12.2** Power of the test vs percent reduction in treatment response relative to the control mean at various CV's (8 replicates, alpha = 0.05 (one-tailed)).



**Figure 12.3** Power of the test vs percent reduction in treatment response relative to the control mean at various CV's (5 replicates, alpha = 0.05 (one-tailed)).



**Figure 12.4** Power of the test vs percent reduction in treatment response relative to the control mean at various CV's (8 replicates, alpha = 0.10 (one-tailed)).

12.2.3.8 The effect of the choice of  $\alpha$  and  $\beta$  on number of replicates for various CV's is illustrated in Figure 12.5 in which the combined total probability of Type I and Type II statistical errors is fixed and assumed to be 0.25. An  $\alpha$  of 0.10 therefore establishes a  $\beta$  of 0.15. In Figure 12.5, if  $\alpha = \beta = 0.125$ , the number of replicates required to detect a difference of 20% relative to the control is at a minimum. As  $\alpha$  or  $\beta$  decrease, the number of replicates required to detect the same 20% difference relative to the control increases. However, the curves are relatively flat over the range of 0.05 to 0.20 and that the curves are very dependent upon the choice of the combined total of  $\alpha + \beta$ . Limiting the total of  $\alpha + \beta$  to 0.10 greatly increases the number of replicates necessary to detect a pre-selected percentage reduction in mean treatment response relative to the control mean.

12.2.4 Figure 12.6 outlines a decision tree for analysis of survival and growth 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 non-parametric statistics text such as Conover (1980) may also be helpful.

12.2.4.1 *Mean.* The sample mean ( $\bar{x}$ ) is the average value, or  $\Sigma x/n$ , where:

$n$  = number of observations (replicates)

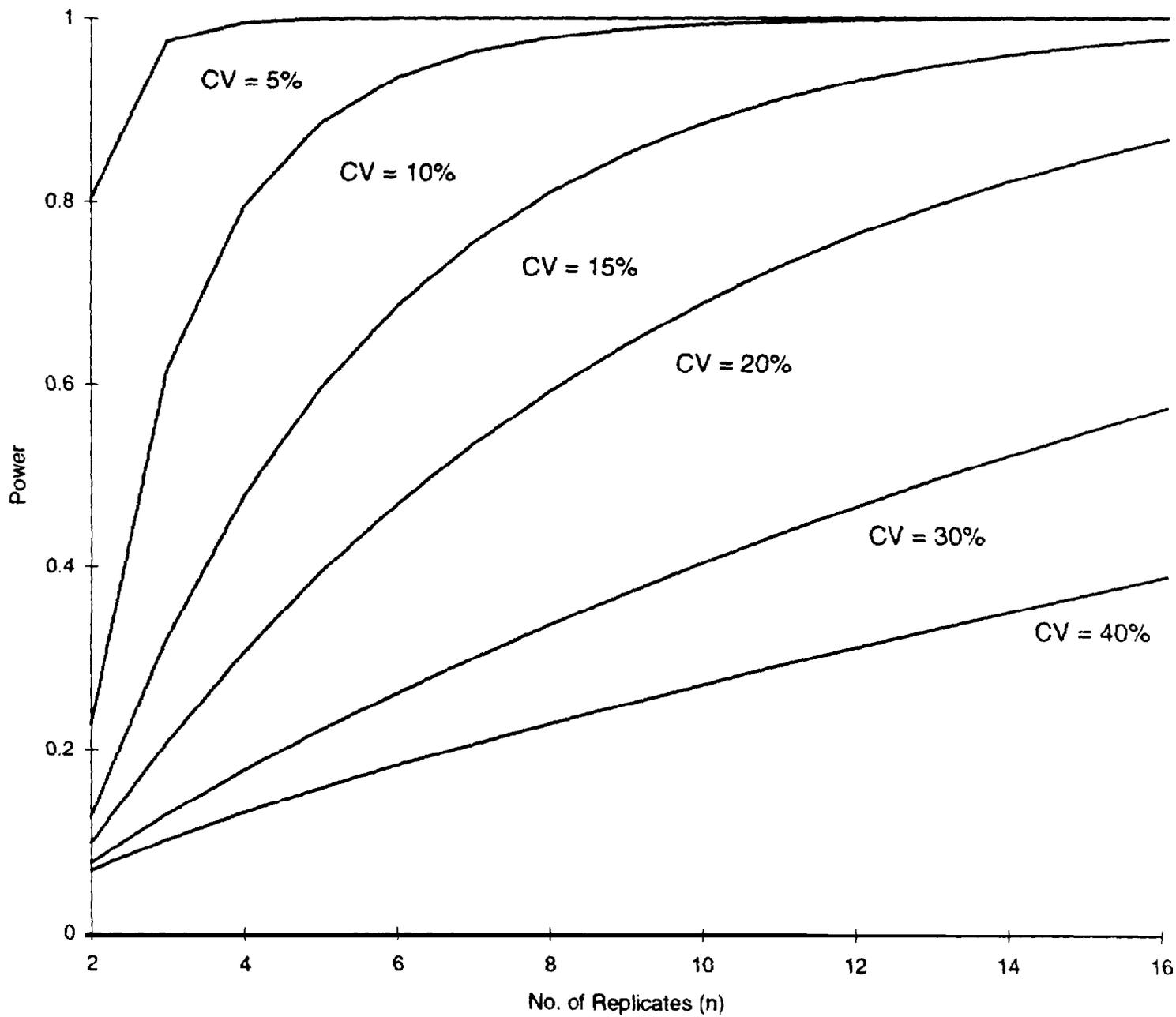
$x_i$  =  $i$ th observation

$\Sigma x_i$  = every  $x$  summed =  $x_1 + x_2 + x_3 + \dots + x_n$

12.2.4.2 *Standard Deviation.* 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:

$$s^2 = \frac{\Sigma x^2 - (\Sigma x)^2/n}{n - 1}$$

12.2.4.3 *Standard Error of the Mean.* 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 SD 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., FDA action level; Lee et al., 1994). 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.



