United States
Environmental Protection
Agency

Office of Wastewater
Management
(4203)

EPA 833-R-00-003
June 2000
http:\\www.epa.gov\owm\npdes.htm

Understanding and Accounting for Method Variability in Whole Effluent Toxicity Applications Under the National Pollutant Discharge Elimination System Program

Understanding and Accounting for Method Variability in Whole Effluent Toxicity Applications Under the National Pollutant Discharge Elimination System Program

June 30, 2000

This page intentionally left blank.

NOTICE AND DISCLAIMER

This document provides guidance to NPDES regulatory authorities and persons interested in whole effluent toxicity testing. This document describes what EPA believes to be sources of variability in the conduct of whole effluent toxicity testing under the Clean Water Act. The document is designed to reflect national policy on these issues. The document does not, however, substitute for the Clean Water Act, an NPDES permit, or EPA or State regulations applicable to permits or whole effluent toxicity testing; nor is this document a permit or a regulation itself. The document does not and cannot impose any legally binding requirements on EPA, States, NPDES permittees, and/or laboratories conducting whole effluent toxicity testing for permittees (or for States in the evaluation of ambient water quality). EPA and State officials retain discretion to adopt approaches on a case-by-case basis that differ from this guidance based on an analysis of site-specific circumstances. This guidance may be revised without public notice to reflect changes in EPA policy.

This page intentionally left blank.

TABLE OF CONTENTS

Acknowledgments ix
Executive Summary xi
List of Acronyms and Abbreviations xv
Glossary xvii
1.0 INTRODUCTION 1-1
1.1 Background 1-1
1.2 Effect of This Guidance 1-2
1.3 Three Goals of This Document 1-2

2.0 DEFINITION AND MEASUREMENT OF METHOD VARIABILITY IN WET TESTING 2-1
2.1 Terms and Definitions 2-1
2.2 Defining WET Test Variability 2-1
2.3 Quantifying WET Test Variability 2-2

3.0 VARIABILITY OF WET TEST METHODS 3-1
3.1 Acquisition, Selection, and Quality Assurance of Data Presented in this Document 3-1
3.2 Variability of EC25, LC50, and NOEC 3-2
3.2.1 Within-Laboratory Variability of EC25, LC50, and NOEC 3-2
3.2.2 Between-Laboratory Variability of EC25, LC50, and NOEC 3-7
3.3 Variability of Endpoint Measurements 3-8
3.4 Conclusions about Variability of WET Methods 3-10
3.4.1 Variability of EC25, LC50, NOEC 3-10
3.4.2 Variability of Endpoint Measurements 3-11

4.0 VARIABILITY IN CONTEXT 4-1
4.1 Society of Environmental Toxicology and Chemistry Pellston WET Workshop 4-1
4.1.1 General Conclusions and Recommendations 4-1
4.1.2 Conclusions about Data Precision 4-2
4.2 Water Environment Research Foundation Study 4-3
4.3 Minimizing Variability by Adhering to WET Methods 4-3
4.4 Conclusion 4-4

5.0 GUIDANCE TO REGULATORY AUTHORITIES, LABORATORIES AND PERMITTEES: GENERATING AND EVALUATING EFFECT CONCENTRATIONS 5-1
5.1 Steps for Minimizing Test Method Variability 5-1
5.2 Collecting Representative Effluent Samples 5-1
5.3 Conducting the Biological Test Methods 5-2
5.3.1 Quality Control Procedures 5-3
5.3.2 Experimental Design 5-5
5.3.3 Test Power to Detect Toxic Effects 5-7
5.4 Test Acceptability Criteria 5-9
5.5 Conducting the Statistical Analysis To Determine the Effect Concentration 5-10
5.6 Chapter Conclusions 5-11

TABLE OF CONTENTS
(continued)

6.0 GUIDANCE TO REGULATORY AUTHORITIES: DETERMINING REASONABLE POTENTIAL AND DERIVING WET PERMIT CONDITIONS 6-1
6.1 Analytical and Sampling Variability in Calculations for Reasonable Potential and Permit Limits 6-1
6.1.1 "Adjusting for Analytical Variability" in Calculations for Reasonable Potential and Permit Limits 6-1
6.1.2 Analytical Variability and Self-monitoring Data 6-2
6.1.3 Precision of WET Measurements and Estimates of Effluent CV 6-2
6.1.4 Between-Laboratory Variability 6-3
6.2 Determining Reasonable Potential and Establishing Effluent Limits 6-4
6.3 Development of a Total Maximum Daily Load for WET 6-5
6.4 Accounting for and Minimizing Variability in the Regulatory Decision Process 6-5
6.4.1 Recommended Additional TACs: Lower and Upper Bounds for PMSD 6-5
6.4.2 How To Determine the NOEC Using the Lower PMSD Bound 6-8
6.4.3 Justification for Implementing the Test Sensitivity Bounds 6-8
6.4.4 Guidance to Testing Laboratories on How to Achieve the Range of Performance for PMSD 6-9
6.5 Additional Guidance That Regulatory Authorities Should Implement to Further Support the WET Program 6-9
6.6 Chapter Conclusions 6-10

7.0 CONCLUSIONS AND GUIDANCE TO LABORATORIES, PERMITTEES, and REGULATORY AUTHORITIES 7-1
7.1 General Conclusions 7-1
7.2 Recommendations for Minimizing Variability and Its Effects 7-2
7.2.1 Guidance to Toxicity Testing Laboratories 7-3
7.2.2 Guidance to NPDES Permittees 7-3
7.3 Guidance to Regulatory Authorities 7-4
7.4 Future Directions 7-5

8.0 BIBLIOGRAPHY 8-1

Appendix A Interim Coefficients of Variation Observed Within Laboratories for Reference Toxicant Samples Analyzed Using EPA's Promulgated Whole Effluent Toxicity Methods

Appendix B Supplementary Information for Reference Toxicity Data

Appendix C Sample Calculation of Permit Limits Using EPA's Statistically-Based Methodology and Sample Permit Language

Appendix D Frequently Asked Questions (FAQs)

Appendix E Examples of Selected State WET Implementation Programs

Appendix F Improvements in Minimizing WET Test Variability by the State of North Carolina

Appendix G Analytical Variability in Reasonable Potential and Permit Limit Calculations

List of Tables

3-1 Promulgated WET Methods Included in This Report 3-3
3-2 Quartiles (25th and 75th) and Median (50th) of the Within-Laboratory Values of CV for EC25 (Chronic Tests) 3-4
3-3 Quartiles (25th and 75th) and Median (50th) of the Within-Laboratory Values of CV for LC50 3-4
3-4 Quartiles (25th and 75th) and Median (50th) of the Within-Laboratory Values of CV for NOEC 3-5
3-5 Estimates of Within-Laboratory and Between-Laboratory Components of Variability 3-7
3-6 Range of Relative Variability for Endpoints of Promulgated WET Methods, Defined by the 10th and 90th Percentiles from the Data Set of Reference Toxicant Tests 3-9
3-7 Number of Laboratories Having a Given Percent of Tests Exceeding the PMSD Upper Bound for the Sublethal Endpoint 3-10
5-1 Tests for Chronic Toxicity: Power and Ability To Detect a Toxic Effect on the Sublethal Endpoint 5-8
5-2 Power to Detect a 25-Percent Difference from the Control at the 90th Percentile PMSD 5-8
6-1 Example of Applying the Lower Bound PMSD for the Chronic *Ceriodaphnia* Test with the Reproduction Endpoint 6-8

List of Figures

5-1 Steps to minimize WET test method variability 5-2
6-1 Paradigm that incorporates the lower and upper percent minimum significant difference 6-6
6-2 Implementing applications of upper and lower PMSD bounds for effluent testing requirements 6-7

This page intentionally left blank.

ACKNOWLEDGEMENTS

This guidance was prepared through the cooperative efforts of the U.S. Environmental Protection Agency's (EPA) Office of Wastewater Management and Office of Science and Technology in the Office of Water, EPA's Office of Research and Development, and EPA's Office of Enforcement and Compliance Assurance. The Cadmus Group, Inc. provided support for the final document production.

EPA Variability Workgroup

Debra Denton, EPA Region 9, San Francisco, CA
John Fox, EPA Office of Science and Technology, Washington, DC
Florence Fulk, EPA Office of Research and Development, Cincinnati, OH
Kathryn Greenwald, EPA Office of Enforcement and Compliance Assurance, Washington, DC
Madonna Narvaez, EPA Region 10, Seattle, WA
Teresa Norberg-King, EPA Office of Research and Development, Duluth, MN
Laura Phillips, EPA Office of Wastewater Management, Washington, DC

EPA Support Outside of the Variability Workgroup

Gregory Currey, EPA Office of Wastewater Management, Washington, DC
Margarete Heber, EPA Office of Wetlands, Oceans, and Watersheds, Washington, DC
Phillip Jennings, EPA Region 6, Dallas, TX
Henry Kahn, EPA Office of Science and Technology, Washington, DC
Marion Kelly, EPA Office of Science and Technology, Washington, DC
James Pendergast, EPA Office of Wetlands, Oceans, and Watersheds, Washington, DC
Stephen Sweeney, EPA Office of General Counsel, Washington, DC
William Telliard, EPA Office of Science and Technology, Washington, DC
Robert Wood, EPA Office of Wastewater Management, Washington, DC
Marcus Zobrist, EPA Region 2, New York, NY

Contractor Support

The Cadmus Group, Inc., Durham, NC

Karalyn Colopy
Susan Conbere
Blanche Dean
Penelope Kellar

DynCorp Information and Enterprise Technology, Inc., Alexandria, VA

Robert Brent

Science Applications International Corporation, Columbia, MD

Sidina Dedah
Ruth Much
Kathleen Stralka

Assistance from States

State Case Example and Chapter 4

EPA especially thanks Larry Ausley and Matt Matthews of the North Carolina Department of Environment and Natural Resources, Division of Water Quality for preparing Chapter 4 (Variability in Context) and Appendix F (Improvements in Minimizing WET Test Variability by the State of North Carolina).

State-Specific WET Program Implementation

EPA appreciates the assistance and support from the following States in providing their State-specific approaches to the WET program implementation used for Appendix E (Examples of Selected State WET Implementation Programs) of this document.

Betty Jane Boros-Russo, State of New Jersey
Kari Fleming, State of Wisconsin
Randall Marshall, State of Washington
Matt Matthews, State of North Carolina
Charlie Roth, State of Kentucky

Additional Assistance

We also thank the EPA Regions, States, and laboratories that provided the WET toxicity data used to develop this guidance document.

EPA Peer Review

Peer review was conducted following the EPA's Science Policy Council Handbook for Peer Review (January 1998). The anonymous review comments are gratefully acknowledged, and the changes were incorporated, as appropriate.

This page intentionally left blank.

EXECUTIVE SUMMARY

Background

The Federal Water Pollution Control Act, commonly known as the Clean Water Act, was enacted in 1972 with the objective of "*restoring the chemical, physical, and biological integrity of the Nation's waters*." Among the U.S. Environmental Protection Agency's (EPA's) efforts toward this objective is the National Pollutant Discharge Elimination System (NPDES) program. This program is designed to control toxic discharges, implement water quality standards, and restore waters to "fishable and swimmable" conditions. Point sources that discharge pollutants must do so under the terms and conditions of an NPDES permit. One approach EPA employs to control toxic pollutants under the NPDES permits program is using whole effluent toxicity (WET) controls.

EPA is issuing this document to both address questions raised on WET test method variability and to satisfy a requirement of a July 1998 settlement agreement with litigants for the Western Coalition of Arid States (WestCAS) and Edison Electric Institute et al. This document was developed by an EPA workgroup consisting of EPA's Office of Water's (OW) Headquarters, Office of Enforcement and Compliance Assurance, Office of Research and Development, and Regional staff. The document was externally peer reviewed in accordance with EPA's peer review guidelines. The document addresses WET test method variability by identifying the potential sources of variance associated with WET testing, discusses how to minimize it and, finally, describes how to address it within the NPDES permitting program. The document cites both Agency and external ongoing research on this topic and scientific findings, particularly technical information that support efforts to minimize WET test result variability.

While the document provides recommendations on how to reduce or minimize WET test variability, the document does not supersede current Agency guidance, policy, or regulation, including EPA's promulgated test methods (40 CFR Part 136), which remain in effect. EPA expects that implementation of the NPDES program and NPDES permits will continue to comply with regulatory requirements and follow applicable EPA guidance and policy.

Why WET Testing?

Whole effluent toxicity is the aggregate toxic effect of an aqueous sample (e.g., effluent, receiving water) measured directly by an aquatic toxicity test. Aquatic toxicity tests are laboratory experiments that measure the biological effect (e.g., growth, survival, and reproduction) of effluents or receiving waters on aquatic organisms. In aquatic toxicity tests, organisms of a particular species are held in test chambers and exposed to different concentrations of an aqueous sample, for example, a reference toxicant, an effluent, or a receiving water, and observations are made at predetermined exposure periods. At the end of the test, the responses of test organisms are used to estimate the effects of the toxicant or effluent.

Whole effluent toxicity test results are an integral tool in the assessment of water quality. For the protection of aquatic life, the integrated strategy includes the use of three control approaches: the chemical-specific control approach, the WET control approach, and the biological criteria/bioassessment/bioassay approach. The primary advantage of using WET controls over individual, chemical-specific controls is that WET integrates the effects of all chemical(s) in the aqueous sample. Reliance solely on chemical-specific numeric criteria or biological criteria would result in only a partially effective State toxics control program. These toxicity tests therefore must be performed using best laboratory practices, and every effort must be made to enhance repeatability of the test method. This document presents EPA's approaches to achieve the goals listed below.

Effect of This Guidance

This document clarifies several issues regarding WET variability and reaffirms EPA's guidance in the *Technical Support Document for Water Quality-Based Toxics Control* (TSD, USEPA 1991a). This document provides NPDES regulatory authorities and all stakeholders, including permittees, with guidance and recommendations on how to address WET variability. EPA's recommendations and conclusions are detailed in Chapter 7, and Appendix C provides sample NPDES permit language reflecting these recommendations.

The most significant recommendation is to use and report the values for the percent minimum significant difference (PMSD) with all WET data results. The minimum significant difference (MSD) represents the smallest difference between the control mean and a treatment mean that leads to the statistical rejection of the null hypothesis (i.e., no toxicity) at each concentration of the WET test dilution series. The MSD provides an indication of within-test variability and test method sensitivity. Using this information, the regulatory authority and permittees can better evaluate WET test results.

This document makes several other recommendations, such as continue to use the TSD statistical approach without adjusting for test method variability, obtain sufficient representative effluent samples, verify effluent toxicity data against reference toxicant data, maintain clear communication between the regulatory authority and permittee, and maintain good laboratory checks and certification programs.

Three Goals of This Document

This document describes three goals EPA has defined to address issues surrounding WET variability. In addition, the document is intended to satisfy the requirements of a settlement agreement to resolve litigation over rulemaking to standardize WET testing procedures.

1. Quantify the variability of promulgated test methods and report a coefficient of variation (CV) as a measure of test method variability (see Chapter 3 and Appendix A).

2. Evaluate the statistical methods described in the *Technical Support Document for Water Quality-Based Toxics Control* (TSD) for determining the need for and deriving WET permit conditions (see Chapter 6 and Appendix G).

3. Suggest guidance for regulatory authorities on approaches to address and minimize test method variability (Chapter 6). In addition, the document is intended to provide guidance to regulatory authorities, permittees, and testing laboratories on conducting the biological and statistical methods and evaluating test effect concentrations (Chapter 5).

Data Evaluated

EPA assembled a comprehensive data base to examine variability in the WET test methods from the EPA Regions, several States, and private laboratories, which represent a widespread sampling of typical laboratories and laboratory practices. EPA applied several criteria to the data before they were accepted, including detailed sample information, strict adherence to published EPA WET test methods, and test acceptability criteria (TAC). The resulting data base contains data from 75 laboratories for 23 methods for tests concluded between 1988 and 1999.

Approach Taken To Evaluate Test Method Variability

The variability that EPA is assessing is associated with replicate tests using reference toxicants and WET testing methods within analytical laboratories. The focus of this guidance is *not* to quantify test variability between laboratories or to quantify the total variability of WET tests conducted on effluents. Rather, the purpose is to quantify method variability within laboratories (repeatability) to enable NPDES

programs to distinguish between variability caused by the testing method and variability associated with toxicity of multiple effluent samples taken from the same facility.

To quantify test method variability within and between laboratories using this data base, EPA examined two key parameters: (1) the effect concentrations [effect concentration (EC25), lethal concentration (LC50), no observed effect concentration (NOEC)] estimated by the test, which are used to derive WET permit limits and evaluate self-monitoring data with those limits; and (2) the minimum significant difference (MSD), which summarizes the variability of organism responses at each test concentration within an individual test. The MSD represents the smallest difference that can be distinguished between the response of the control organisms and the response of the organisms exposed to the aqueous sample. The MSD provides an indication of within-test variability and test method sensitivity.

Principal Conclusions

The principal conclusions of this document follow.

Evaluation of Test Method Variability

- Comparisons of WET method precision with method precision for analytes commonly limited in NPDES permits clearly demonstrate that the variability of the promulgated WET methods is within the range of variability experienced in other types of analyses. Several independent researchers and studies also have concluded that method performance improves when prescribed methods are followed closely by experienced analysts (Section 4.3).

- This document provides interim CVs for promulgated WET methods in Appendix A, Tables A-1 (acute methods) and A-2 (chronic methods), pending completion of between-laboratory studies, which may affect these interim CV estimates.

Evaluation of Approach To Incorporate Test Method Variability

- EPA's TSD presents guidance for developing effluent limits that appropriately protect water quality, regarding both effluent variability and analytical variability, provided that the WET criteria and waste load allocation (WLA) are derived correctly (Section 6 and Appendix G).

- EPA's analysis of data gathered in the development of this document indicates that the TSD approach appropriately accounts for both effluent variability and method variability. EPA does not believe a reasonable alternative approach is available to determine a factor that would discount the effects of method variability using the TSD procedures, because the approach would not ensure adequate protection of water quality (Section 6.1.1 and Appendix G).

Development of Guidance to Regulatory Authorities

- EPA recommends that regulatory authorities implement the statistical approach as described in the TSD to evaluate effluent for reasonable potential and to derive WET limits or monitoring triggers (Section 6.1 and Appendix G).

- EPA recommends that regulatory authorities calculate the facility-specific CVs using point estimate techniques to determine the need for and derive a permit limit for WET, even if self-monitoring data are to be determined using hypothesis testing techniques, for example, to determine a "no effect" concentration ("NOEC"). This document describes such facility-specific calculations (Section 3.4.1 and 6.2).

Additional Recommendations and Guidance

This document also provides recommendations and guidance on minimizing variability in three specific areas in order to generate sound WET test results: (1) obtaining a representative effluent sample; (2) conducting the toxicity tests properly to generate the biological endpoints; and (3) conducting the appropriate statistical analysis to obtain defensible effect concentrations (EC25, LC50, NOEC). If these recommendations are addressed, the reliability of the test endpoint values should improve.

- **Regulatory Authorities:** Design a sampling program that collects representative effluent samples to fully characterize effluent variability for a specific facility over time (Sections 6.1.3 and 6.2).

- **Regulatory Authorities:** Ensure proper application of WET statistical procedures and test methods (Sections 5.2 through 5.5).

- **Regulatory Authorities:** Incorporate both the upper and lower bounds using the percent minimum significant difference (PMSD) to control and to minimize within-test method variability and increase test sensitivity. To achieve the PMSD upper bound, either the replication should increase or within-test method variability should decrease, or both (Section 6.4 and Table 3-6).

- **Testing Laboratories:** Encourage WET testing laboratories to maintain control charts for PMSD and the control mean and report the PMSD with all WET test results (Section 5.3.1.1).

- **Regulatory Authorities:** Participate in the National Environment Laboratory Accreditation Program and routine performance audit inspections to evaluate laboratory performance (Section 5.3.1.1).

- **Regulatory Authorities:** Incorporate EPA's guidance on error rate assumption adjustments, concentration-response relationships, confidence intervals, acceptable dilution waters, how to block by parentage for the chronic *Ceriodaphnia dubia* test, and control of pH drift (USEPA 2000a).

LIST OF ACRONYMS AND ABBREVIATIONS[1]

ACR	acute-to-chronic ratio
AML	average monthly limit
ANOVA	analysis of variance
APHA-AWWA-WEF	American Public Health Association-American Water Works Association-Water Environment Federation
ASTM	American Society for Testing and Materials
BSAB	Biomonitoring Science Advisory Board
CCC	criteria continuous concentration
CFR	Code of Federal Regulations
CMC	criteria maximum concentration
CV	coefficient of variation
CWA	Clean Water Act
DMR	discharge monitoring report
EMS	error mean square [also referred to as mean square error (MSE)]
EPA	U.S. Environmental Protection Agency (also, the Agency)
FR	*Federal Register*
IC	inhibition concentration
IWC	instream waste concentration (sometimes referred to as receiving water concentration)
LC50	lethal concentration, 50 percent
LOEC	lowest observed effect concentration
LTA	long-term average (LTAa = acute LTA; LTAc = chronic LTA; LTAa,c = acute-to-chronic LTA)
MDL	maximum daily limit
MSD	minimum significant difference
MSE	mean square error [also referred to as error mean square (EMS)]
MZ	mixing zone
NELAP	National Environment Laboratory Accreditation Program
NOEC	no observed effect concentration
NPDES	National Pollutant Discharge Elimination System
NTRD	National Toxicant Reference Database
PAI	Performance Audit Inspections
PMSD	percent minimum significant difference

[1] Note: These acronyms and abbreviations may have other meanings in other EPA programs or documents.

QA	quality assurance
QC	quality control
rMSE	square root of the mean square error
RP	reasonable potential
RWC	receiving water concentration (sometimes referred to as instream waste concentration)
SCTAG	Southern California Toxicity Assessment Group
SETAC	Society of Environmental Toxicology and Chemistry
TAC	test acceptability criteria
TIE	toxicity identification evaluation
TMDL	total maximum daily load
TRE	toxicity reduction evaluation
TSD	EPA's *Technical Support Document for Water Quality-based Toxics Control* (March 1991, EPA505/2-90-001)
TU	toxic unit (TUa = acute toxicity; TUc = chronic toxicity)
VF	variability factor
WET	whole effluent toxicity
WLA	waste load allocation
WQBEL	water quality based effluent limit

GLOSSARY

Acute Toxicity Test is a test to determine the concentration of effluent or ambient waters that causes an adverse effect (usually death) on a group of test organisms during a short-term exposure (e.g., 24, 48, or 96 hours). Acute toxicity is measured using statistical procedures (e.g., point estimate techniques or a *t*-test).

Acute-to-Chronic Ratio (ACR) is the ratio of the acute toxicity of an effluent or a toxicant to its chronic toxicity. It is used as a factor for estimating chronic toxicity on the basis of acute toxicity data, or for estimating acute toxicity on the basis of chronic toxicity data.

Ambient Toxicity is measured by a toxicity test on a sample collected from a receiving waterbody.

ANOVA is analysis of variance.

Average Monthly Limit (AML) is the calculated average monthly limit of waste load allocation assigned by a State or EPA for a particular facility.

CCC are water quality criteria for chronic exposure (criteria continuous concentrations).

Chronic Toxicity Test is a short-term test in which sublethal effects (e.g., reduced growth or reproduction) are usually measured in addition to lethality. Chronic toxicity is defined as TUc = 100/NOEC or TUc = 100/ECp or ICp.

CMC are water quality criteria for acute exposures (criteria maximum concentration).

Coefficient of Variation (CV) is a standard statistical measure of the relative variation of a distribution or set of data, defined as the standard deviation divided by the mean. It is also called the relative standard deviation (RSD). The CV can be used as a measure of precision within (within-laboratory) and between (between-laboratory) laboratories, or among replicates for each treatment concentration.

Confidence Interval is the numerical interval constructed around a point estimate of a population parameter.

Effect Concentration (EC) is a point estimate of the toxicant concentration that would cause an observable adverse effect (e.g., death, immobilization, or serious incapacitation) in a given percent of the test organisms, calculated from a continuous model (e.g., Probit Model). EC25 is a point estimate of the toxicant concentration that would cause an observable adverse effect in 25 percent of the test organisms.

Hypothesis Testing is a statistical technique (e.g., Dunnett's test) for determining whether a tested concentration is statistically different from the control. Endpoints determined from hypothesis testing are NOEC and LOEC. The two hypotheses commonly tested in WET are:

> **Null hypothesis (H_o)**: The effluent is not toxic.
> **Alternative hypothesis (H_a)**: The effluent is toxic.

Inhibition Concentration (IC) is a point estimate of the toxicant concentration that would cause a given percent reduction in a non-lethal biological measurement (e.g., reproduction or growth), calculated from a continuous model (i.e., Interpolation Method). IC25 is a point estimate of the toxicant concentration that would cause a 25-percent reduction in a non-lethal biological measurement.

Instream Waste Concentration (IWC) is the concentration of a toxicant in the receiving water after mixing. The IWC is the inverse of the dilution factor. It is sometimes referred to as the receiving water concentration (RWC).

LC50 (lethal concentration, 50 percent) is the toxicant or effluent concentration that would cause death in 50 percent of the test organisms.

Lowest Observed Effect Concentration (LOEC) is the lowest concentration of an effluent or toxicant that results in adverse effects on the test organisms (i.e., where the values for the observed endpoints are statistically different from the control).

Long-term Averages (LTAs) of pollutant concentration or effluent toxicity are calculated from waste load allocations (WLAs), typically assuming that the WLA is a 99th percentile value (or another upper bound value) based on the lognormal distribution. One LTA is calculated for each WLA (typically an acute LTA and a chronic LTA for aquatic life protection). The LTA represents expected long-term average performance from the permitted facility required to achieve the associated WLA.

Maximum Daily Limit (MDL) is the calculated maximum WLA assigned by a State or EPA for a particular facility.

Minimum Significant Difference (MSD) is the magnitude of difference from control where the null hypothesis is rejected in a statistical test comparing a treatment with a control. MSD is based on the number of replicates, control performance, and power of the test.

Mean Square Error (MSE) is the average dispersion of the items around the treatment means. It is an estimate of a common variance, the within variation, or variation among observations treated alike. [Also referred to as error mean square (EMS).]

Mixing Zone is an area where an effluent discharge undergoes initial dilution and is extended to cover the secondary mixing in the ambient waterbody. A mixing zone is an allocated impact zone where water quality criteria can be exceeded as long as acutely toxic conditions are prevented.

No Observed Effect Concentration (NOEC) is the highest tested concentration of an effluent or toxicant that causes no observable adverse effect on the test organisms (i.e., the highest concentration of toxicant at which the values for the observed responses are not statistically different from the controls).

National Pollutant Discharge Elimination System (NPDES) program regulates discharges to the nation's waters. Discharge permits issued under the NPDES program are required by EPA regulation to contain, where necessary, effluent limits based on water quality criteria for the protection of aquatic life and human health.

Power is the probability of correctly detecting an actual toxic effect (i.e., declaring an effluent toxic when, in fact, it is toxic).

Precision is a measure of reproducibility within a data set. Precision can be measured both within a laboratory (within-laboratory) and between laboratories (between-laboratory) using the same test method and toxicant.

Quality Assurance (QA) is a practice in toxicity testing that addresses all activities affecting the quality of the final effluent toxicity data. QA includes practices such as effluent sampling and handling, source and condition of test organisms, equipment condition, test conditions, instrument calibration, replication, use of reference toxicants, recordkeeping, and data evaluation.

Quality Control (QC) is the set of more focused, routine, day-to-day activities carried out as part of the overall QA program.

Reasonable Potential (RP) is where an effluent is projected or calculated to cause an excursion above a water quality standard based on a number of factors.

Reference Toxicant Test is a check of the sensitivity of the test organisms and the suitability of the test methodology. Reference toxicant data are part of a routine QA/QC program to evaluate the performance of laboratory personnel and the robustness and sensitivity of the test organisms.

Significant Difference is defined as a statistically significant difference (e.g., 95 percent confidence level) in the means of two distributions of sampling results.

Statistic is a computed or estimated quantity such as the mean, standard deviation, or coefficient of variation.

Test Acceptability Criteria (TAC) are specific criteria for determining whether toxicity test results are acceptable. The effluent and reference toxicant must meet specific criteria as defined in the test method (e.g., for the *Ceriodaphnia dubia* survival and reproduction test, the criteria are as follows: the test must achieve at least 80 percent survival and an average of 15 young per surviving female in the controls).

Total Maximum Daily Load (TMDL) is a determination of the amount of a pollutant, or property of a pollutant, from point, nonpoint, and natural background sources, including a margin of safety, that may be discharged to a water quality-limited waterbody.

***t*-Test** (formally Student's *t*-Test) is a statistical analysis comparing two sets of replicate observations, in the case of WET, only two test concentrations (e.g., a control and 100 percent effluent). The purpose of this test is to determine if the means of the two sets of observations are different [e.g., if the 100-percent effluent concentration differs from the control (i.e., the test passes or fails)].

Type I Error (alpha) is the rejection of the null hypothesis (H_o) when it is, in fact, true (i.e., determining that the effluent is toxic when the effluent is not toxic).

Type II Error (beta) is the acceptance of the null hypothesis (H_o) when it is not true (i.e., determining that the effluent is not toxic when the effluent is toxic).

Toxicity Test is a procedure to determine the toxicity of a chemical or an effluent using living organisms. A toxicity test measures the degree of effect of a specific chemical or effluent on exposed test organisms.

Toxic Unit-Acute (TUa) is the reciprocal of the effluent concentration (i.e., TUa = 100/LC50) that causes 50 percent of the organisms to die by the end of an acute toxicity test.

Toxic Unit-Chronic (TUc) is the reciprocal of the effluent concentration (e.g., TUc = 100/NOEC) that causes no observable effect (NOEC) on the test organisms by the end of a chronic toxicity test.

Toxic Unit (TU) is a measure of toxicity in an effluent as determined by the acute toxicity units (TUa) or chronic toxicity units (TUc) measured. Higher TUs indicate greater toxicity.

Toxicity Identification Evaluation (TIE) is a set of procedures used to identify the specific chemicals causing effluent toxicity.

Toxicity Reduction Evaluation (TRE) is a site-specific study conducted in a step-wise process designed to identify the causative agents of effluent toxicity, isolate the source of toxicity, evaluate the effectiveness of toxicity control options, and then confirm the reduction in effluent toxicity.

Variance is a measure of the dispersion in a set of values, defined as the sum of the squared deviations divided by their total number.

Whole Effluent Toxicity (WET) is the total toxic effect of an effluent measured directly with a toxicity test.

Waste Load Allocation (WLA) is the portion of a receiving water's total maximum daily load that is allocated to one of its existing or future point sources of pollution.

This page intentionally left blank.

1.0 INTRODUCTION

1.1 Background

The Federal Water Pollution Control Act, commonly known as the Clean Water Act (CWA), was enacted in 1972 with the objective of "*restoring the chemical, physical, and biological integrity of the Nation's waters*." Several goals and policies were established in the Act, including the following:

- Eliminating the discharge of pollutants into navigable waters by 1985;
- Wherever attainable, achieving an interim goal of water quality that provides for the protection and propagation of fish, shellfish, and wildlife, and provides for recreation in and on the water by November 1, 1983; and
- Prohibiting the discharge of toxic pollutants in toxic amounts.

In the 28 years since the CWA was enacted, the U.S. Environmental Protection Agency (EPA) and States authorized to administer EPA's National Pollutant Discharge Elimination System (NPDES) permitting program have made significant progress toward achieving these goals. NPDES is designed to control toxic discharges, implement a water quality standards program, and restore waters to "fishable and swimmable" conditions. A point source that discharges pollutants to waters of the United States must do so under the terms and conditions of an NPDES permit. In setting these terms and conditions, EPA and the States have integrated their control of toxic pollutants through combined use of three approaches [*Technical Support Document for Water Quality-based Toxics Control* (USEPA 1991a, referred to as the TSD)]:

- Chemical-specific controls,
- Whole effluent toxicity (WET) controls, and
- Biological criteria/bioassessments and bioassays.

The WET approach to protection of water quality is the primary subject of this document.

In 1989, EPA defined whole effluent toxicity as "*the aggregate toxic effect of an effluent measured directly by an aquatic toxicity test*" [54 Federal Register (FR) 23868 at 23895, June 2, 1989]. Aquatic toxicity tests are laboratory experiments that measure the biological effect (e.g., growth, survival, and reproduction) of effluents or receiving waters on aquatic organisms. In aquatic toxicity tests, groups of organisms of a particular species are held in test chambers and exposed to different concentrations of an aqueous test sample, for example, a reference toxicant, an effluent, or a receiving water. Observations are made at predetermined exposure periods. At the end of the test, the responses of test organisms are used to estimate the effects of the toxicant or effluent.

In the early 1980s, EPA published methods (USEPA 1985, 1988, 1989) for estimating the short-term acute and chronic toxicity of effluents and receiving waters to freshwater and marine organisms. WET data gathered in the 1980s indicated that approximately 40 percent of NPDES facilities nationwide discharged an effluent with sufficient toxicity to cause water quality problems. Further reductions in the toxicity of wastewater discharges were necessary to achieve compliance with narrative water quality standards expressed as "no toxics in toxic amounts." In response to these findings, EPA implemented a policy to reduce or eliminate toxic discharges. The *Policy for the Development of Water Quality-based Permit Limitations for Toxic Pollutants* (49 FR 9016, March 9, 1984) introduced EPA's integrated toxics control program. To support this policy, EPA developed the TSD (USEPA 1991a). The TSD provides guidance to

regulators in implementing WET testing requirements in NPDES permits. In 1989, EPA promulgated regulations specifying procedures for determining when water quality-based effluent limitations are required in NPDES permits [40 CFR, 122.44(d)]. On October 26, 1995, EPA promulgated WET test methods (USEPA 1993, 1994a, and 1994b) and added them to the list of EPA methods approved under Section 304(h) of the CWA (40 CFR, 136) for use in the NPDES program. Although the rulemaking was challenged in court, that challenge has been stayed pending completion of a settlement agreement. The rulemaking remains in force and effect unless and until EPA takes further action.

1.2 Effect of This Guidance

This document attempts to clarify several issues regarding WET variability and reaffirms EPA's earlier guidance and recommendations published in the TSD (USEPA 1991a). This document is intended to provide NPDES regulatory authorities and all stakeholders, including permittees, with guidance and recommendations on how to understand and account for measurement variability in WET testing. The document's recommendations and conclusions are detailed in Section 7. Appendix C provides sample NPDES permit language reflecting these recommendations.

The most significant recommendation is to use and report the values for the percent minimum significant difference (PMSD) with all WET data results. The minimum significant difference (MSD) is the smallest difference that can be distinguished between the response of control organisms and the response of test organisms at each concentration of the WET test dilution series. The MSD provides an indication of the within-test variability and test method sensitivity. Using this information, the regulatory authority and permittees can better evaluate WET test results.

This document also recommends the following:

- Continue to use the EPA TSD statistical approach for NPDES permit limit development (no test method variability adjustments are needed);
- Collect and evaluate a sufficient number of representative effluent samples;
- Verify effluent toxicity data carefully along with reference toxicant data;
- Maintain good communication between the regulatory authority and permittee throughout all phases of the permitting process;
- Implement the PMSD to evaluate both WET and reference toxicant data to minimize within-test method variability and increase test sensitivity;
- Maintain laboratory checks with good laboratory certification programs to encourage experienced laboratories and skilled analysts for the toxicity testing program for individual WET laboratory performance.

1.3 Three Goals of This Document

EPA prepared this document to achieve the following three goals:

1. Quantify the variability of promulgated test methods and report a coefficient of variation (CV) as a measure of test method variability (see Chapter 3 and Appendix A).

2. Evaluate the statistical methods described in the *Technical Support Document for Water Quality-Based Toxics Control* (TSD) for determining the need for and deriving WET permit conditions (see Chapter 6 and Appendix G).

3. Suggest guidance for regulatory authorities on approaches to address and minimize test method variability (Chapter 6). In addition, the document is intended to provide guidance to regulatory authorities, permittees, and testing laboratories on conducting the biological and statistical methods and evaluating test effect concentrations (Chapter 5).

This document does not address effluent variability. It does, however, discuss how handling effluent samples can affect tests. Chapter 2 provides definitions of terms used and discusses the ways in which variability can be quantified. Chapter 3 describes the variability of the effect concentration estimates (EC25, LC50, and NOEC) and the variability of endpoint measurements (survival, growth, and reproduction). Chapter 4 discusses WET variability in the context of chemical-specific method variability. Chapter 5 provides guidance to permittees, testing laboratories, and regulatory authorities to minimize test method variability. Chapter 6 provides guidance to regulatory authorities on how to determine reasonable potential (RP) and derive permit limits or monitoring triggers and evaluate self-monitoring data. Chapter 7 presents EPA's principal conclusions. Chapter 8 is a bibliography containing a list of documents cited herein and additional reading material.

This page intentionally left blank.

2.0 DEFINITION AND MEASUREMENT OF METHOD VARIABILITY IN WET TESTING

The terms used to express toxicity test results are defined in this chapter, and methods for quantifying WET test method variability are discussed. Additional terms used throughout this document, along with their definitions, are provided in the Glossary as part of the front matter of this document.

2.1 Terms and Definitions

Biological endpoints are the biological observations recorded when conducting toxicity tests. These observations may include the number of surviving organisms or the number of young produced. There are two basic types of biological endpoints: responses recorded as response/no response (e.g., dead or alive) are quantal data; responses recorded as a measured response (e.g., weight) or as a count (e.g., number of young produced) are considered continuous data. For most WET tests, the observations for each tested concentration are combined and then reported as an average or percentage to represent the biological endpoint. For example, the fathead minnow larval survival and growth chronic test method has two biological endpoints (i.e., percent survival and average dry weight for each test concentration).

Effect concentrations are concentrations of a test material (i.e., effluent, referent toxicant, receiving water) derived from the observed biological endpoints followed by data analysis using either hypothesis testing procedures or point estimate techniques. Effect concentrations derived using point estimation techniques represent the concentration of a test material at which a predetermined level of effect occurs. For example, LC50 is the lethal concentration at which 50 percent of the organisms respond. Effect concentrations commonly estimated for WET methods are LC50, EC50 (effect concentration at which a 50-percent effect occurs), and IC25 (inhibition concentration at which a 25-percent effect occurs). Hypothesis test methods are used to determine the no observed effect concentration (NOEC). The NOEC represents the highest effect concentration in the test concentration response that is not significantly different from the control response. Multiple statistical endpoints can be derived for each WET method. For example, the endpoints for the fathead minnow larval survival and growth chronic test can be reported as an EC25 for growth, an NOEC for growth, an LC50 (or EC50) for survival, and an NOEC for survival.

2.2 Defining WET Test Variability

As with any measurement process, WET tests have a degree of variability associated with the test method performance. Three measures of variability related to WET tests are within-test variability, within-laboratory variability, and between-laboratory variability.

- **Within-test (intra-test) variability** is the variability in test organism response within a concentration averaged across all concentrations of the test material in a single test.

- **Within-laboratory (intra-laboratory) variability** is the variability that is measured when tests are conducted using specific methods under reasonably constant conditions in the same laboratory. Within-laboratory variability, as used in this document, includes within-test variability. The American Society for Testing and Materials (ASTM) uses the term “repeatability” to describe within-laboratory variability. Repeatability is estimated (as a sample variance or standard deviation) by repeating a test method under realistically constant conditions within a single laboratory.

- **Between-laboratory (inter-laboratory) variability** is the variability between laboratories. It is measured by obtaining results from different laboratories using the same test method and the same

test material (e.g., reference toxicant). Between-laboratory variability, as used in this document, does *not* include the within-laboratory component of variance. ASTM uses the term "reproducibility" to describe between-laboratory variability. Reproducibility is estimated by having nearly identical test samples (duplicates or splits) analyzed by multiple laboratories using similar standard methods. Although reproducibility is generally synonymous with between-laboratory variability, estimates of reproducibility may combine within-laboratory and between-laboratory components of variance, making between-laboratory variability numerically larger than within-laboratory variability as defined above.

For purposes of consistency, EPA uses the terms within-laboratory and between-laboratory variability throughout this document.

Numerous factors can affect the variability of any toxicity test method. These factors include the number of test organisms, the number of treatment replicates, randomization techniques, the source and health of the test organisms, the type of food used, laboratory environmental conditions, and dilution water quality. The experience of the analyst performing the test, analyzing the data, and interpreting the results may also affect variability (Grothe et al. 1996, Fulk 1996).

2.3 Quantifying WET Test Variability

Historically, information on the variability of toxicity tests has been developed using effect concentrations, such as the NOEC, EC25, EC50, and LC50 for survival, fecundity, and growth. Variability measures should be quantified based on the end use of the data (i.e., effect concentrations) and be directly related to the WET permit requirement. Typically, the effect concentrations are the endpoints used for evaluating self-monitoring results. The variability of the effect concentrations is quantified by obtaining multiple test results under similar test conditions using the same test material. For example, the sample standard deviation and mean for EC25 obtained from multiple monthly reference toxicant tests for the fathead minnow survival and growth chronic test conducted at one laboratory would quantify "within-laboratory" variability for that laboratory. EPA used this approach to evaluate data for the development of this document (see Chapter 3).

Examining variability for each effect concentration of each biological endpoint for each test method is essential. The biological endpoints may be different for various toxicants and effluents. One biological endpoint, such as reproduction, may be more sensitive to a certain toxicant than another endpoint, such as survival. That sensitivity may be reversed for a different toxicant. Alternatively, an endpoint may be more sensitive to one toxicant than another toxicant.

Three other measures of variability (which are not addressed in this document) that have been applied to WET tests are:

1. Determine the variability of the biological endpoint response. For example, the variance of the biological response (e.g., growth and survival) can be calculated. This approach is useful, but does not quantify variability of the WET test effect concentration, which is important in the context of this document.

2. Quantify the uncertainty of each test point estimate (e.g., the EC50, EC25, or LC50) using confidence intervals, which reflect within-test variability.

3. Use the standard deviation to quantify the uncertainty in the mean of the replicate response at each concentration within a particular test. For example, laboratories can compare the standard deviations of the average weight of fathead minnow larvae in four chronic tests at one test concentration, such as 1 mg/L sodium chloride. These standard deviations may be pooled across

all the concentrations when data have been transformed (if necessary) to give similar variances at each concentration. From the pooled variance, one may calculate a minimum significant difference (MSD) value, which is a useful indication of test sensitivity (see Chapters 3 and 5). In this document, the standard deviation at each concentration was not evaluated as a measure of variability. However, the MSD was considered as a measure of WET test variability.

This page intentionally left blank.

3.0 VARIABILITY OF WET TEST METHODS

Chapter 3 describes the variability of effect concentration estimates (EC25, LC50, and NOEC) and endpoint measurements (survival, growth, and reproduction). For definitive studies of the variability of WET methods, readers should also refer to the TSD (USEPA 1991a, Part 1.3.3) and to WET methods manuals (USEPA 1993, 1994a, 1994b). EPA will complete and report on a new between-laboratory study of promulgated methods in 2000 or 2001.

3.1 Acquisition, Selection, and Quality Assurance of Data Presented in This Document

EPA solicited data for reference toxicant tests from laboratories that conduct WET tests and use reference toxicant testing as part of their quality control (QC) program. Reference toxicant testing is required, as specified in EPA toxicity test methods, to document laboratory performance over time for laboratories conducting self-monitoring tests. When laboratories are conducting effluent tests, at least one reference toxicant test must be conducted each month using the same toxicant, test concentrations, dilution water, and data analysis methods. These reference toxicant tests must be conducted using the same test conditions (type of dilution water, temperature, test protocol, and species) that are used for WET tests conducted by the laboratory.

Reference toxicant tests were used to characterize method variability because, in contrast to effluent samples, fixed concentrations of known toxicants are used. Only with this standardization is it possible to conclude that variability of the effect concentration estimates is derived from the sources discussed above, rather than from changes in the toxicant.

EPA received reference toxicant test data from several States, private laboratory sources, and the EPA Regions. Data sources used for these analyses include the EPA National Toxicant Reference Database (NTRD), the EPA Region 9 Toxicity Data Base, and laboratory bench sheets voluntarily submitted by independent sources. Although the data do not represent a random sample of laboratories or tests, they do represent a widespread sampling of typical laboratories and practices.

EPA required that reference toxicant tests included in its data base meet the following four criteria:

1. Test records documented the test method, organism, test date, laboratory, reference toxicant, and individual biological responses in the concentration series.

2. Data for each replicate were provided as required in the published method using the current test method.

3. The test used at least five toxicant concentrations and a control for the most commonly reported chronic toxicity test methods—(1) 1000.0, fathead minnow larval survival and growth; (2) 1002.0, *Ceriodaphnia* survival and reproduction; and (3) 1006.0, inland silverside survival and growth. For other chronic toxicity test methods, the test used at least four toxicant concentrations and a control because the methods permitted, in the recent past, the use of only four concentrations.

4. EPA personnel or an EPA contractor calculated the effect concentration, verified that all test acceptability criteria (TAC) had been met, and verified that the statistical flowchart had been followed correctly. Thus, all summary statistics and estimates were calculated from the replicate data and strictly followed the most current EPA test methods.