**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 ( $\alpha = 0.05$  (One-tailed)).

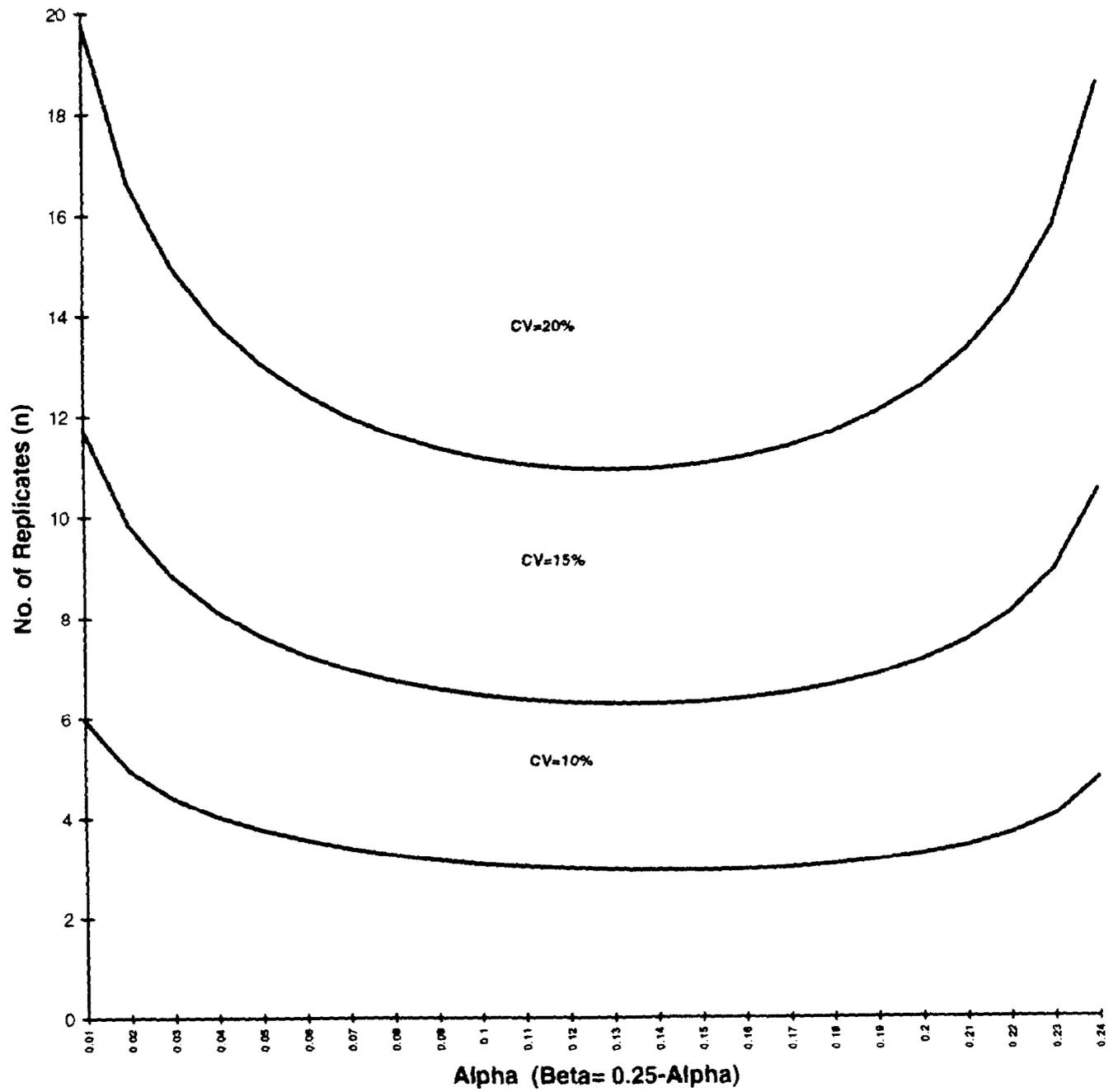


Figure 12.6 Effect of alpha and beta on the number of replicates at various CV's (assuming combined alpha + beta = 0.25).

**12.2.4.4 Tests of Assumptions.** 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 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.

**12.2.4.4.1 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, plotting, or by analysis of residuals. An explanation should be sought for any questionable data points. Without an explanation, data points should only be discarded with extreme caution. 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 non-parametric tests, but careful laboratory practices can reduce the frequency of outliers.

**12.2.4.4.2 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 if 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 normality. Because normality is desired, one looks for a high value of  $W$  with an associated probability greater than the pre-specified  $\alpha$  level.

**12.2.4.4.3** Table 12.2 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  $\alpha$  for the appropriate total 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 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  $\alpha$  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.

12.2.4.4.4 Tables of quantiles of W can be found in Shapiro and Wilk (1965), Gill (1978), Conover (1980), USEPA (1989) 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.

12.2.4.4.5 *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), and any one of these tests is adequate for the analyses in this section. Many software packages for t-tests and analysis of variance (ANOVA) provide at least one of the tests. Bartlett's Test is recommended for routine evaluation of homogeneity of variances (USEPA, 1985; USEPA, 1994b; USEPA, 1994c).

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

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

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).

12.2.4.4.7 Levels of  $\alpha$  for tests of equality of variances are provided in Table 12.2. These levels depend upon number of replicates in a treatment ( $n$ ) and allotment of replicates among treatments. Relatively high  $\alpha$ 's (i.e.,  $\geq 0.10$ ) are recommended because the 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  $\alpha$ .

12.2.4.4 *Transformations of the Data.* 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.

12.2.4.5.1 The arcsine-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:

$$RP = (\text{number of surviving organisms})/(\text{number exposed})$$

2. Transform each RP to arcsine, as follows.

a. For RPs greater than zero or less than one:

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

b. Modification of the arcsine when  $RP = 0$ .

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

where  $n$  = number animals/treatment replicate.

c. Modification of the arcsine when  $RP = 1.0$ .

$$\text{Angle} = 1.5708 \text{ radians} - (\text{radians for } RP = 0)$$

12.2.6.5 Two Samp'le Comparisons (N=2). The true population mean ( $\mu$ ) and standard deviation ( $\sigma$ ) 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  $\sigma$ . 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  $\mu$  and  $\sigma$ .

12.2.4.6.1 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 ( $H_0$ ) 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 control 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.

12.2.4.6.2 Since control organism mortality or tissue residues 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). This is a critical consideration when dealing with a small number of replicates (such as 8/treatment). The other alternative for increasing statistical power is to increase the number of replicates, which increases the cost of the test.

12.2.4.6.3 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 total organic carbon (TOC) content is different (greater or lesser) from the control sediment TOC.

12.2.4.6.4 The t-value for a one-tailed probability may 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  $\alpha = 0.05$  and  $df = 20$  is 1.725, and is found in a two-tailed table using the column for  $\alpha = 0.10$ .

12.2.4.7 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:

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

where  $s_1^2$  and  $s_2^2$  are the 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 t distribution with degrees of freedom (df) given by Satterthwaite's (1946) approximation:

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

This formula can result in fractional degrees of freedom, in which case one should round the degree 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; USEPA, 1994c). When variances are equal, an *F* test for equality is unnecessary.

**12.2.4.8 Nonparametric Tests.** 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 50th 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.

**12.2.4.8.1 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). 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:

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

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 Rohlf and Sokal (1981).

**12.2.4.8.2** 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).** Some experiments are set up to compare more than one treatment with a control while 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 if parametric statistics are appropriate and whether the variances of the treatments are equal. If normality of the data and equal variances are established, then an analysis of variance (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 may 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,  $r_{ik} = o_{ik} -$

( $k^{\text{th}}$  treatment mean). Pooling residuals provides an adequate sample size to test the data for normality.

**12.2.4.9.1** 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 may 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 which 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 may be tested against the other treatment means using the appropriate mean comparison. Equality of variances among the treatments can be evaluated with the Hartley  $F_{\text{max}}$  test or Bartlett's test. The option of using nonparametric statistics on the entire set of data is also an alternative.

**12.2.4.9.2** 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 Procedure, and may 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, 1993a). Wilcoxon's Rank Sum Tests 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, 1993a).

**12.2.4.9.3** Different mean comparison tests are used depending on whether an  $\alpha$  percent comparison-wise error rate or an  $\alpha$  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  $\alpha$  (than 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  $\alpha$  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 was based upon 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 to the control treatment.

12.2.4.9.4 The recommended procedure for pairwise comparisons that have a comparison-wise  $\alpha$  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  $\alpha$  error rate Dunnett's test is recommended for comparisons with the control.

12.2.4.9.5 Dunnett's test has an overall error rate of  $\alpha$ , 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. Dunnett's procedure can only be used when the same number of replicate test chambers have been used at each treatment and the control.