Details of data quality assurance and test acceptance are provided in a separate document, available at EPA's Office of Water docket, located in the Office of Science and Technology ["Whole Effluent Toxicity (WET) Data Test Acceptance and Quality Assurance Protocol"]. An attachment to that document provides a laboratory-by-laboratory listing of quality assurance flags, test dates, and toxicant concentrations, as well as summary statistics by laboratory for the NOEC, EC25, and LC50 estimates and test endpoints (survival, growth, reproduction, etc.). Laboratories are not identified by name.

The data set of reference toxicant tests includes information from 75 laboratories for 23 methods for tests conducted between 1988 and 1999. This document addresses, and provides specific guidance on, the variability of methods promulgated by EPA in 40 CFR Part 136 (Table 3-1). The data are also used to develop between-laboratory interim estimates of method variability for the promulgated methods (Appendix A). The Agency identifies these CVs as "interim;" EPA may revise some or all of these estimates based on between-laboratory studies to evaluate some of the promulgated test methods.

The next section presents summary statistics for the promulgated methods. Summary statistics for all methods in the data set appear in Appendix B. For methods represented by a few laboratories, summary statistics should not be considered representative of method performance. For example, EPA's Office of Water usually relies on acceptable data from at least six laboratories (USEPA 1996b) when it conducts a multi-laboratory study to quantify method performance. The data used here have not been obtained under conditions as rigorous as those applied to a between-laboratory study and for that reason, may overestimate variability, particularly for the extremes.

Coefficients of variation are used as descriptive statistics for NOECs in this document. Because NOECs can take on only values that correspond to concentrations tested, the distribution (and CV) of NOECs can be influenced by the selection of experimental concentrations, as well as additional factors (e.g., within-test variability) that affect both NOECs and point estimates. This makes CVs for NOECs more uncertain than the CVs for point estimates, and the direction of this uncertainty is not uniformly toward larger or smaller CVs. Despite these confounding issues, CVs are used herein as the best available means of expressing the variability of interest in this document and for general comparisons among methods. Readers should be cautioned, however, that small differences in CVs between NOECs and point estimates may be artifactual; large differences are more likely to reflect real differences in variability (a definition of what is "small" or "large" would require a detailed statistical analysis and would depend upon the experimental and statistical details surrounding each comparison). NOECs can only be a fixed number of discrete values; the mean, standard deviation, and CV cannot be interpreted and applied as they are for a continuous variable such as the EC25 or EC50. For instance, the typical reference toxicant test might result in only three observed NOEC values, most of them at one or two concentrations. The mean will fall between tested concentrations, as will the stated confidence intervals; thus, these do not actually represent expected outcomes, only approximations of the expected outcome.

As an alternative to CVs, ratios are used to quantify variability of EC25, EC50, and NOEC measurements in Appendix B. Ratios of measurements have been used previously to quantify and compare variability of NOEC and EC50 (Chapman et al. 1996b, Dhaliwal et al. 1997).

3.2 Variability of EC25, LC50, and NOEC

3.2.1 Within-Laboratory Variability of EC25, LC50, and NOEC

This section characterizes the within-test and within-laboratory variability of effect concentration estimates. Tables 3-2 through 3-4 summarize variation across laboratories of the within-laboratory coefficients of variation (CVs), without respect to reference toxicant tested. Tables showing more extensive summaries appear in Appendix B (Tables B-1 through B-3).

Table 3-1. Promulgated WET Methods Included in This Report

Test Method No.	Test Method	EPA Data Base		
		Toxicants	Tests	Labs
Freshwater Methods for Chronic Toxicity[a]				
1000.0	*Pimephales promelas*, Fathead Minnow Larval Survival and Growth Test	Cd, Cr, Cu, KCl, NaCl, NaPCP, SDS	205	19
1000.0	*Pimephales promelas*, Fathead Minnow Embryo-Larval Survival and Teratogenicity Test		0	0
1002.0	*Ceriodaphnia dubia*, Water Flea Survival and Reproduction Test	Cd, Cu, KCl, NaCl, NaPCP	393	33
1003.0	*Selenastrum capricornutum*,[b] Green Alga Growth Test	Cu, NaCl, Zn	85	9
Marine & Estuarine Methods for Chronic Toxicity[c]				
1004.0	*Cyprinodon variegatus*, Sheepshead Minnow Larval Survival and Growth Test	Cd, KCl	57	5
1005.0	*Cyprinodon variegatus*, Sheepshead Minnow Embryo-larval Survival and Teratogenicity Test		0	0
1006.0	*Menidia beryllina*, Inland Silverside Larval Survival and Growth Test	Cr, Cu, KCl, SDS	193	16
1007.0	*Americamysis (Mysidopsis) bahia*, Mysid Survival, Growth, and Fecundity Test	Cr, Cu, KCl	130	10
1008.0	*Arbacia punctulata*, Sea Urchin Fertilization Test		0	0
1009.0	*Champia parvula*, Red Macroalga Reproduction Test	Cu, SDS	23	2
Methods for Acute Toxicity [d,e]				
2000.0	Fathead Minnow Survival Test	Cd, Cu, KCl, NaCl, NaPCP	217	21
2002.0	*Ceriodaphnia dubia* Survival Test	Cd, Cu, KCl, NaCl, NaPCP	241	23
2004.0	Sheepshead Minnow Survival Test	SDS	65	3
2006.0	Inland Silverside Survival Test	Cd, KCl, SDS	48	5
2007.0	Mysid (*A. bahia*) Survival Test	Cd, Cu, SDS	32	3
2011.0	Mysid (*H. costata*) Survival Test	Cd, SDS	14	2
2019.0	Rainbow Trout Survival Test	Cu, Zn	10	1
2021.0	*Daphnia magna* Survival Test	Cd	48	5
2022.0	*Daphnia pulex* Survival Test	Cu, NaCl, SDS Cd, Cu, NaCl, NaPCP	57	6

[a] See publications EPA/600/4-89-001 (USEPA 1989) and EPA/600/4-91-002 (USEPA 1994b).
[b] The genus and species names for *Selenastrum capricornutum* have been changed to *Raphidocelis subcapitata*. In this document, however, *Selenastrum capricornutum* is used to avoid confusion.
[c] See publication EPA/600/4-91-003 (USEPA 1994a) and EPA/600/4-87/028 (USEPA 1988).
[d] See publications EPA/600/4-85/013 (USEPA 1985) and EPA/600/4-90/027F (USEPA 1993).
[e] EPA did not assign method numbers for acute methods in EPA/600/4-90/027F. The numbers assigned here were created for use in this document and in related materials and data bases.

Reference toxicant codes:

Cd	cadmium	NaCl	sodium chloride
Cr	chromium	NaPCP	sodium pentachlorophenate
Cu	copper	SDS	sodium dodecyl sulfate
KCl	potassium chloride	Zn	zinc

Table 3-2. Quartiles (25th and 75th) and Median (50th) of the Within-Laboratory Values of CV for EC25 (Chronic Tests)

Test Method[a]	Test Method No.	Endpoint[b]	No. of Labs	Percentiles of CV		
				25th	50th	75th
Fathead Minnow Larval Survival & Growth	1000.0	G	19	0.21	0.26	0.38
Fathead Minnow Larval Survival & Growth	1000.0	S	16	0.11	0.22	0.32
Ceriodaphnia (Cd) Survival & Reproduction	1002.0	R	33	0.17	0.27	0.45
Ceriodaphnia (Cd) Survival & Reproduction	1002.0	S	25	0.11	0.23	0.41
Green Alga (*Selenastrum*) Growth	1003.0	G	6	0.25	0.26	0.39
Sheepshead Minnow Larval Survival & Growth	1004.0	G	5	0.09	0.13	0.14
Sheepshead Minnow Larval Survival & Growth	1004.0	S	2	0.15	0.16	0.17
Inland Silverside Larval Survival & Growth	1006.0	G	16	0.18	0.27	0.43
Inland Silverside Larval Survival & Growth	1006.0	S	13	0.22	0.35	0.42
Mysid (Ab) Survival, Growth, & Fecundity	1007.0	F	4	0.30	0.38	0.41
Mysid (Ab) Survival, Growth, & Fecundity	1007.0	G	10	0.24	0.28	0.32
Mysid (Ab) Survival, Growth, & Fecundity	1007.0	S	7	0.17	0.21	0.28
Red Macroalga (*Champia parvula*) Reproduction	1009.0	R	2	0.58	0.58	0.59

[a] Cd = *Ceriodaphnia dubia*, Ab = *Americamysis* (*Mysidopsis*) *bahia*
[b] G = growth, S = survival, R = reproduction, F = fecundity

Table 3-3. Quartiles (25th and 75th) and Median (50th) of the Within-Laboratory Values of CV for LC50

Test Method[a]	Test Method No.	Endpoint[b]	No. of Labs	Percentiles of CV		
				25th	50th	75th
Freshwater Methods for Chronic Toxicity[c]						
Fathead Minnow Larval Survival & Growth	1000.0	S	19	0.15	0.23	0.31
Ceriodaphnia (Cd) Survival & Reproduction	1002.0	S	33	0.10	0.16	0.29
Sheepshead Minnow Larval Survival & Growth	1004.0	S	5	0.07	0.08	0.12
Inland Silverside Larval Survival & Growth	1006.0	S	16	0.16	0.28	0.35
Mysid (Ab) Survival, Growth, & Fecundity	1007.0	S	10	0.16	0.26	0.27
Methods for Acute Toxicity[d,e]						
Fathead Minnow Larval Survival	2000.0	S	21	0.10	0.16	0.19
Ceriodaphnia (Cd) Survival	2002.0	S	23	0.11	0.19	0.29
Sheepshead Minnow Survival	2004.0	S	5	0.12	0.14	0.21
Inland Silverside Larval Survival	2006.0	S	5	0.15	0.16	0.21
Mysid (Ab) Survival	2007.0	S	3	0.17	0.25	0.26
Mysid (Hc) Survival	2011.0	S	2	0.27	0.30	0.34
Rainbow Trout Survival	2019.0	S	1	0.23	0.23	0.23
Daphnia (Dm) Survival	2021.0	S	5	0.07	0.22	0.24
Daphnia (Dp) Survival	2022.0	S	6	0.19	0.21	0.27

[a] Cd = *Ceriodaphnia dubia*, Ab = *Americamysis* (*Mysidopsis*) *bahia*, Hc = *Holmesimysis costata*, Dm = *Daphnia magna*, Dp = *Daphnia pulex*
[b] S = survival
[c] See publications EPA/600/4-89-001 (USEPA 1989) and EPA/600/4-91-002 (USEPA 1994b).
[d] See publications EPA/600/4-85-013 (USEPA 1985 and EPA/600/4-90/027F (USEPA 1993).
[e] EPA did not assign method numbers for acute methods in EPA/600/4-90/027F. The numbers assigned here were created for use in this document and in related materials and data bases.

Table 3-4. Quartiles (25th and 75th) and Median (50th) of the Within-Laboratory Values of CV for NOEC

Test Method[a]	Test Method No.	Endpoint[b]	No. of Labs	Percentiles of CV		
				25th	50th	75th
Freshwater Methods for Chronic Toxicity[c]						
Fathead Minnow Larval Survival & Growth	1000.0	G	19	0.22	0.37	0.53
Fathead Minnow Larval Survival & Growth	1000.0	S	19	0.26	0.39	0.48
Ceriodaphnia (Cd) Survival & Reproduction	1002.0	R	33	0.25	0.33	0.49
Ceriodaphnia (Cd) Survival & Reproduction	1002.0	S	33	0.21	0.30	0.43
Green Alga (*Selenastrum*) Growth	1003.0	G	9	0.40	0.46	0.56
Marine & Estuarine Methods for Chronic Toxicity[d]						
Sheepshead Minnow Larval Survival & Growth	1004.0	G	5	0.34	0.40	0.44
Sheepshead Minnow Larval Survival & Growth	1004.0	S	5	0.14	0.18	0.24
Inland Silverside Larval Survival & Growth	1006.0	G	16	0.31	0.46	0.57
Inland Silverside Larval Survival & Growth	1006.0	S	16	0.30	0.42	0.55
Mysid (Ab) Survival, Growth, & Fecundity	1007.0	F	4	0.17	0.36	0.40
Mysid (Ab) Survival, Growth, & Fecundity	1007.0	G	10	0.35	0.39	0.43
Mysid (Ab) Survival, Growth, & Fecundity	1007.0	S	10	0.28	0.33	0.38
Red Macroalga (*Champia parvula*) Reprod.	1009.0	R	2	0.85	1.00	1.16
Methods for Acute Toxicity[e,f]						
Fathead Minnow Larval Survival	2000.0	S	21	0.18	0.22	0.34
Ceriodaphnia (Cd) Survival	2002.0	S	23	0.18	0.35	0.41
Sheepshead Minnow Survival	2004.0	S	3	0	0.31	0.33
Inland Silverside Larval Survival	2006.0	S	5	0	0.33	0.35
Mysid (Ab) Survival	2007.0	S	3	0.29	0.38	0.43
Mysid (Hc) Survival	2011.0	S	2	0.21	0.26	0.31
Rainbow Trout Survival	2019.0	S	1	0.35	0.35	0.35
Daphnia magna (Dm) Survival	2021.0	S	5	0.09	0.36	0.47
Daphnia pulex (Dp) Survival	2022.0	S	6	0.21	0.38	0.61

[a] Cd = *Ceriodaphnia dubia*, Ab = *Americamysis* (*Mysidopsis*) *bahia*, Hc = *Holmesimysis costata*, Dm = *Daphnia magna*, Dp = *Daphnia pulex*
[b] G = growth, S = survival, R = reproduction, F = fecundity
[c] See publications EPA/600/4-89-001 (USEPA 1989) and EPA/600/4/4-91-002 (USEPA 1994b).
[d] See publication EPA/600/4-91-003 (USEPA 1994a) and EPA/600/4-87/028 (USEPA 1988).
[e] See publications EPA/600/4-85/013 (USEPA 1985) and EPA/600/4-90/027F (USEPA 1993).
[f] EPA did not assign method numbers for acute methods in EPA/600/4-90/027F. The numbers assigned here were created for use in this document and in related materials and data bases.

Effect concentrations having a p-percent effect are symbolized as ECp and may be calculated for sublethal and lethal (survival) endpoints (USEPA 1993,1994a,1994b). Effect concentrations commonly estimated for WET methods are LC50, EC50, IC25, and EC25. The symbol ECp is more general and may be used to represent an LCp, ECp, or ICp endpoint. To simplify presentation of results in this document, the term EC25 is used to represent the concentration at which a 25-percent effect has occurred for either lethal

or sublethal endpoints. The term LC50 is used to represent the concentration at which a 50-percent effect has occurred for lethal endpoints. The EC25 for survival is not routinely used in generating self-monitoring data and is presented here for comparison to the EC25 for sublethal endpoints (i.e., IC25). Estimates of EC25, LC50, and NOEC were calculated for this document as required in the EPA test methods (USEPA 1993, 1994a, 1994b). A CV is reported for NOEC measurements in this document. See Appendix A for further details.

The results in Tables 3-2 through 3-4 were obtained as follows, using as an example the EC25 of the growth endpoint in Method 1000.0 (fathead minnow larval chronic test) on the first row of Table 3-2. The CV of the EC25 estimates was calculated for each laboratory. This calculation resulted in 19 CVs (one per laboratory with each laboratory tested using one toxicant). The sample percentiles were calculated for this set of 19 CVs. In Table 3-2, the column headed "50th" shows the 50th percentile (median value) of CV found across these 19 laboratories; the 50th percentile value is 0.26. In the column headed "75th," the 75th percentile CV is reported as 0.38. When a method is represented by fewer than four laboratories, the minimum and maximum CVs are shown in the columns headed "25th" and "75th," respectively. Note that these CVs represent within-laboratory variability, and that Tables 3-2 through 3-4 show the quartiles and median of the within-laboratory CVs. These tables thus report the typical range of within-laboratory test method variation.

Variation across laboratories in the CV for effect concentration estimates (Tables 3-2 through 3-4) may be summarized as follows, ignoring methods represented by only one or two laboratories. [Refer to the column headed "75th" (the 75th percentile).]

For the EC25 of the growth and reproduction endpoints in chronic toxicity tests, 75 percent of laboratories have a CV no more than 0.14 to 0.45 depending on the method (Table 3-2). For the two most commonly used methods (1000.0, fathead minnow larval chronic test; and 1002.0, *Ceriodaphnia* chronic test), 75 percent of the laboratories have CVs no more than 0.38 and 0.45, respectively.

For the LC50 of the survival endpoint in chronic toxicity tests, 75 percent of laboratories have a CV no more than 0.12 to 0.35, depending on the method. For the two most commonly used methods (1000.0 and 1002.0), 75 percent of laboratories have CVs no more than 0.31 and 0.29, respectively (Table 3-3). For the LC50 in acute toxicity tests, 75 percent of laboratories have a CV no more than 0.19 to 0.29, depending on the method. For the two most commonly used methods (2000.0 and 2002.0), 75 percent of laboratories have CVs no more than 0.19 and 0.29, respectively.

For the NOEC of growth or reproduction endpoints in chronic toxicity tests, 75 percent of laboratories have a CV no more than 0.43 to 0.57, depending on the method. For the two most commonly used methods (1000.0 and 1002.0), 75 percent of laboratories have CVs no more than 0.53 and 0.49, respectively (Table 3-4). For the NOEC of survival in chronic toxicity tests, 75 percent of laboratories have a CV no more than 0.24 to 0.55, depending on the method. For the two most commonly used methods (1000.0 and 1002.0), 75 percent of laboratories have CVs no more than 0.48 and 0.43, respectively. For the NOEC of survival in acute toxicity tests, 75 percent of laboratories have a CV no more than 0.34 to 0.61, depending on the method. For the two most commonly used acute methods (2000.0 and 2002.0), 75 percent of laboratories have CVs no more than 0.34 and 0.41, respectively.

Appendix B discusses the range of toxicant concentrations reported as the NOEC. For chronic toxicity tests, most laboratories report the NOEC to within two to three concentration intervals, and half the laboratories report most NOECs within one to two concentration intervals for reference toxicants. For acute toxicity tests, most laboratories report NOECs at one or two concentrations. This outcome agrees with EPA's expected performance for these methods. The normal variation of the effect concentration estimate in reference toxicant tests has been reported for some EPA WET methods (USEPA 1994a, 1994b) to be plus or minus one dilution concentration for the NOEC and less for LC50.

3.2.2 Between-Laboratory Variability of EC25, LC50, and NOEC

The data set compiled for this document provided reasonable estimates of between-laboratory variability for only a few methods. For many methods and toxicants, there were too few laboratories in the data base. Additional summaries of between-laboratory variability of WET methods are included in the TSD (USEPA 1991a, Part 1.3.3) and the WET methods manuals (USEPA 1994a, 1994b). EPA also intends to provide new data in a forthcoming EPA between-laboratory study of promulgated methods.

Using the data set, credible estimates of between-laboratory variability could be made for a few toxicants and methods having data for six or more laboratories (Table 3-5). The statistical methods are described in Appendix B. Table 3-5 shows values of the square root of within-laboratory and between-laboratory variance components (i.e., standard deviations, σ). The standard deviations and mean are expressed in units of toxicant concentration (e.g., g/L or mg/L). Between-laboratory σ_b estimates the standard deviation for laboratory means of EC25, LC50, and NOEC. The "Mean" column in Table 3-5 shows the mean of the laboratory means, not the mean for all tests. Because the number of tests differed among laboratories, these two means are different. These data suggest that between-laboratory variability (σ_b) is comparable to within-laboratory variability (σ_w) for the methods listed in the table.

In Table 3-5, the ratio of σ_b to the mean is an estimate of the relative variability (CV_b) of laboratory means around their combined mean. The ratio of σ_w to the mean may approach the value of the average within-laboratory CV when the sample of laboratories is large, but to characterize within-laboratory CVs, readers should use Tables 3-2 through 3-4.

Table 3-5. Estimates of Within-Laboratory and Between-Laboratory Components of Variability[a]

Test Method[b]	Test EC Estimate	Toxicant	End-Point[c]	Tests	Labs	Within-lab σ_w	Between-lab σ_b	Mean	CV_w	CV_b
1000.0	EC25	NaCl	G	73	6	0.67	0.44	2.63	0.25	0.17
1000.0	LC50	NaCl	S	73	6	1.14	0.45	4.15	0.27	0.11
1000.0	NOEC	N Cl	G	73	6	0.72	0.35	2.18	0.33	0.16
1000.0	NOEC	NaCl	S	73	6	0.96	0.51	2.43	0.40	0.21
1002.0	EC25	NaCl	R	292	23	0.29	0.27	0.92	0.32	0.29
1002.0	LC50	NaCl	S	285	23	0.48	0.24	1.78	0.27	0.13
1002.0	NOEC	NaCl	G	292	23	0.28	0.18	0.74	0.38	0.24
1002.0	NOEC	NaCl	S	292	23	0.47	0.26	1.42	0.33	0.18
1006.0	EC25	Cu	G	130	9	45.1	52.4	97.4	0.46	0.54
1006.0	LC50	Cu	S	130	9	48.4	70.7	127.0	0.38	0.56
1006.0	NOEC	Cu	G	130	9	51.8	44.4	80.1	0.65	0.55
1006.0	NOEC	Cu	S	130	9	34.2	39.5	65.4	0.52	0.60
2000.0	LC50	NaCl	S	154	14	1.05	1.24	7.46	0.14	0.17
2002.0	LC50	NaCl	S	167	15	0.36	0.38	1.97	0.18	0.19

[a] σ_w = within-laboratory standard deviation, σ_b = between-laboratory standard deviation
CV_w = within-laboratory coefficient of variation, CV_b = between-laboratory coefficient of variation
[b] EPA did not assign method numbers for acute methods in EPA/600/4-90/027F. The numbers assigned here were created for use in this document and in related materials and data bases.
[c] G = growth, S = survival, R = reproduction

3.3 Variability of Endpoint Measurements

This section characterizes the within-laboratory precision of endpoint measurements (e.g., growth, reproduction, and survival). Endpoint variability in methods for chronic toxicity is characterized here using sublethal endpoints. The sublethal endpoint was designed to be more sensitive than the survival endpoint, and it incorporates the effect of mortality (i.e., it incorporates biomass). For example, for the chronic survival and growth fathead minnow larval test, the total dry weight at each replicate is divided by the original number of larvae, rather than the surviving number of larvae.

EPA reports measures of test precision based on the control CV [(control standard deviation)/(control mean)] and the "Percent MSD" [100×MSD/(control mean)], symbolized as PMSD. Recall that MSD, the "minimum significant difference," is calculated as [$d \sqrt{EMS} \sqrt{(2/r)}$], where "d" is the critical value of Dunnett's statistic when comparing "k" treatments to a control, EMS is the error mean square from the analysis of variance of the endpoint responses, and "r" is the number of replicates at each concentration (USEPA 1993, 1994a, 1994b). These measures of test precision quantify within-test variability, or the sensitivity of each test to toxic effects on the biological endpoint.

Measures of variability relative to the control mean are used for two reasons. First, a laboratory having consistently large mean endpoint values for the control will also tend to have larger values of MSD and control standard deviation. Second, PMSD is readily interpreted as the minimum percent difference between control and treatment that can be declared statistically significant in a WET test. A significant effect occurs when (control mean - treatment mean) exceeds the MSD. Dividing by the control mean and multiplying by 100 states this relationship in terms of the percent difference between control and treatment.

To characterize the distribution of values of PMSD, values from all laboratories and toxicants for a given method and endpoint were combined, and sample percentiles reported. Percentiles are also reported for the CV of the control, which also indicates variability among replicates under non-toxic conditions and may be a useful indicator of uniformity of the test organisms. The sample percentiles are reported in more detail in Appendix B; the 10th and 90th percentiles are shown in Table 3-6. Method 1009.0 (red macroalga) is omitted from Table 3-6 because it would be inadvisable to characterize method variability using only 23 tests from only two laboratories.

The 90th percentile may be used as an upper PMSD bound (i.e., a limit on the insensitivity of a test). The 10th percentile may be used as a lower PMSD bound for declaring a significant difference or a lower limit to test sensitivity. The 90th percentile has been used in other WET programs (Chapter 5). The 95th percentile is used as a practical upper limit for the variability of analytical results in well-controlled between-laboratory studies that use a standard protocol and specific quality assurance procedures (ASTM 1992, 1998; USEPA 1993, 1996a, 1996b). The tests summarized here have not been subjected to the rigorous standardization and quality assurance of collaborative studies, and the data have not been screened for outliers as specified by ASTM Practices D2777 and E691 (ATSM 1992, 1998). These considerations justify using the sample 90th percentile to set an upper bound. A lower bound is necessary to avoid creating a disincentive for improving test precision and to objectively specify a limit to the test sensitivity achieved in practice. If no more than ten percent of tests are more precise than this lower bound, then in practice, the analytical method rarely detects toxic effects of this small magnitude.

When comparing values in Table 3-6 to a test result, it is important that the test's MSD be calculated according to procedures described in the EPA method manuals (USEPA 1993, 1994a, 1994b) for Dunnett's test for multiple comparisons with a control (see Section 6.4.1). An analysis of variance (ANOVA) is conducted using several treatments, including the control. EPA methods require excluding from the ANOVA those concentrations for which no organisms survived in any replicate. For a sublethal endpoint, concentrations are excluded from the analysis if they exceed the NOEC for survival. The MSD is calculated

using the square root of the error mean square (rEMS) from the ANOVA, and using Dunnett's critical value (which depends on the number of replicates and concentrations used in the ANOVA).

Table 3-6. Range of Relative Variability for Endpoints of Promulgated WET Methods, Defined by the 10th and 90th Percentiles from the Data Set of Reference Toxicant Tests[a]

Test Method[b]	Endpoint[c]	No. of Labs	No. of Tests	PMSD		Control CV[d]	
				10th	90th	10th	90th
1000.0 Fathead Minnow	G	19	205	9.4	35	0.035	0.20
1002.0 *Ceriodaphnia dubia*	R	33	393	11	37	0.089	0.42
1003.0 Green Alga	G	9	85	9.3	23	0.034	0.17
1004.0 Sheepshead Minnow	G	5	57	6.3	23	0.034	0.13
1006.0 Inland Silverside	G	18	193	12	35	0.044	0.18
1007.0 Mysid	G	10	130	12	32	0.088	0.28
2000.0 Fathead Minnow	S	20	217	4.2	30	0	0.074
2002.0 *Ceriodaphnia*	S	23	241	5.0	21	0	0.11
2004.0 Sheepshead Minnow	S	5	65	0[e]	55	0	0
2006.0 Inland Silverside	S	5	48	7.0	41	0	0.079
2007.0 Mysid (*A. bahia*)	S	3	32	5.1	26	0	0.081
2011.0 Mysid (*H. costata*)	S	2	14	18	47	0	0.074
2021.0 Daphnia (*D. magna*)	S	5	48	5.3	23	0	0.11
2022.0 Daphnia (*D. pulex*)	S	6	57	5.8	23	0	0.11

[a] The precision of the data warrants only three significant figures. When determining agreement with these values, one may round off values to two significant figures (e.g., values >3.45000... and ≤3.5000... are rounded to 3.5). Method 1009.0 (red macroalga) is not reported because it is inadvisable to characterize method variability using only 23 tests from just two laboratories.

[b] EPA did not assign method numbers for acute methods in EPA/600/4-90/027F. The numbers assigned here were created for use in this document and in related materials and data bases.

[c] G = growth, R = reproduction, S = survival

[d] CVs were calculated using untransformed control means for each test.

[e] An MSD of zero will not occur when the EPA flow chart for statistical analysis is followed. In this report, MSD was calculated for every test, including those for which the flow chart would require a nonparametric hypothesis test. EPA recommends using the value 4.2 (the 10th percentile shown for the fathead minnow acute test) in place of zero as the 10th percentile PMSD (lower PMSD bound) for the sheepshead minnow acute test.

The MSD was calculated for all test results reported here, including those for which non-normality and heterogeneity of variance were indicated. Thus, this document presents MSD as an approximate index of test sensitivity. Estimates of power are also approximate. The MSD generally will be related to test sensitivity, even when the assumptions for ANOVA and Dunnett's test are not strictly satisfied.

Table 3-7 shows the number of laboratories in the WET variability data set having tests exceeding the upper PMSD bound reported in Table 3-6. One-half to two-thirds of the laboratories never or infrequently exceeded the bound, and roughly one in five exceeded it in at least 20 percent of their tests. By definition of the 90th percentile, about 10 percent of all the tests exceeded the bound.

Table 3-7. Number of Laboratories Having a Given Percent of Tests Exceeding the PMSD Upper Bound for the Sublethal Endpoint

Test Method	No. Labs	Endpoints[a]	Number of Labs with Various Percentages of Tests Exceeding the PMSD Upper Bound				
			0%	0%-10%	10%-20%	20%-50%	50%-100%
1000.0 Fathead Minnow	19	G	8	2	7	2	0
1002.0 *Ceriodaphnia dubia*	33	R	15	7	5	6	0
1003.0 Green Alga	9	G	6	1	0	2	0
1004.0 Sheepshead Minnow	5	G	3	1	0	1	0
1006.0 Inland Silverside	16	G	6	5	1	4	0
1007.0 Mysid (growth)	10	G	5	2	0	3	0

[a] G = growth, R = reproduction

3.4 Conclusions about Variability of WET Methods

3.4.1 Variability of EC25, LC50, NOEC

For EC25, the quartiles of the within-laboratory CVs ranged across the promulgated methods from 0.09 to 0.45, and the median CV ranged from 0.13 to 0.38. For LC50, the quartiles of the within-laboratory CVs ranged from 0.07 to 0.35, and the median CV ranged from 0.08 to 0.28. For NOEC, the quartiles of the within-laboratory CVs ranged from 0 to 0.61, and the median CV ranged from 0.18 to 0.46. This summary applies to those methods represented by at least 20 tests and three laboratories.

EPA concludes from Tables 3-2 through 3-4 that point estimates are substantially less variable than the NOEC for the same method and endpoint, and that the LC50 for an acute toxicity test usually is less variable than the LC50 for a chronic toxicity test. The estimated NOEC is more variable than ECp *using current experimental designs* because NOEC can take only those values equal to the concentrations tested, while ECp interpolates between tested concentrations (there may be other, more technical reasons as well). In principle, NOEC could be estimated more accurately and precisely by changing the experimental design to use more concentrations at narrower dilution ratios and by using more replicates. The greater variability of the NOEC underscores the desirability of using point estimates to characterize effluent toxicity.

Tables 3-2 through 3-4 may be used as benchmarks for variability, allowing comparison of one laboratory's CV for reference toxicant testing with CVs reported by experienced laboratories reporting tests that passed the TAC. However, CVs for methods represented by too few laboratories in the table may be atypical.

The CVs in Tables 3-2 through 3-4 may be used as an adjunct to the control chart. If the CV for reference toxicant tests is above the 75th percentile in Tables 3-2 through 3-4, variability likely can be reduced, even if the individual EC25 or LC50 values fall within the control limits. If a control chart is constructed using an unreasonably large standard deviation, the control limits will be unreasonable. If a high CV is not fully explained by an unusually small mean, the standard deviation of EC25 or LC50 should be reduced to bring the CV within the normal range. If the CV exceeds the 90th percentile (Appendix B), there is no question that variability is unacceptably large. Detailed guidance is provided in Chapter 5 (Section 5.3.1.1).

Tables 3-2 through 3-4 indicate the magnitude of the analytical variability that becomes part of the variability of effluent test results under certain conditions. This occurs when effluent test results (NOECs, LC50s, or EC25s) fall between the lowest and highest concentrations tested. Under other conditions, these

CVs may not accurately represent analytical variability. If tests give results consistently near or at the lowest or highest concentrations tested, or if the tests often produce "less than" or "greater than" results, Tables 3-2 through 3-4 will not accurately characterize the analytical CV for such tests. To measure the analytical CV under such conditions, reference toxicant tests would have to be designed to have the effect concentration at or near the lowest or highest concentration. The CV and standard deviation measured under such conditions are unknown, but are likely to differ from those for standard reference toxicant tests.

The data set did not contain information supporting an analysis of the causes of between-laboratory variability. Possible causes may include laboratory differences in concentration series, incorrect or ambiguous calculation or reporting of concentrations (e.g., concentration of the metal ion versus the salt), laboratory differences in dilution water (e.g., water hardness or pH), laboratory differences in foods and feeding regimes, and laboratory differences in cultures (genotypic and phenotypic differences in sensitivity to various toxicants).

The lack of a standard or common reference toxicant creates a problem for permittees and regulatory authorities attempting to evaluate and compare laboratories. Real or apparent differences occur between laboratories in the mean values of EC25, LC50, and NOEC. Some of this difference is random and reflects only the within-laboratory variance; some may be systematic. Systematic, between-laboratory differences can be inferred reliably only when laboratories use the same test method, use the same reference toxicants and dilution series, use similar dilution waters, and report a sufficient number of tests.

3.4.2 Variability of Endpoint Measurements

EPA has selected the PMSD to characterize endpoint variability for WET test methods because it integrates variability from several concentrations (always including the control), and it represents the MSD used in the WET hypothesis test. The control CV, by itself, does not fully represent the variability affecting a WET hypothesis test or point estimate. The PMSD also represents the variability affecting point estimates because it is calculated using the EMS for the endpoint measurement. (However, the standard error of a point estimate of an effect concentration may be a complicated function of the EMS.)

PMSD for sublethal endpoints ranged from 6 to 37 across the promulgated chronic methods. For the fathead minnow chronic method, PMSD ranged from 9 to 35; for the *Ceriodaphnia* chronic method, PMSD ranged from 11 to 37. Thus, most chronic tests were able to distinguish a reduction of 37 percent or smaller in the endpoint. Further analysis in Chapter 5 shows that most tests were unable to distinguish consistently a 25-percent reduction. For the survival endpoint of promulgated acute methods, PMSD ranged from 0 to 55. For the two most commonly used acute methods (fathead minnow and *Ceriodaphnia*), PMSD ranged from 4 to 30 and from 5 to 21, respectively. Thus, PMSD varied markedly for some acute methods and not for others.

As shown by the size of PMSD, test sensitivity to detect substantial toxic effects is occasionally insufficient at some laboratories and routinely insufficient at a few laboratories. Inadequate test sensitivity is not always signaled by control charts of EC25, LC50, and NOEC. Laboratories should consider maintaining control charts for MSD or PMSD, and should report MSD and the control mean with all WET tests.

Some portion of MSDs in the WET variability data set could be considered exceptionally large, if not outliers. This observation underscores the importance of a careful review for each WET test, including an examination of means and standard deviations for endpoint responses at each concentration; the plotting of replicate data (not just concentration means); and, when necessary, a search for possible causes of excessive variability. The tables and plots in the promulgated methods (USEPA 1994a, 1994b) provide good examples.

This page intentionally left blank.

4.0 VARIABILITY IN CONTEXT

EPA manages the regulation of WET in the same way it manages the regulation of chemical-specific pollutants in order to determine reasonable potential (RP), derive permit limits, determine data quality control, and evaluate self-monitoring data. Many similarities between chemical-specific toxicant and WET controls can be found in the TSD (USEPA 1991a). Determining RP in both cases uses many of the same strategies. Permit limit derivation makes similar exposure assumptions and relies on nearly identical toxicological data bases.

Considering a value other than the best analytical estimate as a measure for WET or for specific chemical analytes is inappropriate. All analytical results, in either chemical-specific analyses or WET tests, incorporate some estimated range of uncertainty. While infrequently discussed for chemical methods, uncertainty does play a role in the meaning of analytical results. One end of the confidence interval likely will be less protective of aquatic resources than the other. The derived limit and therefore final reported analytical results become the best estimate of the actual ecological need and assessment of the effect.

Significant debate has occurred over assertions that WET data have too much inherent variability for reliable use in the NPDES program. This debate has engendered considerable evaluation of WET precision. Groups of scientists and individual researchers have repeatedly concluded that currently promulgated WET methods are technically sound and that the observed precision is within the range of precision of other analyses frequently required in NPDES permits (Grothe et al. 1996). The findings of some of the significant sources of these conclusions are summarized below.

4.1 Society of Environmental Toxicology and Chemistry Pellston WET Workshop

The 1995 Society of Environmental Toxicology and Chemistry (SETAC) Pellston Workshop on Whole Effluent Toxicity convened 47 experts in the discipline to assess applied methods and their application in the regulatory process. Representation at the workshop was intentionally balanced among government, business, and academic participants. These scientists published consensus conclusions and recommendations, including the following.

4.1.1 General Conclusions and Recommendations

Grothe et al. (1996) state *"Existing WET testing methods (USEPA 1985, USEPA 1988, USEPA 1989) are technically sound, but certain modifications would improve endpoint interpretation. Such changes involve implementing improvements to currently used statistical procedures, establishing acceptable limits for MSD values, and adding confidence limits to WET test endpoints."*

"A number of problems with WET tests are caused by misapplication of the tests, misinterpretation of the data, lack of competence of the laboratories conducting WET testing, poor condition/health of test organisms, and lack of training of laboratory personnel, regulators, and permittees. More widespread use of WET related guidance provided in USEPA's TSD (1991a) would help alleviate some of these problems. In addition, an effective QA/QC program will improve data quality and reduce test variability."

"Increase training opportunities for regulators and permittees to improve the implementation of WET objectives and to promote national consistency in permitting and compliance issues."

"Implement a broadly based and standardized QA/QC program to improve WET testing performance and data quality."

"Quantify the 'confidence' around test endpoints to improve interpretation of WET test results. Specific statistical methods that could improve precision are presented in Chapter 3 of this document and processes to reduce variability are discussed in Chapter 5. In addition, WET tests should be performed using a dilution series of exposure concentrations to establish a dose-response relationship."

4.1.2 Conclusions about Data Precision

Ausley (1996) compared CVs of chemical analyses and aquatic toxicity tests conducted by North Carolina NPDES permittees. Ausley found that CVs of reported values for chemical analytes (including metals, organic analytes, and non-metal inorganic analytes) ranged from 11.8 percent to 291.7 percent. Coefficients of variation for toxicity parameters (acute and chronic *Ceriodaphnia dubia*, acute and chronic *Pimephales promelas*, acute *Daphnia pulex*, and acute *Mysidopsis bahia*) ranged from 14.8 percent to 67.6 percent. From this review, he concluded that "*the precision of toxicity analyses is within the range of that being reported for commonly analyzed and regulated chemical parameters.*" Ausley highlighted the difficulty in comparing precision estimates of chemical analytes and WET analyses (particularly NOECs), noting that while chemical precision is often determined well above analytical detection, WET precision is often based on the minimum detection level. An assumption that WET precision will vary among toxicants is also logical. To establish "inherent variability," considering toxicants that cause minimal variability in the analysis may be appropriate. The high coefficients of variation for some chemical parameters reported by Ausley reflect the fact that, in practice, analytical precision can vary widely in individual studies in which the effects of a single (or a few) poorly operating laboratory can adversely affect precision estimates. In practice, this kind of data must be screened for quality prior to use to evaluate self-monitoring data or estimates of overall method quality.

Ausley's results closely approximate analytical precision of chemical analytes referenced in the TSD (USEPA 1991a, Chapter 1.2). The CVs for metals (aluminum, cadmium, chromium, copper, iron, lead, manganese, mercury, silver, and zinc) ranged from 18 percent to 129 percent at the low end of the measurement detection range. Between-laboratory CVs for organic analytes ranged from greater than 12 percent to 91 percent. The CVs for non-metal analytes (alkalinity, residual chlorine, ammonia nitrogen, Kjeldahl nitrogen, nitrate nitrogen, total phosphorus, biological oxygen demand, chemical oxygen demand, and total organic carbon) ranged from 4.6 percent to 70 percent in between-laboratory studies of precision.

Burton et al. (1996) concluded that "*USEPA-published methods are functional and appropriate in the context of effluent toxicity control programs.*" They recommended developing limits on within-test variability, a quality assurance and audit program, and guidance for permittee procurement of WET analytical services.

Denton and Norberg-King (1996) cited various studies that favorably compare WET methods with chemical analytical methods (Grothe and Kimerle 1985, Rue et al. 1988, Morrison et al. 1989, Grothe et al. 1990). They proposed that improvements in test result consistency could be accomplished by limiting the range of within-test variability through controls of upper and lower statistical power (e.g., limits on test MSD). Three practices to control within-test variability most effectively are (1) controlling within-test sensitivity, (2) following well-defined test methods, and (3) maintaining communication within the regulatory community. For example, the permittee and regulatory authorities should discuss any facility-specific issues to fully characterize the appropriate permit conditions.

4.2 Water Environment Research Foundation Study

Another publication, "*Whole Effluent Toxicity Testing Program: Evaluation of Practices and Implementation*" (DeGraeve et al. 1998), presents the results of a survey of publicly owned treatment works and State regulatory programs about WET issues. The Water Environment Research Foundation (WERF) sponsored this study. Conclusions by DeGraeve et al. (1998) include the following:

> "*The project team believes that the results demonstrate that the test methods can be routinely completed successfully by well-trained, competent WET testing laboratories and that the results, considered collectively, suggest that the test methods that are being used to measure WET are technically sound.*"
>
> "*There is a need for better training/guidance in WET-related issues for both the regulatory staff responsible for implementing WET requirements and for permittees responsible for meeting WET limits*."

DeGraeve et al. (1998) considered the conclusions of the SETAC Pellston WET publication concurring that between-laboratory CV values of toxicity test methods were low, training of regulatory and permittee staff is needed nationally, and strengthened quality assurance (QA)/quality control (QC) practices could improve performance of analyses. Unlike the SETAC Pellston WET conclusions, they found that there are enough laboratories to meet the current market demand for analyses. Like the SETAC effort, DeGraeve et al. (1998) concluded that a national center of expertise on WET issues would be beneficial to provide guidance to regulatory agencies, permittees, and laboratories.

WERF also funded a project entitled "*Whole Effluent Toxicity Testing Methods: Accounting for Variance*" (Warren-Hicks et al. 1999). This study compared within- and between-laboratory results of reference toxicant test variation as measures of reproducibility and comparability, respectively. The authors concluded that some laboratories could consistently reproduce test results, while others could not and inferred that test precision is a factor of laboratory experience and not inherent methodological weakness. The authors recommended that national studies be conducted to evaluate within- and between-laboratory precision of promulgated WET test methods. (EPA has already initiated this study.) They also recommended that additional test acceptability criteria (TAC), such as upper and lower bounds of MSD, be established and incorporated in the NPDES process. The latter recommendation corroborates other researchers' recommendations discussed above.

4.3 Minimizing Variability by Adhering to WET Toxicity Test Methods

Specific factors that affect variability in WET analyses have been described in several papers (Burton et al. 1996, Ausley 1996, Erickson et al. 1998, Davis et al. 1998). The most important initial consideration in developing precise data is a laboratory's experience and success in performing a specific analysis. Most critical reviews of WET data precision emphasize this initial consideration. Experienced professionals most likely will be able to develop the most consistent and reliable information and can interpret anomalous conditions in the testing or results.

An additional factor in considering WET test method variability is whether the prescribed methods (e.g., the EPA toxicity test methods promulgated in 40 CFR Part 136) are being followed appropriately (see Chapter 5). If tests are submitted that do not meet specified TAC or are produced when laboratory QA testing indicates analyses are beyond control limits, these results should not be used in the NPDES process. Tests performed on effluent samples that have not met required temperature maxima or holding times should not be considered for regulatory purposes. Rigorous QA practices are critical to the success of any analytical program. Both the regulatory authority and permittee should strive to ensure that such practices are in place

for any program developing WET data, whether by national laboratory accreditation, State regulatory certification, direct permittee oversight, or specific contractual agreement with the laboratory.

Comparisons of WET method precision with analytes commonly limited in NPDES permits clearly demonstrate that the promulgated WET methods are within the range of variability experienced in other analyses. Several researchers also noted clear indications that method performance improves when prescribed methods are followed closely by experienced analysts (Grothe et al. 1996, DeGraeve et al. 1998).

A review of WET test results confirms that imprecise WET data are being reported. As with any analytical technique, inexperienced individuals can perform analyses incorrectly or fail to follow appropriate methods and quality assurance practices. Using the training that is available for these methods and quality assurance techniques referenced by this document will help ensure that data of maximum reliability are used and that sound decisions are made based on those results. The Western Coalition of Arid States conducted a study in 1997 (Moore et al. 2000), which reported the results of 16 tests with a non-toxic test sample using the *Ceriodaphnia dubia* chronic test. These results indicated that 43 percent of the tests showed toxicity. EPA is in the process of reviewing the paper and the raw data.

Persons interested in WET issues may consult another source of information developed by the SETAC Whole Effluent Toxicity Expert Advisory Panels. This group, established under a cooperative agreement with EPA, provides scientific opinion and training on WET technical issues. This information is available on the Internet at the SETAC web site, **http://www.setac.org**. Appendix D contains frequently asked questions with answers prepared by the SETAC WET Expert Advisory Panels. The expert panels have identified and discussed various factors that affect WET variability.

4.4 Conclusion

When the variability of WET analyses is viewed in the context of the NPDES program, these techniques produce data that are as precise as those from chemical analyses. As with any other analytical system, lack of experience in performing the analyses, adherence to prescribed QA practices, or good laboratory practices will reduce the precision of the results. Studies of these factors by independent researchers from both the regulatory and regulated communities support these conclusions. While examples of poor-quality, highly variable results from chemical analyses have also been publicized, these results are frequently influenced by the shortcomings mentioned above. Permittees that must generate and use WET data should become well-educated in data quality interpretation, and permittees should require that QC practices be followed by laboratories generating the data. Various sources of information presented in this chapter should assist permittees, testing laboratories, and regulatory authorities with this education process. Examples of practices that can further reduce the imprecision of analyses are also discussed in Chapters 5 and 6 of this document. Additional refinements of TAC can likewise improve test power to detect effects (or the lack thereof) and increase the statistical confidence in results.

5.0 GUIDANCE TO REGULATORY AUTHORITIES, LABORATORIES AND PERMITTEES: GENERATING AND EVALUATING EFFECT CONCENTRATIONS

5.1 Steps for Minimizing Test Method Variability

This chapter provides the background and recommendations on WET test procedures related to sampling, conducting the toxicity test methods, and conducting the statistical methods. Implementing these recommendations should decrease or minimize WET test method variability, thereby increasing confidence to make regulatory decisions (see Figure 5-1). EPA stands behind the technical soundness of the current WET test methods. The critical steps in minimizing WET test method variability are (1) obtaining a representative effluent sample, (2) conducting the toxicity tests properly to generate the biological endpoints, and (3) conducting the appropriate statistical analysis to obtain powerful and technically defensible effect concentrations. Minimizing variability at each step increases the reliability of the WET test results. For example, factors that affect variability include sampling procedures; sample representativeness; deviations from standardized test conditions (e.g., temperature, test duration, feeding); test organisms; source of dilution water; and analyst experience and technique in conducting the toxicity tests properly (Burton et al. 1996).

5.2 Collecting Representative Effluent Samples

The goal of effluent sampling is to obtain a representative sample that reflects real-world biological responses. Factors affecting the representativeness of effluent samples may include the sampling location, frequency, and type (e.g., composite or grab), and sample volume, container, preservation methods, and holding time. Burton et al. (1996) concluded that the above factors considerably influence test result variability.

Effluent samples must be collected at a location that represents the entire regulated flow or discharge. Typically, the sampling site is designated in the discharge permit. As with sampling for any parameter, effluent samples should be collected from a location where the flow is turbulent and well-mixed. Additionally, effluent samples should be collected at a frequency that enables adequate characterization of the discharge over time (e.g., accounts for daily to seasonal changes and variations in effluent quality). Major facilities should conduct WET testing monthly or quarterly, while minor facilities should conduct WET testing semi-annually or annually.

Appropriate sample types should be collected to represent the effluent fully. When the effluent is variable, collecting composite samples may be necessary. When the effluent is less variable, grab samples may be sufficient (e.g., from long-term retention pond facilities).

Sample containers should be non-reactive so that they do not affect sample characteristics. Table II of 40 CFR Part 136 requires that toxicity test samples be collected in glass or plastic containers, as specified in the methods. Sufficient sample volume should be collected for the type of test being conducted, including the number of test dilutions. When samples are collected in Cubitainers®, headspace should be minimized.

Samples must be properly preserved. Part 136 of 40 CFR requires that samples for WET testing be cooled to 4°C when shipped off-site and between test sample renewals. Samples must be cooled during all phases of collection, transportation, and storage to minimize physicochemical changes. Samples must be tested within the specified maximum holding times before significant changes occur, such as volatilization or biological or chemical degradation. If samples are not tested within specified maximum holding times, the test is invalid and must be repeated by collecting a new effluent sample and conducting a new toxicity test to comply with the NPDES permit.

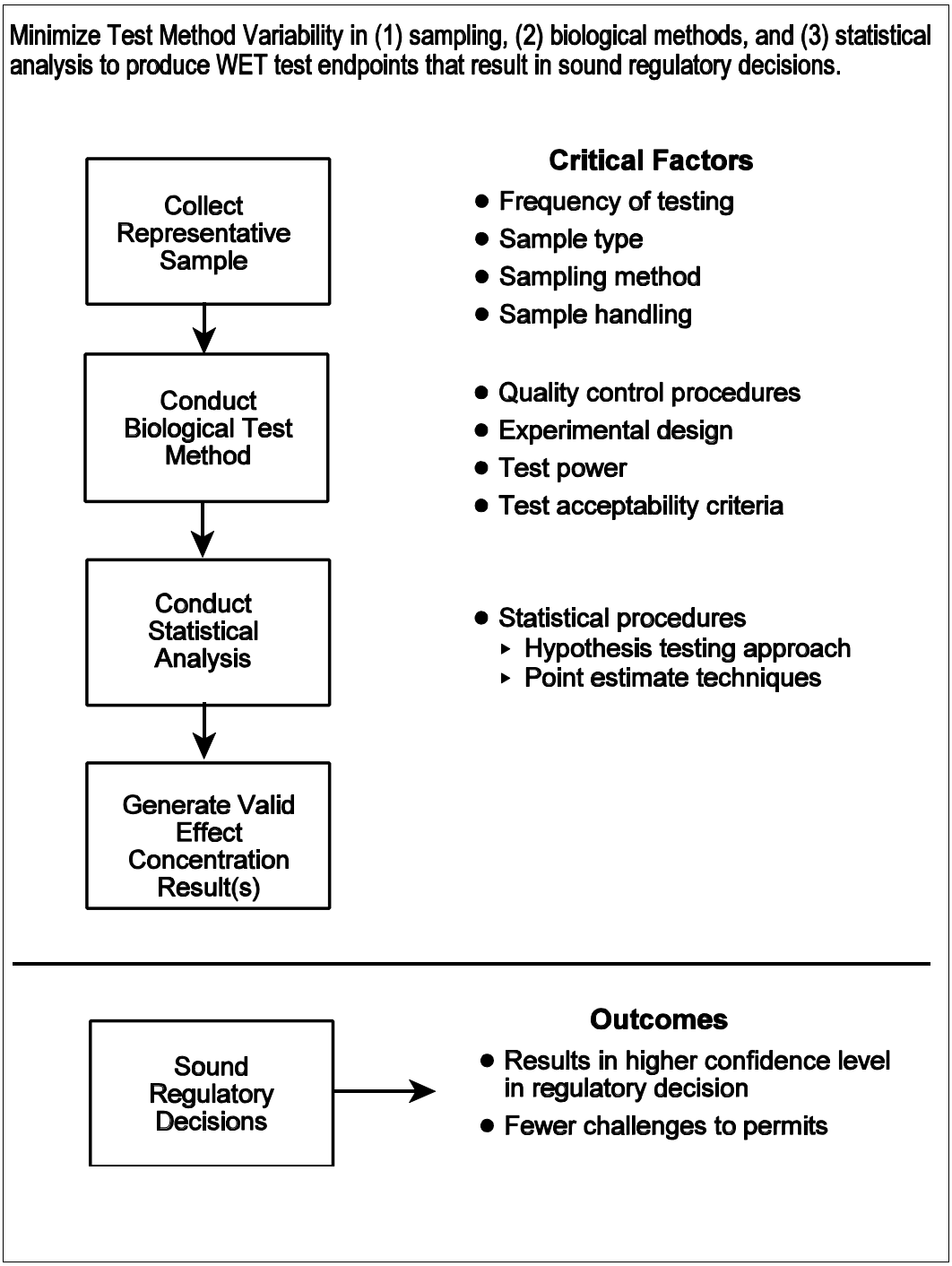

Figure 5-1. Steps to minimize WET test method variability.

5.3 Conducting the Biological Test Methods

Four main components of WET tests afford opportunities to control and minimize variability within tests and within and between laboratories: (1) quality control (QC) procedures; (2) experimental design; (3) test power; and (4) test acceptability criteria (TAC) beyond the minimum requirements specified in EPA's WET test methods.

5.3.1 Quality Control Procedures

Quality assurance (QA) practices for toxicity tests address all aspects of the tests that affect data quality. These practices include effluent sampling and handling, test organism source and condition, equipment condition, test conditions, instrument calibration, replication, use of reference toxicants, recordkeeping, and data evaluation. The EPA WET toxicity testing manuals specify the minimum requirements for each aspect. Regulatory authorities have the discretion to prepare and implement additional guidance beyond the minimum requirements specified in EPA's WET test methods.

An integral part of the QA program is quality control (QC). The QC procedures are the more focused and routine activities conducted under the overall QA program. An important QC component in WET testing is the requirement to conduct reference toxicant tests with effluent tests. The WET test methods outline when reference toxicant tests are to be conducted. (See sections on quality of test organisms in the manuals.) Reference toxicant testing serves two purposes: (1) determine the sensitivity of the test organisms over time; and (2) assess the comparability of within- and between-laboratory test results. Reference toxicant test results can be used to identify potential sources of variability, such as test organism health, differences among batches of organisms, changes in laboratory water or food quality, and performance by laboratory technicians. In the QA section of each promulgated test method (USEPA 1993, 1994a, 1994b), EPA recommends sodium chloride, potassium chloride, cadmium chloride, copper sulfate, copper chloride, sodium dodecyl sulfate, and potassium dichromate as suitable reference toxicants. The methods do not, however, specify a particular reference toxicant or the specific test concentrations for each test method.

The current characterization of WET test method variability is limited by the ability to quantify sources of within- and between-laboratory variability, because laboratories can use different reference toxicants and test concentrations for a particular method. Future evaluations of method variability would be greatly enhanced by having data to analyze from multiple laboratories for the same reference toxicant, the same dilution water at similar pH and hardness, and the same test concentrations. By standardizing reference toxicants, testing laboratories could compare test results, permittees and regulatory authorities could better compare and evaluate laboratories, and the data could be used to further quantify within- and between-laboratory test precision. Specification of the reference toxicant and test concentrations for a method across laboratories would provide a much larger and consistent data base to assess the comparability of within- and between-laboratory test results.

Standardizing reference toxicants and test concentrations has been discussed in the literature. For example, the chronic methods manual for West Coast species (USEPA 1995) specifies the reference toxicant and test concentrations for each test species. The Southern California Toxicity Assessment Group (SCTAG) is comprised of representatives from permittees, testing laboratories, regulatory authorities, and academic institutions that met to discuss technical aspects of WET testing (e.g., standardization of reference toxicants, control charts). The SCTAG (1996) prepared a report to standardize reference toxicants for the chronic freshwater test methods. This report evaluated an extensive data base of reference toxicant data. The report recommended specific reference toxicants and test concentrations for these methods. The SCTAG (1997) also prepared a QA/QC checklist to help toxicity testing laboratories establish and maintain appropriate data quality measures. Regulatory authorities should review these publications when standardizing reference toxicants.

The selection of reference toxicants and test concentrations should be based on specific criteria. The following criteria, recommended in the SCTAG report, provide an excellent basis for selecting standardized reference toxicants:

1. The toxicant should provide precise and reliable measures of toxicological sensitivity.
2. Toxicant disposal should not be legally or environmentally problematic.

3. The toxicant should produce a concentration-response effect for the test organism.
4. The toxicant should be quantifiable.
5. The toxicant should not pose an unacceptable health hazard to laboratory personnel.
6. The toxicant should be readily available.

Most recently, Warren-Hicks et al. (1999) recommended that national acceptance criteria be specified with upper and lower acceptance limits for reference toxicant test results, which all laboratories would need to achieve to obtain accreditation. Variability could decrease nationally if testing laboratories are provided with more detail on the evaluation and interpretation of reference toxicant control charts (APHA-AWWA-WEF 1998). For example, such guidance could describe how to evaluate test results within the warning limits. Both Environment Canada (1990, 2000) and APHA-AWWA-WEF (1998) have prepared guidance on evaluating control chart data. The Environment Canada (2000) report specifies using zinc as an inorganic reference toxicant and phenol as an organic reference toxicant for many aquatic tests. The report also specifies eight criteria for selecting specific reference toxicants.

1. Previous use
2. Availability in a pure form
3. Solubility
4. Stability in solution
5. Stability during storage
6. Ease of analysis
7. Stable toxicity with normal changes in qualities of laboratory water
8. Ability to detect abnormal organisms

Regulatory authorities may want to evaluate the above reports and the SCTAG reference toxicant recommendations for the chronic freshwater test methods. Regulatory authorities may also want to evaluate and recommend a standard reference toxicant and a specific concentration series for each acute and chronic test method using data from this guidance document.

5.3.1.1 Guidance Related to Quality Control Charts and Laboratory Audits

Ausley (1996) recommends some oversight of data quality, such as evaluating tests in meeting QC criteria, using randomization procedures, and operating in allowed reference toxicant ranges to ensure that QC procedures are properly implemented. Another integral component of QC is the maintenance of control charts for reference toxicants and effluents. Laboratories should provide regular review of control charts. EPA suggests keeping a control chart for each combination of test material, test species, test conditions, and endpoints with a maximum of 20 test results. Modern software makes accumulating data and reviewing key test statistics possible with relatively little effort. Elementary methods can identify problems contributing to variability. Laboratories should practice regular control charting of test PMSDs and control performance for all tests along with control charting of effect concentrations such as NOEC and point estimates for reference toxicants tests. Successive tests should be compared occasionally to detect repeated patterns, such as one replicate's being consistently higher or aberrant, or a trend over time. Time sequence plots of concentration means and standard deviations would be useful in this regard. Occasionally, a set of 5 to 20 tests, in which block positions (see Appendix A in USEPA 1994b) have been recorded, should be subjected to ANOVA for block or position effects. If such effects are significant or large, the laboratory should seek advice on randomizing the replicates and concentrations.

If a laboratory's CV exceeds the 75th percentile CV from Tables 3-2 through 3-4, EPA recommends calculating warning and control limits based on the 75th and 90th percentiles, respectively, of CVs for the method and endpoint (Tables 3-2 and 3-3 and Appendix Tables B-1 and B-2). For example, suppose the mean EC25 for a series of *Ceriodaphnia* chronic tests (Method 1002.0 with reproduction as the endpoint) conducted at one laboratory with reference toxicant is 1.34 g/L NaCl. Also suppose that the standard deviation of the EC25s for these tests is 0.85. The CV for this set of EC25s is thus 0.63. In Table 3-2, the 75th percentile of CVs for this test's reproduction endpoint is 0.45. Calculate the standard deviation corresponding to the 75th percentile CV, $S_{A.75} = 1.34 \times 0.45 = 0.60$. In Appendix Table B-1, the 90th percentile of CVs is 0.62 for this method and endpoint. Calculate $S_{A.90} = 1.34 \times 0.62 = 0.83$. Because the CV for this series of EC25s exceeds the 90th percentile reported in Table B-1, EPA recommends the following:

- Set control limits using $S_{A.90} = 0.83$,
- Set warning limits using $S_{A.75} = 0.60$,
- Promptly take actions to bring results within the control limits, and
- Attempt to bring results within the warning limits in 3-12 months.

If the CV for the set of EC25s is less than the 90th percentile reported in Table B-1, use that CV to set control limits. If the CV for the set of EC25s is less than the 75th percentile in Table 3-2, do not set warning limits using the latter value.

In addition, Burton et al. (1996) encourage regulatory programs to have a laboratory audit component to document the existence and effectiveness of a QA/QC program directed at toxicity testing, including analyst training and experience. Regulatory authorities should use the National Environment Laboratory Accreditation Program (NELAP) (USEPA 1999a) and routine performance audit inspections to evaluate individual laboratory performance. Inspections should evaluate the laboratory's performance with QC control charts based on reference toxicants, examine procedures for conducting the toxicity test procedures, and examine procedures for analyzing test results.

Regulatory authorities should develop a QC checklist to assist in evaluating and interpreting toxicity test results. Appendix E presents examples of State WET implementation procedures related to reviewing reference toxicant data and information on additional QA/QC criteria that have been developed and implemented. Regulatory authorities should also provide additional guidance related to the interpretation of QC control charts. This additional guidance could be that laboratories maintain control charts on within-test variability (e.g., PMSD) and use warning and control limits based on the 75th and 90th percentiles of CVs for the test method and endpoint.

5.3.2 Experimental Design

Experimental design includes randomizing the experimental units (i.e., treatments, organisms, replicates); establishing the statistical significance level (i.e., alpha level); and specifying the minimum numbers of replicates, test organisms, and treatments. Oris and Bailer (1993) recommend that test design and TAC be based, not only on a minimum level of control performance, but also on the ability to detect a particular level of effect (i.e., test power).

A Type I error (i.e., "false positive") results in the false conclusion that an effluent is toxic when it is not toxic. A Type II error (i.e., "false negative") results in the false conclusion that an effluent is not toxic when it actually is toxic. Power (1 - beta) is the probability of correctly detecting a true toxic effect (i.e., declaring an effluent toxic when it is in fact toxic). Acceptable values for alpha range from 0.01 to 0.10 (1 to 10 percent). The current EPA test methods recommend an alpha rate of 0.05 or 5 percent in the toxicity

testing manuals. Currently, EPA is preparing guidance on when an alpha rate of 0.01 or 1 percent would be considered acceptable (USEPA 2000a).

5.3.2.1 False Positives in WET Testing

The hypothesis test procedures prescribed in EPA's WET methods provide adequate protection against incorrectly concluding that an effluent is toxic when it is not. The expected *maximum* rate of such errors is the alpha level used in the hypothesis test. The hypothesis test procedure is designed to provide an error rate *no greater than* alpha when the default assumptions are met. The statistical flow chart provided with each EPA WET method identifies cases when default assumptions are not satisfied and, therefore, when data transformations or alternative statistical methods (e.g., a nonparametric test) should be used.

Alpha and beta are related (i.e., as alpha increases, beta decreases), assuming that the sample size (number of treatments, number of replicates), size of difference to be detected, and variance are held constant. The alpha and beta error rates depend on satisfying the assumptions of the hypothesis test. To ensure that statistical assumptions and methods are properly applied, testing laboratories should review the statistical procedures used to produce WET test results and other factors, such as biological and statistical quality assurance, and verify that test conditions and test acceptability criteria were achieved.

If a test is properly conducted and correctly interpreted, identifying any particular outcome as a "false positive" is impossible. An effluent that is deemed toxic may require that the permittee conduct additional toxicity tests to determine if toxicity is re-occurring. Even if no toxicity is demonstrated in follow-up tests, that does not rule out that the original toxic event was a true toxic spike in the effluent. False negatives, however, impact the environment by allowing the discharge of harmful toxicants without identification. This may occur because the toxic effects are not identified as statistically significant due to lack of test sensitivity (see Sections 5.3.3 and 6.4).

Measurement error should not affect the protection against false positives provided by hypothesis tests and confidence intervals when they are appropriately applied. Measurement error, in the case of WET test treatment mean values, likely consists largely of sampling errors (e.g., variability among organisms or containers), although errors in counting, weighing, and other procedures may also occur. Such sources of imprecision are implicitly accounted for in WET test statistical inferences, because the sample variance among the replicates within each treatment (dilution) is used for inference. The test "size" 1 - alpha will protect adequately against false positives. A larger variance among replicates, however, could make detecting real toxicity (i.e., false negatives) more difficult unless the number of replicates is increased to provide more test sensitivity and power, which will reduce the rate of false negatives.

5.3.2.2 False Negatives in WET Testing

For a given alpha, beta decreases (power increases) as the sample size increases and the variance decreases. Decreasing alpha from 0.05 to 0.01 without otherwise changing the hypothesis test will reduce the ability of the test to detect toxicity, that is, will reduce the power of the test. Thus, as alpha for the hypothesis test is decreased, there is an inevitable trade-off between the rate of false positives when toxicity is not present and the ability to detect toxicity when it is present (i.e., statistical power).

To limit within-test variability and thus increase power, EPA developed a minimum significant difference (MSD) criterion that must be achieved in the chronic West Coast marine test methods (USEPA 1995). The MSD is a measure of the within-test variability and represents differences between treatments and the control that can be detected statistically. Distributions of the MSD values of multiple tests for a specific reference toxicant and test method can be used to determine the level of test sensitivity achievable by a certain percentage of tests. Denton and Norberg-King (1996) analyzed several chronic test methods to quantify the effect size based on the existing toxicity test method experimental design and MSD distributions.

Denton and Norberg-King found when setting the beta error rate at 0.20 (power = 0.80), the effect size detected varies from at least a 15-percent reduction from the control response for the chronic red abalone larval development test to a 40-percent reduction from the control response for the chronic *Ceriodaphnia dubia* test. In this document, EPA has calculated power for each test method (see Section 5.3.3).

5.3.3 Test Power To Detect Toxic Effects

This section describes the statistical power and ability to detect toxic effects achieved by the current WET methods, as inferred from the WET variability data set used to develop this document. These inferences are approximate, because assumptions of normality and homogeneity of variance were not always satisfied.

Power can be characterized only by repeated testing. Power is an attribute not of a single test, but of a sequence of many tests conducted under similar conditions and with the same test design. Therefore, in this document, EPA used the sample averages for each laboratory's data set to characterize each laboratory. The following two parameters were required: (1) the mean endpoint response in the control (growth, reproduction, survival); and (2) the mean value of the error mean square (EMS) for tests.

EPA evaluated the ability to detect toxic effects using three approaches for each test method: (1) number of replicates required to detect a 25-percent difference from the control with power of 0.80; (2) percent difference from the control that can be detected with power of 0.80; and (3) power to detect a 25-percent difference from the control. All calculations are based on a one-sided, two-sample t-test at a level of 0.95 (alpha of 0.05). The power for a multiple comparison (Dunnett's or Steel's test) will be less than the power for this two-sample t-test.

Table 5-1 summarizes the results for this evaluation. Depending on the method, between 30 percent and 80 percent of the laboratories were able to detect a 25-percent effect for the sublethal endpoint consistently. Between 60 percent and 100 percent of the laboratories were able to detect a 33-percent effect.

To examine whether the upper bounds presented in Table 3-6 provide adequate test precision, EPA calculated an estimate of the power to detect a 25-percent effect on a sublethal endpoint when the PMSD equals the upper bound reported in Table 3-6. The upper bounds of the PMSD are shown in Table 3-6 in Chapter 3. At the lower PMSD bound, the power always exceeded 0.98. Tests with PMSD equaling the upper bound are not often able to detect a 25-percent effect. This finding does not mean that the upper bound is ineffective. The PMSD varies between tests, and each laboratory has a distribution of PMSDs. To avoid exceeding this upper bound often, a laboratory would have to achieve substantially lower PMSDs in most tests.

5.3.3.1 Attainment of the PMSD Related to Power

The power of the current experimental design could be reevaluated by comparing it to alternative designs that use increased number of replicates or number of test concentrations (Chapman et al. 1996). In this document, EPA found that about half of the laboratories in the data set were able routinely to detect a 25-percent difference between control and treatment. About two-thirds of the laboratories could routinely detect a 33-percent difference (Table 5-2). For example, mere attainment of the 90th percentile PMSD values shown in Table 3-6 will not ensure the ability to detect a 25-percent effect (Table 5-2). If every acceptable test has a PMSD below that upper bound, however, the average PMSD will be lowered. Based on the within-laboratory variability of PMSD,[1] the average PMSD likely will be substantially lower than the upper bound in Table 3-6, if *most* tests conducted by a laboratory are to have acceptable PMSDs.

[1] The average CV for PMSD is one-third to one-half its mean in commonly used methods.

Table 5-1. Tests for Chronic Toxicity: Power and Ability To Detect a Toxic Effect on the Sublethal Endpoint

Test Method	No. Labs	No. Labs with Power		Power (Range)	No. Labs Having Power at Least 0.8 To Detect Effect of		Effect Detected with Power 0.8 as Percent of Control Mean (Range)
		0.8	0.5		25%	33%	
1000.0 Fathead Minnow	19	6	14	0.21 - 1.00	6	13	8.2 - 62
1002.0 *Ceriodaphnia*	33	10	29	0.38 - 1.00	10	19	14 - 45
1003.0 Green Alga	9	7	8	0.33 - 0.99	7	8	13 - 49
1004.0 Sheepshead Minnow	5	4	5	0.77 - 1.00	4	5	8.6 - 26
1006.0 Inland Silverside	16	7	13	0.23 - 0.97	7	12	17 - 59
1007.0 Mysid (growth)	10	5	8	0.21 - 0.91	5	8	21 - 70

Note: Power was calculated for a two-sample, one-sided t-test at alpha = 0.05, for a 25-percent difference from the control. Effect size detected was calculated for the same test using power 0.80. Calculations used the average EMS from all tests at each laboratory and the minimum number of replicates reported for those tests. Calculations assumed that the parametric mean and variance equal the corresponding sample estimates. They also assumed approximate normality of means and homogeneity of variance. Because these assumptions may be violated, the results here are approximate. By saying "detect a 25-percent difference from control," this alternative hypothesis is intended: (control mean - treatment mean) > 0.25 × control mean.

Table 5-2. Power To Detect a 25-Percent Difference from the Control at the 90th Percentile PMSD

Chronic Method	Replicates	90th Percentiles of PMSD	Three Treatments		Four Treatments		Five Treatments	
			alpha = 0.05	alpha = 0.05/3	alpha = 0.05	alpha = 0.05/4	alpha = 0.05	alpha = 0.05/5
1000.0 Fathead Minnow	3	35	0.39	0.25	0.39	0.19	0.39	0.15
	4	35	0.41	0.30	0.42	0.26	0.43	0.23
1002.0 *Ceriodaphnia*	10	37	0.39	0.31	0.41	0.30	0.43	0.30
1003.0 Green Alga	3	35	0.39	0.25	0.39	0.19	0.39	0.15
	4	35	0.41	0.30	0.42	0.26	0.43	0.23
1004.0 Sheepshead Minnow	3	23	0.72	0.69	0.72	0.62	0.73	0.55
	4	23	0.73	0.71	0.74	0.68	0.75	0.66
1006.0 Inland Silverside	3	23	0.72	0.69	0.72	0.62	0.73	0.55
	4	23	0.73	0.71	0.74	0.68	0.75	0.66
1007.0 Mysid	8	32	0.48	0.41	0.50	0.40	0.52	0.40

Notes: Values are rounded to two significant figures. Number of treatments is the number of concentrations compared with the control in the hypothesis test. The calculations assumed (1) the usual assumptions of the test are satisfied (approximate normality, homogeneity of variances); and (2) equal replication in treatments and control. Because these assumptions may be violated, the results here are approximate. Because the MSD/mean implies a value for [root (error mean square)/mean], the latter could be calculated from the MSD, Dunnett's critical value, and the number of replicates, and then used in a calculation of power. Calculations apply to a one-sided, two-sample t-test of equal means, assuming equal variances and equal replication, with hypotheses H_o:{control mean - treatment mean = 0} versus H_a:{control mean - treatment mean > 0.25 × control mean}. The power achieved by Dunnett's multiple comparison procedure will lie between the two-sample power at alpha = 0.05 and that for alpha = 0.05/(no. of treatments).

Testing laboratories and permittees can examine the EMS or MSD in Tables B-14 and B-15 (Appendix B) to estimate the ability of a WET test to detect toxic effects. Some regulatory authorities may require a comparison between the control and the receiving water concentration, which requires a two-sample, one-sided t-test. Others may require the multiple comparisons procedure described in the EPA WET methods (Dunnett's or Steel's tests, one-sided, with alpha of 0.05). The power of Dunnett's procedure falls between the power of the one-sided, two-sample t-test with alpha of 0.05 and alpha of 0.01, assuming that no more than five toxicant concentrations are compared to a control. The power of Steel's procedure will be related to, and should usually increase with, the power of Dunnett's procedure and the t-tests. Tables B-14 and B-15 in Appendix B also provide an appropriate guide to achieving power using a nonparametric test.

Recently, the State of Washington (1997) issued guidance specifying an acute and chronic statistical power standard to be achieved for compliance testing. EPA's sediment toxicity testing manuals (USEPA 1994c, USEPA 2000) include power curves for various numbers of experimental units, CV ranges, and associated alpha and beta levels. Sheppard (1999) is a good source to provide a simple explanation of how power helps determine how large a sample should be. Additional information on power may be obtained at: http://www.psychologie.uni-trier.de:8000/projects/gpower/literature.html.

EPA recommends that regulatory authorities specify in their State WET implementation procedures that individual test results achieve a level of within-test sensitivity by using the upper and lower PMSD test sensitivity bounds (see Section 6.4). To achieve the test sensitivity bounds, testing laboratories may need to minimize within-test variability (e.g., EMS) or increase the number of replicates tested, or both. If laboratories cannot achieve PMSD values of less than 25 percent for the toxicity test methods that require a minimum of only three replicates (Methods 1000.0, 1004.0, 1006.0), then the numbers of replicates may need to be increased. Appendix B (Section B.4) provides information related to the number of replicates needed and discusses the relationship between test power and effect size achieved. The magnitude of the effect size achieved relates to the test sensitivity.

5.4 Test Acceptability Criteria

EPA test methods have specific TAC that the effluent and reference toxicant tests must meet. A test is considered invalid if the TACs are not met. The recommended test conditions for each test method specify the minimum requirements and the TAC. For example, control survival must be 80 percent or greater and average control reproduction at least 15 young per surviving female in the chronic *Ceriodaphnia dubia* survival and reproduction test.

The chronic West Coast marine methods (USEPA 1995) require additional TAC. For example, to limit the degree of within-test variability, the methods specify a maximum allowable value for PMSD (see Section 5.3.2 on experimental design). Some States have additional TAC in their State WET implementation policies. North Carolina (1998) for example, requires that the chronic *Ceriodaphnia dubia* analyses meet an additional TAC of complete third brood neonate production by at least 80 percent of the control organisms and that the control reproduction CV be less than 40 percent.

Additional TAC might be specified to minimize variability among replicates. Variability of any toxicity test result is influenced by the number of replicates used, number of organisms tested, and variability among replicates at each test concentration and the control. Variability among replicates has been quantified by treatment CV, EMS, or MSD. The application of a maximum acceptable value for CV or MSD helps ensure adequate laboratory QA/QC and increases the reliability of submitted data. One benefit of requiring a maximum allowable within-test variability limit is that laboratories will improve culturing, test handling, and housekeeping, which are usually incorporated into the laboratories' standard operating procedures. For example, the CV requirement might be incorporated directly into the NPDES permit. Sample EPA Region 6 permit language reads:

1. *The coefficient of variation between replicates shall be less than or equal to 40 percent in the control.*

2. *The coefficient of variation between replicates shall be less than or equal to 40 percent at the instream waste concentration (IWC).*

3. *Test failure may not be construed or reported as invalid due to a CV of greater than 40 percent. A repeat test shall be conducted within the required reporting period if any test is determined to be invalid.*

Occasionally, statistical analyses indicate a test failure when as little as 15-percent mortality has occurred in a test dilution. Permit language has been developed to address this occurrence, as in the following example:

If all TAC conditions are met and the percent survival of the test organism is greater than or equal to 80 percent (in a chronic test) or 90 percent (in an acute test) in the critical dilution concentration and all lower dilution concentrations, the test shall be considered to be a valid test, and the PERMITTEES shall report an NOEC of not less than the critical dilution for the discharge monitoring report (DMR) reporting requirements.

Regulatory authorities may consider providing guidance or imposing additional TAC, such as those implemented by EPA Region 6 or like some States have implemented (North Carolina 1998, Washington 1997). Appendix E provides additional examples of States that have implemented further guidance on WET QA/QC procedures and TAC. Warren-Hicks (1999) also recommended that additional national TAC be established for each test method (e.g., upper and lower bounds on the MSD). Therefore, EPA recommends that regulatory authorities require that additional TACs be implemented in permits to minimize within-test variability and increase test sensitivity (see Section 6.4 and Appendix C for sample permit language).

5.5 Conducting the Statistical Analysis To Determine the Effect Concentration

EPA test methods currently recommend two statistical approaches to estimate a chemical or effluent concentration for each biological effect endpoint (e.g., survival, growth, and reproduction). One approach is to derive the NOEC by hypothesis testing, which equates biological significance with statistical significance. The second approach is to estimate an effect concentration that reduces the control response by 25 percent for chronic methods. The expanded use of WET tests in the NPDES program has brought increased attention to the statistical analysis of toxicity test data. A common goal for both regulatory authorities and permittees is to confirm that the effect concentrations were derived correctly using the appropriate analysis approaches. Reliable effect concentrations lead to increased confidence in the data used for making regulatory decisions, such as determining reasonable potential, deriving a permit limit or monitoring trigger, and generating self-monitoring test results.

Another important consideration in conducting statistical analyses is the inconsistent use of statistical programs. The proliferation of statistical packages has been helpful in data analysis; however, these packages also can result in the misapplication of the statistical methods. APCA-AWWA-WEF (1998) cautions the user to confirm the results of each analysis with each package before accepting them. The data user is responsible for evaluating all data submitted to the regulatory authorities.

The 1995 SETAC Pellston Workshop discussed unresolved scientific issues and highlighted significant research needs associated with WET testing. The attendees recommended the following:

Immediately instigate studies to evaluate improvements in the statistical analysis of WET test data. These studies should include, but not necessarily be limited to, the following activities:

(a) investigate the implications of concurrent application of NOEC/MSD, tests of bioequivalence, and ECp estimators (Chapman et al. 1996a).

In response to this recommendation, EPA began projects to evaluate the bioequivalence approach and additional point estimate models for the WET program. At present, two test methods are being used for this evaluation: (1) the chronic *Ceriodaphnia dubia* survival and reproduction tests and (2) the giant kelp germination and germ-tube length test with reference toxicants.

The bioequivalence approach poses the following question: Do the mean responses of the effluent concentration and the control differ by more than some amount? For example, the control response and the response at the critical effluent concentration (i.e., instream waste concentration) must differ by no more than a fixed value in order to accept the hypothesis of no significant difference (i.e., no toxicity). This approach could address the concern that an imprecise test might not detect toxicity when toxicity is present or that a small but statistically significant effect would detect toxicity that may not be biologically important. Some researchers have suggested that the bioequivalence approach could provide a positive incentive for dischargers to produce test results with lower within-test variability to demonstrate that no toxicity occurs at a level greater than a biologically (bioequivalence approach) significant amount (Shukla et al. 2000, Wang et al. 2000).

Bailer et al. (2000) evaluated the proposed regression-based estimators with the current EPA point estimate models. They found that it appears reasonable to incorporate parametric estimation models in the WET program. Bailer et al. (2000) concluded that these models are appropriate for all response scales (i.e., dichotomous, count, and continuous) and can incorporate monotonicity without bias. However, confidence intervals still need to be developed for these parametric models.

In this document, EPA has not recommended either the bioequivalence or additional point estimate models to supplement the current statistical approaches as described in the testing manuals. An independent, peer-reviewed workshop should be convened to evaluate the benefits of these alternative statistical approaches to enhance the statistical approaches currently applied.

5.6 Chapter Conclusions

In this chapter, EPA provides guidance to permittees and testing laboratories on collecting representative effluent samples, conducting the biological test methods, and evaluating the statistical analyses. EPA recommends that States implement the lower and upper PMSD test sensitivity bounds to achieve an acceptable level of test sensitivity and minimize within-test variability (see Section 6.4). EPA also provides guidance to permittees and testing laboratories on the number of replicates required to achieve the PMSD bounds. Testing laboratories should maintain and evaluate both effluent and reference toxicant data using a measure of within-test variability such as the PMSD.

Permittees and toxicity testing laboratories may need to increase replication in order to reduce PMSD below the upper bound. Table B-15 can be used for initial planning of replication, given knowledge of typical values of the error mean square (EMS) or MSD and the number of concentrations used in the multiple comparison hypothesis test. To ensure that all PMSD values fall below the upper bound in Table 3-6, a laboratory would select the largest EMS value experienced in its past testing.

EPA recommends that testing laboratories require a minimum of four replicates for the fathead minnow, sheepshead minnow, and inland silverside chronic test methods (Methods 1000.0, 1004.0, and 1006.0, respectively). Four replicates are needed to execute the statistical flow chart when a nonparametric test is needed. Three replicates are also sometimes insufficient to keep PMSD below the recommended upper bound. In addition, four replicates are needed to help achieve the upper PMSD bound.

This page intentionally left blank.

6.0 GUIDANCE TO REGULATORY AUTHORITIES: DETERMINING REASONABLE POTENTIAL AND DERIVING WET PERMIT CONDITIONS

EPA developed the TSD (USEPA 1991a) to support implementation of national policy to control the discharge of toxic pollutants. The TSD presents a statistical approach for determining the need for and the method of deriving water quality-based effluent limits (WQBELs) based on aquatic life (including WET), human health, and wildlife criteria. This approach accounts for the uncertainty associated with small data sets and data variability by assuming a statistical distribution of effluent data (usually lognormal) and calculating a CV or using a default CV to describe data variability.

6.1 Analytical and Sampling Variability in Calculations for Reasonable Potential and Permit Limits

Section 6.1 discusses use of the CV of sample measurements of toxicity to make a reasonable potential determination and to calculate permit limits. Two points must be understood: (1) this CV is to be calculated using toxic unit (TU) values (USEPA 1991a) (see Section 6.2); and (2) EPA strongly recommends that point estimates (not NOEC or LOEC values) be used to calculate the TU values (USEPA 1994a, 1994b).

Water quality-based effluent limits are required when a discharge causes, has reasonable potential to cause, or contributes to an instream excursion above a water quality standard. Throughout this document, EPA uses the commonly understood, shorthand reference "reasonable potential" to refer to this standard for determining the need for a water quality-based effluent limit.

6.1.1 "Adjusting for Analytical Variability" in Calculations for Reasonable Potential and Permit Limits

Adjustment approaches (see Appendix G.3) have been suggested to "adjust for analytical variability" when deriving permit limits and determining the need for a WET limit in the first place. EPA does not recommend these adjustment approaches (Appendix G.3) and strongly reaffirms the statistical approach and methods for calculating permit limits provided in the TSD (USEPA 1991a). *EPA recommends that regulatory authorities use the statistical approach and calculation methods in the TSD*. The TSD methods were designed to provide a reasonable degree of protection for water quality (i.e., to avoid exceedances of water quality criteria), while providing a reasonable degree of protection from the variability of effluent toxicity and analytical variability. The various "adjustment" approaches would undermine these objectives.

The TSD limit calculation for a point source can be divided into two steps: first, convert the wasteload allocation (WLA) to a long-term average (LTA), and then convert the LTA to effluent limits (maximum daily, average weekly, and average monthly limits). WET limit calculations include an intermediate step in which the acute WLA is converted to a WLAa,c. These calculations employ a facility-specific CV based upon effluent sampling data. The TSD approach uses this CV in both steps.

Adjustment approaches intended to account for analytical variability, discussed in detail in Appendix G, would inappropriately use different CVs in these two steps. The first step would use an estimate of the CV of "true" effluent toxicity, which is smaller than the CV for measured toxicities. This approach would result in a larger calculated LTA. The second step would use the CV for the measured toxicities, which is the same CV used in both steps of the TSD approach.

Use of such adjustment approaches would frequently result in setting an average monthly permit limit (AML) that exceeds the chronic WLA. Appendix G demonstrates that such outcomes (i.e., the AML exceeds

the chronic WLA) generally can be expected to occur when various adjustment approaches are used. Appendix G, Table G-1, presents a numerical example of how an adjustment approach would allow calculation of an AML exceeding the chronic WLA (a four-day average value), even when sampling frequency for the calculation is set at the recommended minimum of four samples per month. [It is acceptable for the maximum daily limit (MDL), which applies to a single sample, to exceed the chronic WLA. It is also acceptable for the AML to exceed the chronic WLA, if the AML calculation is based on fewer than four samples per month. Note, however, that the TSD recommends always assuming at least four samples per month when calculating the AML.]

The TSD reasonable potential calculation first calculates the percentile represented by the maximum observed TU value. For example, the maximum of 10 reported TU values is identified with the 63rd percentile. Then the sample CV is used to project the 95th or 99th percentile TU value, using a table of reasonable potential multiplying factors. This value is combined with the appropriate mixing-zone dilution to project a maximum receiving water toxicity, which is compared with the applicable water-quality criterion. If an adjustment were applied to the reasonable potential calculation, the CV would be adjusted downward and the maximum projected receiving water toxicity would be smaller. This would make a determination of need for a permit limit less likely.

Because of these considerations, EPA strongly recommends that no adjustment be made to the CV or variance of toxicity, either for reasonable potential or permit limit calculations. The TSD statistical approaches already account for analytical variability appropriately. EPA continues to recommend the TSD approach, which ensures that effluent limits and, thereby, *measured* effluent toxicity or pollutant parameter concentrations are consistent with calculated WLAs.

6.1.2 Analytical Variability and Self-monitoring Data

EPA determines compliance with permit limits on the basis of self-monitoring data, and these data include some measure of analytical variability. The influence of analytical variability is accounted for in the TSD statistical procedures used to set water-quality limits and determine the potential for toxicity, as explained in Appendix G.

The permittee is responsible for ensuring that measured discharge toxicity never exceeds the permit limits. No special allowance is made for analytical variability in assessing compliance. The maximum discharge toxicity should incorporate a margin of safety, which will account for sampling and analytical variability. In other words, to avoid exceeding permit limits, the facility's treatment system should be designed so that the maximum toxicity is somewhat lower than its permit limits.

6.1.3 Precision of WET Measurements and Estimates of Effluent CV

Single measurements on effluent involve some uncertainties about the true concentration or toxicity related to representativeness of the sample, including sample holding time and conditions, and the analytical measurement system. Like all analytical measurements, WET measurements (NOEC, EC25, LC50) are inexact. That is, the exact toxicity of an analyte in a sample can be specified only within some range. This imprecision can be reduced by using a suitable number of organisms and replicates for each test (see Section 5.3.2 on experimental design).

The numbers of organisms and replicates required for EPA WET method test acceptability are specified as minimums. Test precision will be approximately proportional to the square root of the number of replicates. Thus, doubling the number of replicates may decrease the MSD to approximately 70 percent of its former value. Increased replication also tightens the confidence interval for a point estimate of the effect concentration (e.g., EC25 and LC50).

EPA strongly recommends that toxicity measurements of an effluent be obtained at least quarterly for three years to provide a good basis for determining the need for limits and for calculating limits. One year should be regarded as the minimum duration needed to characterize effluent variability (due to seasonal, stream flow, or process fluctuations), and ten the minimum number of measurements, unless scientific and technical knowledge supports a shorter period as representative of the distribution of pollutant types and concentrations of toxicity.

Estimates based on multiple measurements involve the same uncertainties that apply to single measurements. They also may involve larger uncertainties related to sampling error, that is, the chance that typical levels of toxicity or concentrations of pollutant may not be encountered during the sampling program. The sampling program may not fully characterize effluent variability if too few samples are taken, the sampling times and dates are not representative, or the duration of the sampling program is not long enough to represent the full range of effluent variability. When determining the need for limits and calculating limits, the variance or the CV of toxicity measurements is key. The larger the number of samples, the more precise is the estimate. Confidence intervals for the variance and CV can be calculated and carried through the calculations for reasonable potential and effluent limits (Appendix G). Even when assumptions are not strictly met, confidence intervals provide a useful perspective on the uncertainty of the results and the need for more samples. The *minimum* number of measurements recommended for calculating estimates of the CV for effluent toxicity is 10.

6.1.4 Between-Laboratory Variability

Between-laboratory variability may increase the CV as discussed in Section 6.1.1, if the toxicity tests were conducted by more than one laboratory for a specific facility. A concern to permittees is that this may increase the likelihood of making a finding of reasonable potential.

Within-laboratory variability is the component of analytical variability that should be reflected in regulatory calculations. If the data used for reasonable potential or permit limit calculations are effluent measurement data reported by at least two laboratories, there are ways to appropriately estimate the variance to be used in TSD statistical calculations.

For example:

- If the same laboratories continue to be used in the same proportion or frequency and the measurements from the individual laboratories represent different sampling dates, the measurement data can be treated as if they were generated by a single laboratory. This approach may increase the estimated variance and the AML, which is not in the interest of the permittee. Selecting one laboratory for future monitoring, based on the variance of its reported reference toxicant test results, should mitigate this problem.

- If only one laboratory has reported data on each sampling date, and the other laboratories report over different time spans or over the same time span on alternating dates, EPA recommends forming a pooled estimate of variance. Calculate the sample variance (S^2) of log(TU) for each laboratory separately, and combine these using the formula:

  $$\text{pooled variance of } \log(X) = [(N_1 - 1)S_1^2 + (N_2 - 1)S_2^2] / [(N_1 - 1) + (N_2 - 1)]$$

  An analogous formula is used for more than two laboratories. The same result can be obtained by conducting a one-way analysis of variance on log(TU) (with laboratories treated as the groups or classes) and using the reported EMS.

Changing a laboratory may change analytical (within-laboratory) variability of measurements and test sensitivity (i.e., PMSD values). That is, the average effect concentration may change (e.g., Warren-Hicks et al. 1999). Ideally, the permittee will anticipate and plan for a change of testing laboratory. Permittees should compare reference toxicant test data for current and candidate replacement laboratories, selecting one with acceptable variability and a similar average effect concentration.