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

$$t_i = \frac{(\bar{Y}_i - \bar{Y}_c)}{S_w \sqrt{(1/n_i) + (1/n_c)}}$$

where  $\bar{Y}_i$  = Mean for each treatment

$\bar{Y}_c$  = Mean for the control

$S_w$  = Square root of the within mean square

$n_c$  = 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:

$$MSD = d S_w \sqrt{(1/n_i) + (1/n_c)}$$

where d = Critical value for the Dunnett's Procedure

$S_w$  = The square root of the within mean square

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

**n<sub>i</sub>** = 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 (USEPA, 1985; USEPA, 1989b; USEPA, 1994b; USEPA, 1994c) discuss in detail the mechanics of calculating LC50 (or EC50) or ICp values using the most current methods. The most commonly used methods are the Graphical, Probit, Trimmed Spearman-Kärber and the Linear Interpolation Methods. In general, results from these methods should yield similar estimates. Each method is outlined below and recommendations 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. A computer program to estimate the LC50 values and associated 95% confidence intervals with the methods discussed below (except for the Graphical Method) was developed by USEPA and can be obtained by sending a diskette with a written request to USEPA, Environmental Monitoring Systems Laboratory (EMSL), 26 W. Martin Luther King Drive, Cincinnati, OH 45268 or call 513/569-7076.

**12.2.5.4 The Graphical Method.** 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%.

**12.2.5.4.1** 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  $p_0, p_1, \dots, p_k$  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  $p_0 \leq p_1 \leq \dots \leq p_k$ . The smoothing process replaces any adjacent  $p_i$ 's that do not conform to  $p_0 \leq p_1 \leq \dots \leq p_k$  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  $i$ .

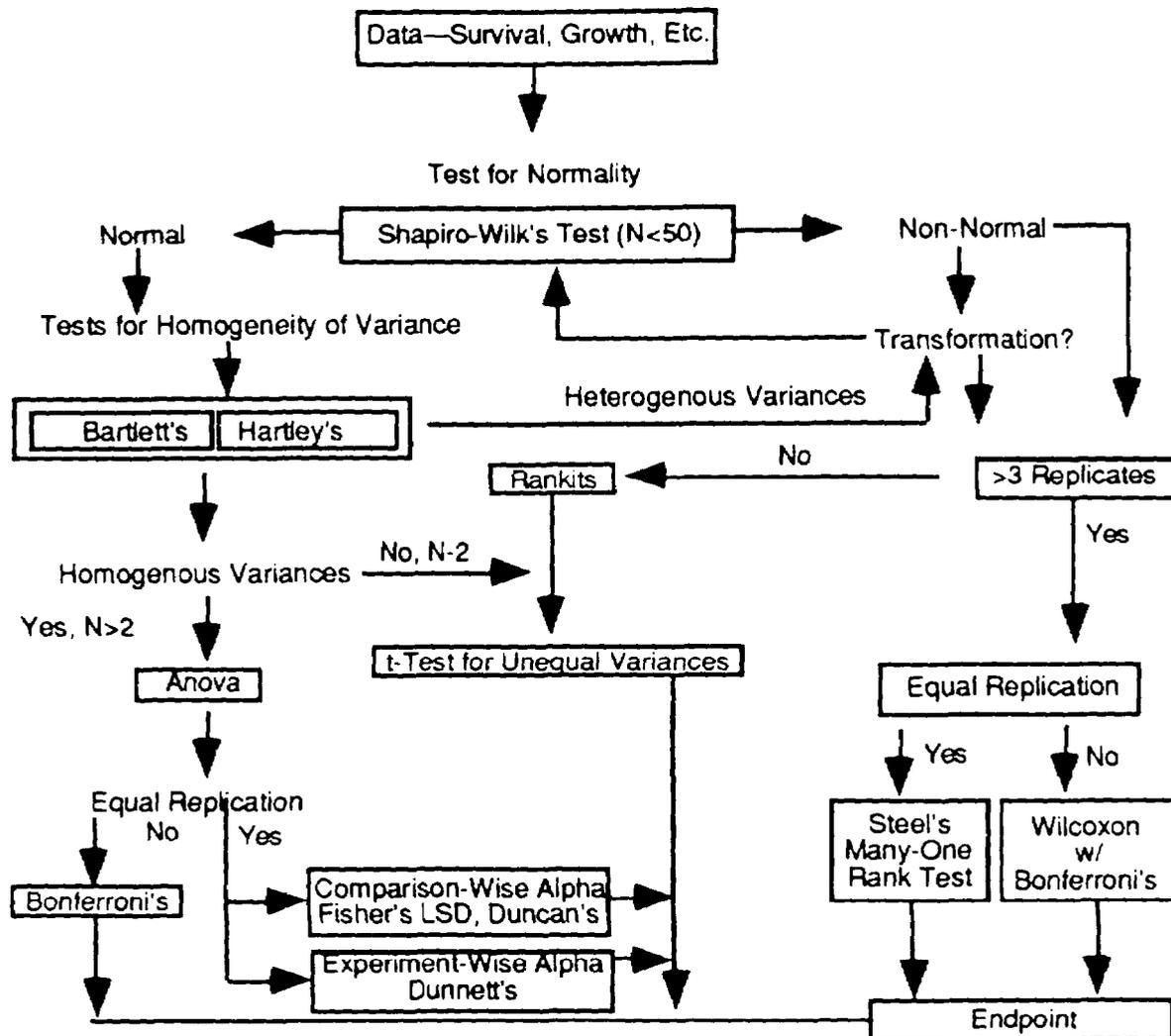


Figure 12.7 Decision tree for analysis survival and growth data subjected to hypothesis testing.

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:

$$p_i^a = (p_i^s - p_0^s)/(1 - p_0^s)$$

where  $p_0^s$  = 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.** The Probit Method is a parametric statistical procedure for estimating the LC50 (or EC50) and the associated 95% confidence interval (Finney, 1978). 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  $\approx .52$ . 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 mortalities should be corrected for control mortality using Abbott's formula before the Probit transformation is applied to the data.

**12.2.5.5.1** A goodness-of-fit procedure with the Chi-square statistic is used to determine if the data fit the Probit model. If many data sets are to be compared to one another, the probit method is not recommended because it may not be appropriate for many of the data sets. This method also is only appropriate for mortality data sets and should not be used for estimating endpoints that are a function of the control response, such as inhibition of growth. Most computer programs that generate probit estimates also generate confidence interval estimates for the LC50. These confidence interval estimates on the LC50 may not be correct if replicate mortalities are pooled to obtain a mean treatment response. 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-Kärber Method.** The Trimmed Spearman-Kärber Method is a modification of the Spearman-Kärber, non-parametric 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-Kärber Method is only appropriate when the requirements for the Probit Method are not met (USEPA, 1994b; USEPA, 1994c). This method is only appropriate for lethality data sets.

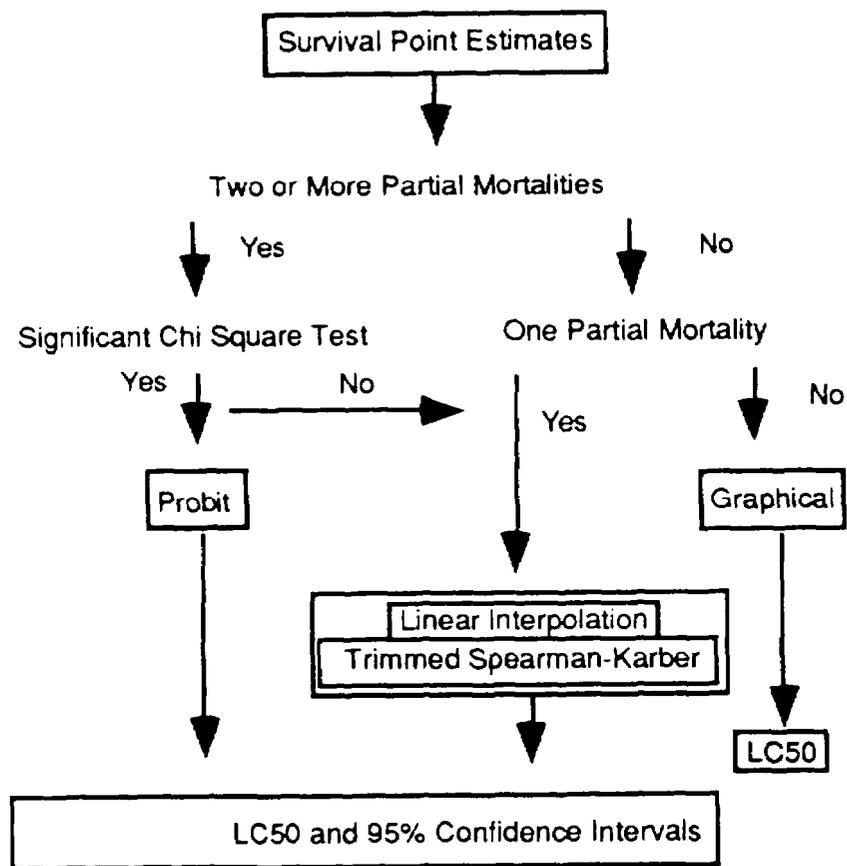


Figure 12.8 Decision tree for analysis of point estimate data.

**12.2.5.6.1** To calculate the LC50 estimate with the Trimmed Spearman-Kärber 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.

**12.2.5.6.2** 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). Calculate the amount of trim to use in the estimation of the LC50 as follows:

$$\text{Trim} = \max(p_1^a, 1 - p_k^a)$$

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.** This 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 IC<sub>p</sub> value (IC = Inhibition Concentration; where  $p$  = the percent effect). The procedure was designed for general applicability in the analysis of data from chronic toxicity tests, and 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, and a mean and coefficient of variation for the endpoints of multiple tests.

**12.2.5.7.1** As described in USEPA (USEPA, 1994b; USEPA, 1994c), the Linear Interpolation Method of calculating an IC<sub>p</sub> 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 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 IC<sub>p</sub> 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.

**12.2.5.7.2** 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.

12.2.5.7.3 If the assumption of monotonicity of test results is met, the observed response means ( $\bar{Y}_i$ ) 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 to 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 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 may require an additional step of smoothing. Where  $\bar{Y}_i$  decrease monotonically, the  $\bar{Y}_i$  become  $M_i$  without smoothing.

12.2.5.7.4 To obtain the IC<sub>p</sub> estimate, determine the concentrations  $C_J$  and  $C_{J+1}$  which 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:

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

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.

$IC_p$  = estimated concentration at which there is a percent reduction from the smoothed mean control response.

12.2.5.7.5 Standard statistical methods for calculating confidence intervals are not applicable for the IC<sub>p</sub>. The bootstrap method, as proposed by Efron (1982), is used to

obtain the 95% confidence interval for the true mean. In the bootstrap method, the test data  $Y_{ji}$  is 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  $IC_p$  ( $IC_p^*$ ). This process is repeated at least 80 times (Marcus and Holtzman, 1988) resulting in multiple "data" sets, each with an associated  $IC_p^*$  estimate. The distribution of the  $IC_p^*$  estimates derived from the sets of resampled data approximates the sampling distribution of the  $IC_p$  estimate. The standard error of the  $IC_p$  is estimated by the standard deviation of the individual  $IC_p^*$  estimates. Empirical confidence intervals are derived from the quantiles of the  $IC_p^*$  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  $IC_p^*$  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**

**14.3.1 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 (median lethal concentration), an EC50 (median effect concentration), an IC50 (inhibition concentration), or as an NOEC (no observed effect concentration) or LOEC (lowest observed effect concentration; Section 3). Consistent spiking procedures should be followed in order to make interlaboratory comparisons (Section 8.3).**

**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 sediments (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 non-ionic 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 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 may be useful for establishing effect concentrations.**

**12.3.3 Toxic units can be used to help interpret the response of organisms to multiple contaminants 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., 1991a). Toxicity expressed as toxic units may be summed and this may provide information on the toxicity of chemical mixtures (Ankley et al., 1991a).

**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. 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; Burton, 1991).

**12.3.6** Toxicity Identification Evaluation (TIE) procedures can be used to help provide insights as to specific contaminants responsible for toxicity in sediment (USEPA, 1991a; 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 or test sediment, 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 the 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, holding procedures.

12.4.1.6 Source and composition of food, concentrations of test material and other contaminants, procedure used to prepare food, 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, dissolved oxygen concentration (as percent saturation) 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 randomly selected test results. 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 which affect organism response. Determining the accuracy of a sediment test using field samples is not possible since the true values are not known. Since 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 (intralaboratory or repeatability; Section 13.5.1) precision or multi-laboratory (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 how reproducible a method is when conducted by a large number of laboratories using the same method, organism, and samples. Generally, intralaboratory results are less variable than interlaboratory results (USEPA, 1991b; USEPA, 1993a; USEPA, 1994b; USEPA, 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  $CV\% = \text{standard deviation}/\text{mean} \times 100$ ) 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 effect levels derived from statistical analyses of hypothesis testing. The CVs may 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 may 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 differences (MDD) compared to control survival or growth. Several factors can affect the precision of the test, including test organism age, condition, sensitivity, handling 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: (1) single laboratory precision determinations using reference toxicants for each of the test organisms which are used to determine the ability of the laboratory personnel to obtain precise results--these determinations should be made before conducting a sediment test and should be**

routinely performed as long as whole sediment tests are being conducted; (2) control charts (Section 13.3) should be prepared for each reference toxicant and test organism to determine if the test results are within prescribed limits; and (3) tests must meet the minimum criteria of test acceptability specific for each test organism (Table 11.3; USEPA, 1991b).

13.1.4 Intralaboratory precision data are routinely calculated for test organisms using water-only 96-h exposures to a reference toxicant, such as CdCl<sub>2</sub>. Intralaboratory precision data should be tracked using a control chart. Each laboratory's reference toxicant data will reflect conditions unique to that facility, including dilution water, culturing, and other variables (Section 9). However, each laboratory's reference toxicant CVs should reflect good repeatability.

13.1.5 To date, two interlaboratory precision (round-robin) tests have been completed using 10-d whole sediment tests, one with *Rhepoxynius abronius* (Mearns et al., 1986), and the other with *Ampelisca abdita*, *Eohaustorius estuarius*, and *Leptocheirus plumulosus* (C. Schlegel, SAIC, Narragansett, RI, unpublished data). The results of these round-robin study are described in Section 13.5.1.

## 13.2 Accuracy

13.2.1 The accuracy of toxicity tests cannot be determined since 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 probability levels (alpha and beta) selected, and the type of statistical analysis. For a given 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 It is the responsibility of a laboratory to demonstrate its ability to obtain precise results with reference toxicants before it performs sediment tests (Section 9.16). Intralaboratory precision, expressed as a coefficient of variation (CV), of the range for each type of test to be used in a laboratory should 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. This should be done to gain experience for the toxicity tests and a point of reference for future testing. A reference toxicant concentration series (0.5 or higher) should be selected that will consistently provide partial mortalities at two or more concentrations of the test chemical (Section 9.14, Table 9.1).

**13.4.2** The quality of test organisms obtained from an outside source, regardless of whether they are from culture or collected from the field, must be verified by conducting a reference-toxicity test concurrently with the sediment test. 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 water salinity and temperature at the time of collection, and collection site sediment for holding and acclimation purposes. If the supplier has not conducted five reference toxicity tests with the test organism, it is the responsibility of the testing laboratory to conduct these five reference toxicity tests before starting a sediment test (Section 13.4.2).