6.2 Determining Reasonable Potential and Establishing Effluent Limits

Effluent characterization is an essential step in determining the need for an NPDES permit limit. NPDES regulations under 40 CFR Part 122.44(d)(1)(ii) specify that reasonable potential include "*whether a discharge causes, has the reasonable potential to cause, or contributes to an instream excursion above a State water quality standard*." Calculations for reasonable potential determination and for permit limits should follow EPA guidance in the TSD (USEPA 1991a). In particular, the TSD statistical methods should be used. Such calculations should use TUs for WET data, not effect concentrations (percent whole effluent). Toxic units are defined (USEPA 1991a, Chapter 1.3.1, page 6) as the reciprocal of the effect concentration times 100, where the effect concentration is expressed as a percentage of whole effluent, thus TUa = 100/LC50 and TUc = 100/ECp.

When characterizing an effluent to determine whether a permit limit is necessary, permit writers can use the available effluent WET data and a water-quality model to perform a reasonable potential analysis. The TSD outlines the statistical approach. This approach uses existing effluent data to project a maximum pollutant concentration or a maximum toxicity in the effluent (USEPA 1991a). The projected maximum concentration or toxicity is used as an input in the water quality model to determine whether the effluent has the reasonable potential to cause or contribute to an excursion of ambient water quality criteria. If reasonable potential exists, the permit writer must derive a WET permit limit for that facility.[1]

The variability of the existing effluent data, as measured by the CV, has a significant effect on the projected maximum pollutant concentration or toxicity. The higher the CV, the higher the projected maximum, and the more likely that there is reasonable potential and a limit is needed. EPA recommends that regulatory authorities use all valid, relevant, and representative data in making reasonable potential determinations. EPA is developing a national policy clarifying use of the TSD procedures for determining reasonable potential for WET. Important components of this policy include specifying the minimum number of valid WET tests necessary to calculate facility-specific CVs,[2] as well as recommending a step-wise approach to determining the need for WET permit limits. This approach reflects a strong preference by EPA and its stakeholders to rely on facility-specific WET testing, based on adequate frequency and duration of effluent sampling, for making reasonable potential determinations for toxicity.

EPA recommends that point estimates be used to estimate effluent variability, to determine the need for limits, and to set permit limits. This is recommended whether the self-monitoring test results will be determined using hypothesis tests or point estimates. Point estimates have less analytical variability than NOECs using current experimental designs, as shown in Chapter 3. Point estimates make the best use of the WET test data for purposes of estimating the CV, LTA, and RP factor and calculating the permit limit.

---

[1] When the State has narrative criteria for toxicity and the TIE/TRE identifies a specific chemical that is the source of toxicity, the permit writer may include a chemical-specific limit for that parameter instead of a WET permit limit in accordance with 40 CFR Part 122.44(d)(v).

[2] If fewer than ten data points are available, the regulatory authority must use a default CV. As a result, the need for a WET permit limit may be based on a default value rather than actual data.

6.3 Development of a Total Maximum Daily Load for WET

Total maximum daily loads (TMDLs) may be indicated when there is acute or chronic toxicity in a waterbody, leading to the listing of the waterbody as impaired under CWA Section 303(d), and when there are multiple sources of the toxicity. EPA believes that TMDL calculations should be performed on the pollutants causing toxicity whenever possible. In these situations, EPA suggests that the first step of the analysis is to conduct ambient toxicity identification evaluations to identify the pollutant(s) and the source(s) causing the toxicity. Once the pollutant(s) and source(s) causing toxicity have been identified for the waterbody, then a TMDL should be developed for the individual pollutant(s).

6.4 Accounting for and Minimizing Variability In the Regulatory Decision Process

A common goal for the permittee and the regulatory authority is to have confidence in the test results from the biological and statistical procedures. Both permittees and regulatory authorities would then have more confidence in taking regulatory actions, such as evaluating multiple effluent samples to determine reasonable potential and derive permit conditions (e.g., permit limits, monitoring triggers). If steps such as collecting a representative effluent sample to conducting the toxicity tests properly, as discussed in Sections 5.2 through 5.4, and requiring additional TACs (Section 6.4.1) are used to reduce or minimize within-test variability, then the reliability of the WET test results increases.

6.4.1 Recommended Additional TACs: Lower and Upper Bounds for PMSD

Reference toxicant data from a large number of tests and laboratories were used to generate PMSD values; percentiles of these values are reported in Table 3-6. The MSD represents the smallest difference between the control mean and a treatment mean that leads to the statistical rejection of the null hypothesis (i.e., no toxicity) using Dunnett's multiple comparison test. MSD values are divided by the control mean and multiplied by 100 to produce a "percent MSD" (PMSD) value. The PMSD allows comparison of different tests and represents the smallest significant difference from the control as a percentage of the control mean. Thus, it represents the smallest significant value of the relative difference [100 (control mean - treatment mean)/control mean]. The MSD is often expressed as a percentage of the biological endpoint in the control response.

The following formula is used to calculate MSD (as recommended by USEPA 1995):

$$\text{MSD} = d s_w \sqrt{(1/n_1) + (1/n)}$$

where

- d = critical value for the Dunnett's procedure
- s_w = the square root of the error mean square (EMS)
- n_1 = number of experimental units in the control treatment
- n = the number of experimental units per treatment, assuming an equal number at all other treatments

Percent MSD is calculated as follows:

$$\text{PMSD} = \frac{\text{MSD}}{\text{control mean}} \times 100$$

EPA recommends that regulatory authorities implement both the lower and upper PMSD bound approach to minimize within-test variability when using hypothesis testing approaches to report an NOEC. The implementation of the upper PMSD bound should also apply when using point estimate techniques. There are five possible outcomes for regulatory decisions (see Figure 6-1). Two outcomes imply unqualified acceptance of the WET test statistical result:

1. **Unqualified Pass**–The test's PMSD is within bounds and there is no significant difference between the means for the control and the instream waste concentration (IWC) treatment. The regulatory authority would conclude that there *is no toxicity at the IWC concentration*.

2. **Unqualified Fail**–The test's PMSD is larger than the lower bound (but not greater than the upper bound) in Table 3-6 and there is a significant difference between the means for the control and the IWC treatment. The regulatory authority would conclude that there *is toxicity at the IWC concentration*.

3. **Lacks Test Sensitivity**–The test's PMSD exceeds the upper bound in Table 3-6 and there is no significant difference between the means for the control and the IWC treatment. The test is considered invalid. A new effluent sample must be collected and another toxicity test must be conducted.

4. **Lacks Test Sensitivity**–The test's PMSD exceeds the upper bound in Table 3-6 and there is a significant difference between the means for the control and the IWC treatment. The test is considered valid. The regulatory authority would conclude that there *is toxicity at the IWC concentration.*

5. **Very Small but Significant Difference**–The relative difference (see Section 6.4.2, below) between the means for the control and the IWC treatment is smaller than the lower bound in Table 3-6 and this difference is statistically significant. The test is acceptable. The NOEC is determined as described in Section 6.4.2 below.

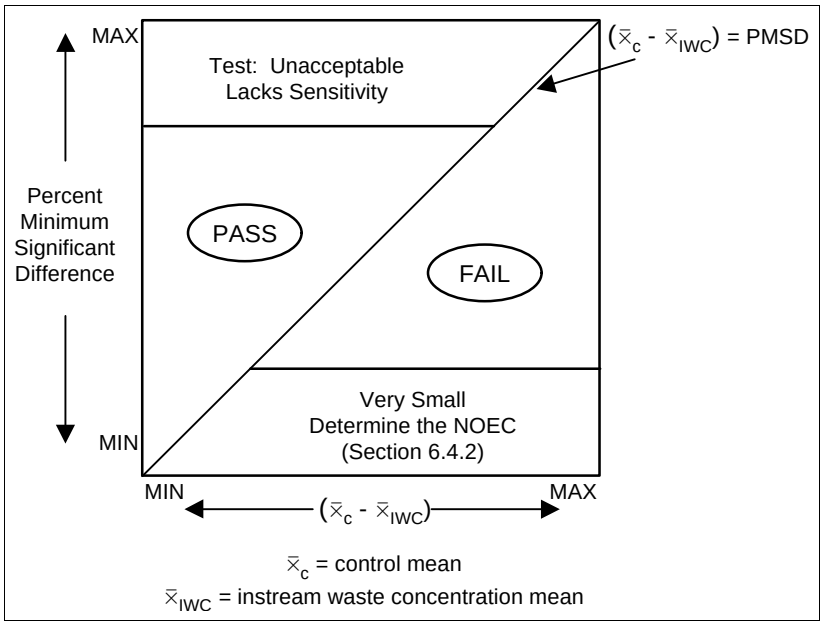

Figure 6-1. Paradigm that incorporates the lower and upper percent minimum significant difference.

Regulatory authorities should examine the sample permit language as provided in Appendix C, for incorporation of the PMSD bound language in a NPDES permit.

Note that "unqualified acceptance" of a WET test result requires that all of the following must be achieved: (1) collect the effluent sample properly; (2) conduct the toxicity test methods as specified in the toxicity manuals; (3) meet the required TACs; (4) meet the proper water quality parameters (e.g., temperature, pH); and (5) conduct the proper statistical calculations. All these conditions must be reviewed and deemed acceptable before a test is evaluated for self-monitoring data and reporting.

Figure 6-2 provides a decision tree that regulatory authorities can use when implementing the lower and upper PMSD bounds.

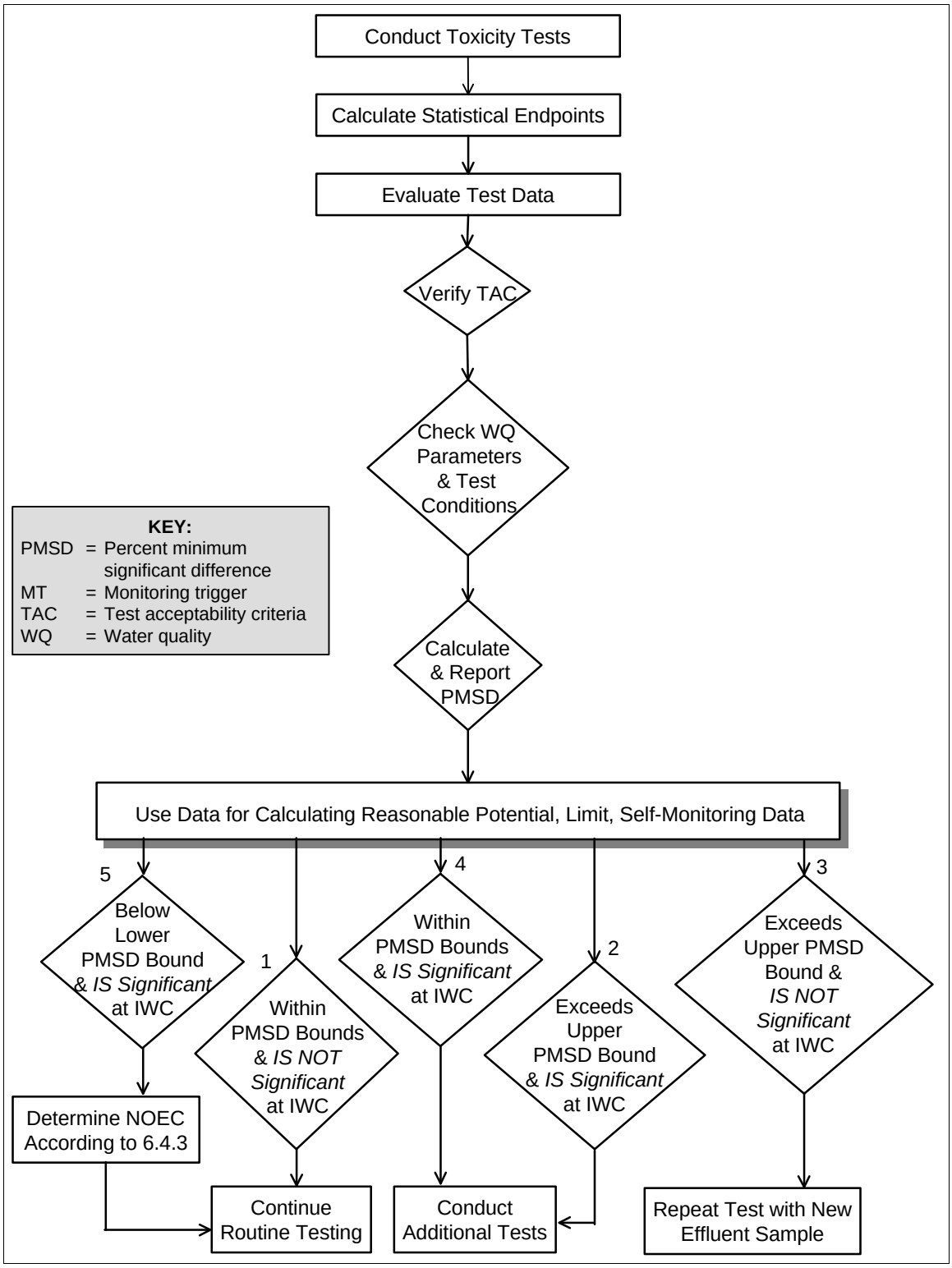

Figure 6-2. Implementing applications of upper and lower PMSD bounds for effluent toxicity testing requirements.

6.4.2 How to Determine the NOEC Using the Lower PMSD Bound

If the permit specifies that self-monitoring data are to be generated using hypothesis testing approaches, then the analyst should report the NOEC as the following. Find the smallest concentration for which (a) the treatment mean differs significantly from the control mean and (b) the relative difference (see example below) is not smaller than the 10th percentile in Table 3-6. Therefore, the NOEC is the next smaller test concentration.

In other words, concentrations having a very small relative difference with control (smaller than the lower PMSD bound) would be treated as if they do not differ significantly from control (even if they do so), for the purpose of determining the NOEC.

Table 6-1 illustrates the application of the lower PMSD bound for the reproduction endpoint of a *Ceriodaphnia* chronic test. In this example, the test's PMSD was 9.9, smaller than the 10th percentile value of 11 found in Table 3-6. The IWC concentration differed significantly from the control. The test falls under outcome number 5, a significant but very small difference at the IWC. The first step is to calculate the relative differences from control (Table 6-1) as [(control mean - treatment mean) divided by (control mean)] × 100. The next step is to determine which relative differences exceed the PMSD lower bound, 11 in this case (see the last column of Table 6-1). Finally, the NOEC is determined as described above. The NOEC is 12.5 percent effluent for this example.

Table 6-1. Example of Applying the Lower PMSD Bound for the Chronic *Ceriodaphnia* Test with the Reproduction Endpoint

Concentration (percent effluent)	Reproduction (mean of ten replicates)	Relative Difference from Control	Does Relative Difference Exceed 11?
100%	5.08 *	82	Yes
50%	12.4 *	56	Yes
25%	23.4 *	17	Yes
IWC = 12.5%	25.3 *	10	No
6.25%	26.1	7.4	No
Control	28.2	0	No

NOTE: The lower PMSD bound for this method and endpoint is 11 (Table 3-6). In this example, the NOEC is 6.25 percent effluent using the test's (very small) PMSD. Therefore, the reported NOEC should be 12.5 percent effluent after applying the lower PMSD bound.

* Differs statistically from the control as determined by MSD = 2.8 neonates. Thus, treatment means that are less than 28.2 - 2.8 = 25.4 would be statistically significant. These correspond to relative differences greater than 100 (2.8 / 28.2) = 9.9 percent.

6.4.3 Justification for Implementing the Test Sensitivity Bounds

A lower bound is needed to avoid penalizing laboratories that achieve unusually high precision. The 10th percentile PMSD represents a practical limit to the sensitivity of the test method because few laboratories are able to achieve such precision on a regular basis and most do not achieve it even occasionally. Several independent researchers have evaluated and provide support for using the MSD approach as additional TAC for the toxicity test methods. Thursby et al. (1997) advocate and provide reasons for using an empirical data base of minimum significant differences to provide TAC using statistical performance assessment. The State of California (Hunt et al. 1996, Starrett et al. 1993) and the West Coast marine toxicity test methods (USEPA 1995) have implemented an upper PMSD bound to minimize insensitive tests. Also the State of North Carolina has implemented additional requirements for the *Ceriodaphnia* chronic tests that reduced method

variability. North Carolina's evaluation of these additional TACs and subsequent improvements in test sensitivity appears in Appendix F.

The North Carolina data base affords the opportunity to evaluate the effectiveness of additional TAC and changes to the toxicity test procedures as they relate to the variability of WET test results (see Appendix F). For example, for PMSD, the median value decreased from 21 percent to 16 percent, while the 90th percentile decreased from 39 percent to 31 percent, indicating an overall increase in test sensitivity. The range in median values across all laboratories before adopting additional TACs was 12 percent to 36 percent. After adopting additional TACs, the range in median values was 10 percent to 27 percent, indicating a decrease in the overall spread between laboratories. The range in control CVs within a laboratory was from 21 percent to 79 percent before adopting TACs, compared to the range in control CVs within a laboratory after adopting TACs, which was narrowed to 17 percent to 36 percent. Overall, laboratories are generating data with more consistency within and between laboratories, after implementation of the additional TACs and additional method guidance provided by the State for the chronic *Ceriodaphnia dubia* test method.

6.4.4 Guidance to Testing Laboratories on How to Achieve the Range of Performance for PMSD

EPA recommends that regulatory authorities use the upper bounds (90th percentiles for PMSD in Table 3-6) to identify tests that are insufficiently sensitive. If PMSD exceeds this upper bound more often than occasionally, the laboratory should thoroughly investigate ways to reduce variability. There are three principal ways to reduce PMSD: (1) decrease within-test variability (that is, decrease the error mean square and therefore the standard deviation at each concentration); (2) increase the control mean; and (3) increase the number of replicates. The number of replicates required could be determined by trial-and-error calculations using the error mean square values obtained from a series of WET tests. At least 20 tests are recommended. The number "n" in the formula for MSD (number of replicates) would be increased and MSD re-calculated for each error mean square value. This approach uses a sample of tests specific to a particular laboratory and reveals the variation among tests. This approach would demonstrate how many replicates would be needed to achieve the upper PMSD bound, as required in Table 3-6.

6.5 Additional Guidance That Regulatory Authorities Should Implement to Further Support the WET Program

As discussed in Section 5.3, regulatory authorities have the discretion to develop and implement additional WET program requirements and guidance to ensure that WET test method variability is reduced by specifying additional guidance beyond the minimum requirements of EPA's WET test method's QA/QC and TACs. Appendix E provides a snapshot of State approaches to implementing NPDES WET programs to minimize WET test variability.

These State approaches include WET information to assist the regulated community with the following:

- Guidance regarding the evaluation of reference toxicant and effluent test results
- Guidance regarding how the State reviews reference toxicant data for laboratory performance
- Guidance regarding additional QA/QC criteria the State has developed and implemented
- Guidance regarding efforts the State has made to minimize test method variability
- Description of how the State reviews or conducts performance laboratory audits
- Description of specific implementation guidance that the State has developed to assist permit writers
- Description of how the State provides or uses toxicity test training

States contemplating such changes should consult with EPA to ensure the changes will be appropriate in the context of the State's overall NPDES WET program. In addition, States should implement a step-wise approach to address toxicity when the permit limit or monitoring trigger is exceeded in their State WET implementation plans.

For example, when an effluent is deemed toxic, then the permittee should take appropriate steps to demonstrate the magnitude, frequency, and potential source(s) of the toxicity. The components of the step-wise approach could include increased frequency of toxicity testing to characterize the magnitude and frequency of toxicity. If continued toxicity is demonstrated, then the permittee could conduct a Toxicity Reduction Evaluation/Toxicity Identification Evaluation (TRE/TIE) with toxic effluent sample(s) (USEPA 1991b, 1992). For example, EPA Regions 9 and 10 have prepared WET implementation guidance to assist their States (Denton and Narvaez 1996). This guidance provides sample permit language for a step-wise approach to address toxic samples (see Appendix C).

6.6 Chapter Conclusions

The TSD statistical approach to reasonable potential determination and permit limit derivation considers combined effluent and analytical variability through the CV of measured effluent values. Because determination of effluent variability is based on empirical measurements, the variability estimated for effluent measurements includes the variability of pollutant levels, sampling variability, and a smaller component owed to method variability. Steps should be taken to reduce these sources of variability. EPA believes that the TSD statistical procedures are appropriately protective in considering both effluent and analytical variability in reasonable potential and effluent limit calculations.

EPA recommends that regulatory authorities use a sampling program that conducts at least ten representative WET tests over a period of three years to represent the full range of effluent variability. Regulatory authorities should use recommended procedures in the TSD to determine when numeric WET limits or WET monitoring triggers are needed. Other permit conditions may include monitoring triggers, such as increased toxicity testing, TREs/TIEs, and follow-up actions initiated because a permit limit is exceeded or a monitoring trigger is not met. Regulatory authorities should implement the additional test sensitivity requirements by requiring that each test result not exceed the upper PMSD bound. In addition, regulatory authorities should determine the appropriate NOEC for test results below the lower PMSD bound as described in Section 6.4.2. These efforts should lead to increased confidence in the effect concentrations that are generated to evaluate self-monitoring data.

7.0 CONCLUSIONS AND GUIDANCE TO LABORATORIES, PERMITTEES, AND REGULATORY AUTHORITIES

This document was prepared to address whole effluent toxicity (WET) test variability. The document has three goals: (1) quantify the variability of promulgated test methods and report a coefficient of variation (CV) as a measure of test method variability; (2) evaluate the statistical methods described in the TSD for determining the need for and deriving WET permit conditions; and (3) suggest guidance for regulatory authorities on approaches to address and minimize test method variability. This document quantified the variability of toxicity test methods based on the end use of the data, that is, the effect concentrations (e.g., NOEC, LC50, EC25). The within-laboratory variability of these effect concentrations was quantified by obtaining multiple test results under similar test conditions using the same reference toxicant. The major conclusions of this document are discussed below.

7.1 General Conclusions

- EPA's *Technical Support Document for Water Quality-based Toxics Control* (referred to as the TSD) presents guidance for developing effluent limits based on three key components: (1) water quality criteria; (2) a calculated dilution factor used to derive a waste load allocation (WLA) from the criteria; and (3) a statistical calculation procedure that uses a CV based on effluent data to calculate effluent limits from the WLA. EPA's TSD statistical approach is appropriately protective, regarding both effluent and analytical variability, provided that the criteria and WLA are derived correctly. It is inappropriate to adjust the TSD statistical methodology for determining when water quality-based effluent limits are needed and for calculating such limits (Section 6 and Appendix G).
- EPA's analysis indicates that the TSD approach appropriately accounts for both effluent variability and method variability. EPA does not believe a reasonable alternative approach is available to determine a factor that would discount the effects of method variability using the TSD procedures (Section 6.1.1 and Appendix G).
- Interim CVs are identified for promulgated WET test methods [Appendix A, Table A-1 (acute methods) and Table A-2 (chronic methods)], pending completion of between-laboratory studies, which may affect these interim CV estimates.
- Comparisons of WET method precision with method precision for analytes commonly limited in the National Pollutant Discharge Elimination System (NPDES) permits clearly demonstrate that the variability of the promulgated WET methods is within the range of variability experienced in other types of analyses. Several researchers also noted that method performance improves when prescribed methods are followed closely by experienced analysts (Section 4.3).
- The hypothesis test procedures prescribed in EPA's WET methods will provide adequate protection against false conclusions that an effluent is toxic. However, the incidence of false negatives can be high because of high within-test variability, making it difficult to detect toxicity when toxicity is truly present. Therefore, evaluating the power of current experimental designs is desirable. EPA expects that regulatory authorities will make prompt and measurable progress toward the goal of requiring all WET tests to detect a toxic effect of 25 percent to 33 percent with power of 0.80 (Section 5.3.3 and Appendix B.4).
- Quality assurance problems became apparent when evaluating the data for this study, especially for the metal reference toxicants and sodium dodecyl sulfate (SDS). Standardizing the choice of reference toxicant and the concentrations to be tested may be appropriate, as well as establishing bounds on the range of acceptable effect concentrations for each test method. As a result,

quantifying between-laboratory variability will be difficult unless these issues can be resolved (Section 5.3.1 and Appendix G.2.5).

- The data analysis did not reveal the potential sources and causes of variability, such as using different sources of test organisms, dilution water, and food. To assess the sources of variability fully, experimenters must carefully design new studies (Section 5.3.1).

7.2 Recommendations for Minimizing Variability and Its Effects

Three critical areas are identified to minimize WET test method variability:

- Obtaining a representative effluent sample,
- Conducting the toxicity tests properly to generate biological endpoints, and
- Calculating the appropriate statistical endpoints to have confidence in the effect concentration.

This document provides guidance to toxicity testing laboratories, permittees, and regulatory authorities in conducting biological and statistical methods and evaluating test effect concentrations. It also develops guidance for regulatory authorities on approaches to address and minimize test method variability. The principal aspects of the guidance are presented in Table 3-6 and re-presented here.

Range of Relative Variability for Endpoints of Promulgated WET Methods, Defined by the 10th and 90th Percentiles from the Data Set of Reference Toxicant Tests[a]

Test Method[b]	Endpoint[c]	No. of Labs	No. of Tests	PMSD		Control CV[d]	
				10th	90th	10th	90th
1000.0 Fathead Minnow	G	19	205	9.4	35	0.035	0.20
1002.0 *Ceriodaphnia dubia*	R	33	393	11	37	0.089	0.42
1003.0 Green Alga	G	9	85	9.3	23	0.034	0.17
1004.0 Sheepshead Minnow	G	5	57	6.3	23	0.034	0.13
1006.0 Inland Silverside	G	18	193	12	35	0.044	0.18
1007.0 Mysid	G	10	130	12	32	0.088	0.28
2000.0 Fathead Minnow	S	20	217	4.2	30	0	0.074
2002.0 *Ceriodaphnia*	S	23	241	5.0	21	0	0.11
2004.0 Sheepshead Minnow	S	5	65	0[e]	55	0	0
2006.0 Inland Silverside	S	5	48	7.0	41	0	0.079
2007.0 Mysid (*A. bahia*)	S	3	32	5.1	26	0	0.081
2011.0 Mysid (*H. costata*)	S	2	14	18	47	0	0.074
2021.0 Daphnia (*D. magna*)	S	5	48	5.3	23	0	0.11
2022.0 Daphnia (*D. pulex*)	S	6	57	5.8	23	0	0.11

[a] The precision of the data warrants only three significant figures. When determining agreement with these values, one may round off values to two significant figures (e.g., values >3.45000... and ≤3.5000... are rounded to 3.5). Method 1009.0 (red macroalga) is not reported because it is inadvisable to characterize method variability using only 23 tests from just two laboratories.

[b] EPA did not assign method numbers for acute methods in EPA/600/4-90/027F. The numbers assigned here were created for use in this document and in related materials and data bases.

[c] G = growth, R = reproduction, S = survival

[d] CVs were calculated using untransformed control means for each test.

[e] An MSD of zero will not occur when the EPA flow chart for statistical analysis is followed. In this report, MSD was calculated for every test, including those for which the flow chart would require a nonparametric hypothesis test. EPA recommends using the value 4.2 (the 10th percentile shown for the fathead minnow acute test) in place of zero as the 10th percentile PMSD (lower PMSD bound) for the sheepshead minnow acute test.

7.2.1 Guidance to Toxicity Testing Laboratories

- Testing laboratories should maintain quality assurance/quality control (QA/QC) control charts for percent minimum significant difference (PMSD) along with the statistical endpoints such as NOEC, LC50, and EC25. Testing laboratories should regularly plot the individual raw test data and the average treatment responses to examine possible causes of excessive variability (Section 5.3.1.1).

- The minimum number of replicates for the chronic toxicity tests should be four for the chronic fathead minnow, sheepshead minnow, and inland silverside test methods (Sections 5.3.3.1 and 5.6).

- Testing laboratories should take steps to ensure that the test PMSD does not exceed the upper bound provided in the table above (Sections 3.3, 5.3.3, and 6.4 and Table 3-6). This may require ensuring more uniformity among test organisms and/or using more replicates. Tables are provided to aid in choosing the number of replicates (Tables B-14 and B-15).

- Testing laboratories should examine the power tables to ensure that test results will meet the recommended test sensitivity criteria. These tables can be used to make decisions about replication, given the knowledge of typical values for error mean square (EMS) and number of tested concentrations (Section 5.3.3 and Tables B-9 through B-15).

7.2.2 Guidance to NPDES Permittees

- Permittees should select and conduct all data analyses with one qualified toxicity testing laboratory to determine reasonable potential, derive permit limits, and generate self-monitoring test results. Conducting all effluent testing consistently using one reference toxicant is also prudent (Section 6.1.4 and Appendix G.2.5).

- Permittees should generate WET data (n = 10) that have been accumulated over a year or more to fully characterize effluent variability over time. The sampling dates and times should span a sufficient duration to represent the full range of effluent variability (Sections 6.1.3 and 6.2 and Appendix G.2.4).

- Permittees should examine testing laboratories' QA/QC control charts. If the CV for reference toxicant tests is greater than the 75th percentile in Tables 3-2 through 3-4, variability can likely be reduced, even if the individual EC25 and LC50 values fall within the control limits (Section 5.3.1.1).

- Permittees should examine toxicity test data to ensure that data being submitted to regulatory authorities meet specified effluent holding times, temperature, laboratory control limits, and test acceptability criteria, such as requirements for test sensitivity lower and upper PMSD bounds (Sections 5.2 through 5.4).

- Permittees should anticipate and plan for a change if switching to a different testing laboratory. The permittee should compare reference toxicant test data from the current laboratory with data from the candidate replacement laboratory in order to ensure acceptable variability and a similar average effect (Section 6.1.4).

7.3 Guidance to Regulatory Authorities

Guidance to Regulatory Authorities Related to Determining Reasonable Potential and Deriving Permit Limits:

- Regulatory authorities should use EPA's recommended statistical approach in deriving permit limitations. The statistical approach outlined in the TSD represents an effective and appropriately protective approach to effluent limit development (Section 6.1 and Appendix G.1).

- Regulatory authorities should calculate the facility-specific CV using point estimate techniques to determine the need for and derive a permit limit, even if the self-monitoring test results will be determined using hypothesis test procedures (Sections 3.4.1 and 6.2).

- Regulatory authorities that need to cite a characteristic CV for a promulgated method may use Tables A-1 and A-2 in Appendix A, which show the median CV from Tables 3-2 through 3-4, pending completion of between-laboratory studies.

- EPA recommends that regulatory authorities implement a step-wise approach to address toxicity. This approach can determine the magnitude and frequency of toxicity and appropriate follow-up actions for test results that indicate exceedance of a monitoring trigger or a permit limit (Section 6.5).

Guidance to Regulatory Authorities Related to Collecting Effluent Samples, Conducting the Toxicity Test, and Evaluating the Effect Concentrations:

- Regulatory authorities should design a sampling program that collects representative effluent samples to fully characterize effluent variability for a specific facility over time. At least 10 samples are needed to estimate a variance or CV with acceptable precision for a specific facility (Sections 6.1.3 and 6.2).

- Regulatory authorities should ensure that statistical procedures and test methods have been properly applied to produce WET test results. Evaluating other factors and data, such as biological and statistical quality assurance, and ensuring that test conditions and test acceptability criteria (TAC) have been met would be prudent (Sections 5.2 through 5.5).

- Regulatory authorities should apply both the upper and lower bounds using the PMSD as an additional TAC (Section 6.4 and Table 3-6). The State of North Carolina implemented an effective WET program that required additional TAC and guidance for test methods that served to minimize test method variability (Appendix F).

- Regulatory authorities should develop a QC checklist to assist in evaluating and interpreting toxicity test results (Section 5.3.1.1). See Appendix E for examples of State WET implementation procedures.

- Regulatory authorities should consider participation in the National Environment Laboratory Accreditation Program and should conduct routine performance audit inspections to evaluate individual laboratory performance. Inspections should evaluate the laboratory's performance with QC control charts based on reference toxicants, examine procedures for conducting the toxicity test procedures, and examine procedures for analyzing test results (Section 5.3.1.1).

- Regulatory authorities should incorporate revised technical guidance recently published by EPA captioned "Method Guidance and Recommendations for Whole Effluent Toxicity (WET) Testing"

(40 CFR Part 136) (USEPA 2000a). The guidance addresses: (1) error rate assumption adjustments; (2) concentration-response relationships; (3) incorporation of confidence intervals; (4) acceptable dilution waters for testing; (5) guidance on blocking by parentage for the chronic *C. dubia* test method; and (6) procedures for controlling pH drift.

7.4 Future Directions

- An independent peer-reviewed workshop should be convened to evaluate alternatives to the statistical approaches currently used in EPA's WET test methods. Such a workshop might suggest alternatives regarding (1) WET statistical flowcharts, (2) WET statistical methods used to estimate effect concentrations, and (3) test data interpretation and review guidelines (Section 5.5).

- Such a workshop might also evaluate additional QC requirements and recommendations regarding the specification of a reference toxicant and the concentrations to be tested for each test method (Section 5.3.1).

This page intentionally left blank.

8.0 BIBLIOGRAPHY

Anderson, B.S., J.W. Hunt, S.L. Turpen, A.R. Coulon, M. Martin, D.L. Denton, and F.H. Palmer. 1990. *Procedures Manual for Conducting Toxicity Tests Developed by the Marine Bioassay Project.* California State Water Resources Control Board. No. 90-10WQ.

Anderson, S.L., and T.J. Norberg-King. 1991. Precision of short-term chronic toxicity tests in the real world. *Environ. Toxicol. Chem.* 10(2):143–145.

APHA-AWWA-WEF. 1998. *Standard Methods for the Examination of Water and Wastewater* (20th Edition). L.S. Clesceri, A.E. Greenberg, and A. Eaton, eds. American Public Health Association, American Water Works Association, and Water Environment Federation. Washington, DC. ISBN 0-87553-235-7/WB.

American Society for Testing and Materials (ASTM). 1992. *Standard Practice for Conducting an Interlaboratory Study To Determine the Precision of a Test Method.* E691-92.

ASTM. 1998. *Standard Practice for Determination of Precision and Bias of Applicable Test Methods of Committee D-19 on Water*. D2777.

Ausley, L.W. 1996. Effluent toxicity testing variability. In *Whole Effluent Toxicity Testing: An Evaluation of Methods and Prediction of Receiving System Impacts*. D.R. Grothe, K.L. Dickson, and D.K. Reed-Judkins, eds. Pensacola, FL: SETAC Press, 157–171.

Bailer, A.J., M.R. Hughes, D.L. Denton, and J.T. Oris. 2000. An empirical comparison of effective concentration estimators for evaluating aquatic toxicity test responses. *Environ. Toxicol. Chem*. 19(1): 141-150.

Biomonitoring Science Advisory Board (BSAB). 1994. *West Coast Marine Species Chronic Protocol Variability Study.* Washington Department of Ecology. Olympia, WA.

Burton, G.A., A. Raymon, L.A. Ausley, J.A. Black, G.M. DeGraeve, F.A. Fulk, J.F. Heltshe, W.H. Peltier, J.J. Pletl, and J.H. Rodgers. 1996. Session 4: Effluent toxicity test variability. In *Whole Effluent Toxicity Testing: An Evaluation of Methods and Prediction of Receiving System Impacts.* D.R. Grothe, K.L. Dickson, and D.K. Reed-Judkins, eds. Pensacola, FL: SETAC Press, 131–156.

Chapman, G.A. 1992. *Sea Urchin (Strongylocentrotus purpuratus) Fertilization Test Method.* USEPA ERL-Narragansett, Pacific Ecosystems Branch. Newport, OR.

Chapman, G.A., B.S. Anderson, A.J. Bailer, R.B. Baird, R. Berger, D.T. Burton, D.L. Denton, W.L. Goodfellow, M.A. Heber, L.L. McDonald, T.J. Norberg-King, and P.L. Ruffier. 1996a. Discussion synopsis, methods and appropriate endpoints. In *Whole Effluent Toxicity Testing: An Evaluation of Methods and Prediction of Receiving System Impacts.* D.R. Grothe, K.L. Dickson, and D.K. Reed-Judkins, eds. Pensacola, FL: SETAC Press, 51-82.

Chapman, P.F., M. Crane, J. Wiles, F. Noppert, and E. McIndoe. 1996b. Improving the quality of statistics in regulatory ecotoxicity tests. *Ecotoxicol*. 5:169-186.

Chapman, P.M., R.S. Caldwell, and P.F. Chapman. 1996c. A warning: NOECs are inappropriate for regulatory use. *Environ. Toxicol. Chem.* 15:77–79.

Collett, D. 1991. *Modelling Binary Data.* London: Chapman & Hall.

Commonwealth of Virginia. 1993. *Toxics Management Program Implementation Guidance*.

Davis, R.B., A.J. Bailer, and J.T. Oris. 1998. Effects of organism allocation on toxicity test results. *Environ. Toxicol. Chem*. 17(5): 928-931.

DeGraeve, G.M., J.D. Cooney, B.H. Marsh, T.L. Pollock, N.G. Reichenbach, J.H. Dean, and M.D. Marcus. 1991. Variability in the performance of the seven-day fathead minnow (*Pimephales promelas*) larval survival and growth test: A within- and among-laboratory study. *Environ. Toxicol. Chem*. 10(9):1189–1203.

DeGraeve, G.M., J.D. Cooney, B.H. Marsh, T.L. Pollock, and N.G. Reichenbach. 1992. Variability in the performance of the 7-d *Ceriodaphnia dubia* survival and reproduction test: A within- and among-laboratory study. *Environ. Toxicol. Chem*. 11(6):851–866.

DeGraeve, G.M., G. Smith, W. Clement, D. McIntyre, and T. Forgette. 1998. *WET Testing Program: Evaluation of Practices and Implementation.* Water Environment Research Foundation. Project 94-HHE-1. Alexandria, VA.

Denton, D.L., and T.J. Norberg-King. 1996. Whole effluent toxicity statistics: A regulatory perspective. In *Whole Effluent Toxicity Testing: An Evaluation of Methods and Prediction of Receiving System Impacts.* D.R. Grothe, K.L. Dickson, and D.K. Reed-Judkins, eds. Pensacola, FL: SETAC Press, 83–102.

Denton, D.L, and M. Narvaez. 1996. *Regions 9 and 10 Guidance for Implementing Whole Effluent Toxicity Testing Programs*. U. S. Environmental Protection Agency, Regions 9 and 10. May.

Denton, D.L., A.L. Suer, B.S. Anderson, and J.W. Hunt. 1992. Precision of marine critical life stage tests with west coast species (abstract). In *Society of Environmental Toxicology and Chemistry (SETAC) Abstracts, 13th Annual Meeting, November 8-12, 1992*, Cincinnati, OH. Pensacola, FL: SETAC Press, 184.

Dhaliwal, B.S., R.J. Dolan, C.W. Batts, J.M. Kelly, R.W. Smith, and S. Johnson. 1997. Warning: replacing NOECs with point estimates may not solve regulatory contradictions. *Environ. Toxicol. Chem.* 16:124–126.

Dinnel, P.J., J. Link, and Q. Stober. 1987. Improved methodology for sea urchin sperm cell bioassay for marine waters. *Arch. Environ. Contam. Toxicol.* 16:23–32.

Dunnett, C.W. 1964. New tables for multiple comparisons with a control. *Biometrics* 20:482-491.

Eagleson, K.W., S.W. Tedder, and L.W. Ausley. 1986. Strategy for whole effluent toxicity evaluations in North Carolina. In *Aquatic Toxicology and Environmental Fate: Ninth Volume, ASTM STP 921*. T.M. Poston, R. Purdy, eds. American Society for Testing and Materials. Philadelphia, PA, 154-160.

Environment Canada. 1990. *Guidance Document on Control of Toxicity Test Precision Using Reference Toxicants.* Ottawa, Ontario: Environmental Protection, Conservation and Protection. Report EPSI/RM/12.

Environment Canada. 2000. *Guidance Document on Application and Interpretation of Single-species Tests in Environmental Toxicology.* Ottawa, Ontario: Environmental Technology Centre, Method Development and Application Section. Report EPS 1/RM/34.

Erickson, R.J., L.T. Brooke, M.D. Kahl, F. Vende, S.L. Venter, T. Harting, P. Markee, and R.L. Spehar. 1998. Effects of laboratory test conditions on the toxicity of silver to aquatic organisms. *Environ. Toxicol. Chem*. 17(4): 572-578.

Fulk, F.A. 1996. Whole effluent toxicity testing variability: A statistical perspective. In *Whole Effluent Toxicity Testing: An Evaluation of Methods and Prediction of Receiving System Impacts*. D.R. Grothe, K.L. Dickson, and D.K. Reed-Judkins, eds. Pensacola, FL: SETAC Press, 172–179.

Grothe, D.R., and R.A. Kimerle. 1985. Inter- and intralaboratory variability in *Daphnia magna* effluent toxicity test results. *Environ. Toxicol. Chem*. 4(2):189–192.

Grothe, D.R., R.A. Kimerle, and C.D. Malloch. 1990. A perspective on biological assessments. *Water Environ. Technol*. pp. 1707–1710.

Grothe, D.R., K.L. Dickson, and D.K. Reed-Judkins, eds. 1996. *Whole Effluent Toxicity Testing: An Evaluation of Methods and Prediction of Receiving System Impacts*. Pensacola, FL: SETAC Press.

Hunt, J.W., B.S. Anderson, S. Tudor, M.D. Stephenson, H.M. Puckett, F.H. Palmer, and M.W. Reeve. 1996. *Marine Bioassay Project Eighth Report.* Sacramento, CA: State Water Resources Control Board. Report 85-102.

Kahn, H.D., and M.B. Rubin. 1989. Use of statistical methods in industrial water pollution control regulations in the United States. *Environ. Monitor. Assess.* 12:129–148.

Moore, T.F., S.P. Canton, and M. Grimes. 2000. Investigation of the incidence of type I errors for chronic whole effluent toxicity testing using *Ceriodaphnia dubia*. *Environ. Toxicol. Chem.* 19(1):118-122.

Morrison, G.E., E. Torelo, R. Comeleo, R. Walsh, A. Kuhn, R. Burgess, M. Tagliabue, and W. Green. 1989. Interlaboratory precision of saltwater short-term chronic toxicity tests. *Res. J. Wat. Pollut. Control Fed*. 61:1707-1710.

New Jersey Department of Environmental Protection. 1994. *The Use of Chronic Whole Effluent Toxicity Testing in the New Jersey Pollutant Discharge Elimination System—An Assessment of Compliance Data*.

North Carolina Department of Environment and Natural Resources. 1998. *North Carolina Biological Laboratory Certification/Criteria Procedures Document.* Division of Water Quality, Water Quality Section. Raleigh, NC.

Oris, J.T., and A.J. Bailer. 1993. Statistical analysis of the *Ceriodaphnia* toxicity test: Sample size determination for reproductive effects. *Environ. Toxicol. Chem*. 12(1):85–90.

Rosebrock, M.M., N.W. Bedwell, and L.W. Ausley. 1994. Indicators of *Ceriodaphnia dubia* chronic toxicity test performance and sensitivity. Poster presentation, Society of Environmental Toxicology and Chemistry 15th Annual Meeting. Denver, CO.

Rue, W.J., J.A. Fava, and D.R. Grothe. 1988. A review of inter- and intralaboratory effluent toxicity test method variability. In *Aquatic Toxicology and Hazard Assessment* (10th Volume). W.J. Adams, G.A. Chapman, and W.G. Landis, eds. ASTM STP 971.

SAS Institute. 1990. SAS/STA *User's Guide* (4th Edition), Version 6. Cary, NC.

Sheppard, C.R. 1999. How large should my sample be? Some quick guides to sample size and the power of tests. *Mar. Pollut. Bull.* 38(6):439-477.

Shukla, R., Q. Wang, F.A. Fulk, C. Deng, and D.L. Denton. 2000. Bioequivalence approach for whole effluent toxicity testing. *Environ. Toxicol. Chem*. 19(1):169-174.

Southern California Toxicity Assessment Group (SCTAG). 1996. *Reference Toxicant Standardization and Use in Toxicity Testing*. J.R. Gully, R.B. Baird, P.J. Markle, and J.P. Bottomley, eds. First Report. Fountain Valley, CA.

SCTAG. 1997. *Laboratory Practices Checklist Toxicity Testing* (3rd Edition). Fountain Valley, CA.

Starrett, G.L., D.L. Denton, and R.W. Smith. 1993. Sensitivity of toxicity test protocols and the need for additional quality assurance requirements (abstract). In *Society of Environmental Toxicology and Chemistry (SETAC) Abstracts, 14th Annual Meeting, November 14-18, 1993*, Houston, TX. Pensacola, FL: SETAC Press, P106, p. 183.

Thursby, G.B., J. Heltshe, and K.J. Scott. 1997. Revised approach to toxicity test acceptability criteria using a statistical performance assessment. *Environ. Toxicol. Chem*. 16:1322-1329.

TOXIS® [computer software]. Ojai, CA: EcoAnalysis, Inc.

TOXCALC® [computer software]. McKinleyville, CA: TidePool Scientific Software.