**13.4.3** Before conducting tests with contaminated sediment, the laboratory should demonstrate its ability to conduct tests by conducting five exposures in control sediment as outlined in Table 11.1. It is recommended that these five exposures with control sediment be conducted concurrently with the five reference toxicity tests described in Section 9.14.1.

**13.4.4** A control chart should be prepared for each combination of reference toxicant and test organism. 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 endpoint (LC50, NOEC, ICp) are examined to determine if they are within prescribed limits. Control charts as described in USEPA (1994a) and USEPA (1994b) are used to evaluate the cumulative trend of results from a series of samples. The mean and upper and lower control limits ( $\pm 2$  SD) are re-calculated with each successive test result. After two years of data collection, or a minimum of 20 data points, the control chart should be maintained using only the 20 most recent data points.

**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 alpha 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 total previous 20 fall outside the control limits, the sediment toxicity tests conducted during the that 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 toxicant test fall outside the expected ranges (Section 9). Specifically, a sediment test should not automatically 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.

**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 are suspect (USEPA, 1993a). 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  $\pm 2$  SD, by definition, will be exceeded 5% of the time, regardless of how well a laboratory performs. Highly proficient laboratories which develop a very narrow control limit may be unfairly penalized if a test which 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 which 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-Kärber Method, or 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 Sediment Toxicity Test Methods

### 13.5.1 Intralaboratory Precision

13.5.1.1 Intralaboratory precision has not been evaluated for any of the four species.

### 13.5.2 Interlaboratory Precision

13.5.2.1.1 Interlaboratory precision for *R. abronius* using 10-d whole sediment toxicity tests using the methods described in this manual (Table 11.1) is described by Mearns et al. (1986). Details of this study are described here. Five laboratories participated in the study, including federal and state government laboratories, a contract laboratory, and an academic laboratory. The laboratories were chosen because each had demonstrated experience in sediment toxicity tests with *R. abronius*. The experimental design required each laboratory to conduct 10-d whole sediment tests on a total of 7 sediment treatments. One control sediment was tested. Three sediment treatments consisted of control sediment that was amended with CdCl<sub>2</sub> to result in the following measured concentrations: 4, 8, and 12 mg Cd/kg dry weight. Three field-collected sediments were also used. They were collected from the following locations in Puget Sound, WA: Central Basin (Metro Seattle Station A600E), inner Sinclair Inlet, and Slip No. 1 in City Waterway, Commencement Bay.

13.5.2.1.2 Amphipods were collected from a depth of 6 m off West Beach, Whidbey Island, WA, and distributed to each participating laboratory. Each laboratory used its own source of clean seawater.

13.5.2.1.3 All five laboratories had >90% survival in control sediment, and thereby met the performance criteria for the test. Mean survival in control sediment was 96.4%, the CV was 3.7%, and the range was from 92 to 100% (Table 13.1). Of the cadmium-spiked sediments, survival was the least variable in the 4 mg/kg Cd treatment. Mean survival was 96.2%, the CV was 4.2%, and the range was from 89 to 98%. The most variable response was in the 12 mg/kg Cd sediment. Mean survival was 19%, the CV was 79.1%, and the range was from 6 to 41%. City Waterway showed the least variability among the field-collected sediments, with a mean survival of 83%, a CV of 6.4%, and a range from 74 to 87%. Sinclair Inlet showed the greatest variability among the field-collected sediments, with a mean survival of 78.8%, a CV of 11.3%, and a range from 67 to 88%.

13.5.2.2.1 Interlaboratory precision for *A. abdita*, *E. estuarius*, and *L. plumulosus* using 10-d whole sediment toxicity tests is described by C. Schlekat (SAIC, Narragansett, RI, unpublished data). Details of this study are described below. The number of participating laboratories varied with the test species: six for *A. abdita*, eight for *E. estuarius*, and seven for *L. plumulosus*. Laboratories were chosen on the basis of demonstrated experience with the particular test species. Each laboratory conducted 10-d sediment toxicity tests on 4 sediment treatments. Sediment treatments were selected for each species to include one negative control sediment and three contaminated sediments. Highly contaminated sediment from Black Rock Harbor, CT, was diluted with species-specific, non-contaminated control sediment, creating test sediments that ranged in relative contamination from low to high.

13.5.2.2.2 Independent suppliers distributed amphipods to each laboratory. *Ampelisca abdita* and *Eohaustorius estuarius* were field-collected from locations in Narragansett, RI, and Newport, OR, respectively. *Leptocheirus plumulosus* were obtained from cultures located at the University of Maryland, Queenstown, MD. Each laboratory used its own supply of clean seawater.

13.5.2.2.3 Mean survival of *A. abdita* in control sediment ranged from 85% to 100% (Table 13.2). Five of the six laboratories achieved greater than 90% survival in control sediment, which is the minimum survival that must be obtained in control sediment in order for the test to be accepted. The grand mean was 94.5%, and the CV was 5.5. A dose response was exhibited with decreasing survival with increasing proportions of BRH sediment. Test sediments (i.e., 7%, 25, and 33% BRH dilutions) exhibited a higher degree of variability than in control sediment. In 7% BRH sediment, mean survival ranged from 20% in Laboratory 5 to 97% in Laboratory 6 (Table 13.2). Twenty-percent BRH exhibited the greatest magnitude of variability, with a range of 1% to 90%. Thirty-three percent BRH also exhibited considerable variability. The overall rank of sediment toxicity as measured by absolute mortality was consistent among laboratories. One hundred percent of laboratories were in agreement for in

**Table 13.1 Inter-laboratory precision for survival of *Rhepoxynius abronius* in 10-d whole sediment toxicity tests using seven sediments**

Percent Survival (SD) in Sediment Samples							
Lab	Control	4 mg/kg Cd	8 mg/kg Cd	12 mg/kg Cd	Central Basin	Sinclair Inlet	City Waterway
1	92 (7)	89 (7)	87 (9)	8 (3)	83 (11.5)	78 (13)	74 (11.5)
2	96 (4)	98 (3)	90 (10)	41 (11)	69 (7.5)	67 (11)	87 (12)
3	100 (0)	97 (3)	78 (10.5)	12 (7.5)	90 (8)	87 (7.5)	83 (12.5)
4	94 (7)	99 (2)	50 (15)	6 (5.5)	92 (5.5)	88 (3)	84 (11)
5	100 (0)	98 (4.5)	77 (3)	28 (11.5)	80 (3.5)	74 (9)	87 (3)
Mean	96.4 (3.6)	96.2 (4.1)	76.4 (15.8)	19 (15.5)	82.8 (9.1)	78.8 (8.9)	83 (5.3)
CV (%)	3.7	4.2	20.7	79.1	11.0	11.3	6.4

Note: From Mearns et al., 1986

**Table 13.2 Inter-laboratory precision for survival of *Ampelisca abdita* in 10-d whole sediment toxicity tests using four sediments**

<b>Percent Survival (SD) in Sediment Samples</b>				
<b>Lab</b>	<b>Control</b>	<b>7% Black Rock Harbor</b>	<b>20% Black Rock Harbor</b>	<b>33% Black Rock Harbor</b>
1	97.0 (4.5)	63.0 (19.6)	10.0 (7.9)	6.0 (4.2)
2	94.0 (8.9)	75.0 (6.1)	7.0 (4.5)	0.0 (0)
3	97.0 (4.5)	90.0 (3.5)	36.0 (9.6)	38.0 (14.4)
4	94.0 (8.9)	79.0 (17.8)	7.0 (4.5)	3.0 (6.7)
5	85.0 (7.1)	20.0 (12.7)	1.0 (2.2)	1.0 (2.2)
6	100.0 (0)	97.0 (4.5)	90.0 (5.0)	72.0 (13.0)
<b>Mean</b>	<b>94.5 (5.2)</b>	<b>70.7 (13.0)</b>	<b>25.2 (34.0)</b>	<b>20.0 (29.2)</b>
<b>CV (%)</b>	<b>5.5</b>	<b>38.9</b>	<b>135.1</b>	<b>146.2</b>

Note: From: C. Schlekat et al., SAIC, Narragansett, RI, unpublished data.

ranking control and 7% BRH sediments as the first and second least toxic sediments, respectively (Table 13.2).