USEPA. 1985. *Methods for Measuring the Acute Toxicity of Effluents to Freshwater and Marine Organisms* (3rd Edition). W. Peltier and C.I. Weber, eds. Environmental Monitoring Systems Laboratory. Cincinnati, OH. EPA/600/4-85/013.

USEPA. 1988. *Short-term Methods for Estimating the Chronic Toxicity of Effects of Receiving Water to Marine/Estuarine Organisms* (1st Edition). C.I. Weber, W.B. Horning, D.J. Klemm, T.W. Neiheisel, P.A. Lewis, E.L. Robinson, J.R. Menkedick, F.A. Kessler, eds. Office of Research and Development. Cincinnati, OH. EPA/600/4-87/028.

USEPA. 1989. *Short-term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Freshwater Organisms* (2nd Edition). C.I. Weber, W.H. Peltier, T.J. Norberg-King, W.B. Homing, II, F.A. Kessler, J.R. Menkedick, T.W. Neiheisel, P.A. Lewis, D.J. Klemm, Q.H. Pickering, E.L. Robinson, J.M. Lazorchak, L.J. Wymer, R.W. Freyberg, eds. Office of Research and Development. Cincinnati, OH. EPA/600/4-89/001.

USEPA. 1989a. *Generalized Methodology for Conducting Industrial Toxicity Reduction Evaluations (TREs).* Office of Research and Development. Cincinnati, OH. EPA/600-2-88-070.

USEPA. 1989b. *Toxicity Reduction Evaluation Protocol for Municipal Wastewater Treatment Plants.* Office of Research and Development. Washington, DC. EPA/600/2-88-062.

USEPA. 1989c. *Methods for Aquatic Toxicity Identification Evaluations: Phase II Toxicity Identification Procedures.* Office of Research and Development. Washington, DC. EPA/600/3-88-035.

USEPA. 1989d. *Methods for Aquatic Toxicity Identification Evaluations: Phase III Toxicity Confirmation Procedures.* Office of Research and Development. Washington, DC. EPA/600/3-88-036.

USEPA.1991a. *Technical Support Document for Water Quality-based Toxics Control*. Office of Water. Washington, DC. EPA/505/2-90-001.

USEPA. 1991b. *Methods for Aquatic Toxicity Identification Evaluations: Phase I Toxicity Characterization Procedures* (2nd Edition). T.J. Norberg-King, D.I. Mount, E.J. Durhan, G.T. Ankley, L.P. Burkhard, J.R. Amato, M.T. Lukasewycz, M.K. Schubauer-Berigan, and L. Anderson-Carnahan, eds. Office of Research and Development, Washington, DC. EPA/600/6-91-003.

USEPA. 1992. *Toxicity Identification Evaluation: Characterization of Chronically Toxic Effluents, Phase I.* T.J. Norberg-King, D.I. Mount, J.R. Amato, D.A. Jensen, and J.A. Thompson, eds. Office of Research and Development. Washington, DC. EPA/600/6-91-005F.

USEPA. 1993. *Methods for Measuring the Acute Toxicity of Effluents and Receiving Waters to Freshwater and Marine Organisms* (4th Edition). C.I. Weber, ed. Office of Research and Development. Cincinnati, OH. EPA/600/4-90/027F.

USEPA. 1993a. *Methods for Aquatic Toxicity Identification Evaluations: Phase II Toxicity Identification Procedures for Samples Exhibiting Acute and Chronic Toxicity.* Office of Research and Development. Washington, DC. EPA/600/R-92-080.

USEPA. 1993b. *Methods for Aquatic Toxicity Identification Evaluations: Phase III Toxicity Identification Procedures for Acutely and Chronically Toxic Samples.* U.S. Environmental Protection Agency, Duluth, MN. EPA/600/R-92-081.

USEPA. 1994a. *Short-term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Marine and Estuarine Organisms* (2nd Edition). D.J. Klemm, G.E. Morrison, T.J. Norberg-King, W.H. Peltier, and M.A. Heber, eds. Office of Research and Development. Cincinnati, OH. EPA/600/4-91/003.

USEPA. 1994b. *Short-term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Freshwater Organisms* (3rd Edition). P.A. Lewis, D.J. Klemm, J.M. Lazorchak, T.J. Norberg-King, W.H. Peltier, and M.A. Heber, eds. Office of Research and Development. Cincinnati, OH. EPA/600/4-91/002.

USEPA. 1994c. *Short-term Methods for Estimating the Sediment Toxicity of Effluents and Receiving Waters to Freshwater and Estuarine Organisms*. Office of Research and Development. Duluth, MN. EPA/600/R-94-001.

USEPA. 1995. *Short-term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to West Coast Marine and Estuarine Organisms*. G.A. Chapman, D.L. Denton, and J.M. Lazorchak, eds. Office of Research and Development. Cincinnati, OH. EPA/600/R-95-136.

USEPA. 1996a. *NPDES Permit Writer's Manual*. Office of Water. Washington, DC. EPA/833/B-96-003.

USEPA 1996b. *Guide to Method Flexibility and Approval of EPA Water Methods.* EPA/821/D-96-004.

USEPA. 1999a. *National Environmental Laboratory Accreditation Conference: Constitution, Bylaws and Standards*. Office of Research and Development. Washington, DC. EPA/600/R-99-068.

USEPA. 1999b. *Toxicity Reduction Evaluation Guidance for Municipal Wastewater Treatment Plants* (2nd Edition). Office of Water. Washington, DC. EPA/833/B-99-002.

USEPA. 1999c. *Errata for Effluent and Receiving Water Toxicity Test Manuals: Acute Toxicity manuals: Acute Toxicity of Effluents and Receiving Waters to Freshwater and Marine Organisms; Short-Term methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Freshwater Organisms; and Short-Term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Marine and Estuaraine Organisms*. Office of Research and Development. Duluth, MN.

USEPA. 2000a. *Method Guidance and Recommendations for Whole Effluent Toxicity (WET) Testing* (40 CFR Part 136). Office of Water, Office of Science and Technology. Washington, DC. EPA/821/B-00-004.

USEPA. 2000b. *Methods for Measuring the Toxicity and Bioaccumulation of Sediment-Associated Contaminants with Freshwater Invertebrates* (2nd Edition). Office of Research and Development. Duluth, MN. EPA/600/R-99/064.

Wang, Q., D.L. Denton, and R. Shukla. 2000. Applications and statistical properties of minimum significant difference-based criterion testing in a toxicity testing program. *Environ. Toxicol. Chem*. 19(1):113-117.

Warren-Hicks, W., B.R. Parkhurst, D. Moore, and S. Teed. 1999. *Whole Effluent Toxicity Testing Methods: Accounting for Variance*. Water Environment Research Foundation. Project 95-PQL-1. ISBN 1-893664-01-5.

Washington State Department of Ecology. 1997. *Laboratory Guidance and Whole Effluent Toxicity Test Review Criteria*. Publication No. WQ-R-95-80.

Washington State Department of Ecology. 1998. *Whole Effluent Toxicity (WET) Program Evaluation*. Publication No. 98-03.

Whitehouse, P., M. Crane, C.J. Redshaw, and C. Turner. 1996. Aquatic toxicity tests for the control of effluent discharges in the UK: The influence of test precision. *Ecotoxicol.* 5:155-168.

APPENDIX A

INTERIM COEFFICIENTS OF VARIATION OBSERVED WITHIN LABORATORIES FOR REFERENCE TOXICANT SAMPLES ANALYZED USING EPA'S PROMULGATED WHOLE EFFLUENT TOXICITY METHODS

This page intentionally left blank.

INTERIM COEFFICIENTS OF VARIATION OBSERVED WITHIN LABORATORIES FOR REFERENCE TOXICANT SAMPLES ANALYZED USING EPA'S PROMULGATED WHOLE EFFLUENT TOXICITY METHODS

Tables A-1 and A-2 identify interim coefficients of variation for each promulgated WET method. The Agency identifies these as “interim” because EPA may revise some or all of these estimates based on between-laboratory studies currently underway to evaluate some of the test methods. For the acute toxicity methods, only “primary” organisms identified in the EPA method manuals (USEPA 1994a, 1994b) are reported in the tables. The primary data used to calculate these CVs were estimated effect concentrations (EC25, LC50, and NOEC) in units of concentration (e.g., mg/L of toxicant). Most CVs in Tables A-1 and A-2 come directly from Tables 3-2 through 3-4. Those data were supplemented as necessary with data from EPA publications (USEPA 1991, 1994a, 1994b). In Table 3-2, the NOEC values are reported separately for each test endpoint. In Tables A-1 and A-2, however, the NOEC values are reported as the most sensitive test endpoint. The data for a given method represent a variety of toxicants. In general, laboratories reported data for only one toxicant for a given method. Some of the data taken from EPA publications involved tests using different toxicants but conducted at one laboratory. In such cases, CVs were calculated separately for each toxicant.

Tables A-1 and A-2 report a default value when results were available from fewer than three laboratories and a similar species could be used as a basis for the default value of the CV. The sources of default values are identified in the footnotes to Tables A-1 and A-2. For methods and endpoints represented by fewer than three laboratories, the interim CV should be regarded as highly speculative.

Coefficients of variation are used as descriptive statistics for NOECs in this document. Because NOECs can take on only values that correspond to concentrations tested, the distribution (and CV) of NOECs can be influenced by the selection of experimental concentrations, as well as additional factors (e.g., within-test variability) that affect both NOECs and point estimates. This makes CVs for NOECs more uncertain than those of point estimates, and the direction of this uncertainty is not uniformly toward larger or smaller CVs. Despite these confounding issues, CVs are used herein as the best available means of expressing the variability of interest in this document and for general comparisons among methods. Readers should be cautioned, however, that small differences in CVs between NOECs and point estimates may be artifactual; large differences are more likely to reflect real differences in variability (a definition of what is “small” or “large” would require a detailed statistical analysis and would depend upon the experimental and statistical details surrounding each comparison).

These results are based on tests conducted using reference toxicants. These CVs may not apply to tests conducted on effluents and receiving waters unless the effect concentration (i.e., the EC25, LC50, or NOEC) happens to fall in the middle of the range of concentrations tested. More often, tests of effluents and receiving waters show smaller effects at the middle concentrations. Many effluent tests also demonstrate that the effect concentration equals or exceeds the highest concentration tested. In such cases, the sample standard deviation and CV tend to be smaller than reference toxicant CVs.

Table A-1. Interim Coefficients of Variation for EPA's Promulgated Whole Effluent Toxicity Methods for Acute Toxicity

Test Method No. [a]	Test Organism	Estimate	CV	No. of Laboratories
2002.0	*Ceriodaphnia dubia*	LC50	0.19[b]	23
2021.0	*Daphnia magna*	LC50	0.22[b]	5
2022.0	*Daphnia pulex*	LC50	0.21[b]	6
2000.0	*Pimephales promelas*	LC50	0.16[b]	21
2019.0	*Oncorhynchus mykiss*	LC50	0.16[c]	na[c]
NA	*Salvelinus fontinalis*	LC50	0.16[c]	na[c]
2004.0	*Cyprinodon variegatus*	LC50	0.14[b]	5
2006.0	*Menidia beryllina*	LC50	0.16[b]	5
2007.0	*Mysidopsis bahia*	LC50	0.25[b]	3

[a] These codes for acute methods were developed specifically for this document.
[b] From Table 3-3.
[c] Default values. These values are identified for methods represented by fewer than three laboratories. Default values for the trout (*Salvelinus fontinalis*) are based on Method 2000.0. Default values for *Menidia menidia* and *M. penisulae* (not shown) are based on the median for *M. beryllina*.

NOTE: CVs represent the median coefficient of variation observed within laboratories for WET tests conducted on reference toxicant samples. The test endpoint is survival.

Table A-2. Interim Coefficients of Variation for EPA's Promulgated Whole Effluent Toxicity Methods for Short-Term Chronic Toxicity

Test Method No.	Test Organism	Endpoint	Estimate	CV	No. of Laboratories
1000.0	*Pimephales promelas*	Growth Survival Most sensitive	EC25 LC50 NOEC	0.26[a] 0.23[a] 0.31[a]	19 19 19
1001.0	*Pimephales promelas* Embryo-larval	Mortality + Teratogenicity Mortality + Teratogenicity Mortality + Teratogenicity	EC01 LC50 NOEC	0.52[b] 0.07[c] 0.22[c]	1 na na
1002.0	*Ceriodaphnia dubia*	Reproduction Survival Most sensitive	EC25 LC50 NOEC	0.27[a] 0.16[a] 0.35[a]	33 33 33
1003.0	*Selenastrum capricornutum*[d]	Cell count Cell count	EC25 NOEC	0.26[a] 0.46[a]	6 9 9
1004.0	*Cyprinodon variegatus*	Growth Survival Most sensitive	EC25 LC50 NOEC	0.13 0.08 0.38[c]	5 5 5
1005.0	*Cyprinodon variegatus* Embryo-larval	Mortality + Teratogenicity Mortality + Teratogenicity Mortality + Teratogenicity	EC10 LC50 NOEC	0.19[e] 0.07[e] 0.22[e]	1 1 1
1006.0	*Menidia beryllina*	Growth Survival Most sensitive	EC25 LC50 NOEC	0.27[a] 0.28[a] 0.46[a]	16 16 16
1007.0	*Mysidopsis bahia*	Growth Survival Most sensitive	EC25 LC50 NOEC	0.28[a] 0.26[a] 0.40[a]	10 10 10
1008.0	*Arbacia punctulata*	Fertilization Fertilization	EC25 NOEC	0.36[e] 0.50[c]	2 na
1009.0	*Champia parvula*	Cystocarp production Cystocarp production	EC25 NOEC	0.59[a, e] 0.85[a, e]	3 3

[a] Tables 3-2 through 3-4.
[b] USEPA 1994b, USEPA 1991.
[c] Default values. These values are identified, when possible, for methods represented by fewer than three laboratories. The default value for *Cyprinodon* is based on *Pimephales*. Default values for *Menidia menidia* and *M. penisulae* (not shown) are based on the median for *Menidia beryllina*. Default values for Method 1001.0 were based on Method 1005.0. The default value for Method 1008.0 was based on Method 1016.0 of Table B-3 in Appendix B.
[d] Genus and species recently changed to *Raphidiopsis subcapitata*.
[e] USEPA 1994a, USEPA 1991.

NOTE: CVs represent the median coefficient of variation observed within laboratories for WET tests conducted on reference toxicant samples. NOEC estimates are reported for the most sensitive endpoint. This means that, for each test, the NOEC value was recorded for the endpoint that produced the lowest NOEC test result.

This page intentionally left blank.

APPENDIX B

SUPPLEMENTARY INFORMATION FOR REFERENCE TOXICITY DATA

This page intentionally left blank.

SUPPLEMENTARY INFORMATION FOR REFERENCE TOXICITY DATA

Appendix B contains technical and explanatory notes, and supplementary tables pertaining to the statistical analyses of reference toxicant test results presented in Chapters 3 and 5.

B.1 Acquisition, Selection, and Quality Assurance of Data

Details of data quality assurance and test acceptance are provided in a separate document, available from the EPA Office of Water's Office of Science and Technology ("Whole Effluent Toxicity (WET) Data Test Acceptance and Quality Assurance Protocol"). On request, EPA will also make available a list by laboratory of quality assurance (QA) flags, test dates, toxicant concentration, and summary statistics for the NOEC, EC25, and EC50 estimates and the test endpoints (survival, growth, reproduction, etc.). Laboratories are not named. Data were obtained as data sets from the data base and statistical software packages TOXIS® and TOXCALC® (see Chapter 8 for citations).

TOXIS® software produces an acceptability criterion field code based on the TAC specified by the EPA WET test methods. The tests having "I" (Incomplete) or "F" (Failed) values in this field were eliminated from consideration. TOXCALC® data were examined at the individual test level. The first step, before data entry, consisted of examining the test for TAC from bench sheets. The data were then imported into TOXCALC® for analysis. However, TOXCALC®, unlike TOXIS®, does not generate error codes but issues a warning on the screen. These messages were examined and decisions were made case-by-case following EPA test methods. In the second step, a QA program code was written in SAS® to check the TAC listed in the WET test methods for acute and chronic toxicity tests.

The effect concentration values produced using TOXCALC® or TOXIS®, along with related test information, were exported to spreadsheets and then imported into a SAS® data set. All statistical analyses, other than calculations of effect concentration estimates, were conducted using SAS®. Various data QA tests were conducted. Checks were made to ensure that data were within acceptable concentration-response ranges. Also, the frequency of tests, laboratories, and toxicants were compared for initial and final data sets to ensure that the data were properly imported and exported. Furthermore, TOXIS® effect concentrations having unacceptable error codes such as 905 (i.e., exposure concentrations for LC/EC values unrealistically high due to small slope and estimates well beyond the highest concentration used) and 904 (i.e., non-homogeneity of variance for a Probit estimate) were rejected. The TAC were not verified independently of TOXIS®, although the data used passed the required TAC. Because TOXIS® does not export the qualifier for censored endpoint values (i.e., ">" for greater than and "<" for less than), these qualifiers were later added to cases in which the point estimate equaled the maximum or minimum concentration in the dilution series. The methods having two biological endpoints per test method (e.g., survival and reproduction) had to pass both endpoint TACs to be included in the data analysis.

Non-standard laboratory codes were investigated by follow-up with the data provider; such cases were resolved either by reconfirming the laboratory identity or in a few cases by flagging the data as unusable. Duplicate data sets were identified and eliminated; this involved comparing the test methods, organisms, laboratory codes, test dates, test codes, concentration series, and replicate endpoint means. Concentration units were standardized for each toxicant. Errors in concentration units (e.g., µg versus mg) were identified and resolved. The number of organisms and number of replicates were not used to select or reject tests. For example, the minimum number of replicates was three for Method 1000.0 (which applied to only a few tests, since most tests used four replicates, but some used three) and seven for Method 1002.0 (which was exceptional since most tests used ten replicates).

Only the 20 most recent tests were used if more were submitted. Only laboratories having at least six data points were reported for the toxicants potassium chloride (KCl) and sodium chloride (NaCl) for two common methods: Method 1000.0 (fathead minnow larval survival and growth) and Method 1002.0

(*Ceriodaphnia* survival and reproduction). For other toxicants and methods, the minimum number of data points per laboratory was set at four. The within-laboratory statistics based on only four tests can be imprecise and should be regarded with caution.

In past protocols, the growth and reproduction effect values for the fathead minnow test (Method 1000.0), inland silverside test (Method 1006.0), and mysid test (Method 1007.0) were determined by dividing the weight or reproduction by the number of survivors. In contrast, the currently promulgated methods require that the weight or reproduction values be divided by the original (starting) number of organisms. All such results herein were calculated as currently required, using the weight or reproduction divided by the original number of organisms.

Note that data for Method 1016.0 (purple urchin fertilization test) and Method 1017.0 (sand dollar fertilization) included three different test methods with primary method differences including different sperm-egg ratios, sperm collection procedures, and sperm exposure time. This method has since been standardized and included in the West Coast chronic marine test methods manual (USEPA 1995).

A large percentage of data from a few laboratories was censored (i.e., recorded as "<" or ">") because the effect concentration was outside the range of the concentration series. In some cases, the data were censored because of the number or range of toxicant concentrations tested. When many data are censored, a reversal in the most sensitive endpoint can occur. For example, in the data for Method 1006.0 (*Menidia beryllina* larval survival and growth test), the NOEC for the survival endpoint indicated a more sensitive response than the sublethal endpoint for some tests.

B.2 Summary Statistics for IC25, LC50, and NOEC

B.2.1 Within-Laboratory Variability of EC25, EC50, and NOEC

Test data were not screened for outliers as provided for in ASTM Practices D2777 and E691 (ASTM 1992, 1998). Thus, maximum and minimum values for the laboratory statistics summarized in Tables B-1 through B-6 may be distorted by outliers. Therefore, EPA concluded that the maximum and minimum values are not necessarily reliable and has not reported them in these tables. EPA recommends that the 10th and 90th percentiles reported in Tables B-1 through B-6 be used to characterize the range of test variability.

Tables B-1 through B-3 show percentiles of the within-laboratory coefficients of variation (CVs) for EC25, EC50, and NOEC for all methods in the variability data set. However, when a method is represented by few laboratories, this summary cannot be considered typical or representative. When there were fewer than ten laboratories for a method, the 10th and 90th percentiles could not be estimated in an unbiased manner. Columns P10 and P90 show the minimum and maximum in such cases. Similarly, when there were fewer than four laboratories, columns P10 and P25 show the minimum and columns P75 and P90 show the maximum. An unbiased estimate of the median is always shown.

These percentiles are found by interpolation between two sample order statistics. The kth sample order statistic has an expected probability estimated by $P_k = (k - 0.375)/(N + 0.25)$. Linear interpolation between two order statistics (X_k and X_{k+1}) having expected probabilities $P_k < P < P_{k+1}$ provides the estimate of the Pth quantile.

Tables B-4 through B-6 summarize variation across laboratories for the within-laboratory normal ratio of extremes for the EC25, EC50, and NOEC estimates. Instead of using the ratio of largest-to-smallest observations, which is vulnerable to outliers, the ratio of the 90th to the 10th percentiles (symbolized P90:P10) was used to provide some robustness to outliers. This ratio is a measure of variability in terms of concentration ratio. About 80 percent of observations are expected to fall between these percentiles. Thus, if P90:P10 equals 4, about 80 percent of observations are expected to fall within a dilution ratio of 4 (e.g., 0.25 mg/L to 1.00 mg/L).

The ratio is dimensionless and a more useful measure of the "range" of test results than the concentration range. For example, NOECs may vary at one laboratory between 0.5 mg/L and 2.0 mg/L (giving a range of 1.5 mg/L) and at another laboratory between 0.25 mg/L and 1.0 mg/L (giving a range of 0.75 mg/L), yet both NOECs span two standard concentrations having a ratio of 1:4. Also, using a ratio allows direct comparison among different toxicants having different concentration units. Further, toxicity tests often require a log scale (that is, a ratio scale) of concentration to provide an approximately linear curve of endpoint response (Collett 1991). Environment Canada (2000) expects that plotting and statistical estimation for WET tests will employ a logarithmic scale. In EPA publications, logarithmic (constant-ratio) graphical scales are used for concentrations (USEPA 1994a,1994b).

Tables B-4 through B-6 provide an easy way to quantify the ratio among effect concentrations expected for 80 percent of tests. For example, in Table B-6 under the NOEC for the growth endpoint of Method 1000.0, the median laboratory has a ratio of 2.0. This means that for half of the laboratories, repeated reference toxicant tests gave NOECs, 80 percent of which differed by no more than one standard dilution. That is, most NOECs occurred at only one concentration or at two adjacent concentrations at half of the laboratories. Note that most tests used 1:2 dilutions, so for the NOEC, the only exact ratios possible for each test are 1:1, 1:2, 1:4, 1:8, and 1:16. Thus, for NOECs, the results presented in the tables may be interpreted by rounding to these ratios.

The ratios P90:P10 in Tables B-4 through B-6 can be summarized as follows. For the NOEC in most of the promulgated WET methods, 75 percent of laboratories achieve a ratio of no more than 1:4, and half of the laboratories routinely achieve ratios of 1:1 or 1:2. For the LC50 (survival endpoint) for most methods, 75 percent of laboratories have ratios no more than 1:3, and half the laboratories have ratios no more than 1:2. For the IC25 (growth and reproduction endpoints), 75 percent of laboratories have ratios no more than 1:4, and half of laboratories have ratios no more than 1:2.5. The ratio for acute methods is usually somewhat less than that for chronic methods.

Note that two laboratories having the same ratio P90:P10 do not necessarily have similar NOECs; between-laboratory variation also occurs. For example, consider three laboratories that reported data for the growth endpoint of Method 1000.0 tested with NaCl. Each has a ratio P90:P10 of 2.0. One laboratory reported 11 tests, with the NOEC ranging from 0.4 mg/L to 3.2 mg/L. The 10th and 90th percentile estimates were 1.6 and 3.2. A second laboratory reported 8 tests, with the NOEC ranging from 1.0 mg/L to 2.0 mg/L. The 10th and 90th percentile estimates were 1.0 and 2.0. A third laboratory reported 12 tests, with the NOEC ranging from 1.0 mg/L to 4.0 mg/L. The 10th and 90th percentile estimates were 1.0 and 2.0.

B.2.2 Between-Laboratory Variability of EC25, EC50, and NOEC

The estimates of within- and between-laboratory variability for WET tests in Table 3-5 (Chapter 3) are based on Type-I analysis of variance and expected mean squares for random effects. Within-laboratory variability is estimated as the square root of the error mean square (column "Within-lab σ_w"), that is, the pooled standard deviation for all tests and all laboratories available for a given method, toxicant, and endpoint. Column "Between-lab σ_b" is the square root of the between-laboratory variance term, calculated as shown below. The column headed "Mean" shows the mean of the (unweighted) laboratory means. Sample sizes (numbers of laboratories) are insufficient for credible estimates of between-laboratory variability for most methods. The expected mean squares assume that the population of laboratories is large. Finite population estimates would be more accurate for some combinations of method and toxicant.

Table B-1. Percentiles of the Within-Laboratory Values of CV for EC25

Test Method[a]	Test Method No.[b]	End-point[c]	No. of Labs	CV				
				P10	P25	P50	P75	P90
Chronic, Promulgated								
Fathead Minnow Larval Survival & Growth	1000.0	G	19	0.12	0.21	0.26	0.38	0.45
Fathead Minnow Larval Survival & Growth	1000.0	S	16	0.03	0.11	0.22	0.32	0.52
Ceriodaphnia (Cd) Survival & Reproduction	1002.0	R	33	0.08	0.17	0.27	0.45	0.62
Ceriodaphnia (Cd) Survival & Reproduction	1002.0	S	25	0.07	0.11	0.23	0.41	0.81
Green Alga (*Selenastrum*)[d] Growth	1003.0	G	6	0.02	0.25	0.26	0.39	0.51
Sheepshead Minnow Larval Survival & Growth	1004.0	G	5	0.03	0.09	0.13	0.14	0.18
Sheepshead Minnow Larval Survival & Growth	1004.0	S	2	0.15	0.15	0.16	0.17	0.17
Inland Silverside Larval Survival & Growth	1006.0	G	16	0.05	0.18	0.27	0.43	0.55
Inland Silverside Larval Survival & Growth	1006.0	S	13	0.15	0.22	0.35	0.42	0.62
Mysid (Ab) Survival, Growth, & Fecundity	1007.0	R	4	0.22	0.03	0.38	0.41	0.42
Mysid (Ab) Survival, Growth, & Fecundity	1007.0	G	10	0.21	0.24	0.28	0.32	0.04
Mysid (Ab) Survival, Growth, & Fecundity	1007.0	S	7	0.17	0.17	0.21	0.28	0.32
Red Macroalga *(Champia parvula)* Reprod	1009.0	R	2	0.58	0.58	0.58	0.59	0.59
West Coast								
Topsmelt Larval Survival & Growth	1010.0	G	1	0.25	0.25	0.25	0.25	0.25
Topsmelt Larval Survival & Growth	1010.0	S	1	0.20	0.20	0.20	0.20	0.20
Pacific Oyster Embryo-Larval Survival & Dev.	1012.0	D	1	0.25	0.25	0.25	0.25	0.25
Mussel Embryo-Larval Survival & Dev.	1013.0	D	3	0.14	0.14	0.27	0.42	0.42
Red Abalone Larval Development	1014.0	D	10	0.13	0.15	0.25	0.35	0.36
Sea Urchin Fertilization	1016.0	F	12	0.18	0.26	0.41	0.58	0.68
Sand Dollar Fertilization	1017.0	F	7	0.25	0.35	0.43	0.51	0.60
Giant Kelp Germination & Germ-Tube Length	1018.0	G_e	11	0.33	0.34	0.40	0.43	0.60
Giant Kelp Germination & Germ-Tube Length	1018.0	L	11	0.22	0.25	0.31	0.36	0.36
Acute								
Fathead Minnow Larval Survival	2000.0	S	7	0.05	0.09	0.15	0.21	0.44
Ceriodaphnia (Cd) Survival	2002.0	S	8	0.04	0.09	0.10	0.19	0.33
Sheepshead Minnow Survival	2004.0	S	3	0.08	0.08	0.13	0.46	0.46
Inland Silverside Larval Survival	2006.0	S	4	0.03	0.09	0.20	0.40	0.55
Mysid (Ab) Survival	2007.0	S	1	0.26	0.26	0.26	0.26	0.26
Mysid (Hc) Survival	2011.0	S	1	0.20	0.20	0.20	0.20	0.20
Rainbow Trout Survival	2019.0	S	1	0.11	0.11	0.11	0.11	0.11
Daphnia (Dm) Survival	2021.0	S	1	0.19	0.19	0.19	0.19	0.19
Daphnia (Dp) Survival	2022.0	S	3	0.06	0.06	0.41	0.48	0.48

[a] Cd = *Ceriodaphnia dubia*, Ab = *Americamysis (Mysidopsis) bahia*, Hc = *Holmesimysis costata*, Dm = *Daphnia magna*, Dp = *Daphnia pulex*

[b] EPA did not assign method numbers for acute methods in EPA/600/4-90/027F. The numbers assigned here were created for use in this document and in related materials and data bases.

[c] D = development, F = fertilization, G = growth, G_e = Germination, L = length, R = reproduction or fecundity, S = survival

[d] Genus and species recently changed to *Raphidocelis subcapitata*.

Table B-2. Percentiles of the Within-Laboratory Values of CV for EC50[a]

Test Method[b]	Test Method No.[c]	End-point[d]	No. of Labs	CV				
				P10	P25	P50	P75	P90
Chronic, Promulgated								
Fathead Minnow Larval Survival & Growth	1000.0	G	19	0.10	0.15	0.24	0.26	0.46
Fathead Minnow Larval Survival & Growth	1000.0	S	19	0.12	0.15	0.23	0.31	0.44
Ceriodaphnia (Cd) Survival & Reproduction	1002.0	R	33	0.06	0.12	0.23	0.29	0.46
Ceriodaphnia (Cd) Survival & Reproduction	1002.0	S	33	0.04	0.10	0.16	0.29	0.46
Green Alga (*Selenastrum*)[e] Growth	1003.0	G	9	0.16	0.19	0.27	0.30	0.63
Sheepshead Minnow Larval Survival & Growth	1004.0	G	5	0.02	0.04	0.06	0.11	0.13
Sheepshead Minnow Larval Survival & Growth	1004.0	S	5	0.02	0.07	0.08	0.12	0.13
Inland Silverside Larval Survival & Growth	1006.0	G	16	0.03	0.16	0.26	0.37	0.50
Inland Silverside Larval Survival & Growth	1006.0	S	16	0.05	0.16	0.28	0.35	0.49
Mysid (Ab) Survival, Growth, & Fecundity	1007.0	R	4	0.06	0.17	0.30	0.37	0.43
Mysid (Ab) Survival, Growth, & Fecundity	1007.0	G	10	0.15	0.19	0.22	0.27	0.31
Mysid (Ab) Survival, Growth, & Fecundity	1007.0	S	10	0.12	0.16	0.26	0.27	0.28
Red Macroalga *(Champia parvula)* Reprod	1009.0	R	2	0.35	0.35	0.36	0.38	0.38
West Coast Methods								
Topsmelt Larval Survival & Growth	1010.0	G	1	0.25	0.25	0.25	0.25	0.25
Topsmelt Larval Survival & Growth	1010.0	S	1	0.17	0.17	0.17	0.17	0.17
Pacific Oyster Embryo-Larval Survival & Dev.	1012.0	D	1	0.21	0.21	0.21	0.21	0.21
Mussel Embryo-Larval Survival & Dev.	1013.0	D	3	0.25	0.25	0.35	0.35	0.35
Red Abalone Larval Development	1014.0	D	10	0.13	0.16	0.21	0.28	0.33
Sea Urchin Fertilization	1016.0	F	12	0.24	0.30	0.35	0.52	0.61
Sand Dollar Fertilization	1017.0	F	7	0.28	0.33	0.34	0.50	0.79
Giant Kelp Germination & Germ-Tube Length	1018.0	G_e	11	0.18	0.20	0.30	0.37	0.40
Giant Kelp Germination & Germ-Tube Length	1018.0	L	11	0.17	0.18	0.25	0.32	0.32
Acute								
Fathead Minnow Larval Survival	2000.0	S	21	0.08	0.10	0.16	0.19	0.33
Ceriodaphnia (Cd) Survival	2002.0	S	23	0.06	0.11	0.19	0.29	0.34
Sheepshead Minnow Survival	2004.0	S	5	0.11	0.12	0.14	0.21	0.37
Inland Silverside Larval Survival	2006.0	S	5	0.07	0.15	0.16	0.21	0.44
Mysid (Ab) Survival	2007.0	S	3	0.17	0.17	0.25	0.26	0.26
Mysid (Hc) Survival	2011.0	S	2	0.27	0.27	0.30	0.34	0.34
Rainbow Trout Survival	2019.0	S	1	0.23	0.23	0.23	0.23	0.23
Daphnia (Dm) Survival	2021.0	S	5	0.05	0.07	0.22	0.24	0.46
Daphnia (Dp) Survival	2022.0	S	6	0.15	0.19	0.21	0.27	0.48

[a] EC50 is a more general term than LC50 and may be used to represent an LC50 endpoint (such as survival).
[b] Cd = *Ceriodaphnia dubia*, Ab = *Americamysis (Mysidopsis) bahia*, Hc = *Holmesimysis costata*, Dm = *Daphnia magna*, Dp = *Daphnia pulex*
[c] See footnote b on Table B-1.
[d] D = development, F = fertilization, G = growth, G_e = Germination, L = length, R = reproduction or fecundity, S = survival
[e] Genus and species recently changed to *Raphidocelis subcapitata*.

Table B-3. Percentiles of the Within-Laboratory Values of CV for NOEC

Test Method[a]	Test Method No.[b]	End-point[c]	No. of Labs	CV				
				P10	P25	P50	P75	P90
Chronic, Promulgated								
Fathead Minnow Larval Survival & Growth	1000.0	G	19	0	0.22	0.37	0.53	0.65
Fathead Minnow Larval Survival & Growth	1000.0	S	19	0.13	0.26	0.39	0.48	0.59
Ceriodaphnia (Cd) Survival & Reproduction	1002.0	R	33	0.20	0.25	0.33	0.49	0.60
Ceriodaphnia (Cd) Survival & Reproduction	1002.0	S	33	0.09	0.21	0.30	0.43	0.55
Green Alga (*Selenastrum*)[d] Growth	1003.0	G	9	0.30	0.40	0.46	0.56	0.82
Sheepshead Minnow Larval Survival & Growth	1004.0	G	5	0.20	0.34	0.40	0.44	0.52
Sheepshead Minnow Larval Survival & Growth	1004.0	S	5	0	0.14	0.18	0.24	0.38
Inland Silverside Larval Survival & Growth	1006.0	G	16	0.14	0.31	0.46	0.57	0.63
Inland Silverside Larval Survival & Growth	1006.0	S	16	0.19	0.30	0.42	0.55	0.66
Mysid (Ab) Survival, Growth, & Fecundity	1007.0	R	4	0	0.17	0.36	0.40	0.41
Mysid (Ab) Survival, Growth, & Fecundity	1007.0	G	10	0.22	0.35	0.39	0.43	0.67
Mysid (Ab) Survival, Growth, & Fecundity	1007.0	S	10	0.13	0.28	0.33	0.38	0.41
Red Macroalga *(Champia parvula)* Reprod	1009.0	R	2	0.85	0.85	1.00	1.16	1.16
West Coast Methods								
Topsmelt Larval Survival & Growth	1010.0	G	1	0.31	0.31	0.31	0.31	0.31
Topsmelt Larval Survival & Growth	1010.0	S	1	0.42	0.42	0.42	0.42	0.42
Pacific Oyster Embryo-Larval Survival & Dev.	1012.0	D	1	0.45	0.45	0.45	0.45	0.45
Mussel Embryo-Larval Survival & Dev.	1013.0	D	3	0	0	0.39	0.43	0.43
Red Abalone Larval Development	1014.0	D	10	0.24	0.25	0.29	0.31	0.38
Sea Urchin Fertilization	1016.0	F	12	0.31	0.40	0.50	0.69	0.76
Sand Dollar Fertilization	1017.0	F	7	0.40	0.41	0.53	0.75	0.81
Giant Kelp Germination & Germ-Tube Length	1018.0	G_e	11	0.36	0.40	0.54	0.65	0.81
Giant Kelp Germination & Germ-Tube Length	1018.0	L	11	0.39	0.48	0.59	0.68	0.76
Acute								
Fathead Minnow Larval Survival	2000.0	S	21	0.15	0.18	0.22	0.34	0.61
Ceriodaphnia (Cd) Survival	2002.0	S	23	0.07	0.18	0.35	0.41	0.57
Sheepshead Minnow Survival	2004.0	S	3	0.0	0	0.31	0.33	0.33
Inland Silverside Larval Survival	2006.0	S	5	0.0	0	0.33	0.35	0.72
Mysid (Ab) Survival	2007.0	S	3	0.29	0.29	0.38	0.43	0.43
Mysid (Hc) Survival	2011.0	S	2	0.21	0.21	0.26	0.31	0.31
Rainbow Trout Survival	2019.0	S	1	0.35	0.35	0.35	0.35	0.35
Daphnia (Dm) Survival	2021.0	S	5	0	0.09	0.36	0.47	0.83
Daphnia (Dp) Survival	2022.0	S	6	0.20	0.21	0.38	0.61	0.67

[a] Cd = *Ceriodaphnia dubia*, Ab = *Americamysis (Mysidopsis) bahia*, Hc = *Holmesimysis costata*, Dm = *Daphnia magna*, Dp = *Daphnia pulex*

[b] See footnote b on Table B-1.

[c] D = development, F = fertilization, G = growth, G_e = germination, L = length, R = reproduction or fecundity, S = survival

[d] Genus and species recently changed to *Raphidocelis subcapitata*.

Table B-4. Variation Across Laboratories in the Within-Laboratory Value of P90:P10 for EC25

Test Method[a]	Test Method No.[b]	End-point[c]	No. of Labs	CV				
				P10	P25	P50	P75	P90
Chronic, Promulgated								
Fathead Minnow Larval Survival & Growth	1000.0	G	19	1.3	1.7	2.1	3.6	4.1
Fathead Minnow Larval Survival & Growth	1000.0	S	16	1.0	1.3	1.7	2.3	3.5
Ceriodaphnia (Cd) Survival & Reproduction	1002.0	R	33	1.2	1.4	2.2	3.6	6.3
Ceriodaphnia (Cd) Survival & Reproduction	1002.0	S	25	1.1	1.3	1.6	2.6	4.8
Green Alga (*Selenastrum*)[d] Growth	1003.0	G	6	1.7	1.8	2.0	2.5	3.8
Sheepshead Minnow Larval Survival & Growth	1004.0	G	5	1.1	1.1	1.4	1.4	1.4
Sheepshead Minnow Larval Survival & Growth	1004.0	S	2	1.3	1.3	1.3	1.3	1.3
Inland Silverside Larval Survival & Growth	1006.0	G	16	1.1	1.5	2.0	2.5	4.2
Inland Silverside Larval Survival & Growth	1006.0	S	13	1.3	1.7	2.2	3.2	4.3
Mysid (Ab) Survival, Growth, & Fecundity	1007.0	R	4	1.7	2.1	2.4	2.7	2.9
Mysid (Ab) Survival, Growth, & Fecundity	1007.0	G	10	1.4	1.8	2.2	2.6	3.0
Mysid (Ab) Survival, Growth, & Fecundity	1007.0	S	7	1.5	1.5	1.8	2.4	2.5
Red Macroalga *(Champia parvula)* Reprod	1009.0	R	2	6.7	6.7	10.2	13.7	13.7
West Coast								
Topsmelt Larval Survival & Growth	1010.0	G	1	1.7	1.7	1.7	1.7	1.7
Topsmelt Larval Survival & Growth	1010.0	S	1	1.8	1.8	1.8	1.8	1.8
Pacific Oyster Embryo-Larval Survival & Dev.	1012.0	D	1	2.0	2.0	2.0	2.0	2.0
Mussel Embryo-Larval Survival & Dev.	1013.0	D	3	1.4	1.4	2.2	4.0	4.0
Red Abalone Larval Development	1014.0	D	10	1.3	1.5	2.0	2.9	3.1
Sea Urchin Fertilization	1016.0	F	12	1.6	1.8	3.0	6.7	14.9
Sand Dollar Fertilization	1017.0	F	7	2.4	3.1	3.8	3.9	6.1
Giant Kelp Germination & Germ-Tube Length	1018.0	G_e	11	2.1	2.1	3.3	4.1	5.9
Giant Kelp Germination & Germ-Tube Length	1018.0	L	11	1.7	1.8	2.3	2.5	3.1
Acute								
Fathead Minnow Larval Survival	2000.0	S	7	1.1	1.2	1.4	1.5	3.7
Ceriodaphnia (Cd) Survival	2002.0	S	8	1.1	1.1	1.3	1.4	1.6
Sheepshead Minnow Survival	2004.0	S	3	1.2	1.2	1.3	5.2	5.2
Inland Silverside Larval Survival	2006.0	S	4	1.0	1.3	1.7	2.6	3.4
Mysid (Ab) Survival	2007.0	S	1	1.7	1.7	1.7	1.7	1.7
Mysid (Hc) Survival	2011.0	S	1	1.5	1.5	1.5	1.5	1.5
Rainbow Trout Survival	2019.0	S	1	1.2	1.2	1.2	1.2	1.2
Daphnia (Dm) Survival	2021.0	S	1	1.9	1.9	1.9	1.9	1.9
Daphnia (Dp) Survival	2022.0	S	3	1.1	1.1	2.5	2.8	2.8

[a] Cd = *Ceriodaphnia dubia*, Ab = *Americamysis (Mysidopsis) bahia*, Hc = *Holmesimysis costata*, Dm = *Daphnia magna*, Dp = *Daphnia pulex*

[b] See footnote b on Table B-1.

[c] D = development, F = fertilization, G = growth, G_e = germination, L = length, R = reproduction or fecundity, S = survival

[d] Genus and species recently changed to *Raphidocelis subcapitata*.

Table B-5. Variation Across Laboratories in the Within-Laboratory Value of P90:P10 for EC50[a]

Test Method[b]	Test Method No.[c]	End-point[d]	No. of Labs	CV				
				P10	P25	P50	P75	P90
Chronic, Promulgated								
Fathead Minnow Larval Survival & Growth	1000.0	G	19	1.3	1.5	1.8	2.4	3.3
Fathead Minnow Larval Survival & Growth	1000.0	S	19	1.4	1.5	1.8	2.3	3.0
Ceriodaphnia (Cd) Survival & Reproduction	1002.0	R	33	1.2	1.3	1.7	2.3	3.7
Ceriodaphnia (Cd) Survival & Reproduction	1002.0	S	33	1.1	1.3	1.5	2.2	3.5
Green Alga (*Selenastrum*)[e] Growth	1003.0	G	9	1.2	1.5	1.7	2.4	9.4
Sheepshead Minnow Larval Survival & Growth	1004.0	G	5	1.0	1.1	1.1	1.2	1.3
Sheepshead Minnow Larval Survival & Growth	1004.0	S	5	1.0	1.1	1.1	1.2	1.3
Inland Silverside Larval Survival & Growth	1006.0	G	16	1.1	1.5	1.8	2.7	3.5
Inland Silverside Larval Survival & Growth	1006.0	S	16	1.2	1.5	1.9	2.8	2.9
Mysid (Ab) Survival, Growth, & Fecundity	1007.0	R	4	1.2	1.5	1.9	2.4	2.9
Mysid (Ab) Survival, Growth, & Fecundity	1007.0	G	10	1.4	1.5	1.8	2.2	2.4
Mysid (Ab) Survival, Growth, & Fecundity	1007.0	S	10	1.4	1.6	1.9	2.0	2.3
Red Macroalga *(Champia parvula)* Reprod	1009.0	R	2	2.3	2.3	4.9	7.6	7.6
West Coast								
Topsmelt Larval Survival & Growth	1010.0	G	1	1.7	1.7	1.7	1.7	1.7
Topsmelt Larval Survival & Growth	1010.0	S	1	1.5	1.5	1.5	1.5	1.5
Pacific Oyster Embryo-Larval Survival & Dev.	1012.0	D	1	2.0	2.0	2.0	2.0	2.0
Mussel Embryo-Larval Survival & Dev.	1013.0	D	3	2.0	2.0	2.0	2.8	2.8
Red Abalone Larval Development	1014.0	D	10	1.4	1.4	1.8	2.4	2.6
Sea Urchin Fertilization	1016.0	F	12	1.8	2.0	2.9	4.2	6.5
Sand Dollar Fertilization	1017.0	F	7	2.4	2.6	2.8	4.4	6.0
Giant Kelp Germination & Germ-Tube Length	1018.0	G_e	11	1.7	1.8	2.1	3.3	3.6
Giant Kelp Germination & Germ-Tube Length	1018.0	L	11	1.6	1.6	1.8	2.5	2.7
Acute								
Fathead Minnow Larval Survival	2000.0	S	21	1.2	1.3	1.5	1.7	2.6
Ceriodaphnia (Cd) Survival	2002.0	S	23	1.1	1.2	1.7	2.0	2.4
Sheepshead Minnow Survival	2004.0	S	5	1.1	1.2	1.4	1.7	2.8
Inland Silverside Larval Survival	2006.0	S	5	1.2	1.4	1.6	1.7	2.7
Mysid (Ab) Survival	2007.0	S	3	1.7	1.7	2.1	2.1	2.1
Mysid (Hc) Survival	2011.0	S	2	1.8	1.8	2.5	3.1	3.1
Rainbow Trout Survival	2019.0	S	1	1.8	1.8	1.8	1.8	1.8
Daphnia (Dm) Survival	2021.0	S	5	1.2	1.2	1.8	2.2	4.1
Daphnia (Dp) Survival	2022.0	S	6	1.4	1.5	1.9	2.1	2.2

[a] EC50 is a more general term than LC50 and may be used to represent an LC50 endpoint (such as survival).
[b] Cd = *Ceriodaphnia dubia*, Ab = *Americamysis (Mysidopsis) bahia*, Hc = *Holmesimysis costata*, Dm = *Daphnia magna*, Dp = *Daphnia pulex*
[c] See footnote b on Table B-1.
[d] D = development, F = fertilization, G = growth, G_e = germination, L = length, R = reproduction or fecundity, S = survival
[e] Genus and species recently changed to *Raphidocelis subcapitata*.