13.5.2.2.4 Every laboratory surpassed the minimum survival criteria of 90% survival in control sediment with *E. estuarius*. The range was from 96 to 100%, with a Grand Mean of 98.2% and a CV of 1.5 (Table 13.3). Grand Mean survival decreased with increasing proportions of BRH. BRH sediment dilutions exhibited greater variability than control sediment, with 25% BRH displaying the highest coefficient of variation. All eight laboratories ranked survival of *E. estuarius* for control and 9% BRH as the least and second least toxic, respectively (Table 13.2). With the exception of Laboratories 1 and 8, the rank for 25% and 42% BRH were appropriately third and fourth least toxic, respectively.

13.5.2.2.5 *Leptocheirus plumulosus* exhibited a range of survival in control sediment from 86% to 99% (Table 13.4). The Grand Mean was 91.8%, and the CV was 4.7. Two laboratories, 3 and 5, failed to meet the minimum control sediment survival criteria of 90%. Grand means displayed a dose response of decreasing survival with increasing proportion of BRH sediment. Coefficients of variation were uniformly higher in BRH sediment dilutions as compared to control sediment, but did not vary greatly among BRH sediments (Table 13.4). Laboratory 1 appeared to be an outlier with respect to survival in BRH sediment dilutions, as survival of *L. plumulosus* was the lowest for all three BRH sediments for any laboratory. The rank of sediments according to their toxicity was generally consistent among laboratories. Agreement was 100% for control and the highest BRH sediment; these were appropriately ranked 1 and 4, respectively (Table 13.4). Laboratories 4 and 5 anomalously ranked 10% and 28% BRH as 3 and 2, respectively, whereas the remaining laboratories ranked these sediments appropriately according to the proportion of BRH.

13.5.2.3 These tests exhibited similar or better precision than many chemical analyses and effluent toxicity test methods (USEPA, 1991b). The success rate for test initiation and completion of this round-robin evaluation is a good indication that a well equipped and trained staff will be able to successfully conduct this test. This is an important consideration for any test performed routinely in any regulatory program.

**Table 13.3 Interlaboratory precision for survival of *Eohaustorius estuarius* in 10-d whole sediment toxicity tests using four sediments**

Lab	Percent Survival (SD) in Sediment Samples			
	Control	9% Black Rock Harbor	25% Black Rock Harbor	42% Black Rock Harbor
1	96.0 (6.5)	45.0 (19.7)	6.0 (6.5)	16.0 (9.6)
2	98.0 (2.7)	76.0 (10.8)	46.0 (13.9)	25.0 (7.1)
3	97.0 (2.7)	89.0 (4.2)	59.0 (10.8)	45.0 (10.0)
4	98.8 (2.7)	59.0 (23.0)	47.2 (23.2)	45.8 (27.0)
5	100.0 (0)	75.0 (19.7)	36.0 (12.4)	16.0 (9.6)
6	100.0 (0)	69.0 (12.9)	56.0 (18.8)	38.0 (14.4)
7	99.0 (2.2)	79.0 (6.5)	61.0 (10.8)	50.0 (7.9)
8	97.0 (6.7)	53.0 (14.4)	24.0 (14.7)	29.0 (15.6)
Mean	98.2 (1.5)	68.1 (14.7)	41.9 (19.1)	33.1 (13.5)
CV (%)	1.5	21.6	45.5	40.9

Note: From: C. Schlekat et al., SAIC, Narragansett, RI, unpublished data.

**Table 13.4 Inter-laboratory precision for survival of *Leptocheirus plumulosus* in 10-d whole sediment toxicity tests using four sediments**

Lab	Percent Survival (SD) in Sediment Samples			
	Control	10% Black Rock Harbor	28% Black Rock Harbor	47% Black Rock Harbor
1	91.3 (4.8)	6.0 (4.2)	5.0 (3.5)	2.5 (2.9)
2	91.0 (8.9)	62.0 (11.0)	51.0 (15.6)	33.0 (11.5)
3	88.0 (8.4)	34.0 (15.2)	22.0 (13.0)	7.0 (5.7)
4	92.0 (7.6)	48.0 (23.9)	59.0 (21.6)	27.0 (10.4)
5	86.0 (10.2)	20.0 (9.4)	28.0 (4.5)	12.0 (9.1)
6	95.0 (6.1)	76.0 (10.2)	65.0 (14.6)	38.0 (17.5)
7	99.0 (2.2)	78.0 (13.0)	56.0 (4.2)	26.0 (6.5)
Mean	91.8 (4.3)	46.3 (27.7)	40.9 (22.6)	20.8 (13.6)
CV (%)	4.7	59.8	55.2	65.5

Note: From: C. Schlekat et al., SAIC, Narragansett, RI, unpublished data.

## 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., Philadelphia, PA: American Society for Testing and Materials, pp. 429-453.
- ASTM. 1988. Standard guide for conducting acute toxicity tests with fishes, macroinvertebrates, and amphipods. ASTM 1993 Annual Book of Standards Vol. 11.04, E729-88, Philadelphia, PA.
- ASTM. 1992. Guide for conducting 10-day static sediment toxicity tests with marine and estuarine amphipods. ASTM 1992 Annual Book of Standards Vol. 13.04, E1367-92, Philadelphia, PA.
- ASTM. 1993a. Standard guide for designing biological tests with sediments. ASTM 1994 Annual Book of Standards Vol. 11.04, E1525-93, Philadelphia, PA.
- ASTM. 1993b. Standard terminology relating to biological effects and environmental fate. ASTM 1994 Annual Book of Standards Vol. 11.04, E943-93, Philadelphia, PA.
- ASTM. 1994a. Standard guide for collection, storage, characterization, and manipulation of sediments for toxicological testing. ASTM 1993 Annual Book of Standards Vol. 11.04, E1391-90, Philadelphia, PA.
- ASTM. 1994b. Draft standard guide for the use of light in laboratory testing. Under development by B.M. Greenburg, Department of Biology, University of Waterloo, Waterloo, Ontario. ASTM Subcommittee E47.11, January, 1994.
- Ankley, G.T., M.K. Schubauer-Berigan, and R.A. Hoke. 1992. Use of toxicity identification evaluation techniques to identify dredged material disposal options: A proposed approach. *Environ. Management*. 16:1-6.
- Ankley, G. and Thomas, N. 1992. Interstitial water toxicity identification evaluation approach. In: Sediment classification methods compendium, pp. 5-1 to 5-14. EPA-823/R-92-006. Washington, DC.
- 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. and M. K. Schubauer-Berigan. 1994. Comparison of techniques for the isolation of pore water for sediment toxicity testing. *Arch. Environ. Contam. Toxicol.*: In press.