Table B-6. Variation Across Laboratories in the Within-Laboratory Value of P90:P10 for NOEC

Test Method[a]	Test Method No.[b]	End-point[c]	No. of Labs	CV				
				P10	P25	P50	P75	P90
Chronic, Promulgated								
Fathead Minnow Larval Survival & Growth	1000.0	G	19	1.0	1.5	2.0	4.2	8.0
Fathead Minnow Larval Survival & Growth	1000.0	S	19	1.0	1.7	2.0	4.0	5.0
Ceriodaphnia (Cd) Survival & Reproduction	1002.0	R	33	1.3	1.9	2.2	4.0	4.0
Ceriodaphnia (Cd) Survival & Reproduction	1002.0	S	33	1.0	1.5	2.0	3.0	5.3
Green Alga (*Selenastrum*)[d] Growth	1003.0	G	9	1.8	2.0	2.7	4.0	10.0
Sheepshead Minnow Larval Survival & Growth	1004.0	G	5	1.3	2.0	2.0	4.0	4.0
Sheepshead Minnow Larval Survival & Growth	1004.0	S	5	1.0	1.0	1.3	2.0	2.0
Inland Silverside Larval Survival & Growth	1006.0	G	16	1.3	2.0	4.0	4.2	7.8
Inland Silverside Larval Survival & Growth	1006.0	S	16	1.8	2.0	2.9	4.0	4.1
Mysid (Ab) Survival, Growth, & Fecundity	1007.0	R	4	1.0	1.5	2.0	2.0	2.0
Mysid (Ab) Survival, Growth, & Fecundity	1007.0	G	10	1.9	2.0	2.0	4.0	7.6
Mysid (Ab) Survival, Growth, & Fecundity	1007.0	S	10	1.4	2.0	2.0	2.0	3.4
Red Macroalga *(Champia parvula)* Reprod	1009.0	R	2	5.6	5.6	12.8	20.0	20.0
West Coast								
Topsmelt Larval Survival & Growth	1010.0	G	1	1.8	1.8	1.8	1.8	1.8
Topsmelt Larval Survival & Growth	1010.0	S	1	3.2	3.2	3.2	3.2	3.2
Pacific Oyster Embryo-Larval Survival & Dev.	1012.0	D	1	4.0	4.0	4.0	4.0	4.0
Mussel Embryo-Larval Survival & Dev.	1013.0	D	3	1.0	1.0	3.2	4.0	4.0
Red Abalone Larval Development	1014.0	D	10	1.2	1.8	1.8	1.8	3.2
Sea Urchin Fertilization	1016.0	F	12	1.8	2.0	4.0	6.9	9.4
Sand Dollar Fertilization	1017.0	F	7	2.1	3.1	4.0	6.0	17.8
Giant Kelp Germination & Germ-Tube Length	1018.0	G_e	11	1.8	2.3	3.2	5.7	5.7
Giant Kelp Germination & Germ-Tube Length	1018.0	L	11	3.1	3.1	5.6	5.7	10.0
Acute								
Fathead Minnow Larval Survival	2000.0	S	21	1.3	1.5	1.6	2.0	4.0
Ceriodaphnia (Cd) Survival	2002.0	S	23	1.0	1.3	2.0	3.3	5.0
Sheepshead Minnow Survival	2004.0	S	3	1.0	1.0	2.0	2.0	2.0
Inland Silverside Larval Survival	2006.0	S	5	1.0	1.0	1.8	2.0	4.0
Mysid (Ab) Survival	2007.0	S	3	2.7	2.7	3.2	5.0	5.0
Mysid (Hc) Survival	2011.0	S	2	1.8	1.8	1.9	2.1	2.1
Rainbow Trout Survival	2019.0	S	1	2.0	2.0	2.0	2.0	2.0
Daphnia (Dm) Survival	2021.0	S	5	1.0	1.3	2.0	4.0	6.1
Daphnia (Dp) Survival	2022.0	S	6	1.3	1.7	2.0	2.0	10.0

[a] Cd = *Ceriodaphnia dubia*, Ab = *Americamysis (Mysidopsis) bahia*, Hc = *Holmesimysis costata*, Dm = *Daphnia magna*, Dp = *Daphnia pulex*

[b] See footnote b on Table B-1.

[c] D = development, F = fertilization, G = growth, G_e = germination, L = length, R = reproduction or fecundity, S = survival

[d] Genus and species recently changed to *Raphidocelis subcapitata*.

Estimation formulas were:

Expected mean square for error (within-laboratory): σ_w^2

Expected mean square between-laboratories: $\sigma_w^2 + U\,\sigma_b^2$

$U = [\sum n_i - (\sum n_i^2 / \sum n_i)] / (L-1)$

L is the number of laboratories and n_i the number of tests within the ith laboratory (i = 1, ... L).

B.3 Variability of Endpoint Measurements

Dunnett's critical value, needed for the minimum significant difference (MSD), was computed using the SAS function "PROBMC," for a one-sided test at the 0.95 level ($\alpha = 0.05$). Note that Dunnett's test can be applied when the number of replicates differs among treatments (Dunnett 1964), and that the SAS function "PROBMC" can calculate an appropriate critical value for the case of unequal replication.

The MSD was calculated for sublethal endpoints using untransformed values of "growth" (larval biomass) and "reproduction" (number of offspring in the *Ceriodaphnia* test, or cells per mL in the *Selenastrum* test), and for lethal endpoints using the arc sine transform (arc sine ($\sqrt{p}$)) of the proportion surviving. The CV was calculated for all endpoints using the untransformed mean control response.

Tables B-7 and B-8 show percentiles of CV and of the percent minimum significant difference (PMSD), which is [100×MSD/(control mean)]. These are the sample percentiles for all tests in the data set (see row "No. of tests"). Data for all laboratories and toxicants for a given method and endpoint were combined.

Methods in Tables B-1 through B-3 that are represented by fewer than three laboratories or fewer than 20 tests are not shown in Tables B-7 and B-8, because characterizing method variability using so few tests and laboratories would be inadvisable.[1]

B.4 Test Power to Detect Toxic Effects

Power can be characterized only by repeated testing. It is an attribute, not of a single test, but of a sequence of many tests conducted under similar conditions and the same test design. Therefore, the sample averages for each laboratory's data set are used in this analysis to characterize each laboratory. The key parameters required were the (a) mean endpoint response in the control (growth, reproduction, survival) and (b) the mean value of the error mean square (EMS) for tests.

Power is reported in this section for single two-sample, one-sided t-tests at $1-\alpha = 0.95$, and for a set of k such tests (comparing k treatments to a control) at level $1 - \alpha/k = 1 - 0.05/k$. Some permitting authorities may require a comparison between control and the receiving water concentration, which requires a two-sample, one-sided test. Others may require the multiple comparisons procedure described in the EPA WET methods (Dunnett's or Steel's tests, one-sided, with $\alpha = 0.05$). The power of Dunnett's procedure (using $\alpha = 0.05$ as recommended in EPA effluent test methods) will fall between the power of the one-sided, two-sample t-test with $\alpha = 0.05$ and that with $\alpha = 0.05/k$, when k toxicant concentrations are compared to a control. The power of Steel's procedure will be related to and should usually increase with the power of Dunnett's procedure and the t-tests, so the following tables will also provide an inexact guide to power achieved by the nonparametric test.

Tables B-9 through B-13 illustrate the ability of the sublethal endpoint for the chronic toxicity promulgated methods to detect toxic effects using a two-sample, one-sided hypothesis test (t-test) at two

[1] Tables B-7 through B-18 begin on page B-14.

significance levels, $\alpha = 0.05$ and $\alpha = 0.01$. Data for Method 1009.0 (red macroalga) are not presented, because characterizing method performance using data from only two laboratories and 23 tests is inadvisable.

Table B-14 shows the power and PMSD to be expected for various combinations of (1) number of replicates; (2) k, number of treatments compared with a control; and (3) value of the square root of the error mean square (rEMS) divided by the control mean, when the t-test can be used.

Table B-15 shows the value of PMSD for various combinations of number of replicates, number of treatments compared with a control, and rEMS/(Control Mean). (For definitions and explanations of the terms used here, see Chapters 2 and 3.) This table can be used as a guide to planning the number of replicates needed to achieve a given PMSD. The number of replicates needed can be determined by calculating MSD using the average EMS for a series of tests (at least 20 tests are recommended) and experimenting with various choices of number of replicates (the same number for each concentration and test). This approach is recommended because it uses a sample of test EMSs specific to a particular laboratory. This approach also reveals variation by test, showing how frequently PMSD exceeds the upper bound in Table 3-6 if the number of replicates is increased.

The number of replicates needed to achieve a given value of PMSD will depend on the variability among replicates (rEMS). Table B-16 shows percentiles of the rEMS divided by the control mean, for each promulgated method for chronic toxicity, pooling all tests available in the WET variability data set. The data for Method 1009.0 (red macroalga, *Champia parvula*) are based on only two laboratories and 23 tests and therefore cannot be considered representative.

Table B-15 can be used to infer the number of replicates needed to make the MSD a certain percentage of the control mean (25 percent and 33 percent are used here) for any particular value of rEMS. Table B-17 shows the number of replicates needed to do the same for the 90th and 85th percentiles of rEMS found in Table B-16, in which three or four treatments are compared to a control. These percentiles represent rather extreme examples of imprecision. The precision achieved in most tests and by most laboratories is within the bounds set by these percentiles. The exact number of replicates was not determined beyond “>15” (*Ceriodaphnia* chronic test).

Table B-17 agrees with conclusions drawn from Table 5-1: For most methods, most laboratories can detect a 33 percent effect most of the time, but many laboratories are unable to detect a 25 percent difference between treatment and control in many tests.

B.5 NOEC for Chronic Toxicity Test Methods (Calculated Using the Most Sensitive Endpoint)

NOEC for chronic toxicity methods is calculated using the most sensitive endpoint in each test (meaning the smallest NOEC among those for the two or three endpoints). Table B-18 shows percentiles of within-laboratory CVs in a format like that for Tables B-1 through B-6, and similar calculations were used.

Table B-7a. Percentiles of Control CV for Sublethal Endpoints of Chronic WET Tests, Using Data Pooled Across All Laboratories and Toxicants[a]

	Test Method					
	1000.0 Fathead Minnow	**1002.0 *Cerio-daphnia***	**1003.0 Green Alga**	**1004.0 Sheepshead Minnow**	**1006.0 Inland Silverside**	**1007.0 Mysid (*A. bahia*)**
No. of tests	205	393	85	57	193	130
No. of labs	19	33	9	5	16	10
Endpoint[b]	G	R	G	G	G	G
Percentile	**Control CV**					
5%	0.03	0.08	0.03	0.03	0.03	0.07
10%	0.04	0.09	0.03	0.03	0.04	0.09
15%	0.05	0.10	0.04	0.04	0.05	0.09
20%	0.06	0.11	0.05	0.04	0.06	0.10
25%	0.06	0.12	0.05	0.04	0.06	0.11
50%	0.10	0.20	0.08	0.07	0.10	0.15
75%	0.14	0.33	0.12	0.09	0.14	0.20
80%	0.16	0.36	0.14	0.09	0.14	0.22
85%	0.17	0.39	0.16	0.10	0.16	0.25
90%	0.20	0.42	0.17	0.13	0.18	0.28
95%	0.23	0.52	0.18	0.17	0.23	0.37

[a] Methods in Table B-1 having fewer than three laboratories or fewer than 20 tests are not shown here because so few results may not be representative of method performance.
[b] G = growth, R = reproduction

Table B-7b. Percentiles of Control CV for Endpoints of Chronic WET Tests, Using Data Pooled Across All Laboratories and Toxicants (West Coast Methods)[a]

	Test Method					
	1013.0 Mussel Embryo-Larval Survival & Development	**1014.0 Red Abalone Larval Development**	**1016.0 Sea Urchin Fertilization**	**1017.0 Sand Dollar Fertilization**	**1018.0 Giant Kelp Germination & Germ-Tube Length**	**1018.0 Giant Kelp Germination & Germ-Tube Length**
No. of tests	34	137	159	67	159	159
No. of labs	3	10	11	7	11	11
Endpoint[b]	S	L	F	F	G_e	L
Percentile	**Control CV**					
5%	0.01	0.01	0.01	0.01	0.01	0.02
10%	0.01	0.01	0.01	0.01	0.02	0.03
15%	0.01	0.01	0.01	0.02	0.02	0.03
20%	0.01	0.01	0.02	0.02	0.02	0.04
25%	0.01	0.02	0.02	0.03	0.02	0.05
50%	0.02	0.03	0.04	0.04	0.04	0.07
75%	0.04	0.05	0.07	0.06	0.06	0.09
80%	0.05	0.05	0.08	0.07	0.06	0.11
85%	0.06	0.05	0.10	0.08	0.07	0.11
90%	0.07	0.06	0.12	0.08	0.08	0.12
95%	0.07	0.08	0.18	0.12	0.10	0.14

[a] Methods in Table B-1 having fewer than three laboratories or fewer than 20 tests are not shown here because so few results may not be representative of method performance.

[b] G_e = germination, F = fertilization, L = length, S = survival

Table B-7c. Percentiles of Control CV for Survival Endpoint of Acute WET Tests, Using Data Pooled Across All Laboratories and Toxicants

	Test Method							
	2000.0 Fathead Minnow	**2002.0 *Cerio-daphnia***	**2004.0 Sheepshead Minnow**	**2006.0 Inland Silverside**	**2007.0 Mysid (*A. bahia*)**	**2011.0 Mysid (*H. costata*)**	**2021.0 Daphnia (*D. magna*)**	**2022.0 Daphnia (*D. pulex*)**
No. of tests	217	241	65	48	32	14	48	57
No. of labs	20	23	5	5	3	2	5	6
Percentile	**Control CV**							
5%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
15%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
20%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
25%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
50%	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00
75%	0.00	0.00	0.00	0.07	0.07	0.07	0.00	0.00
80%	0.00	0.00	0.00	0.07	0.07	0.07	0.00	0.00
85%	0.07	0.00	0.00	0.07	0.07	0.07	0.10	0.07
90%	0.07	0.11	0.00	0.08	0.08	0.07	0.11	0.11
95%	0.11	0.11	0.06	0.10	0.11	0.10	0.12	0.11

Table B-8a. Percentiles of PMSD for Sublethal Endpoints of Chronic WET Tests, Using Data Pooled Across All Laboratories and Toxicants[a,b]

	Test Method					
	1000.0 Fathead Minnow	**1002.0 *Cerio-daphnia***	**1003.0 Green Alga**	**1004.0 Sheepshead Minnow**	**1006.0 Inland Silverside**	**1007.0 Mysid (*A. bahia*)**
No. of tests	205	393	85	57	193	130
No. of labs	19	33	9	5	16	10
Endpoint[c]	G	R	G	G	G	G
Percentile	**PMSD**					
5%	6.8	10	8.2	5.5	10	10
10%	9	11	9.3	6.3	12	12
15%	11	13	10	6.8	12	14
20%	13	15	11	7.9	13	16
25%	14	16	11	8.4	14	16
50%	20	23	14	13	18	20
75%	25	30	19	18	25	25
80%	28	31	20	19	27	26
85%	29	33	21	21	31	28
90%	35	37	23	23	35	32
95%	44	43	27	26	41	34

[a] PMSD = Percent MSD [100×MSD/(Control Mean)]

[b] Methods in Table B-1 having fewer than three laboratories or fewer than 20 tests are not shown here because so few results may not be representative of method performance.

[c] G = growth, R = reproduction

Table B-8b. Percentiles of PMSD for Endpoints of Chronic WET Tests, Using Data Pooled Across All Laboratories and Toxicants (West Coast Methods)[a, b]

	Test Method					
	1013.0 Mussel Embryo-Larval Survival & Development	**1014.0 Red Abalone Larval Development**	**1016.0 Sea Urchin Fertilization**	**1017.0 Sand Dollar Fertilization**	**1018.0 Giant Kelp Germination & Germ-Tube Length**	**1018.0 Giant Kelp Germination & Germ-Tube Length**
No. of tests	34	137	159	67	159	159
No. of labs	3	10	11	7	11	11
Endpoint[c]	S	L	F	F	G_e	L
Percentile	**PMSD**					
5%	3.9	3.1	3.7	6.5	5.7	6.6
10%	5.5	3.8	5.1	6.9	6.5	7.9
15%	6.2	4.6	6.5	8.0	7.0	8.8
20%	7.1	5.0	7.3	8.5	7.4	9.2
25%	8.5	5.3	8.1	9.0	8.2	9.6
50%	11	7.9	12	12	10	11
75%	16	12	18	17	14	15
80%	19	13	19	19	15	16
85%	20	15	21	21	17	18
90%	42	16	25	26	18	21
95%	49	20	29	30	20	24

[a] PMSD = Percent MSD [100×MSD/(Control Mean)]

[b] Methods in Table B-1 having fewer than three laboratories or fewer than 20 tests are not shown here because so few results may not be representative of method performance.

[c] G_e = germination, F = fertilization, L = length, S = survival

Table B-8c. Percentiles of PMSD for Survival Endpoint of Acute WET Tests, Using Data Pooled Across All Laboratories and Toxicants[a]

	Test Method							
	2000.0 Fathead Minnow	**2002.0 *Cerio-daphnia***	**2004.0 Sheepshead Minnow**	**2006.0 Inland Silverside**	**2007.0 Mysid (*A. bahia*)**	**2011.0 Mysid (*H. costata*)**	**2021.0 *Daphnia* (*D. magna*)**	**2022.0 *Daphnia* (*D. pulex*)**
No. of tests	217	241	65	48	32	14	48	57
No. of labs	20	23	5	5	3	2	5	6
Percentile	**PMSD**							
5%	0	4.6	0	4.5	3.9	14	4.5	4.3
10%	4.2	5.0	0	7.0	5.1	18	5.3	5.8
15%	5.0	5.6	0	8.9	6.9	21	6.4	6.8
20%	6.6	5.9	0	10	8.4	22	6.9	7.5
25%	7.4	7.1	6.1	12	8.9	23	8.4	8.3
50%	13	11	16	20	15	30	13	14
75%	21	16	32	26	23	38	19	20
80%	23	18	36	29	24	40	20	21
85%	26	19	49	36	24	42	20	22
90%	30	21	55	41	26	47	23	23
95%	51	25	67	46	33	58	27	27

[a] PMSD = Percent MSD [100×MSD/(Control Mean)]

Table B-9. Test Method 1000.0, Fathead Minnow Chronic Toxicity Test, Growth Endpoint: Power and Effect Size Achieved

Lab	No. of Tests	No. of Reps Per Test	Average Control Mean	Average Control Std Dev	Square Root of Variance of Control Mean	Square Root of Average EMS	Average PMSD	Power of Hypothesis Test (2-sample, 1-sided t-test)							
								α = 0.05				α = 0.01			
								N (Reps)	Delta	100×Delta/ Mean	Power	N (Reps)	Delta	100×Delta/ Mean	Power
1	9	4	0.38	0.040	0.081	0.043	19	4	0.09	23	0.85	6	0.12	33	0.48
2	13	4	0.32	0.013	0.028	0.013	6	2	0.03	8	1.00	3	0.04	12	1.00
3	11	3	0.55	0.066	0.117	0.069	25	5	0.17	31	0.62	7	0.26	48	0.13
4	18	4	0.45	0.051	0.107	0.066	21	6	0.13	30	0.67	9	0.19	42	0.25
5	8	4	0.41	0.041	0.115	0.064	26	6	0.13	31	0.63	10	0.18	44	0.21
6	10	3	0.60	0.081	0.189	0.082	28	5	0.20	34	0.54	8	0.31	52	0.10
7	7	4	0.39	0.063	0.064	0.073	31	9	0.15	38	0.47	14	0.21	54	0.12
8	20	4	0.55	0.053	0.109	0.065	17	4	0.13	24	0.82	7	0.19	34	0.43
9	5	4	0.46	0.054	0.217	0.044	17	3	0.09	20	0.93	5	0.13	28	0.68
10	11	3 to 4	0.34	0.047	0.042	0.043	20	5	0.11	32	0.60	7	0.16	49	0.13
11	11	3 to 4	0.54	0.074	0.101	0.084	21	6	0.21	39	0.44	10	0.32	59	0.08
12	11	4	0.59	0.083	0.142	0.076	20	5	0.15	26	0.77	7	0.22	37	0.35
13	10	4	0.42	0.046	0.080	0.044	16	4	0.09	21	0.90	6	0.13	30	0.58
14	11	3 to 4	0.39	0.055	0.063	0.063	26	7	0.16	41	0.40	11	0.24	63	0.07
15	8	3 to 4	0.48	0.048	0.108	0.051	18	4	0.13	27	0.76	6	0.19	41	0.22
16	11	3 to 4	0.35	0.041	0.056	0.052	23	6	0.13	37	0.48	9	0.20	57	0.08
17	6	3	0.40	0.050	0.055	0.098	31	13	0.25	62	0.21	22	0.38	95	0.03
18	20	4	0.40	0.061	0.095	0.064	27	6	0.13	32	0.60	10	0.18	46	0.19
19	6	4	0.54	0.061	0.177	0.060	19	4	0.12	22	0.87	6	0.17	32	0.51

NOTE: Column "N (Reps)" shows the number of replicates needed to detect a 25 percent difference from control with power 0.8, given the observed averages for EMS and control mean. Column "Delta" gives the effect size of the endpoint in milligrams that can be detected with power 0.8, given the observed averages for EMS and control mean. Column "100×Delta/Mean" gives the effect size as a percent of the control mean. Column "Power" gives the power to detect a 25 percent difference from control, given the observed averages for EMS and control mean. PMSD = 100 × MSD / (Control Mean); EMS = error mean square.

Table B-10. Test Method 1002.0, *Ceriodaphnia* Chronic Toxicity Test, Reproduction Endpoint: Power and Effect Size Achieved

Lab	No. of Tests	No. of Reps Per Test	Average Control Mean	Average Control Std Dev	Square Root of Variance of Control Mean	Square Root of Average EMS	Average PMSD	Power of Hypothesis Test (2-sample, 1-sided t-test)							
								α = 0.05				α = 0.01			
								N (Reps)	Delta	100×Delta/ Mean	Power	N (Reps)	Delta	100×Delta/ Mean	Power
1	11	10	34	3.3	2.9	4.6	13	5	5.3	16	0.99	8	7.0	21	0.94
2	9	10	25	7.2	2.6	7.1	29	18	8.2	33	0.59	28	10.8	44	0.28
3	13	10	17	2.6	1.4	3.6	18	10	4.1	24	0.82	16	5.4	32	0.55
4	20	7 to 10	28	8.8	9.5	7.2	25	15	10.2	37	0.51	24	13.6	49	0.20
5	15	10 to 15	19	6.1	4.0	6.6	32	24	7.7	40	0.46	39	10.1	52	0.19
6	20	9 to 10	22	8.5	3.4	7.8	32	26	9.5	44	0.40	42	12.6	58	0.15
7	20	9 to 10	34	11.8	9.7	10.3	31	19	12.7	37	0.50	31	16.8	49	0.21
8	18	10	22	8.6	6.3	7.4	31	23	8.6	39	0.48	37	11.3	51	0.20
9	13	10	25	4.9	3.0	4.8	17	8	5.6	22	0.88	13	7.3	29	0.66
10	12	10	20	2.1	0.8	2.4	12	4	2.8	14	1.00	6	3.6	18	0.98
11	13	10	17	1.5	0.5	3.2	15	8	3.7	21	0.90	13	4.8	28	0.68
12	12	10	31	4.8	2.8	5.0	15	6	5.8	19	0.95	10	7.6	24	0.82
13	8	10	24	5.1	2.5	5.3	22	11	6.2	25	0.79	17	8.1	33	0.51
14	8	10	24	9.2	5.0	6.7	27	17	7.8	33	0.59	28	10.2	43	0.28
15	12	10	18	5.2	2.7	4.8	24	15	5.6	31	0.65	24	7.4	40	0.34
16	20	10	21	5.4	4.6	4.9	22	12	5.7	27	0.74	19	7.5	36	0.44
17	10	9 to 10	24	6.1	4.5	6.9	29	18	8.5	35	0.54	29	11.2	47	0.23
18	10	10	20	5.8	3.7	5.5	24	15	6.4	31	0.64	25	8.4	41	0.32
19	6	9 to 10	23	10.9	3.9	8.4	36	28	10.3	45	0.38	45	13.6	60	0.13
20	12	10	23	3.3	4.7	4.9	21	10	5.7	24	0.81	16	7.5	32	0.54
21	9	10	28	5.3	3.0	6.0	20	11	6.9	25	0.79	17	9.1	33	0.51
22	10	10	17	4.5	2.2	4.9	26	17	5.7	33	0.59	28	7.6	43	0.28
23	9	9 to 10	27	6.9	3.6	7.4	27	16	9.1	33	0.58	25	12.0	44	0.27
24	10	10	18	4.4	1.4	4.5	23	13	5.3	29	0.70	21	6.9	38	0.39
25	12	10	20	6.4	3.6	6.0	30	19	7.0	35	0.55	30	9.2	46	0.25
26	12	10	27	4.4	3.2	4.2	14	6	4.9	18	0.96	10	6.5	24	0.84
27	10	10	21	6.0	4.0	6.1	27	19	7.0	34	0.56	30	9.3	45	0.26
28	6	10	20	6.1	5.2	4.7	23	12	5.5	27	0.74	20	7.3	36	0.43
29	14	10	31	5.6	3.0	5.9	19	9	6.8	22	0.87	14	9.0	29	0.64
30	5	10	16	4.7	0.3	4.9	28	20	5.7	36	0.53	32	7.4	47	0.24
31	12	10	24	5.4	5.9	6.1	25	14	7.1	30	0.67	23	9.3	39	0.35
32	4	10	32	5.9	6.3	5.6	17	8	6.5	21	0.91	12	8.6	27	0.72
33	18	10	24	6.9	5.6	6.8	28	17	7.9	32	0.61	27	10.3	42	0.30

NOTE: See note at bottom of Table B-9.

Table B-11. Test Method 1004.0, Sheepshead Minnow Chronic Toxicity Test, Growth Endpoint: Power and Effect Size Achieved

Lab	No. of Tests	No. of Reps Per Test	Average Control Mean	Average Control Std Dev	Square Root of Variance of Control Mean	Square Root of Average EMS	Average PMSD	Power of Hypothesis Test (2-sample, 1-sided t-test)							
								α = 0.05				α = 0.01			
								N (Reps)	Delta	100×Delta/ Mean	Power	N (Reps)	Delta	100×Delta/ Mean	Power
1	12	4	0.88	0.040	0.11	0.037	6.6	2	0.08	8.6	1.00	3	0.11	12	1.00
2	11	4	0.68	0.051	0.11	0.071	16	4	0.14	21	0.90	6	0.20	30	0.59
3	16	4	0.65	0.088	0.091	0.084	20	5	0.17	26	0.77	7	0.24	37	0.34
4	14	4	1.00	0.074	0.13	0.076	12	3	0.15	15	0.98	4	0.22	22	0.91
5	4	4	0.86	0.048	0.12	0.066	11	3	0.13	16	0.98	4	0.19	22	0.90

NOTE: See note at bottom of Table B-9.

Table B-12. Test Method 1006.0, Inland Silverside Chronic Toxicity Test: Power and Effect Size Achieved

Lab	No. of Tests	No. of Reps Per Test	Average Control Mean	Average Control Std Dev	Square Root of Variance of Control Mean	Square Root of Average EMS	Average PMSD	Power of Hypothesis Test (2-sample, 1-sided t-test)							
								α = 0.05				α = 0.01			
								N (Reps)	Delta	100×Delta/ Mean	Power	N (Reps)	Delta	100×Delta/ Mean	Power
1	10	4	2.3	0.18	0.58	0.26	18	4	0.53	23	0.86	6	0.75	32	0.50
2	15	4	0.94	0.10	0.24	0.17	20	8	0.34	36	0.52	12	0.48	51	0.15
3	19	4	2.1	0.24	0.86	0.27	19	5	0.54	25	0.79	7	0.76	36	0.38
4	12	3	1.4	0.20	0.56	0.22	32	7	0.56	42	0.40	11	0.86	63	0.07
5	6	3 to 4	1.8	0.25	0.57	0.43	31	12	1.07	59	0.23	20	1.6	90	0.04
6	19	4	0.85	0.11	0.23	0.10	20	4	0.20	24	0.83	7	0.29	34	0.43
7	20	3 to 4	1.4	0.15	0.53	0.31	31	11	0.79	56	0.24	18	1.2	86	0.04
8	4	4 to 5	1.1	0.10	0.20	0.11	15	4	0.23	21	0.91	5	0.33	29	0.62
9	20	4	2.4	0.23	0.47	0.25	17	4	0.51	22	0.89	6	0.73	31	0.56
10	20	3 to 4	0.91	0.088	0.35	0.11	22	4	0.27	30	0.65	7	0.42	46	0.15
11	9	4	1.2	0.13	0.19	0.11	14	3	0.22	18	0.96	5	0.31	25	0.79
12	7	4	2.1	0.22	0.38	0.25	17	4	0.50	24	0.84	6	0.72	34	0.45
13	14	4	0.76	0.095	0.12	0.11	22	5	0.22	28	0.70	8	0.31	40	0.27
14	5	4	1.5	0.12	0.33	0.12	13	3	0.25	17	0.97	4	0.35	24	0.84
15	8	4	0.77	0.10	0.22	0.12	25	6	0.24	31	0.64	9	0.34	44	0.22
16	5	3	1.2	0.11	0.20	0.14	20	4	0.35	30	0.67	6	0.53	45	0.16

NOTE: See note at bottom of Table B-9.

Table B-13. Test Method 1007.0, Mysid Chronic Toxicity Test, Growth Endpoint: Power and Effect Size Achieved

Lab	No. of Tests	No. of Reps Per Test	Average Control Mean	Average Control Std Dev	Square Root of Variance of Control Mean	Square Root of Average EMS	Aver-age PMSD	Power of Hypothesis Test (2-sample, 1-sided t-test)							
								α = 0.05				α = 0.01			
								N (Reps)	Delta	100×Delta/ Mean	Power	N (Reps)	Delta	100×Delta/ Mean	Power
1	18	8	0.25	0.040	0.042	0.041	17	7	0.054	22	0.89	11	0.072	29	0.66
2	19	8	0.37	0.15	0.13	0.11	25	20	0.15	41	0.44	33	0.20	54	0.16
3	7	4	0.36	0.042	0.065	0.047	21	5	0.094	26	0.77	7	0.13	37	0.35
4	12	8	0.25	0.044	0.035	0.13	37	58	0.18	70	0.21	94	0.23	94	0.06
5	10	8	0.37	0.073	0.049	0.075	22	9	0.098	26	0.76	15	0.13	35	0.45
6	14	8	0.23	0.034	0.059	0.040	20	7	0.053	22	0.87	11	0.070	30	0.62
7	18	8	0.28	0.075	0.056	0.067	26	13	0.089	32	0.62	20	0.12	42	0.30
8	12	8	0.30	0.048	0.070	0.053	19	8	0.070	23	0.85	12	0.093	31	0.58
9	16	8	0.38	0.041	0.048	0.060	16	7	0.079	21	0.90	10	0.11	28	0.68
10	4	8	0.30	0.041	0.018	0.047	14	6	0.061	21	0.91	10	0.081	27	0.71

NOTE: See note at bottom of Table B-9.

Table B-14. Power to Detect a 25% Difference Between Two Means in a Two-sample, One-sided Test (continued)

N (Reps)	k	df	rEMS / Control Mean = 0.10			rEMS / Control Mean = 0.20			rEMS / Control Mean =0.30			rEMS / Control Mean = 0.40		
				Power With			Power With			Power With			Power With	
			PMSD	α= 0.05	α= 0.05/k	PMSD	α= 0.05	α= 0.05/k	PMSD	α= 0.05	α= 0.05/k	PMSD	α= 0.05	α= 0.05/k
3	2	4	21	0.80	0.66	43	0.29	0.17	64	0.16	0.09	85	0.12	0.07
3	3	6	21	0.80	0.68	42	0.29	0.18	63	0.16	0.10	84	0.12	0.07
3	4	8	21	0.80	0.68	42	0.29	0.18	63	0.16	0.10	83	0.12	0.07
3	5	10	21	0.80	0.68	42	0.29	0.18	63	0.16	0.10	84	0.12	0.07
4	2	6	17	0.92	0.86	33	0.43	0.29	50	0.24	0.15	66	0.17	0.10
4	3	9	17	0.92	0.86	34	0.43	0.28	50	0.24	0.14	67	0.17	0.09
4	4	12	17	0.92	0.85	34	0.43	0.27	51	0.24	0.13	68	0.17	0.09
4	5	15	17	0.92	0.84	35	0.43	0.26	52	0.24	0.13	69	0.17	0.08
5	2	8	14	0.97	0.94	28	0.55	0.41	42	0.30	0.20	56	0.20	0.13
5	3	12	14	0.97	0.93	29	0.55	0.38	43	0.30	0.18	58	0.20	0.12
5	4	16	15	0.97	0.93	30	0.55	0.36	44	0.30	0.17	59	0.20	0.11
5	5	20	15	0.97	0.92	30	0.55	0.35	45	0.30	0.16	60	0.20	0.10
6	2	10	12	0.98	0.97	25	0.63	0.51	37	0.36	0.25	50	0.24	0.16
6	3	15	13	0.98	0.97	26	0.63	0.47	39	0.36	0.22	52	0.24	0.14
6	4	20	13	0.98	0.96	27	0.63	0.45	40	0.36	0.20	53	0.24	0.12
6	5	25	14	0.98	0.96	27	0.63	0.43	41	0.36	0.19	54	0.24	0.12

Table B-14. Power to Detect a 25% Difference Between Two Means in a Two-sample, One-sided Test

			rEMS / Control Mean = 0.10			rEMS / Control Mean = 0.20			rEMS / Control Mean =0.30			rEMS / Control Mean = 0.40		
				Power With			Power With			Power With			Power With	
N (Reps)	k	df	PMSD	α= 0.05	α= 0.05/k	PMSD	α= 0.05	α= 0.05/k	PMSD	α= 0.05	α= 0.05/k	PMSD	α= 0.05	α= 0.05/k
7	5	30	12	0.99	0.98	25	0.71	0.50	37	0.41	0.23	50	0.28	0.13
8	2	14	10	1.00	0.99	21	0.76	0.66	31	0.46	0.34	42	0.31	0.21
8	3	21	11	1.00	0.99	22	0.76	0.62	33	0.46	0.31	44	0.31	0.18
8	4	28	11	1.00	0.99	23	0.76	0.59	34	0.46	0.28	45	0.31	0.16
8	5	35	12	1.00	0.99	23	0.76	0.57	35	0.46	0.26	46	0.31	0.15
9	2	16	10	1.00	1.00	19	0.81	0.72	29	0.51	0.39	39	0.34	0.24
9	3	24	10	1.00	1.00	20	0.81	0.68	31	0.51	0.35	41	0.34	0.21
9	4	32	11	1.00	1.00	21	0.81	0.65	32	0.51	0.32	42	0.34	0.18
9	5	40	11	1.00	1.00	22	0.81	0.63	33	0.51	0.30	44	0.34	0.17
10	2	18	9	1.00	1.00	18	0.85	0.77	27	0.55	0.43	36	0.37	0.26
10	3	27	10	1.00	1.00	19	0.85	0.73	29	0.55	0.39	39	0.37	0.23
10	4	36	10	1.00	1.00	20	0.85	0.71	30	0.55	0.36	40	0.37	0.21
10	5	45	10	1.00	1.00	21	0.85	0.69	31	0.55	0.33	41	0.37	0.19
11	2	20	9	1.00	1.00	17	0.88	0.81	26	0.59	0.47	35	0.40	0.29
11	3	30	9	1.00	1.00	18	0.88	0.78	27	0.59	0.42	37	0.40	0.25
11	4	40	10	1.00	1.00	19	0.88	0.75	29	0.59	0.39	38	0.40	0.23
11	5	50	10	1.00	1.00	20	0.88	0.73	29	0.59	0.37	39	0.40	0.21
12	2	22	8	1.00	1.00	16	0.90	0.85	25	0.63	0.51	33	0.43	0.32
12	3	33	9	1.00	1.00	17	0.90	0.82	26	0.63	0.46	35	0.43	0.27
12	4	44	9	1.00	1.00	18	0.90	0.79	27	0.63	0.43	36	0.43	0.25
12	5	55	9	1.00	1.00	19	0.90	0.78	28	0.63	0.40	37	0.43	0.23
13	2	24	8	1.00	1.00	16	0.92	0.87	24	0.66	0.55	32	0.45	0.34
13	3	36	8	1.00	1.00	17	0.92	0.85	25	0.66	0.50	33	0.45	0.30
13	4	48	9	1.00	1.00	17	0.92	0.83	26	0.66	0.46	35	0.45	0.27
13	5	60	9	1.00	1.00	18	0.92	0.81	27	0.66	0.44	36	0.45	0.25
14	2	26	8	1.00	1.00	15	0.94	0.90	23	0.69	0.58	30	0.48	0.37
14	3	39	8	1.00	1.00	16	0.94	0.88	24	0.69	0.53	32	0.48	0.32
14	4	52	8	1.00	1.00	17	0.94	0.86	25	0.69	0.50	33	0.48	0.29
14	5	65	9	1.00	1.00	17	0.94	0.84	26	0.69	0.47	34	0.48	0.27
15	2	28	7	1.00	1.00	15	0.95	0.92	22	0.72	0.61	29	0.50	0.39
15	3	42	8	1.00	1.00	15	0.95	0.90	23	0.72	0.56	31	0.50	0.34
15	4	56	8	1.00	1.00	16	0.95	0.88	24	0.72	0.53	32	0.50	0.31
15	5	70	8	1.00	1.00	17	0.95	0.87	25	0.72	0.50	33	0.50	0.29

NOTE: Power is reported for tests with two values of α, 0.05 and 0.05/k. Power for Dunnett's multiple comparison test will fall between these two values. All numbers have been rounded to two significant figures. The number of treatments tested (k) and used to calculate EMS and MSD for a sublethal endpoint will vary depending on the NOEC for survival. k = number of treatments in Dunnett's test; df = degrees of freedom; PMSD = 100 × MSD / (Control Mean); EMS = error mean square; rEMS = square root of the error mean square.

Table B-15. Values of PMSD in Dunnett's Test in Relation to the Square Root of the Error Mean Square (rEMS) for the Test

				Value of PMSD When rEMS / (Control Mean) Equals These Values			
Reps	**k**	**df**	**d**	**0.1**	**0.2**	**0.3**	**0.4**
3	2	4	2.61	21	43	64	85
4	2	6	2.34	17	33	50	66
5	2	8	2.22	14	28	42	56
6	2	10	2.15	12	25	37	50
7	2	12	2.11	11	23	34	45
8	2	14	2.08	10	21	31	42
9	2	16	2.06	10	19	29	39
10	2	18	2.04	9	18	27	37
11	2	20	2.03	9	17	26	35
12	2	22	2.02	8	16	25	33
13	2	24	2.01	8	16	24	32
14	2	26	2.00	8	15	23	30
15	2	28	1.99	7	15	22	29
3	3	6	2.56	21	42	63	84
4	3	9	2.37	17	34	50	67
5	3	12	2.29	14	29	43	58
6	3	15	2.24	13	26	39	52
7	3	18	2.21	12	24	35	47
8	3	21	2.19	11	22	33	44
9	3	24	2.17	10	20	31	41
10	3	27	2.16	10	19	29	39
11	3	30	2.15	9	18	27	37
12	3	33	2.14	9	17	26	35
13	3	36	2.13	8	17	25	33
14	3	39	2.13	8	16	24	32
15	3	42	2.12	8	15	23	31
3	4	8	2.55	21	42	63	83
4	4	12	2.41	17	34	51	68
5	4	16	2.34	15	30	44	59
6	4	20	2.30	13	27	40	53
7	4	24	2.28	12	24	37	49
8	4	28	2.26	11	23	34	45
9	4	32	2.25	11	21	32	42
10	4	36	2.24	10	20	30	40
11	4	40	2.23	10	19	29	38
12	4	44	2.22	9	18	27	36

Table B-15. Values of PMSD in Dunnett's Test in Relation to the Square Root of the Error Mean Square (rEMS) for the Test

				Value of PMSD When rEMS / (Control Mean) Equals These Values			
Reps	**k**	**df**	**d**	**0.1**	**0.2**	**0.3**	**0.4**
13	4	48	2.22	9	17	26	35
14	4	52	2.21	8	17	25	33
15	4	56	2.21	8	16	24	32
3	5	10	2.56	21	42	63	84
4	5	15	2.44	17	35	52	69
5	5	20	2.39	15	30	45	60
6	5	25	2.36	14	27	41	54
7	5	30	2.34	12	25	37	50
8	5	35	2.32	12	23	35	46
9	5	40	2.31	11	22	33	44
10	5	45	2.30	10	21	31	41
11	5	50	2.29	10	20	29	39
12	5	55	2.29	9	19	28	37
13	5	60	2.28	9	18	27	36
14	5	65	2.28	9	17	26	34
15	5	70	2.28	8	17	25	33

NOTE: The number of treatments tested (k) and used to calculate EMS and MSD for a sublethal endpoint will vary depending on the NOEC for survival. k = number of treatments in Dunnett's test; df = degrees of freedom; d = Dunnett's statistic ($\alpha = 0.05$); PMSD = 100 × MSD / (Control Mean); EMS = error mean square; rEMS = square root of the error mean square.

Table B-16. Percentiles of the rEMS/Control Mean, for the Growth or Reproduction Endpoint of Chronic WET Tests, Using Data Pooled Across All Laboratories and Toxicants[a]

	Test Method						
	1000.0 Fathead Minnow	**1002.0 *Cerio-daphnia***	**1003.0 Green Alga**	**1004.0 Sheepshead Minnow**	**1006.0 Inland Silverside**	**1007.0 Mysid (*A. bahia*)**	**1009.0 Red Macroalga**
No. of tests	206	393	85	57	193	130	23
No. of labs	19	33	9	5	16	10	2
Endpoint	G	R	G	G	G	G	R
Percentile	**rEMS/Control Mean**						
25%	0.09	0.17	0.06	0.05	0.09	0.15	0.11
50%	0.12	0.24	0.08	0.08	0.11	0.18	0.18
75%	0.16	0.31	0.10	0.11	0.15	0.23	0.25
80%	0.17	0.32	0.11	0.12	0.16	0.24	0.26
85%	0.18	0.34	0.12	0.13	0.18	0.27	0.27
90%	0.21	0.39	0.13	0.14	0.21	0.29	0.27
95%	0.26	0.44	0.16	0.15	0.26	0.33	0.34

[a] rEMS = square root of the error mean square
[b] G = growth, R = reproduction

Table B-17. Number of Replicates Needed to Provide PMSD of 25% and 33% for Some Less Precise Tests in Each Chronic Test Method (that is, for 85th and 90th Percentiles from Table B-17) for the Sublethal Endpoints in Table B-16

		rEMS / Control Mean		Number of Replicates to Make PMSD = 25		Number of Replicates to Make PMSD = 33	
Test Method	**Required No. of Replicates**	**85th Percentile**	**90th Percentile**	**For 85th Percentile**	**For 90th Percentile**	**For 85th Percentile**	**For 90th Percentile**
1000.0 Fathead Minnow	4 (3)	0.18	0.21	6	8 (7)	4	5
1002.0 *Ceriodaphnia*	10	0.34	0.39	19 (17)	24 (22)	11	14 (13)
1003.0 Green Alga	4 (3)	0.12	0.13	4	4	3	3
1004.0 Sheepshead Minnow	4 (3)	0.13	0.14	4	4	3	3
1006.0 Inland Silverside	4 (3)	0.18	0.21	6	8 (7)	4	5
1007.0 Mysid	8	0.27	0.29	12 (11)	14 (13)	7	9 (8)
1009.0 Red Macroalga	4 (3)	0.27	0.27	12 (11)	12 (11)	7	7

NOTE: The number for k = 3 treatments appears in parentheses if it differs from the number needed when four treatments are compared with the control; rEMS = square root of the error mean square; PMSD = percent minimum significant difference.