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

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

Barnard, J.L., and C.M. Barnard. The genus *Rhepoxynius* (Crustacea: Amphipoda: Phoxocephalidae) in American Seas. *Smithsonian Contributions to Zoology*, 357:1-49.

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

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* (Conrad): A mass balance approach. *Environ. Toxicol. Chem.* 9:221-231.

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

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. 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.

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.

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. Ann Arbor, MI: Lewis Publishers. pp. 313-340.

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

Chapman, P.M., R.N. Dexter, and E.R. Long. 1987. Synoptic measurements of sediment contamination, toxicity, and infaunal community composition (the Sediment Quality Triad) in San Francisco Bay. *Marine Ecology Progress Series* 37:75-96.

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

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 13<sup>th</sup> Annual Meeting, Society of Environmental Toxicology and Chemistry, Cincinnati, OH, Nov. 8-12.

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

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.

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, W.A. 1984. *Fate and effects of sediment-bound chemicals in aquatic systems*. 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.

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. 678 p.

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

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.

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.

Food and Drug Administration. 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.

Green, R.H. 1979. Sampling design and statistical methods for environmental biologists. New York, NY: Wiley Interscience. 257 p.

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

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

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

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., A.M. Haddon, and L.A. Wiechert. 1990. Guild structure and foraging impact of blue crabs and epibenthic fish in a subestuary of Chesapeake Bay. *Mar. Ecol. Prog. Ser.*, 67:105-126.
- 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.
- Hopkins, D.R. 1986. Atlas of the distributions and abundances of common benthic species in San Francisco Bay, California. Water Resources Investigations Report 86-4003. Menlo Park, CA: U.S. Geological Survey. 16 pp.
- Hyland, J.L. 1981. Comparative structure and response to (petroleum) disturbance in two nearshore infaunal communities. PhD. Thesis, University of Rhode Island, 141p.
- 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. and M.K. Nelson. 1990. Testing sediment toxicity with *Hyaella azteca* (Amphipoda) and *Chironomus riparius* (Diptera). In: W.G. Landis and W.H. van der Schalie (eds.), Aquatic toxicology and risk assessment, 13th volume. ASTM STP 1096. Philadelphia, PA, pp. 93-109, 1990.
- 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, C.G., 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.*, in press.
- Kemp, P.F., F.A. Code, and R.C. Swartz. 1985. Life history and production of the Phoxocephalid amphipod *Rhepoxynius abronius* (Barnard). *J. Crustacean Biology* 5:449-464.
- 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.

Lamberson, J.O., T.H. DeWitt, and R.C. Swartz. 1992. Assessment of sediment toxicity to marine benthos. In: Sediment toxicity assessment, G.A. Burton, Jr., Ed. Ann Arbor, MI: Lewis Publishers. pp. 183-211.

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. New York, NY: John Wiley and Sons. pp. 257-279.

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, H., II, 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, 232 p.

Lee, D.R. 1980. Reference toxicants in quality control of aquatic bioassays. In: A.L. Buikema and J. Cairns Jr. (eds.), Aquatic invertebrate bioassays. ASTM STP 715, 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).

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.

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. and M.F. Buchman. 1989. An evaluation of candidate measures of biological effects for the National Status and Trends Program. NOAA Technical Memorandum NOS OMA 45, Seattle, WA. 105 pp.

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, DC, 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. Schlekot, D.M. Boward, and T.L. Wade. 1994. Sediment contamination and biological effects in a Chesapeake Bay marina. *Ecotoxicology*. In Press.

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

McLeay, D.J., S. Yee, K. Doe, and S. Wade. 1991. Phase-II and Phase-III studies by Environment Canada laboratories of 10-day tests for sediment toxicity using marine or estuarine infaunal amphipods. Report prepared for Environment Canada (EP, Conservation and Protection) and IGATG by McLeay Associates Ltd., West Vancouver, B.C.

Mearns, A.J., R.C. Swartz, J.M. Cummins, P.A. Dinnel, P. Plesha, and P.M. Chapman. 1986. Inter-laboratory comparison of a sediment toxicity test using the marine amphipod, *Rhepoxynius abronius*. *Mar. Environ. Res.* 19:13-37.

Mills, E.L. 1967. The biology of an amphipod crustacean sibling species pair. *J. Fisheries Res. Board Canada*, 24:305-355.

National Research Council (NRC). 1989. Contaminated marine sediments--Assessment and remediation, NRC, National Academy Press, Washington, DC.

Nelson, W.G. 1980. Reproductive patterns of gammaridean amphipods. *Sarsia*, 65:61-71.

Nichols, F.H., and J.K. Thompson. 1985. Persistence of and introduced mudflat community in south San Francisco Bay, California. *Mar. Ecol. Prog. Ser.*, 24:83-97.

Nkedi-Kizza, P., P.S.C. Rao, and A.G. Hornsby. 1985. Influence of organic cosolvents on sorption of hydrophobic organic chemicals by soils. *Environ. Sci Tech.*, 19:975-979.

Oakden, J.M. 1984. Feeding and sediment selection in five species of Central Californian phoxocephalid amphipods. *J. Crustacean Biol.* 4:233-247.

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

Pastorak, R.A., and D.S. Becker. 1990. Comparative sensitivity of sediment bioassays at three Superfund sites in Puget Sound. In: W.G. Landis and W.H. van der Schalie, Eds., Aquatic toxicology and risk assessment: Thirteenth volume, STP 1096, Philadelphia, PA: American Society for Testing and Materials. pp. 123-139.

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. 478 p.

Ramsdell, K.A., J.A. Strand, and V.I. Cullinan. 1989. Amphipod bioassay of selected sediments from Sequim Bay, Washington. In: Oceans 1989, Volume 2: Ocean Pollution. IEEE Publication No. 89CH2780-5, pp. 443-448.

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: D.F. Soule and G.S. Kleppel (eds.), Marine organisms as indicators. Springer-Verlag, New York. pp. 213-245.

Richards, S.W. The demersal fish populations of Long Island Sound: III Food for juveniles from a mud locality. *Bull. Bingham Oceanogr. Coll.*, 18:32-101, 1963.

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.

- Satterthwaite, F.W. 1946. An approximate distribution of estimates of variance components. *Biom. Bull.* 2:110-114.
- Schaeffer, D.J. and K.G. Janardan. 1978. Theoretical comparison 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., B.L. McGee, D.M. Bowan, R. Reinharz, D.J. Velinsky and T.L. Wade. 1994. Tidal river sediments in the Washington, DC area. III. Biological effects associated with sediment contamination. *Estuaries* 17: In press.
- 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, 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: 12'th Volume, ASTM STP 1027. U.M. Cowgill and L.R. Williams, Eds. Philadelphia, PA: American Society for Testing and Materials. pp. 289-303.
- 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, DC. pp. 132-154.
- 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, DC.
- 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.

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. Ann Arbor, MI: Lewis Publishers. pp. 341-370.

Steel, R.G.D., and J.H. Torrie. 1960. Principles and procedures of statistics. New York, NY: McGraw Hill Book Co.

Stein, J.R. 1973. Handbook of psycholical methods: Culture methods and growth measurements. New York, NY: Cambridge University Press. 445 pp.

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

Swartz, R.C., Code, F.A., Lamberson, J.O, Ferraro, S.P., Schults, D.W., DeBen, W.A., Lee II, H., and Ozretich, R.J. 1994. Sediment toxicity, contamination, and amphipod abundance at a DDT and dieldrin-contaminated site in San Francisco Bay. *Environ. Toxicol. Chem.* 13:949-962.