Table B-18. Percentiles of the Within-Laboratory Values of CV for NOEC (using NOEC for the Most Sensitive Endpoint in Each Test)

Method No.	Method	No. Labs	P10	P25	P50	P75	P90
1000.0	Fathead Minnow Larval Survival & Growth	19	0	0.22	0.31	0.52	0.65
1002.0	*Ceriodaphnia* Survival & Reproduction	33	0.20	0.25	0.35	0.49	0.60
1003.0	Green Alga Growth	9	0.30	0.40	0.46	0.56	0.82
1004.0	Sheepshead Minnow Larval Survival & Growth	5	0.20	0.36	0.38	0.44	0.52
1006.0	Inland Silverside Larval Survival & Growth	16	0.19	0.35	0.46	0.59	0.66
1007.0	Mysid Survival, Growth, & Fecundity	10	0.28	0.32	0.40	0.50	0.60
1009.0	Red Macroalga Reprod	2	0.85	0.85	1.00	1.16	1.16
1010.0	Topsmelt Larval Survival & Growth	1	0.22	0.22	0.22	0.22	0.22
1012.0	Pacific Oyster Embryo-Larval Survival & Dev.	1	0.45	0.45	0.45	0.45	0.45
1013.0	Mussel Embryo-Larval Survival & Dev.	3	0	0	0.39	0.43	0.43
1014.0	Red Abalone Larval Development	10	0.24	0.25	0.29	0.31	0.38
1016.0	Sea Urchin Fertilization[a]	12	0.31	0.40	0.50	0.69	0.76
1017.0	Sand Dollar Fertilization[a]	7	0.40	0.41	0.53	0.75	0.81
1018.0	Giant Kelp Germination & Germ-Tube Length	11	0.33	0.36	0.59	0.68	0.72

[a] These two test species include previous test method procedures (Dinnel 1987, Chapman 1992). However, EPA (USEPA 1995) has standardized these two methods to provide further guidance and therefore minimize within-test variability.

APPENDIX C

SAMPLE CALCULATION OF PERMIT LIMITS USING EPA'S STATISTICALLY-BASED METHODOLOGY AND SAMPLE PERMIT LANGUAGE

This page intentionally left blank.

SAMPLE CALCULATIONS OF PERMIT LIMITS USING EPA'S STATISTICALLY-BASED METHODOLOGY AND SAMPLE PERMIT LANGUAGE

The NPDES regulation (40 CFR Part 122.44(d)(1)) implementing section 301 (b)(1)(C) of the CWA requires that permits include limits for all pollutants or parameters that "*are or may be discharged at a level which will cause, have the reasonable potential to cause, or contribute to an excursion above any State water quality standard, including State narrative criteria for water quality*." Once it has been established that a permit limit is needed, Federal regulations at 40 CFR Part 122.45(d) require that limits be expressed as maximum daily discharge limits (MDL) and average monthly discharge limits (AML) for all dischargers other than publicly owned treatment works (POTWs), and as average weekly and average monthly discharge limits for POTWs, unless impracticable. EPA does not believe that it is impracticable to express WET permit limits as MDLs and AMLs.

C.1 Sample Calculations

To set MDLs and AMLs based on acute and chronic wasteload allocations (WLAs), use the following four steps.

1. Convert the acute wasteload allocation to chronic toxic units.

2. Calculate the long-term average wasteload that will satisfy the acute and chronic wasteload allocations.

3. Determine the lower (more limiting) of the two long-term averages.

4. Calculate the maximum daily and average monthly permit limits using the lower (more limiting) long-term average.

Step 1 - Determine the Wasteload Allocation

The acute and chronic aquatic life criteria are converted to acute and chronic wasteload allocations (WLAa or WLAc) for the receiving waters based on the following mass balance equation:

$$Q_d C_d = Q_e C_e + Q_u C_u \qquad \textbf{(Eq. 1)}$$

where

Q_d = downstream flow = $Q_u + Q_e$
C_d = aquatic life criteria that cannot be exceeded downstream
Q_e = effluent flow
C_e = concentration of pollutant in effluent = WLAa or WLAc
Q_u = upstream flow
C_u = upstream background concentration of pollutant.

Rearranging Equation 1 to determine the effluent concentration (C_e) or the wasteload allocation (WLA) results in the following:

$$C_e = WLA = \frac{Q_d C_d - Q_u C_u}{Q_e} \qquad \textbf{(Eq. 2)}$$

When a mixing zone[1] is allowed, this equation becomes:

$$C_e = WLA = \left[\frac{C_d\left(Q_u \times \%MZ\right) + C_d Q_e}{Q_e}\right] - \left[\frac{Q_u C_u\left(\%MZ\right)}{Q_e}\right] \qquad \textbf{(Eq. 2a)}$$

where %MZ is the mixing zone allowable by State standards. In this example, the State authorized a mixing zone of 50 percent of river volume for WET. The effluent limits were derived using the State's guidelines. Establishing a mixing zone, however, is a discretionary function of the State. If the State does not certify a mixing zone in the 401 certification process, the effluent limits must be recalculated without a mixing zone.

There is an additional step for WET. The WLAa needs to be converted from acute toxic units (TUa) to chronic toxic units (TUc). The acute WLA is converted into an equivalent chronic WLA by multiplying the acute WLA by an acute-to-chronic ratio (ACR). Optimally, this ratio is based on effluent data. A default value of 10, however, can be used based on the information presented in Chapter 1 and Appendix A of the TSD.

WLAa,c = WLAa × ACR, where
ACR = acute-to-chronic ratio

For this example, the following information applies:

	C_d	Q_e	Q_u	%MZ	Q_{umix} [a]	Q_d	C_u	CV[b]
Acute	0.3 TUa	15.5 cfs	109 cfs	50	54.5 cfs	70 cfs	0 TU_a	0.6
Chronic	1.0 TUc	15.5 cfs	170 cfs	50	85 cfs	100.5 cfs	0 TU_c	0.6

[a] Q_{umix} is the upstream flow in the mixing zone ($Q_{umix} = Q_u \times \%MZ$)
[b] Only 7 valid data points were available, so a default coefficient of variation was used in the calculations.

$$WET\ WLAa = \left[\frac{\left(0.3TUa\right)\times\left(109\times 0.50\right)+\left(0.3\times 15.5\right)}{15.5}\right] - \left[\frac{109\times 0\times 0.25}{15.5}\right] = 1.35TUa$$

$$WET\ WLAa,c = 10\times 1.35TUa = 13.5TUa,c$$

$$WET\ WLAc = \left[\frac{1.0TUc\times\left(170\times 0.50\right)+\left(1.0\times 15.5\right)}{15.5}\right] - \left[\frac{170\times 0\times 0.50}{15.5}\right] = 6.5TUc$$

Step 2 - Determine the Long-Term Average (LTA)

The acute WLA is converted to a long-term average concentration (LTAa,c) using the following equation:

$$LTAa,c = WLAa,c \times e^{\left[0.5s^2 - zs\right]} \qquad \textbf{(Eq. 3)}$$

where,

σ^2 = $\ln(CV^2 + 1) = \ln(0.6^2 + 1) = 0.307$; $\sigma = 0.555$
z = 2.326 for 99th percentile probability basis
CV = coefficient of variation = standard deviation/mean = 0.6
Acute multiplier = $e^{(0.5 \times 0.307 - (2.326 \times 0.555))} = 0.321$.

$$LTAa,c = 13.5TUa,c \times 0.321 = 4.33TUa,c$$

[1] A mixing zone is an allocated impact zone where water quality criteria can be exceeded if acutely toxic conditions are prevented. Only the State has the regulatory authority to grant the establishment of a mixing zone.

The chronic WLA is converted to a long-term average concentration (LTAc) using the following equation:

$$LTAc = WLAc \times e^{\left[0.5s^2 - zs\right]} \qquad \textbf{(Eq. 4)}$$

where,

σ^2 = $\ln(CV^2/4 + 1) = \ln(0.6^2/4+1)$ 0.086; $\sigma = 0.294$
z = 2.326 for 99th percentile probability basis
CV = coefficient of variation = standard deviation/mean = 0.6
Chronic multiplier = $e^{(0.5 \times 0.086 - 2.326 \times 0.294)} = 0.542$.

$$LTAc = 6.5\,TUc \times 0.542 = 3.43\,TUc$$

Step 3 - Determine the More Limiting Long-Term Average

To protect a waterbody from both acute and chronic effects, the more limiting of the calculated LTAa and LTAc is used to derive the effluent limits. The TSD recommends using the 95th percentile for the AML and the 99th percentile for the MDL. As shown above, the LTAc value was less than the LTAa value.

Step 4 - Determine the Permit Limits

The MDL and the AML are calculated as follows.

$$MDL = LTAc \times e^{\left[zs - 0.5s^2\right]} \qquad \textbf{(Eq. 5)}$$

where,

σ^2 = $\ln(CV^2 + 1) = 0.307$; $\sigma = 0.555$
z = 2.326 for 99th percentile probability basis
CV = coefficient of variation = 0.6

$$AML = LTAc \times e^{\left[zs - 0.5s^2\right]} \qquad \textbf{(Eq. 6)}$$

where,

σ^2 = $\ln(CV^2/n + 1) = 0.086$; $\sigma = 0.294$
z = 1.645 for 95th percentile probability basis
CV = coefficient of variation = 0.6
n = number of sampling events required per month for WET = 1
n = 4 for calculations[2]

The following table lists the effluent limits for this example:

Parameter	CV	LTA_c	$e^{[z\sigma-0.5\sigma^2]}$ (for MDL)	$e^{[z\sigma-0.5\sigma^2]}$ (for AML)	MDL	AML
WET	0.6	3.43	3.11	2.13	10.7 TU_c	7.3 TU_c

[2] When the sample frequency is monthly or less than monthly, the TSD recommends that "n" be set equal to 4.

C.2 Sample Chronic Toxicity Permit Language

Sample chronic toxicity permit language is provided in the following paragraphs. Alternative wording, as appropriate for a specific permit, is provided in redline typeface for the regulatory authority to decide.

The permittee shall conduct monthly/quarterly/semi-annual/annual toxicity tests on grab/24-hour composite effluent samples. Samples shall be taken at the NPDES sampling location. In addition, a split of each sample collected must be analyzed for the chemical and physical parameters required in Part 1.A below. When the timing of sample collection coincides with timing of the sampling required in Part I.A, analysis of the split sample will fulfill the requirements of Part I.A. as well.

1. Test Species and Methods

NOTE: CHOOSE EITHER FRESHWATER OR MARINE LANGUAGE

Freshwater

a. The permittee shall conduct short-term tests with the cladoceran, water flea, *Ceriodaphnia dubia* (survival and reproduction test), the fathead minnow, *Pimephales promelas* (larval survival and growth test), and the green alga, *Selanastrum capricornutum* (growth test) for the first three suites of tests. After this screening period, monitoring shall be conducted using the most sensitive species.

b. Every year, the permittee shall re-screen once with the three species listed above and continue to monitor with the most sensitive species. Re-screening shall be conducted at a different time of year from the previous year's re-screening. Note to permit writers: If testing is annual or less than annual, omit this step.

c. The presence of chronic toxicity shall be estimated as specified in EPA's methods (USEPA 1994b).

Marine and Estuarine

a. The permittee shall conduct tests as follows with a vertebrate, an invertebrate, and a plant for the first three suites of tests. After the screening period, monitoring shall be conducted using the most sensitive species.

b. Every year, the permittee shall re-screen once with the three species listed above and continue to monitor with the most sensitive species. Re-screening shall be conducted at a different time of year from the previous year's re-screening. Note to permit writers: If testing is annual or less, omit this step.

For West Coast only:

c. The presence of chronic toxicity shall be estimated as specified using West Coast marine organisms according to EPA's methods (USEPA 1995).

or

For East Coast only:

c. The presence of chronic toxicity shall be estimated as specified using East Coast marine organisms according to EPA's methods (USEPA 1994c).

2. Toxicity Limits/Toxicity Monitoring Trigger

a. Chronic toxicity measures a sublethal effect (e.g., reduced growth, reproduction) to experimental test organisms exposed to an effluent or ambient waters compared to that of the control organisms. When a permit limit is appropriate, the chronic toxicity limitation is written based on State Water Quality Standards. If a permit limit is not appropriate, then this section should be called "Toxicity Monitoring Trigger."

b. Results shall be reported in TUc, where TUc = 100/NOEC or 100/ICp or ECp (in percent effluent). The no observed effect concentration (NOEC) is the highest concentration of toxicant to which organisms are exposed in a chronic test that causes no observable adverse effect on the test organisms (e.g., the highest concentration of toxicant to which the values for the observed responses are not statistically significantly different from the controls). The inhibition concentration, IC, is a point estimate of the toxicant concentration that causes a given percent reduction (p) in a non-quantal biological measurement (e.g., reproduction or growth) calculated from a continuous model (the EPA Interpolation Method). The effective concentration, EC, is a point estimate of the toxicant concentration that would cause a given percent reduction (p) in quantal biological measurement (e.g., larval development, survival) calculated from a continuous model (e.g., Probit).

3. Quality Assurance

a. A series of at least five dilutions and a control will be tested. The series shall include the instream waste concentration (IWC) (permit writer should insert the actual value of the IWC), two dilutions above the IWC, and two dilutions below the IWC. The IWC is the concentration of effluent at the edge of the mixing zone. If there is no mixing zone, then the dilution series would be the following concentrations: 12.5, 25, 50, 75, and 100 percent effluent.

b. If organisms are not cultured in-house, concurrent testing with a reference toxicant shall be conducted. Where organisms are cultured in-house, monthly reference toxicant testing is sufficient. Reference toxicant tests also shall be conducted using the same test conditions as the effluent toxicity tests (e.g., same test duration, etc).

c. If either the reference toxicant test or effluent test does not meet all test acceptability criteria (TAC) as specified in the manual, then the permittee must re-sample and re-test within 14 days or as soon as possible.

d. The reference toxicant and effluent tests must meet the upper and lower bounds on test sensitivity as determined by calculating the percent minimum significant difference (PMSD) for each test result. The test sensitivity bound is specified for each test method (see variability document EPA/833-R-00-003, Table 3-6). There are five possible outcomes based on the PMSD result:

1. **Unqualified Pass**–The test's PMSD is within bounds and there is no significant difference between the means for the control and the IWC treatment. The regulatory authority would conclude that there *is no toxicity at the IWC concentration*.

2. **Unqualified Fail**–The test's PMSD is larger than the lower bound (but not greater than the upper bound) in Table 3-6 and there is a significant difference between the means for the control and the IWC treatment. The regulatory authority would conclude that there *is toxicity at the IWC concentration*.

3. **Lacks Test Sensitivity**–The test's PMSD exceeds the upper bound in Table 3-6 and there is no significant difference between the means for the control and the IWC treatment. The test

is considered invalid. An effluent sample must be collected and another toxicity test must be conducted. The permittee must re-sample and retest within fourteen (14) days or as soon as possible.

4. **Lacks Test Sensitivity**–The test's PMSD exceeds the upper bound in Table 3-6 and there is a significant difference between the means for the control and the IWC treatment. The test is considered valid. The regulatory authority will conclude that the *is toxicity at the IWC concentration*.

5. **Very Small but Significant Difference**–The relative difference (see Section 6.4.2, below) between the means for the control and the IWC treatment is smaller than the lower bound in Table 3-6 and this difference is statistically significant. The test is acceptable. The NOEC is determined as described in Sections 6.4.2 and 6.4.3 (below).

e. Control and dilution water should be receiving water or laboratory water, as appropriate, as described in the manual. If the dilution water used is different from the culture water, a second control using culture water shall be used.

4. Preparing the Initial Investigation of the TRE Workplan

The permittee shall submit to EPA a copy of the permittee's initial investigation Toxicity Reduction Evaluation (TRE) workplan (1-2 pages) within 90 days of the effective date of this permit. This plan shall describe the steps the permittee intends to follow if toxicity is detected, and should include, at least the following items:

a. A description of the investigation and evaluation techniques that would be used to identify potential causes and sources of toxicity, effluent variability, and treatment system efficiency.

b. A description of the facility's methods of maximizing in-house treatment efficiency and good housekeeping practices.

c. If a toxicity identification evaluation (TIE) is necessary, an indication of the person who would conduct the TIEs (i.e., an in-house expert or an outside contractor).

5. Accelerated Testing

a. If the initial investigation indicates the source of toxicity (for instance, a temporary plant upset), then only one additional test is necessary. If toxicity is detected in this test as specified in Section 2a, then Section 6 shall apply.

b. If chronic toxicity/the chronic toxicity monitoring requirements as defined in Section 2a are triggered, then the permittee shall conduct six more tests, approximately every two weeks, over a twelve-week period. Testing shall commence within two weeks of receipt of the sample results of the exceedance of the WET monitoring trigger.

c. If none of the six tests indicate toxicity as specified in Section 2a, then the permittee may return to the normal testing frequency.

6. Toxicity Reduction Evaluation (TRE) and Toxicity Identification Evaluation (TIE)

a. If chronic toxicity (defined as either the toxicity permit limit or monitoring trigger specified in Section 2a) is detected in any of the six additional tests, then, in accordance with the facility's initial investigation according to the TRE workplan, the permittee shall initiate a TRE within

fifteen (15) days of the exceedance to reduce the cause(s) of toxicity. At a minimum, the permittee shall use EPA manuals EPA/600/2-88/070 (industrial) or EPA/833B-99/002 (municipal) as guidance. The permittee will expeditiously develop a more detailed TRE workplan, which includes:

(1) Further actions to investigate and identify the cause of toxicity
(2) Actions the permittee will take to mitigate the impact of the discharge and prevent the recurrence of toxicity
(3) A schedule for these actions

b. The permittee may initiate a TIE as part of the TRE process to identify the cause(s) of toxicity. The permittee shall use the EPA acute and chronic manuals, EPA/600/6-91/005F (Phase I)/EPA/600/R-96-054 (for marine), EPA/600/R-92/080 (Phase II), and EPA-600/R-92/081 (Phase III) as guidance.

7. Reporting

a. The permittee shall submit the results of the toxicity tests, including any accelerated testing conducted during the month, in TUs with the discharge monitoring reports (DMR) for the month in which the test is conducted. If an initial investigation indicates the source of toxicity and accelerated testing is unnecessary, pursuant to Section 5, then those results also shall be submitted with the DMR for the quarter in which the investigation occurred.

b. The full report shall be submitted by the end of the month in which the DMR is submitted.

c. The full report shall consist of (1) the results; (2) the dates of sample collection and initiation of each toxicity test; (3) the monthly average limit or trigger and daily maximum limit or trigger as described in Section 2a.

d. Test results for chronic tests also shall be reported according to the chronic manual chapter on Report Preparation and shall be attached to the DMR.

e. The permittee shall notify EPA in writing 15 days after the receipt of the results of a monitoring limit or trigger. The notification will describe actions the permittee has taken or will take to investigate and correct the cause(s) of toxicity. It may also include a status report on any actions required by the permit, with a schedule for actions not yet completed. If no actions have been taken, the reasons shall be given.

8. Reopener

a. This permit may be modified in accordance with the requirements set forth at 40 CFR Parts 122 and 124 to include appropriate conditions or limits to address demonstrated effluent toxicity based on newly available information.

This page intentionally left blank.

APPENDIX D

FREQUENTLY ASKED QUESTIONS (FAQS)

This page intentionally left blank.

FREQUENTLY ASKED QUESTIONS (FAQS)

Appendix D contains some of the frequently asked questions regarding WET and WET testing. These questions and answers were prepared by and appear on a web site maintained by the Society of Environmental Toxicology and Chemistry (SETAC) (http://www.setac.org). The SETAC WET Expert Advisory Panels provide scientific opinion and training on WET technical issues under a cooperative agreement with EPA (WET Cooperative Agreement No. CX 824845-01-0). EPA's inclusion of these questions and answers in this document is not an endorsement of the Panels' opinions or responses to the FAQs, but rather provides readers with an additional source of information in issues commonly raised with regard to WET and WET testing. This information was prepared in response to questions received by SETAC about WET. It was generated by the WET Expert Advisory Panels (EAP) Steering Committee (SC), all volunteers and all member of the Society of Environmental Toxicology and Chemistry. Each person is considered an expert in some aspect of WET, and the information provide in these FAQs represents the consensus of the Committee's collective expertise at the time this summary was written (Feb., 1999).

This information is intended to stimulate further discussion about WET, WET-related research, and the science underlying WET. The information is not to be construed as representing an official position of SETAC, the SETAC Foundation for Environmental Education, or the U.S. Environmental Protection Agency. Any questions, comments, and requests should be sent to: Society of Environmental Toxicology and Chemistry (SETAC), 1010 North 12th Avenue, Pensacola, FL 32501-3367, Telephone: 850-469-1500, Facsimile: 850-469-9778, e-mail: setac@setac.org. All materials copyright Society of Environmental Toxicology and Chemistry (SETAC), 2000, and may not be used without written permission.[1,2]

Whole effluent toxicity tests rely on the assumption that test organisms used are representative of a normal and healthy population. What indicators of test organism health are utilized in testing programs?

Both subjective and objective (e.g., test acceptability criteria) indicators of organism health are available, some described within the methods manuals. Some national indicators exist which allow comparison of analytical results between laboratories (i.e., the DMRQA program for major NPDES facilities) or regional activities such as State WET certification programs which provide round-robin validation of test practice including organism health (e.g., North Carolina's Biological Laboratory Certification program). Other national programs like the National Environmental Laboratory Accreditation Program (NELAP) are being followed by the WET EAP SC. Commonly used indicators of organism health are the required reference toxicity analyses and individual test acceptability criteria. Tests properly utilizing randomization procedures along with required and suggested quality control standards retain many built-in checks of typical organism response.

What are the definitions of acceptability criteria for reference toxicant tests?

Reference toxicant tests should meet the same test acceptability criteria as those of compliance test. With regard to assessment of organism health and the overall test practice, USEPA has recommended that routine reference toxicant tests be performed to establish a CUSUM or cumulative summation chart of testing results. Normal results should lie within plus or minus two standard deviations of the cumulative mean value

---

[1] Reprinted with permission of SETAC.

[2] Note that the terms, abbreviations, and acronyms used in this appendix may differ from their usage throughout the rest of this document. EPA consciously chose not to edit this SETAC-supplied information so that the actual nomenclature and terminology as used by SETAC on their web site would be reflected here.

of point estimate endpoints. Values falling outside of those ranges should result in careful scrutiny of the data and testing systems. Data produced during these "out of control" conditions should be considered suspect.

How does increasing the difference in test concentration dilutions affect the prediction of response?

Better resolution around threshold effect concentrations provide better input to mathematical models to predict point estimations of effect and reduce uncertainty in hypothesis tests of effect. Reducing the distance between effluent dilutions should be encouraged. There may be some confusion about USEPA's specification of dilution series in these cases. The methods specify a minimum set of dilutions, i.e., no wider than 0.5 dilution between concentrations. No limitations on added concentrations within that range exist. Experimental design should account for concentrations of concern and should attempt to maximize resolution in that range. Test design should maximize test concentrations around the effect concentration of concern, i.e., the instream waste concentration or limited concentration of a discharging facility, in order to minimize the need for interpolation of effects between tested concentrations.

What are the different types of variability in whole effluent toxicity tests?

Variability is inherent in any analytical procedure. The precision of a method describes the closeness of agreement between test results obtained from repeated testing of a prescribed method. WET test precision can be categorized by: 1) intratest (within-test) variability, 2) intralaboratory (within-laboratory) variability, and 3) interlaboratory (between-laboratory) variability. Intratest variability can be attributed to variables such as the number of treatment replicates, the number of test organisms exposed per replicate, and the sensitivity differences between individual organisms (i.e., genetic variability). Intralaboratory variability is that which is measured when tests are conducted under reasonably constant conditions in the same laboratory (e.g., reference toxicant or effluent sample tested over time). Sources of intralaboratory variability include those factors described for intratest variability, as well as differences: 1) in test conditions (e.g., seasonal differences in dilution water quality, differences in environmental conditions), 2) from test to test in organism condition/health, and 3) in analyst performance from test to test. Interlaboratory variability reflects the degree of precision that is measured when the same sample or reference toxicant is analyzed by multiple laboratories using the same methods. Variability measured between laboratories is a consequence of variability associated with both intratest and intralaboratory variability factors, as well as differences allowed within the test methods themselves (e.g., source of dilution water), technician training programs, sample and organism culturing/shipping effects, testing protocols, food quality, and testing facilities.

Two general categories of variability are of greatest concern: 1) analyst experience, and 2) test organism condition/health. The experience and qualifications of the analyst who actually performs the toxicity test in the laboratory will dictate how well the culture and test methods are followed and the extent to which good judgment is exercised when difficulties/issues arise in the process of conducting the test, analyzing the data, and interpreting the results. Improper utilization of WET methods can have a substantial impact on test result variability. Guidance for specific test conditions and standard methods to control many causes of variability are found in the USEPA (U.S. Environmental Protection Agency) methods manuals (USEPA 1993, USEPA 1994a, USEPA 1994b, USEPA 1995). Strict adherence to these methods can greatly reduce variability.

USEPA. 1993. Methods for measuring the acute toxicity of effluents and receiving waters to freshwater and marine organisms. 4th ed. Weber C.I., editor. Cincinnati: U.S. Environmental Protection Agency (USEPA) Office of Research and Development. EPA/600/4-90/027F. 293 p.

USEPA. 1994a. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to marine and estuarine organisms. 2nd ed. Klemm, D.J., Morrison, G.E., Norberg-King, T.J., Peltier, W.H. and Heber, M.A., editors.

Cincinnati: U.S. Environmental Protection Agency (USEPA) Office of Research and Development. EPA/600/4-91/003. 341 p.

USEPA. 1994b. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to freshwater organisms. 3rd ed. Lewis, P.A., Klemm, D.J., Lazorchak, J.M., Norberg-King, T.J., Peltier, W.H. and Heber, M.A., editors. Cincinnati: U.S. Environmental Protection Agency (USEPA) Office of Research and Development. EPA/600/4-91/002. 341 p.

USEPA. 1995. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to west coast marine and estuarine organisms. Chapman, G.A., Denton, D.I., Lazorchak, J.M., editors. Cincinnati: U.S. Environmental Protection Agency (USEPA) Office of Research and Development. EPA/600/R-95-136. 661 p.

What specific factors influence WET test variability?

There are a number of factors that can meaningfully influence the variability of test results. These factors include, but are not limited to, those listed below.

Sample Characteristics

The nature of the sample collected can have a significant influence on the outcome of a WET test. Care must be exercised to collect the most representative sample possible during the time frame of interest. Sample volume can influence the outcome of a toxicity test. For example, if the sample-to-container-wall ratio is small, or if the sample-container contact time is especially long before the sample is refrigerated; certain particulate-active constituents such as zinc (Chapter 5 in Grothe et al. 1996), polymeric substances, charged materials, or hydrophobic chemicals in a sample can interact with the container. Samples too small in volume may also increase the potential of collecting a non-representative fraction of a non-homogenous sample stream. The type of sample (i.e., grab or composite) may influence the outcome of a WET test and contribute to variability. Grab samples may hit or miss toxicity spikes thus possibly increasing the variability between samples taken at different times at the same outfall. Composite samples will average concentrations over the entire collection period, possibly smoothing peaks and valleys of toxicity in variable water media. The various USEPA method manuals review the importance of using appropriate sample types for different types of effluents. Storage and handling can affect the toxicity and variability of samples. The general assumption is that the toxicity of a sample is most likely to decrease with holding time due to factors such as biodegradation, hydrolysis, and adsorption. These factors are minimized by “cold” storage and shipment on ice as well as test initiation within the specified USEPA guidelines. Water samples for WET testing may be manipulated in a variety of ways to comply with special requirements or circumstances. This applies, for example, when freshwater effluents are discharged to a saline receiving stream and marine or estuarine organisms are used for testing. Care must be taken, in this case, that ionic strength and composition are within levels tolerated by the specific test organisms or results may not be representative of actual toxicity or comparable between labs.

Abiotic Conditions

Abiotic conditions can strongly influence the variability of WET test results. For that reason, most of the abiotic conditions that should be standardized during WET testing (DO, light, hardness, alkalinity, etc.) are specified in protocols contained in the USEPA methods manuals. While these factors may not be problematic sources of variability within tests, they may be of major concern across tests (both within and among laboratories). Very small ranges of temperatures are specified for WET testing. Test solution pH can influence the bioavailability and toxicity of chemical constituents, such as some metals (e.g., Cu, Zn) and ammonia. Careful use of dilution waters, salinity adjustments, aeration, feeding, and other factors causing shifts in pH will help to reduce variability.

Exposure

In WET testing, we seek a balance between realistically mimicking exposure scenarios and evaluating effluents with sufficient testing while controlling testing costs. Variability in test results can be greatly influenced by the method of exposure chosen (i.e., static, static renewal, and flow-through). For example, tests of samples with nonpersistent toxicants or with chambers with high loading rates will be influenced to a greater degree using a static design rather than a flow-through design. As the number of variables which influence test results increases, overall test variability increases unless those variables are controlled. However, flow-through tests are much more costly than static tests. The number of concentrations and dilution series may influence variability of the test results. Point estimate models will more precisely estimate the statistical endpoint if the test concentrations are near the actual LCx (concentration that is lethal to x percent of organisms), ECx (concentration that affects x percent of organisms), or ICx (concentration that inhibits response by x percent). In contrast, as the NOEC approaches the concentration at which effects begin to be observed (i.e., LOEC), estimates may show greater variation. Many NPDES permits include a test dilution that is consistent with the Instream Waste Concentration (IWC) based upon dilution in the receiving system. The minimum number of tested dilutions recommended can be increased, particularly in the range of expected effects (if known), in order to improve resolution of the acute or chronic endpoint. Costs of increased dilutions testing are incremental to the cost of a typical test, but such testing is cost effective in cases where small changes in organism responses may affect compliance.

The WET endpoint is a function of test duration, in most cases (percent mortality after a period of time, for example). Test duration can be a function of the endpoint that is to be assessed. In at least one situation, the *C. dubia* survival and reproduction test, exposure duration is governed by the amount of time needed for 60 percent of the control organisms to produce a third brood (up to 8 days), at which time the test is repeated if the control performance is not acceptable (USEPA 1994b). The timing for test termination can therefore vary between 6 and 8 days. This introduces the possibility of intertest variability in terms of both number of young produced and test sensitivity due to exposure duration. The cost of reducing test duration variability is small; the corresponding reduction in test results variability could, however, be significant.

Sample Toxicity

The exposure-response relationship can be affected by the sensitivity of the test species to the individual and combined chemicals of a sample as well as the concentrations of those chemicals in that sample. Testing of samples which exhibit high slopes in their concentration-response curves at the test statistical endpoint (LCx, ECx, and ICx) tends to provide less variable (intratest and inter-test) results than tests of samples exhibiting low slopes in their concentration-response curves. The sensitivity of different species to any single chemical or mixture of chemicals can also be quite different, even when all variables are held constant. For example, rainbow trout are approximately an order of magnitude more acutely sensitive to cadmium than daphnids (USEPA 1985a) while daphnids are approximately 2.5 times more acutely sensitive to chlorine than rainbow trout (USEPA 1985b). Herbicides (e.g., atrazine) are more acutely toxic to plants than fish (Solomon et al. 1996). This is why vertebrates, invertebrates, and plants are recommended for testing effluents in the NPDES program.

Food

Food quality can vary in a number of ways. Organisms whose diets vary in nutritional quality and size, before and during testing, may respond differently to the same sample under identical test conditions. For example, brine shrimp nauplli that are less than 24 hours old are required in all tests using these organisms as food to maintain the nutritional quality of the nauplii and to keep their size at the optimum for consumption by test organisms. The YCT and algal diet for *C. dubia* should contain specific concentrations of solids and algal cells as outlined in the manual. The quantity of food available can affect dissolved oxygen and pH levels within a test chamber and act as a substrate for the absorption and adsorption of toxic chemicals from the tested sample, thus reducing bioavailability.

Dilution Water

Optimally, the dilution water should replicate the quality of the receiving water. However, if the objective of the test is to estimate the absolute toxicity of the sample (effluent), which is the primary objective of NPDES permit-related toxicity testing, then a synthetic (standard) dilution water is used (USEPA 1993, USEPA 1994a, USEPA 1994b). If the objective is to estimate the toxicity of the sample in uncontaminated receiving water, then the test may be conducted using non-toxic receiving water. Dilution water quality can affect the toxicity of effluent, surface water, and stormwater dilutions by modifying the bioavailability of toxic chemicals in the sample. In addition, parameters such as TDS (hardness, salinity, conductivity), turbidity, DO, pH, micronutrients, and bacteria counts can impact test organism physiology, sensitivity, and biological response. Therefore, test variability at all levels can be affected by variability in dilution water quality. Synthetic dilution water quality can also vary with the age of the prepared water in relation to the exposure of test organisms and with the source and quality of the base water.

Organism History and Handling

Perhaps one of the most important considerations in controlling WET variability is an organism's pretest history of health and maintenance, which consists of four factors: collection, culture, acclimation, and handling specific to the test. Organism history can be evaluated through charting performance of laboratory controls with a reference toxicant over time. All practical attempts should be made to avoid use of field-collected animals for WET testing. The most common sources of test organisms for WET tests are in-house cultures and/or organism suppliers. Organisms to be tested, whether field-collected or cultured, may require acclimation to test conditions. Variation in acclimation practices between tests can result in the use of organisms of varying sensitivity between tests. The importance of analyst technique is most pronounced when the analyst handles organisms before and during the test.

Randomization

Results will be variable in all analytical techniques, not just WET, despite all efforts to eliminate and reduce sources of variability. The randomization approach used to assign test replicates within an incubator or water bath and the approach used to assign test organisms to test replicates are attempts to evenly distribute this variability within the testing environment and between organisms. All test methods include procedures for randomization which must be followed.

Organism Numbers

The number of organisms exposed in a toxicity test has a direct and calculable bearing on the ability of that test to detect and estimate effects resulting from that exposure. Generally, as the total number of organisms increases in a test, the ability to detect effects (i.e., statistical power in a hypothesis test) and the certainty in point estimates increases. Differences in number of organisms per replicate and treatment can be due to the loss of individuals or replicates through analyst errors or to the death or lack of response of all organisms in one or more replicates. The former reduces power or effect-estimate certainty (point estimate confidence intervals) by reducing sample size. The latter may reduce power or effect-estimate certainty by increasing variation in response relative to other replicates and treatments. Intra- and interlaboratory variability can include the factors discussed above, as well as possible differences in study design (total number of organisms and total number of replicates).

Organism Age and Quality

The recommended ages of test organisms for established protocols have two general considerations: (1) relative physical sensitivity of different life stages to the test conditions, independent of the challenges of a toxicant and, (2) relative sensitivity of different life stages to toxic constituents. Young organisms are often considered more sensitive to toxic and physical stressors than their older counterparts. For this reason, the use of early life stages, such as first instars of daphnids and juvenile mysids and fish, is recommended for all tests.

The effects of organism age on WET variability are potentially greatest between tests and between laboratories where age differences may be greater. As examples, all *C. dubia* used in a reproduction test must be within 8 hours of age but can be up to 24 h old; and fathead minnow larvae used in the growth test must be within 24 hours of age in a single test but could range between 1 to 2 days depending on whether the organisms are cultured in-house or shipped from an off-site culture facility. In the acute tests with fathead and sheepshead minnows, the age difference between tests can range from <24 h to 14 d.

Grothe, D. R., K. L. Dickson, and D. K. Reed-Judkins, eds.1996. Whole Effluent Toxicity Testing: An Evaluation of Methods and Prediction of Receiving System Impacts, SETAC Press, Pensacola, FL, USA. 340 p.

Solomon, K.R., D.B. Baker, R.P. Richards, K.R. Dixon, S.J. Klaine, T.W. LaPoint, R.J. Kendall, J.M. Giddings, J.P. Giesy, L.W. Hall, Jr. and W.M. Williams. 1996. Ecological risk assessment of atrazine in North America surface waters. Environ. Toxicol. Chem. 15:31-76.USEPA. 1985a. Ambient water quality criteria for cadmium - 1984. EPA 440/5-84-032. Office of Regulations and Standards, Washington, DC.

USEPA. 1985b. Ambient water quality criteria for chlorine - 1984. EPA 440/5-84-030. Office of Regulations and Standards, Washington, DC.

USEPA. 1993. Methods for measuring the acute toxicity of effluents and receiving waters to freshwater and marine organisms. 4th ed. Weber C.I., editor. Cincinnati: U.S. Environmental Protection Agency (USEPA) Office of Research and Development. EPA/600/4-90/027F. 293 p.

USEPA. 1994a. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to marine and estuarine organisms. 2nd ed. Klemm, D.J., Morrison, G.E., Norberg-King, T.J., Peltier, W.H. and Heber, M.A., editors. Cincinnati: U.S. Environmental Protection Agency (USEPA) Office of Research and Development. EPA/600/4-91/003. 341 p.

USEPA. 1994b. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to freshwater organisms. 3rd ed. Lewis, P.A., Klemm, D.J., Lazorchak, J.M., Norberg-King, T.J., Peltier, W.H. and Heber, M.A., editors. Cincinnati: U.S. Environmental Protection Agency (USEPA) Office of Research and Development. EPA/600/4-91/002. 341 p.

How can WET variability be quantified?

Intratest Variability

Intratest variability is the variability of the responses (survival, growth, or reproduction), both among and between concentrations of the test material for a given test. Hypothesis test intratest variability is derived for an individual test by pooling the variability at each concentration including the control to obtain an estimate of the random error for the test. The intratest variability is used to determine the amount of difference from the control that can be detected statistically. When adjusted for the control mean, the minimum significant difference (MSD) represents the amount of difference expressed as a percentage of the control response (MSD%). Intratest variability for the point estimate approach is also represented by an estimate of the random error for the test, the mean square error (MSE). The MSE is one component in the calculation of confidence intervals for a point estimate, thus the width of a 95 percent confidence interval provides an indication of the magnitude of the intratest variability.

The intratest variability is the foremost single measure used to indicate the statistical sensitivity of a WET test analyzed with the hypothesis test approach. Statistical sensitivity, in this case, equates to a test's ability to distinguish a difference between an exposure concentration and the control. Controlling or reducing the amount of variability within a single test will increase the power of the test and therefore the ability of the test to detect responses that differ from the control response (decrease MSD). Increased power will also increase certainty in the determination of a difference from controls, which is important to regulators and the regulated community. However, minimal variability in all treatments of a test may lead

to such high statistical power that detected differences may not be biologically significant. Such tests should be interpreted with caution. Although there is no specific guidance from the USEPA on statistical versus biological significance, various States and USEPA Regions have developed some guidelines (e.g., see SETAC FAQ on addressing variability). Close attention to the factors described under the FAQ on factors affecting variability will tend to decrease heterogeneity among replicates and decrease intratest variability. In addition, increasing the number of replicates will also lead to an increase in the sensitivity of the test by decreasing the MSD.

Intratest variability is also important in representing the uncertainty associated with point estimates of toxicity. As the 95 percent confidence intervals of the point estimate increases, the uncertainty in that estimate of the statistical endpoint increases. The confidence intervals for chronic endpoints are directly influenced by the variability of response between replicates in each treatment and the model used to interpolate the point estimate. The confidence intervals for acute test results using a point estimate approach, however, are not influenced by variability between replicates but by the characteristics of the dose-response relationship. As discussed before, the certainty in point estimates is also a function of the dilutions tested and their proximity to the actual statistical endpoint being calculated. One will get a better estimate of the LC50 (tighter confidence intervals) if dilutions are tested near the concentration which actually results in 50 percent mortality.

Evaluation of a number of existing data sets by members of the Pellston workgroup (Sessions 3 and 4) (Grothe, et al, 1996) seemed to indicate that, for most WET test methods, MSDs of <40 percent were achievable. MSD's for most methods examined ranged from 18 percent to 40 percent. The consensus of the workgroup is that an additional study is necessary to determine the acceptable level of intratest variability for each USEPA recommended toxicity method, although some participants proposed that sufficient data exists to select MSD criteria. In the proposed study, data would be used to establish variability limits from laboratories that document data quality and adhere to USEPA method guidelines. Study data from each assay evaluation would include expected CVs, MSD, MSD%, MSE, and American Society for Testing and Materials (ASTM, 1992) "h" and "k" statistics. The "h" statistic represents a measure of the reproducibility between laboratories while the "k" statistic represents the repeatability within laboratories. Distributions of these values would be examined to determine criterion levels for intratest variability, and probabilities of laboratories exceeding the criterion levels would be calculated. The direct advantages of an acceptability criterion for intratest variability are 1) establishing a minimum protection level, 2) setting the power of a test to detect a toxic sample for each method, and 3) decreasing intra- and interlaboratory variability. Acceptability criteria will also allow users of WET data to better evaluate test acceptability, laboratory performance, and program effectiveness.

Intertest and Interlaboratory Variability

The scientific community familiar with analytical procedures, not just WET, recognizes that tests performed on presumably identical materials in presumably identical circumstances do not typically yield identical results. An indication of a test method's consistency is its repeatability and its reproducibility with repeatability defined as the variability between independent test results obtained from the same laboratory in a short period of time and reproducibility defined as the variability between test results obtained from different laboratories.

Several measures of repeatability and reproducibility have been proposed. The simplest of these is the intra- and interlaboratory CV (standard deviation (s) of repeated test results, divided by the mean (m) of the repeated test results, multiplied by 100 (CV = (s/m) x 100). The intralaboratory CV is generated by test results from repeated tests performed in the same laboratory, while the interlaboratory CV is obtained from test results from several different laboratories. The use of the CV removes from consideration the units of the measurement and allows the analyst to compare variability of different types of test methods (i.e., WET tests with analytical chemistry tests). It also allows analysts to compare tests that use different scales of measurement.

However, CVs alone cannot be used as diagnostic tools to help identify unusual test values or outliers. Since the CV is a function of the standard deviation of a set of test results, the measure suffers from the same problems associated with standard deviations, and there is no common agreement on what is an acceptable standard deviation. For instance, the range of test values is an easier descriptive statistic to understand. In addition, the value of the standard deviation is affected by extreme values in the data set; single large or small test values inflate the standard deviation. The CV also ignores the 95 percent confidence intervals (uncertainty) associated with each point estimate and can only be calculated for point estimates. CVs are not appropriate for hypothesis test endpoint comparisons since the effect levels are fixed by the choice of test concentrations.

Quality Management Considerations. Reference toxicant tests are typically used to monitor a laboratory's performance. Charting the performance of a laboratory's controls relative to its reference toxicant test results is a good way to track the laboratory's performance and to identify when the laboratory's performance is not acceptable. The width of a control chart's limits is an indication of a laboratory's capability to reproduce the desired endpoints of a reference toxicant test. However, control chart limits are a function of the reference toxicant, test species, test type (acute or chronic) and biological endpoint (survival, growth, etc.). These factors must be considered before drawing conclusions regarding laboratory performance. Performance on reference toxicant tests as recorded by control charts should be a criterion that is used by permittees in selecting which laboratories to use for WET tests.

Laboratories with very wide control limits, and/or many points outside of the control limits, should investigate problems related to the quality of the data being produced. Laboratories should monitor at a minimum, using control charts, the calculated endpoints for each test type/species combination. Laboratories can also monitor the control treatment mean response for survival, growth, and reproduction. In addition, laboratories can chart the control treatment replicate variance, or standard deviation. Reference toxicant tests are very important to track analyst technique and the health and condition of the test organisms. It is particularly important when performing these tests (as with all compliance toxicity tests) that the analysts precisely follow the published test methods, without deviation between tests.

ASTM-American Society for Testing and Materials. 1992. Standard practice for conducting an interlaboratory study to determine precision of a test method, E691-92. In: *Annual Book of ASTM Standards*, Vol. 14.02. Philadelphia, PA.

Grothe, D. R., K. L. Dickson, and D. K. Reed-Judkins, eds.1996. Whole Effluent Toxicity Testing: An Evaluation of Methods and Prediction of Receiving System Impacts, SETAC Press, Pensacola, FL, USA. 340 p.

APPENDIX E

EXAMPLES OF SELECTED STATE WET IMPLEMENTATION PROGRAMS

This page intentionally left blank

EXAMPLES OF SELECTED STATE WET IMPLEMENTATION PROGRAMS

Appendix E contains summaries of approaches that States have taken in implementing their NPDES whole effluent toxicity (WET) programs and efforts instituted to reduce or ensure minimal test variability when conducting WET tests. Preceding the State responses is a matrix (Table E-1) that briefly summarizes the common approaches or program themes for the States that responded. The respondent States are a geographic sampling across the United States. EPA's inclusion of the various State approaches in this document is not an endorsement of their approaches, but a snapshot of additional steps that a permitting authority could consider taking beyond the minimum requirements (i.e., test acceptability criteria) outlined in EPA guidance. This sample of State approaches also responds to recommendations EPA received on the initial draft document to consider and provide reference to other State approaches.[1]

[1] Note that the terms, abbreviations, and acronyms used in this appendix may differ from their usage throughout the rest of this document. EPA consciously chose not to edit the State-supplied information so that the actual States' nomenclature and terminology as used in their NPDES programs would be reflected here.