Swartz, R.C., W.A. DeBen, K.A. Secru, and J.O. Lamberson. 1982. Sediment toxicity and the distribution of amphipods in Commencement Bay, Washington, USA. *Mar. Poll. Bull.* 13:359-364.

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., Philadelphia, PA: American Society for Testing and Materials. 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.

Swartz, R.C. 1989. Marine sediment toxicity tests. In: Contaminated Marine Sediments - Assessments and Remediation, Washington, D.C.: National Academy Press. pp. 115-129.

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 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. 1986a. Quality criteria for water. EPA-440/5-86-001, Washington, DC.**

**USEPA. 1986b. Occupational health and safety manual. Office of Administration, EPA, Washington, DC.**

**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, DC.**

**USEPA. 1989b. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to freshwater organisms. EPA-600/4-89/001, Cincinnati, OH.**

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

**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, DC.**

**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, DC.**

**USEPA. 1990c. Analytical procedures and quality assurance plan for determination of xenobiotic chemical contaminants in fish. EPA-600/3-90/023, Duluth, MN.**

**USEPA. 1990d. Analytical procedures and quality assurance plan for determination of PCDD/PCDF in fish. EPA-600/3-90/022, Duluth, MN.**

USEPA. 1991a. 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. 1991b. Technical support document for water quality-based toxic control. EPA-505/2-90/001, Washington, DC.

USEPA. 1992a. Proceedings from workshop on tiered testing issues for freshwater and marine sediments. Washington, DC, 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, DC.

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

USEPA. 1993a. Methods for measuring the acute toxicity of effluents to freshwater and marine organisms. Fourth edition. EPA 600/4-90/027F, Cincinnati, OH, 1993a.

USEPA. 1994a. Procedures for conducting sediment toxicity and bioaccumulation tests using freshwater invertebrates. EPA-600/R-94/024. June 1994. USEPA, Office of Research and Development, Duluth, MN.

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

USEPA-USCOE (U.S. Army Corps of Engineers). 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-USCOE (U.S. Army Corps of Engineers). 1994. Evaluation of dredged material proposed for discharge in inland and near coastal waters. EPA- 000/0-93/000, Washington, DC.

USGS (U.S. Geological Survey). 1969. Techniques of water-resources investigations of the USGS. Chapter C1, Harold P. Guy, p. 58, Laboratory Theory and Methods for Sediment Analysis; Book 5, Laboratory Analysis, U.S.G.S. Arlington, VA

Walters, D.B. and C.W. Jameson. 1984. Health and safety for toxicity testing. Butterworth Publications, Woburn, MA.

Weisberg, S.B., J.B. Frithsen, A.F. Holland, J.F. Paul, K.J. Scott, J.K. Summers, H.T. Wilson, R. Valente, D.G. Heimbuch, J. Gerritsen, S.C. Schimmel, and R.W. Latimer. 1992. EMAP-Estuaries Virginian Province 1990 Demonstration Project Report. EPA 600/R-92/100. USEPA, Environmental Research Laboratory, Narragansett, RI.

Wiederholm, T., A.M. Wiederholm, and M. Goran. 1987. Fresh-water oligochaetes. *Water Air Soil Pollut.* 36:131-154.

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.Q., J.A. Ward, B. Brown, B.D. Walls, and S. Lemlich. 1989. Relative sensitivity and cost of amphipod bioassays. In: Oceans '89 Volume 2: Ocean Pollution. IEEE Publication No. 89CH2780-5. pp. 467-473.

Zar, J.H. 1984. *Biostatistical Analysis.* Second edition. Prentice-Hall, Inc., Englewood Cliffs, NJ. 717 p.

## **APPENDIX A**

### **Sample Data Sheets**



96hr Reference Toxicant Test--Daily Data Sheet

Species: \_\_\_\_\_  
 Toxicant: \_\_\_\_\_

Experiment #: \_\_\_\_\_  
 Test Start Date: \_\_\_\_\_

Set up and Daily Survival Checks

Test Level mg/L	Rep	Set-up Carboy #		Renewal Carboy #		24hr		48hr		72hr		96hr	
		date:		date:		date:		date:		date:		date:	
		Total No. (live+dead)	Today's No. Live										
Control 0 mg/L	A												
	B												
	A												
	B												
	A												
	B												
	A												
	B												
	A												
	B												
Temp °C													
Observer													

Physical Data

Test Level mg/L	Rep	date:			date:		
		pH	DO ppm	Salinity ppt	pH	DO ppm	Salinity ppt
Control 0 mg/L	A						
	B						
	A						
	B						
	A						
	B						
	A						
	B						
	A						
	B						
Observer							

Comments

Figure A.2 Data sheet for 96 h reference toxicant test.

10 Day Solid Phase Test--Daily Data Sheet

Project \_\_\_\_\_

Experiment #: \_\_\_\_\_

Day \_\_\_\_\_ Date: \_\_\_\_\_

Organism: \_\_\_\_\_

Time/Initials: \_\_\_\_\_

Jar #	Observations*				Cum # Dead	Jar #	Observations*				Cum # Dead	Jar #	Observations*				Cum # Dead
	E	M	NMT	D			E	M	NMT	D			E	M	NMT	D	
1						31						61					
2						32						62					
3						33						63					
4						34						64					
5						35						65					
6						36						66					
7						37						67					
8						38						68					
9						39						69					
10						40						70					
11						41						71					
12						42						72					
13						43						73					
14						44						74					
15						45						75					
16						46						76					
17						47						77					
18						48						78					
19						49						79					
20						50						80					
21						51						81					
22						52						82					
23						53						83					
24						54						84					
25						55						85					
26						56											
27						57											
28						58											
29						59											
30						60											

Animals/rep. \_\_\_\_\_  
 Temp \_\_\_\_\_  
 Thermometer #. \_\_\_\_\_  
 Previous day's Cumulative  
 number dead Q&A'd by \_\_\_\_\_

\*E=emerged M=molt NMT=neuromuscular twitch D=dead  
 Comments: \_\_\_\_\_

Figure A.3 Data sheet for daily observations during the 10-d solid phase test.



10 Day Solid Phase Test--Breakdown Data Sheet

Project: \_\_\_\_\_ Experiment #: \_\_\_\_\_  
 Species: \_\_\_\_\_ Date: \_\_\_\_\_

First Pick							QA - RePick				Final Count				
							Recount					Recount			
Initials	Time	Jar #	Dead During Test	Day 10 Emerged	# Live	# Dead	Initials	# Live	# Animals Missing **	Initials	# Live	Initials	# Live	Final # Live ***	
		1													
		2													
		3													
		4													
		5													
		6													
		7													
		8													
		9													
		10													
		11													
		12													
		13													
		14													
		15													
		16													
		17													
		18													
		19													
		20													
		21													
		22													
		23													
		24													
		25													
		26													
		27													
		28													
		29													
		30													

Comments: \* # Live = the # live emerged on day 10 + the # live found by the picker  
 \*\* If >10% of the animals are missing (ie. >2 of 20), the sample must be QA'd  
 \*\*\* Final # Live = the # live from the first pick + recount + the # live from the QA recount

Figure A.5 Data sheet for test summary.

United States  
Environmental Protection Agency  
Center for Environmental Research Information  
Cincinnati, OH 45268

Official Business  
Penalty for Private Use  
\$300

EPA/600/R-94/025

Please make all necessary changes on the below label,  
detach or copy, and return to the address in the upper  
left-hand corner.

If you do not wish to receive these reports CHECK HERE   
detach, or copy this cover, and return to the address in the  
upper left-hand corner.

BULK RATE  
POSTAGE & FEES PAID  
EPA  
PERMIT No. G-35