Table E-1. Overview of Selected State WET Implementation Programs

ST	How do you evaluate reference toxicant & effluent test results?	How do you review reference toxicant test data for laboratory performance?	Describe additional QA/QC criteria developed & implemented.
KY	Acute—point-estimation Chronic—linear interpolation	Labs submit annual summary of RTT data, used to determine consistency & conformance with expected values.	1. Monthly acute/chronic RTT within 30 days of each WET test. 2. RTT conducted on each batch of purchased test organisms unless supplier provides information. 3. Culturing & testing in different incubators. 4. Chronic toxicity tests with CV > 40% evaluated on case-by-case basis.
NJ	Acute—point-estimation Chronic—linear interpolation	RTT results reported on standardized form, with UCL & LCL. Control charts submitted annually. RTT data reviewed in on-site audit.	1. C. *dubia* test: Number of males in surviving organisms over all concentrations $\leq$10%; number of males in controls $\leq$20%. 2. For all tests, no sporadic moralities present; $\leq$10% variation per concentration in start count. 3. For tests indicating permit violation, review raw data & test results (data trend, MSD, chain-of-custody, sample handling/holding time).
NC	Acute—point-estimation Chronic—linear interpolation Acute pass/fail, chronic pass/fail, & chronic multi-concentration effluent tests—hypothesis tests	RTT data reviewed during annual lab inspection. Lab provides bench sheets, water quality data, calculations, control charts, etc. Information reviewed for test frequency, test conditions, test result validity, & responses to out-of-control events.	1. Dilution water pH 6.5 -8.5, total hardness 30 - 50 ppm. 2. Biweekly acute RTT or within 7 days of any NPDES test. 3.Test organisms identified to species once/quarter. 4. Culturing & testing in different incubators. 5.Chronic *C. dubia* test: 3rd brood neonate production $\geq$ 80% of control; neonate reproduction from 1st - 3rd broods only; % male control organisms $\leq$ 20%; control reproduction CV $\leq$ 40%; solution DO $\geq$ 5.0 mg/l; exposure duration at least 7 d ± 2 h. 6. Acute tests terminated w/i 1 h of stated length.
WA	Acute—point estimation Chronic pass/fail, & chronic multi-concentration effluent tests—hypothesis tests	1. Review data in conjunction with effluent tests. 2. Lab provides bench sheets, water quality data, calculations, control charts, etc. Information reviewed for test frequency, test conditions, test result validity, & responses to out-of-control events. 3. If reference CV does not meet certain criteria, test is rejected.	1. Minimum % difference in survival between IWC & control (or NOEC for chronic) that is statistically significant: acute–30%, chronic–40%. 2. Tests failing must be repeated with more replicates. 3. Specific requirements for *Ceriodaphnia* & bivalve chronic tests.
WI	Acute & chronic—point-estimation Certified labs perform monthly reference toxicant tests	RTT data reviewed by auditor before on-site lab inspections. Lab provides bench sheets, water quality data, control charts, etc. Information reviewed for test frequency, proper test conditions, test result validity, & proper responses to out-of-control events.	1. Testing & culturing in separate rooms or incubators 2. Additional or revised TAC for Static Renewal Acute & Chronic Tests will be included in revised methods.

Abbreviations: CV = coefficient of variation; DO = dissolved oxygen; IWC = instream waste concentration; LCL = lower control limit; MSD = minimum significant difference; NOEC = no observed effect concentration; NPDES = National Pollutant Discharge Elimination System; ppm = parts per million; RTT = reference toxicity test; SETAC = Society of Environmental Toxicology and Chemistry; SOP = standard operating procedures; TAC = technical acceptability criteria; UCL = upper control limit; WET = whole effluent toxicity

Table E-1. Overview of Selected State WET Implementation Programs (continued)

ST	Describe efforts to minimize test method variability.	Explain how you review or conduct lab performance audits.	Describe specific implementation guidance developed for permit writers. How is the guidance available to the public?	Describe how you provide or use toxicity test training.
KY	1. Reporting on standardized form. 2. Labs submit culturing/testing SOP for State review. 3. Tests must comply with all EPA & State test manuals. 4. Dilution water moderately hard-reconstituted water or dilute mineral water. 5. Follow written protocol for splits. 6. Lab audits by State or EPA Region.	EPA Region or State conducts lab audits, following procedures in EPA inspection manual.	1. Several guidance documents developed by State. 2. Face-to-face training as needed, also available to the public. 3. Some documents available on web or through newsletters.	1. State communicates program changes & guidance on culturing & testing issues through newsletter & web page. 2. Training sessions for State personnel. 3. Participate in Wastewater Operator's Conference to discuss issues with regulated community & consultants. 4.Teach in SETAC's WET training & statistical analysis courses.
NJ	1. Standardized checklist for screening submitted tests. 2. Lab certification program, on-site audits. 3. Labs report premature cancellation of test & reason. 4. Quarterly meetings of State/lab representatives to discuss current test developments.	1. Inspections announced or unannounced. Lab SOPs reviewed for adherence to NJ & EPA protocols. Subsets of data reviewed, technician summarizes problems with test reports. 2. Inspections consist of opening conference, lab walk-through, closing conference. SOP review, problematic test results review. Auditor examines equipment, documentation, cultures, lab procedures, chain-of-custody, sample handing. Review of inspection results during closing conference.	1. Training sessions to permit writers & public. 2. Written guidance is copies of past training sessions on shared drive for permit writers. Generally not available to public.	Staff attend USEPA or SETAC-sponsored training.

Abbreviations: CV = coefficient of variation; DO = dissolved oxygen; IWC = instream waste concentration; LCL = lower control limit; MSD = minimum significant difference; NOEC = no observed effect concentration; NPDES = National Pollutant Discharge Elimination System; ppm = parts per million; RTT = reference toxicity test; SETAC = Society of Environmental Toxicology and Chemistry; SOP = standard operating procedures; TAC = technical acceptability criteria; UCL = upper control limit; WET = whole effluent toxicity

Table E-1. Overview of Selected State WET Implementation Programs (continued)

ST	Describe efforts to minimize test method variability.	Explain how you review or conduct lab performance audits.	Describe specific implementation guidance developed for permit writers. How is the guidance available to the public?	Describe how you provide or use toxicity test training.
NC	1. Review submitted test results against TAC. 2. Implement lab certification program. 3. Document investigations of differing test results from splits of effluent samples. 4. Test protocol modifications.	1. Inspections announced or unannounced. Lab SOPs reviewed for adherence to NJ & EPA protocols. Subsets of data reviewed, technician summarizes problems with test reports. 2. Inspections consist of opening conference, lab walk-through, closing conference. SOP review, problematic test results review. Auditor examines equipment, documentation, cultures, lab procedures, chain-of-custody, sample handing. Review of inspection results during closing conference.	1. Written guidance established by memo. Face-to-face training sessions as needed. 2. Written guidance available to public upon request, also sent to permit holders with permit & subsequent renewals. Also available on the web.	1. Participate in aquatic toxicologists group. Communicate program changes & guidance on culturing & testing issues through meetings. 2. Workshops held for Division's regional office personnel 3. Attend SETAC's WET training & statistical analysis courses.
WA	1. Develop, distribute *Laboratory Guidance and Whole Effluent Toxicity Test Review Criteria*. 2. Review tests for compliance with method & canary book. 3. Fish/mysid growth tests with SD of proportion alive > 0.25 in effluent concentration analyzed for original growth endpoint, not combined endpoint. 4. Permit requirements will lower alpha level for hypothesis testing when differences in test organism response are small. 5. Anomalous test identification criteria established to make WET test results fair & enforceable.	1. Inspections announced or unannounced. Lab SOPs reviewed for adherence to NJ & EPA protocols. Subsets of data reviewed, technician summarizes problems with test reports. 2. Inspections consist of opening conference, lab walk-through, closing conference. SOP review, problematic test results review. Auditor examines equipment, documentation, cultures, lab procedures, chain-of-custody, sample handing. Review of inspection results during closing conference. 3. Audit report prepared within 30 days of audit 4. Performance audits required twice/year, system audits every three years.	Develop & update language for use in NPDES permits & fact sheets for POTWs & industry. Language is part of templates for permits & fact sheets that permit writers use as they draft permits. Manual available to the public for cost of printing & also on Web.	Extensive training in all offices early in 1990s. WET test review & technical assistance are centralized functions, permit writing guidance available in *Permit Writer's Manual* & suggested permit language.

Abbreviations: CV = coefficient of variation; DO = dissolved oxygen; IWC = instream waste concentration; LCL = lower control limit; MSD = minimum significant difference; NOEC = no observed effect concentration; NPDES = National Pollutant Discharge Elimination System; ppm = parts per million; RTT = reference toxicity test; SETAC = Society of Environmental Toxicology and Chemistry; SOP = standard operating procedures; TAC = technical acceptability criteria; UCL = upper control limit; WET = whole effluent toxicity

Table E-1. Overview of Selected State WET Implementation Programs (continued)

ST	Describe efforts to minimize test method variability.	Explain how you review or conduct lab performance audits.	Describe specific implementation guidance developed for permit writers. How is the guidance available to the public?	Describe how you provide or use toxicity test training.
WI	1. Review submitted test results against TAC. 2. Lab certification program 3. Document investigations of differing test results from splits of effluent samples. 4. Strict adherence to clearly specified methods, such as sampling procedures, holding times, test duration. 5. Revising methods to require that labs verify staff training & qualifications.	1. Inspections announced or unannounced. Auditor reviews laboratory SOPs & recent RTT results for adherence to WDNR protocols. 2. Inspections consist of opening conference, lab walk-through, closing conference. SOP review, problematic test results review. Auditor examines equipment, documentation, cultures, lab procedures, chain-of-custody, sample handing. Review of inspection results during closing conference. 3. Auditor reviews reference toxicant data after inspection, generates inspection letter. Lab has 60 days to respond. Significant deficiencies may result in decertification.	Written guidance & clarification on existing rules & methods for State staff, permittees, labs, consultants, others.	1. One-on-one training for State staff & permittees. 2. University lab provides hands-on WET training to State staff, permittees, labs on request. 3. Attend SETAC's WET training & statistical analysis courses.

Abbreviations: **CV** = coefficient of variation; **DO** = dissolved oxygen; **IWC** = instream waste concentration; **LCL** = lower control limit; **MSD** = minimum significant difference; **NOEC** = no observed effect concentration; **NPDES** = National Pollutant Discharge Elimination System; **ppm** = parts per million; **RTT** = reference toxicity test; **SETAC** = Society of Environmental Toxicology and Chemistry; **SOP** = standard operating procedures; **TAC** = technical acceptability criteria; **UCL** = upper control limit; **WET** = whole effluent toxicity

E.1 RESPONSES FROM KENTUCKY DEPARTMENT FOR ENVIRONMENTAL PROTECTION

E.1.1 Describe How Your State Evaluates Reference Toxicant and Effluent Test Results

Acute reference toxicant test and multi-concentration effluent test results are evaluated using the point-estimate (LC50) technique described in the EPA acute testing manual.

Chronic reference toxicant and multi-concentration effluent test results are evaluated using the linear interpolation method (IC25) as described in the EPA chronic manual and using the TOXCALC statistical program software.

E.1.2 Explain How Your State Reviews Reference Toxicant Data for Laboratory Performance

Consulting laboratories that service permittees are required to annually submit to the Bioassay Section a summary of their reference toxicant test data. This information is used to determine consistency and conformance to the expected values. This serves as a review and audit of all consulting laboratories, measures consistency within a laboratory, and provides a level of reliability and accuracy between laboratories.

A letter of request is sent to each laboratory with a standardized response form. The labs provide the requested information, including test date, dilution series, type of control water, organism age, LC50/IC25, 95 percent confidence interval, and average control reproduction/weight. This information is entered into a laboratory QA data base where it is statistically analyzed.

This information is then compiled into an annual summary report. The compiled information includes the lab name, reference toxicant, test species, test type, test duration, number of tests performed, mean, standard deviation (SD), % coefficient of variation (CV), average reproduction, or growth with SD and % CV.

The results are mailed to each participating laboratory. In addition, the summary results are printed in the Kentucky Biomonitoring Newsletter and are presented on the Bioassay Section's web page (http://water.nr.state.ky.us/wq/bioassay/index.html).

A control chart is prepared for each reference toxicant and organism combination, and successive toxicity values are plotted and examined to determine if the results are within prescribed limits. A minimum of 30 test results are needed for a reliable mean and upper/lower control chart. If the toxicity value from a given test with the reference toxicant does not fall within the expected range for the test organism when using the standard dilution water, then the sensitivity of the organisms and the overall credibility of the test systems are suspect. In this case the test procedure, control water, and reference toxicant are examined.

Missing and/or out-of-range data must be explained and can result in the invalidation of Kentucky Pollution Discharge Elimination System (KPDES) WET test results.

E.1.3 Describe Any Additional QA/QC Criteria Your State Has Developed and Implemented Within Your State

1. Acute and chronic reference toxicant tests are to be conducted monthly. A reference toxicant test must be conducted within 30 days of each KPDES WET test.

2. If test organisms are purchased from a commercial supplier, a reference toxicant test must be conducted on each batch unless the supplier can provide this information.

3. Culturing and testing activities may not be contained within the same incubator.

4. Chronic toxicity tests where the coefficient of variation (CV) is greater than 40 percent will be evaluated on a case-by-case basis to determine if the results will be considered acceptable.

5. All other QA/QC criteria for culturing and testing, as set forth in the most current editions of the EPA manuals, must be followed.

E.1.4 Describe Any Efforts Your State Has Made to Minimize Test Method Variability

1. All KPDES WET test results are submitted using a standardized report form. Each report is closely reviewed by a member of the Bioassay Section to determine if proper test protocols have been followed.

2. Prior to conducting toxicity test for Kentucky permittees, each laboratory must submit its culturing/testing SOP for review by the Bioassay Section. This insures that proper methods and procedures are being followed.

3. Toxicity tests must comply with all conditions as stated in the EPA testing manuals and in the Kentucky Methods for Culturing and Conducting Toxicity Tests with *Pimephales promelas* and *Ceriodaphnia dubia*. (Fourth Edition, 1996). Special attention is paid to sample holding times and temperatures.

4. Dilution water is to be moderately hard-reconstituted water or moderately hard dilute mineral water.

5. If split samples are going to be used, the Biomonitoring Split-Sample Protocol must be followed. This protocol details sample collection and holding procedures as well as test conditions that must be followed.

6. Laboratories must submit all reference toxicant data for the annual summary. This information assists in determining the quality of information being received from these facilities.

7. Laboratories are audited by Kentucky or EPA Region IV to review testing and culturing procedures.

E.1.5 Explain How Your State Reviews or Conducts Performance Lab Audits

Kentucky has been fortunate in having the expertise of EPA Region IV in performing WET laboratory audits. Their experience has proven beneficial in keeping laboratories compliant with the testing requirements. When the services of EPA are not available, the State will conduct its own lab audits. In either case, the procedures are the same and follow those outlined in the EPA inspection manual.

Inspections are usually announced. If EPA is performing the inspection, a representative from the Bioassay Section will accompany the inspectors. Prior to the inspection, the auditor will review the laboratory's SOP for adherence to Kentucky and EPA protocols. Bioassay Section staff will review test reports to document any problems with the subject lab. In addition, the qualifications of the staff will be reviewed at this time. Generally, three test reports will be chosen for which the laboratory will be required to produce supporting documentation.

The inspection consists of an opening conference, a walk-through of the laboratory, and a closing conference. During the opening conference, the auditor discusses the SOP review and general procedures in the laboratory. In addition, information including culturing records, test data, chain of custody records, reference toxicant data, etc., supporting the three test reports selected prior to the inspection will be reviewed. During the walk-through, the auditor examines equipment, log books, written documentation and laboratory procedures. The closing conference serves as a review of observations and comments during the inspection.

The auditor will generate an inspection response letter detailing any deficiencies noted during the audit. All correspondence is addressed to the permittee, whose test results were used for the inspection. The permittee will have usually 60 days to respond to the deficiencies, noting what actions have been taken by the laboratory to correct them. If significant deficiencies are not addressed, then future data from this laboratory may not be accepted by the State.

E.1.6 Describe Any Specific Implementation Guidance That Your State Has Developed to Assist Permit Writers. How Is the Guidance Available to the Public?

Guidance is provided through several documents developed by the Bioassay Section. This section has developed standardized biomonitoring language, which is provided to the KPDES Permitting Branch. This language is incorporated into each permit with a WET limit or monitoring upon permit issuance or reissuance. In addition, a Standard Test Result Report form is provided to each permit holder with WET. The section has another document: Aquatic Toxicity Testing: Questions and Answers, which is available upon request.

The Bioassay Section provides face-to-face training to the KPDES Branch on an as-needed basis. This training is also available to the public if requested.

Some documents are available on the Bioassay Section's web page or through the Biomonitoring newsletter.

E.1.7 Describe How Your State Provides or Utilizes Any Toxicity Testing Training

The Bioassay Section communicates program changes and specific guidance on culturing and testing issues through the newsletter and the web page. The section has held several training sessions for State personnel since the inception of the program. In addition, the section participates in the State's annual Wastewater Operator's Conference to discuss issues with the regulated community and consultants.

Section members have attended and participated as instructors in the Society for Environmental Toxicology and Chemistry's two-day WET training course and statistical analysis course.

E.2 RESPONSES FROM NEW JERSEY DEPARTMENT OF ENVIRONMENTAL PROTECTION

E.2.1 Describe How Your State Evaluates Reference Toxicant and Effluent Test Results

Acute effluent tests are evaluated using the point estimate techniques described in the USEPA acute methods document. New Jersey also uses the NOAEC endpoint set equal to 100 percent effluent when an evaluation of no acute toxicity is required. The hypothesis testing techniques contained in the USEPA manual are used in that case.

Requests have been received from certified laboratories and from permittees that the point estimate techniques be further standardized. Using one version of Probit versus another can result in a different value, sometimes making a difference whether a facility passes or fails.

Chronic effluent and reference toxicant test results are evaluated using the linear interpolation method originally provided by Teresa Norberg King (July 1993). A p value of 25 is selected for all permits and for reference toxicant recording.

E.2.2 Explain How Your State Reviews Reference Toxicant Data For Laboratory Performance

New Jersey Pollution Discharge Elimination System (NJPDES) permits require that in order for chronic toxicity test results to be considered acceptable, there must be an acceptable Standard Reference Toxicant (SRT) result conducted within 30 days of the compliance test result, for the test species and reference toxicant in question. The States standardized report form requires the reporting of the applicable SRT result directly on the compliance test report, along with the applicable upper and lower control limits. Missing or out of range data can result in the invalidation of test results.

Control charts are forwarded to the Department on an annual basis, on the anniversary of the approval for the test species. Many labs have chosen to include copies of applicable control charts with the submittal of compliance test results. SRT data is also reviewed as part of an on-site audit, including a review of procedures, raw data, and data analysis any excluded results.

State methods governing laboratories also require that if a lab produces any SRT test result which is outside the established upper and lower control limits for a test species at a frequency greater than one test in any ten tests, a report shall be forwarded to the Department. That report shall include any identified problem which caused the values to fall outside the expected range and the corresponding actions that have been taken by the laboratory. If a laboratory produces two consecutive SRT test results or three out of any ten test results, which are outside the established upper and lower control limits for a specific test species, the laboratory shall be unapproved to conduct testing. Reapproval is contingent upon the laboratory producing SRT test results within the established upper and lower limits.

The laboratory selects the reference toxicant used. However, the Department recommends using KCl.

E.2.3 Describe Any Additional QA/QC Criteria Your State Has Developed and Implemented With Your State

For Ceriodaphnia testing:

- Number of males in surviving organisms overall concentration $\leq$10 percent [(no. males / total no. surv) x 100].
- Number of males in controls $\leq$20 percent (no. males / total no. organisms in controls).

All test species

- No sporadic mortalities present (Deaths that are not related to sample toxicity, confined to a few test chambers and scattered throughout the test).
- Variation in start count must be $\leq$10 percent per concentration (animals lost or killed by accident).

These items are specifically included on standardized review sheets.

For any tests that would result in the collection of penalties based on violation of an effective toxicity limit, a detailed review of the raw data and test results are conducted, including review of the data trend, minimum significant difference, chain-of-custody, sampling handling, and holding times.

E.2.4 Describe Any Efforts Your State Has Made To Minimize Test Method Variability

Each test that is submitted receives at least a screening using a standardized check list, anywhere from 30 to 40 questions depending upon the test species, dealing will all aspects of the test.

New Jersey maintains a laboratory certification program for toxicity testing, including on-site audits.

A laboratory who cancels a test prior to the scheduled ending time/date must report that cancelled test, including the reason for the cancellation, to the Department. This allows the Department to track a laboratory's ability to run a test to completion. Tests that do not meet USEPA's test acceptability criteria are not submitted to the Department since they are not valid. This way the frequency that this is occurring at a laboratory can be tracked. Frequent test cancellations are addressed during an on-site audit.

New Jersey has a Bioassay Subcommittee that is a subset of the State's Laboratory Advisory Committee. This committee meets quarterly and consists of State and laboratory representatives. The committee discusses problems with the tests, certification, updates from USEPA, SETAC, NELAC, or anything else applicable to toxicity testing. This gives the laboratories and the State an opportunity to discuss either deficiencies that are occurring at laboratories and are showing up in the test data, problems the laboratories are having with regard to any of the methods, and any improvements to the program that should be easily implemented.

E.2.5 Explain How Your State Reviews Or Conducts Performance Lab Audits

Inspections can be announced or unannounced, although generally time is not adequate to perform unannounced inspections. Prior to the inspection, the auditor will review the laboratory's SOPs for adherence to New Jersey and EPA protocols. Subsets of data will also be reviewed and the technician responsible for day to day screening using the standardized check list is asked to summarize any problems with the review of toxicity test reports.

The actual inspections consist of an opening conference, a walk-through of the lab facility, and a closing conference. During the opening conference, the auditor discusses the SOP review and general procedures in the lab. In addition she will request and review-supporting information associated with the any test reports identified prior to the inspection as a concern. During the walk-through, the auditor examines equipment, written documentation, cultures, laboratory procedures, chain-of-custody, and sample handing. The closing conference serves as a review of observations and comments during the inspection.

E.2.6 Describe Any Specific Implementation Guidance That Your State Has Developed To Assist Permit Writers. How Is The Guidance Available To The Public?

The Office of Quality assurance provides training sessions to the permit writer and the public upon request. Written guidance consists of copies of past training sessions, located on the share drive for permit writers. This guidance is not generally available to the public.

E.2.7 Describe How Your State Provides Or Utilizes Any Toxicity Testing Training

When possible, staff will attend any USEPA- or SETAC-sponsored training on the topic.

E.3 RESPONSES FROM NORTH CAROLINA DEPARTMENT OF ENVIRONMENT AND NATURAL RESOURCES

E.3.1 Describe How Your State Evaluates Reference Toxicant and Effluent Test Results

Acute reference toxicant test and multi-concentration effluent test results are evaluated using the point-estimation techniques described in the EPA manual.

Acute pass/fail, chronic pass/fail, and chronic multi-concentration effluent test results are evaluated using hypothesis tests as described in the EPA manuals.

Chronic reference toxicant test results are evaluated using the linear interpolation method (ICp, where p=25) described in the EPA manual.

For both types of chronic *Ceriodaphnia* effluent tests, a reproductive effect is defined by both a statistically significant difference between the treatment and the control and a 20 percent reduction in neonate reproduction of the treatment organisms as compared to the controls. Hypothesis tests for both acute and chronic pass/fail tests are performed at an alpha level of 0.01.

E.3.2 Explain How Your State Reviews Reference Toxicant Data for Laboratory Performance

The data is reviewed in conjunction with the laboratory's annual laboratory inspection. The laboratory provides copies of bench sheets, water quality data, and calculations or printouts from the data analysis for each reference toxicant test performed since the last laboratory inspection:

In addition, the lab submits the current control chart (with data listing) and any explanations of out-of-range test results for each test type and organism combination.

The materials are reviewed for appropriate test frequency, proper test conditions, test result validity, and proper responses to out-of-range events.

Missing or out-of-range data can result in the invalidation of NPDES test results.

E.3.3 Describe Any Additional QA/QC Criteria Your State Has Developed and Implemented Within Your State

— Laboratories must use dilution water in whole effluent toxicity testing with chemical characteristics such that the pH is between 6.5 and 8.5 and total hardness as calcium carbonate is between 30 and 50 µg/l as calcium carbonate.

— Acute and chronic reference toxicant tests must be performed once every two weeks or within one week of any NPDES tests.

— A representative of each test organism cultured shall be taxonomically identified to the species level at a minimum frequency of once per quarter.

— If closed incubators (refrigerator-sized) are utilized for toxicity testing and/or test organism culturing purposes, culturing and testing activities may not be contained within the same incubator.

— Chronic *Ceriodaphnia dubia* analyses will have an additional test acceptability criterion of complete third brood neonate production by at least 80 percent of the control organisms.

— *Ceriodaphnia dubia* neonate reproduction totals from chronic tests shall include only organisms produced in the first through third broods.

— The percentage of male *Ceriodaphnia* control organisms may not exceed 20 percent in chronic *Ceriodaphnia* tests.

— The *Ceriodaphnia* control organism reproduction coefficient of variation (CV) must be less than 40 percent for a chronic *Ceriodaphnia* test to be considered acceptable.

— *Ceriodaphnia* chronic test solutions must maintain dissolved oxygen levels greater than or equal to 5.0 mg/l.

— *Ceriodaphnia* chronic test exposure duration will be no greater than seven days ± 2 hours regardless of control organism reproductive success.

— Acute tests will be terminated within one hour of their stated length.

E.3.4 Describe Any Efforts Your State Has Made to Minimize Test Method Variability

1. Close review of each test result submitted with consistent adherence to test protocol test acceptability criteria.

2. Implementation of a biological laboratory certification program.

3. Paper trail investigations of test results from disagreeing "split" effluent sample analyses.

4. Test protocol modifications.

EPA methods allow for a relatively wide window for termination of the chronic *Ceriodaphnia* test. Tests may be terminated as soon as 60 percent of the control organisms produce three broods of young or as late as eight days after test initiation. Logically, narrowing the termination window will reduce variability and improve precision of test results. The North Carolina Division of Water Quality (NC DWQ) has narrowed the window available for the termination of the chronic *Ceriodaphnia* test by:

— Placing a shorter limit on the exposure period (seven days + two hours)

— Requiring that at least 80 percent of the control organisms produce a third brood prior to test termination

Analysis of a data base of NC chronic *Ceriodaphnia* test results has shown that reducing control organism reproduction variability improves the sensitivity of the reproduction analysis. Logically, holding all labs to a common precision standard with respect to control organism reproduction should reduce between-lab test result variability. The Division has reduced variability of control organism reproduction by:

— Implementing a test acceptability criterion limiting the control organism reproduction coefficient of variation to less than 40 percent

— Requiring that at least 80 percent of the control organisms produce a third brood prior to test termination

— Excluding fourth and subsequent brood neonates from the reproduction effects analysis

DWQ's experience has shown that high quality laboratories can produce extremely sensitive tests that can detect quite small differences between treatment and control reproduction. Unfortunately, this can be a disincentive for laboratories to produce high quality tests, since experience has shown that some clients gravitate toward laboratories that produce compliant test results. Less sensitive tests will be more likely to produce compliant results. Analysis of reproduction data from the same data base described above indicated that tests performed by NC certified labs could routinely detect a difference between the control and a treatment when there was a 20 percent reduction in neonate reproduction by the treatment organisms compared to the controls. Based on this data, NC DWQ has placed a second data evaluation criterion on the *Ceriodaphnia* chronic reproduction analysis. Specifically, for an effluent treatment to be considered producing an effect, the reproduction mean must be both statistically significantly lower than the control mean **and** represent at least a 20 percent reduction from that mean. In effect, this sets a lower limit on test sensitivity and also reduces within-laboratory and between-laboratory test result variability.

E.3.5 Explain How Your State Reviews or Conducts Performance Lab Audits

Inspections may be announced or unannounced. Prior to the inspection, the auditor will review the laboratory's SOP for adherence to North Carolina and EPA protocols. The Aquatic Toxicology Unit member responsible for reviewing test report submittals will be requested to summarize any recurring problems with the target laboratory regarding data submission. Three test reports will be chosen for which laboratory personnel will be asked to produce supporting documentation.

The actual inspection consists of an opening conference, a walk-through of the laboratory facilities, and a closing conference. During the opening conference, the auditor discusses the SOP review and general procedures in the laboratory. In addition he/she will request and review supporting information associated with the three test reports selected prior to the inspection. During the walk-through, the auditor examines equipment, written documentation, cultures, and laboratory procedures. The closing conference serves as a review of observations and comments during the inspection.

The auditor will review reference toxicant data (see question 2 above) after the inspection. Within two weeks, the auditor will generate an inspection response letter, to which the laboratory will be given 60 days to respond. If there are significant deficiencies discovered during the inspection, a laboratory or categorical decertification may occur.

E.3.6 Describe Any Specific Implementation Guidance That Your State Has Developed to Assist Permit Writers. How Is the Guidance Available to the Public?

Written guidance is established by memo from the Water Quality Section Chief to the NPDES Permitting Unit and other affected Water Quality Section Units. The Aquatic Toxicology Unit provides face-to-face training sessions to the NPDES Unit on an as-needed basis.

The written guidance in memo form is available to the public upon request. Parts of the guidance are included in a document called "Aquatic Toxicity Testing: Understanding and Implementing Your Testing Requirement," that is disseminated to each permit holder with a WET limit or monitoring requirement upon permit issuance and subsequent renewals. The document is also available at the Aquatic Toxicology Unit web page, http://www.esb.enr.state.nc.us/ATUwww.default.html.

E.3.7 Describe How Your State Provides or Utilizes Any Toxicity Testing Training

NC DWQ actively participates in the Carolinas Area Aquatic Toxicologists group (CAAT). The Aquatic Toxicology Unit utilizes the meetings of this group to communicate program changes and specific guidance on culturing and testing issues. Additionally, the Unit has held two workshops for the Division's regional office personnel since the inception of the aquatic toxicity testing program. Unit members have

attended The Society of Environmental Toxicology and Chemistry's two-day WET course and statistical analysis course.

E.4 RESPONSES FROM WASHINGTON DEPARTMENT OF ECOLOGY

E.4.1 Describe How Your State Evaluates Reference Toxicant and Effluent Test Results

The State of Washington Department of Ecology reviews every WET test report for compliance with the test method and instructions in the permit. Permit instructions include reference to a document called "Laboratory Guidance and Whole Effluent Toxicity Test Review Criteria" that provides the lab with standard testing instructions and provides the basis for test report review. Reference toxicant tests are not evaluated separately but are evaluated as a part of the review of WET test reports. The Department of Ecology also maintains a data base of WET test raw data and statistical results in order to have comprehensive records for each discharger and to enhance our ability to learn from experience and improve our WET program.

E.4.2 Explain How Your State Reviews Reference Toxicant Data for Laboratory Performance

The minimum reference toxicant testing needed to meet our interpretation of the requirements in the EPA manuals (both sections 4.7 and 4.16) is one per month for every acute and 7-day (short-term) chronic test species used routinely (more than once per month). Because an acute test result can be determined during a 7-day chronic test, acute and chronic reference toxicant testing for a fish or mysid can be combined. If a lab has difficulty establishing a concentration series that produces good results for both a lethal and sublethal endpoint, the lab may focus on lethality, as long as the sublethal endpoint is not completely abandoned in the conduct and analysis of the test.

In addition to the nonroutine tests (test performed once per month or less), all tests conducted with plants are required to have concurrent reference toxicant testing. In addition, brood stock can vary in condition, and the concurrent check on test organism sensitivity is a good precaution. Algal toxicity tests must have concurrent reference toxicant tests for similar reasons. Concurrent reference toxicant testing is also required when test organisms (or the brood stock used to produce the test organisms) have been collected from the wild. Increases in test costs, especially the cost of 7-day chronic tests, are to be avoided if possible. The alternative to concurrent reference toxicant testing in section 4.7 for labs getting test organisms from an outside supplier is reference toxicant testing by the organism supplier, and this alternative seems to be generally believed by testing labs as well as the Department of Ecology to be inferior to monthly reference toxicant testing by the testing lab. We do not accept the use by labs of reference toxicant tests performed by organism suppliers, and apparently labs agree because the vast majority have, to their credit, continued to conduct their own reference toxicant testing. Labs, however, should use organism suppliers that routinely conduct reference toxicant testing and control charting because, as noted in the table below, this information can be useful when deciding the consequences of lab conducted reference toxicant testing.

All labs must conduct ongoing control charting based on reference toxicant testing and report the results, acceptable or unacceptable, of the control charting in the report for each effluent or ambient water test. Acceptability is based on the standard test acceptability criteria for the test and on control charting with the upper and lower control limits set at twice the standard deviation (95 percent confidence) of the point estimates (LC_{50}, EC_{50}, IC_{25}, etc.) accumulated from the last 20 reference toxicant tests. At least five reference toxicant tests are needed to establish a minimally effective control chart for new tests. The reference toxicant test data must be presented with the report for each associated test.

Any reference toxicant test determined to be unacceptable must be repeated either until an acceptable result is obtained or until there have been three consecutive unacceptable test results (the initial unacceptable test plus two repeats). Because about 1/20 reference toxicant test results will fall outside of control limits

due to chance alone, it is necessary to repeat unacceptable reference toxicant tests in order to reduce the role of chance. Assuming no unusual problems with test organisms or lab performance, there is only a 1/400 chance of two unacceptable reference toxicant test results in a row and only a 1/8,000 chance of three unacceptable results in a row. If a lab has no unusual problems, repeating an unacceptable reference toxicant test should quickly produce an acceptable result. If a lab repeatedly produces unacceptable reference toxicant test results, it will give confidence to the conclusion that the lab has problems with test organisms or testing technique.

When the reference toxicant test result is within the 95 percent confidence limits, then the test report must state this fact and present the reference toxicant data at the end of the report. When the reference toxicant test result is outside the 95 percent confidence limits, then the test report must state this fact and present the reference toxicant data at the end of the report. The lab should not delay test reports while waiting for the results of reference toxicant test repeats. The results from the first repeated test might be available in time for inclusion in the test report. If begun promptly, the results of all of the reference toxicant testing in response to an unacceptable reference toxicant test result will be available in time for the review of the test report. The WET Coordinator will contact the lab during the test review for any additional reference toxicant test data not contained in the test report.

When a reference toxicant test result falls outside of the 95 percent confidence limits, a lab must qualify the associated test result for an effluent or ambient water sample by a statement in the test report that the reference toxicant test result was outside control limits. The Department of Ecology WET Coordinator will decide whether these tests are acceptable based on the degree of departure from control limits and the frequency of occurrence. Because it is expected that an average of one out of 20 tests will fall outside of the control limits due to chance alone, the degree of departure from the control limits and frequency of occurrence will be considered before rejecting toxicity tests. Because control limits narrow as laboratory performance improves, the width of the control limits will also be considered before rejecting toxicity test results when the associated reference toxicant test results are just outside the limits.

The Biomonitoring Science Advisory Board (BSAB) criteria for acceptable intralaboratory variability provide values that are useful for considering the width of control limits while deciding whether to reject toxicity tests on the basis of reference toxicant test results. If the coefficient of variation (standard deviation mean toxicity value) from the reference toxicant test data used in control charting falls into the excellent (< 0.35) or good (0.35 to 0.60) range established by the BSAB, then a higher confidence in the test results is justified. If the reference toxicant test data coefficient of variation for the lab falls into the acceptable range (0.61 to 0.85), then a smaller amount of confidence should be applied. If the reference toxicant test data coefficient of variation for the lab falls into the unacceptable range (> 0.85), then none of the lab's test results are acceptable. Labs must report the coefficient of variation for the last 20 reference toxicant tests in every report for the same test conducted on an effluent or environmental sample. (Reference: Biomonitoring Science Advisory Board. BSAB Report #1, ***Criteria for Acceptable Variability of Marine Chronic Toxicity Test Methods.*** Washington Dept. of Ecology. February 1994.) Effluent or ambient water toxicity test results will be accepted or rejected based on the following table. Rejection will occur when any condition in the appropriate "Test Accepted" box was not met or when any condition in the appropriate "Test Rejected" box was met.

Effluent tests and their associated (initial) reference toxicant tests must have start dates separated in time by no more than 18 days. Labs typically take about two weeks to produce a test report. From the point of view of practicality and the most meaningful control charting, it makes sense for a reference toxicant test result to be used retroactively about two weeks. The reference toxicant test result will then be used for control charting for the balance of the monthly time period. A grace period of 7 days will be added to the 18 days for tests begun from December 1st to the following January 10th. Acute tests will be allowed a grace period of 4 days over the 18 day maximum.

Table for Determining Test Rejection Based on Reference Toxicant Test Results

Unacceptable Reftox Tests	Test Accepted	Test Rejected
Only the original reftox test result was outside of control limits (the first repeat reftox test result fell within control limits)	If the organism supplier reftox results were within control limits, and the coefficient of variation for the last 20 reftox tests is ≤ 0.85	If there are notable reporting errors or deviations from test protocol, or if the reftox test result fell outside of control limits to the more sensitive side (point estimate was too low) by 3 or more standard deviations and the effluent test showed toxicity at levels of regulatory concern
Both the original and the first repeat reftox test results were outside of control limits (the second repeat reftox test result fell within control limits)	If the 95 percent confidence interval for the point estimate used in control charting can be calculated and in both failing reftox tests overlapped the control limits in the control chart, organism supplier reftox results were within control limits, and the coefficient of variation for the last 20 reftox tests is ≤ 0.60	If there are notable reporting errors or deviations from test protocol, or if any reftox test result fell outside of control limits to the more sensitive side (point estimate was too low) and the effluent test showed toxicity at levels of regulatory concern
All three reftox tests were outside of control limits	Never	Always
Coefficient of variation for the last 20 reftox tests > 0.85	Never	Always

Because point estimates provide the best basis for control charting, all labs must control chart using point estimates. Point estimates require fewer replicates than NOECs and reference toxicant testing may be done using the minimum number of replicates allowed by the test method.

Another Ecology staff person with primary responsibility for reference toxicant testing requirements is the Advisory Laboratorian in the Quality Assurance Section, who reviews standard operating procedures (SOPs) for toxicity tests and accredits labs. For bioassay labs to maintain Department of Ecology laboratory accreditation, the QA section has begun to require participation in a round-robin test (such as the DMR-QA) or the performance of one reference toxicant test at least once every six months. In the event that a lab does not conduct any tests on environmental samples using a particular species/method within a six-month period, it must perform a reference toxicant or round-robin test. In the event that a lab does not conduct any tests by a particular method within a one-year period, it must do two reference toxicant or round-robin tests for that year. Further, these tests must be done at least four months apart. This is to assure that the labs maintain proficiency with the species and methods for which they are accredited. The Quality Assurance Section can efficiently enforce good reference toxicant testing requirements because it has direct authority over labs to approve SOPs and conduct routine onsite audits.

E.4.3 Describe Any Additional QA/QC Criteria Your State Has Developed and Implemented Within Your State

— Sometimes variability across replicates will prevent a large difference in response (in other words, a toxic effluent) from being detected as statistically significant. False negatives can happen when the number of replicates is low. The acute statistical power standard says that acute toxicity tests must be able to detect a minimum of a 30 percent difference in survival between the IWC and a control as statistically significant. The chronic statistical power standard says that chronic toxicity tests must be able to detect a minimum of a 40 percent difference in response between the IWC (the NOEC if the IWC is unknown) and a control as

statistically significant. Tests which fail to meet the power standard must be repeated with an increased number of replicates.

Ceriodaphnia Chronic Test

— ≤ 10 percent males in the surviving test organisms over all test concentrations.

— ≤ 20 percent males in the surviving test organisms in the IWC or LOEC.

— All surviving *Ceriodaphnia* producing no neonates in the test must be examined to determine gender, and the results of the determination reported. It is not necessary to identify gender when reproduction has been nearly eliminated in any test concentration when this fits an expected concentration-response relationship. It is understood that very young *Ceriodaphnia* can be difficult to sex, and any *Ceriodaphnia* that dies in the first two days of the test may be excluded from calculations for reproduction if gender is difficult to determine and it is one of no more than two mortalities in a concentration. Otherwise, difficult to sex young *Ceriodaphnia* must be considered to be female and included in all calculations.

E.4.4 Describe Any Efforts Your State Has Made to Minimize Test Method Variability

1. Development and distribution to all labs of a document called "Laboratory Guidance and Whole Effluent Toxicity Test Review Criteria" (*canary book*) that lets them know our expectations for an acceptable toxicity test. The *canary book* also narrows testing choices and provides for more consistent testing between labs.

2. Test reviews for compliance with the test method and canary book.

3. Fish or mysid growth tests that have a standard deviation for proportion alive above 0.25 in any effluent concentration (unless the partial mortality occurs at the threshold of toxicity in a good concentration-response relationship) are analyzed for the original growth endpoint instead of the combined ("biomass") endpoint.

4. To reduce the opportunity for WET limit violations due to statistically significant differences in response that are type I errors, permit requirements will lower the *alpha* level for hypothesis testing when differences in test organism response are small. To prevent excessive type I errors, eliminate some interrupted concentration-response relationships**,** and have more fair and enforceable test results, we will set *alpha* = 0.01 for small differences in response. If the difference in survival between the control and the IWC in an acute test is less than 10 percent, the level of significance will be lowered from 0.05 to 0.01. If the difference in test organism response between the control and the IWC in a chronic test is less than 20 percent, the level of significance will be lowered from 0.05 to 0.01.

5. The identification of anomalous tests is a valuable tool for reducing false positives. A concentration-response relationship where response increases with concentration is a good identifier of toxicity as opposed to other sources of organism stress such as disease. Test method variability or lab error will also very rarely produce a good concentration-response relationship. Identifying a test as anomalous does not necessarily mean rejection of the test and a requirement to repeat. If a test result meets one of the criteria for anomalous test identification but has no statistically significant toxicity at concentrations of regulatory concern (IWC), then the test need not be repeated unless other factors contribute to a decision to reject the test.

The anomalous test definitions below must be considered in light of the expectations for the different toxicity tests and endpoints.

Criteria for Identifying Anomalous Test Results

— A WET test result is anomalous if it shows a statistically significant difference in response between the control and the IWC, but no statistically significant difference in response at one or more higher effluent concentrations. The lack of statistical significance must be associated with a lower toxic effect at the higher effluent concentration. Any higher effluent concentration used in this determination must be a part of a dilution series. Labs should not cluster test concentrations just above the IWC in order to increase the opportunity for an anomalous test result.

— A WET test is anomalous if there is a statistically significant difference in response between the control and the IWC which together with other nearby concentrations of effluent, have a zero slope and appear to be nontoxic (performance is typical of healthy test organisms). Another description of this criterion is a test with a control that seems not to belong to the concentration-response relationship because of exceptionally good performance.

— A WET test is anomalous if the overall slope of the line fitted to the concentration-response plot is opposite of normal expectations and there is a statistically significant difference in response at the IWC. A test might be considered acceptable if the slope is opposite over only part of the concentration series.

— A WET test is anomalous if the standard deviation for proportion alive equals or exceeds 0.3 in any test concentration unless the partial mortality fits a good concentration-response relationship. A WET test is anomalous if mortalities occur in any test concentration in excess of the control performance criterion for survival when the concentration-response relationship indicates that the effluent concentration is nontoxic (sporadic mortalities).

E.4.5 Explain How Your State Reviews or Conducts Performance Lab Audits

The Department of Ecology manages an Environmental Laboratory Accreditation Program designed to assure that accredited labs have the capability to provide reliable and accurate environmental data to the department. Applicant labs apply for accreditation for specific parameters and methods. An applicable parameter/method pair for WET testing would be "*Pimephales promelas* by EPA Method 1001.0."

Concurrent with submission of the initial application, the lab submits a quality assurance manual that is given a thorough review by Ecology staff. If there are reasonably-available performance evaluation (also known as "proficiency testing") samples available for the requested tests, the lab is required to submit one set of such PE results for initial accreditation. This is referred to in our program as a "performance audit." There are no PE samples we consider to be "reasonably available" for WET testing.

Following review of the lab's QA manual and PE study results and successful resolution of any noted problems, Ecology and the lab schedule a mutually agreeable date for an on-site, or system, audit. (Although this survey asks about "performance" audits, which could be construed as being synonymous with our required PE studies, we think it rather is synonymous with what we call the on-site, or system, audit). For initial system audits, depending on the scope of tests done by the lab, checksheets may be sent to the lab to be completed and returned to the auditor prior to the audit. The auditor studies the checksheet responses and verifies accuracy of the response during the audit. For subsequent audits, which are routinely scheduled every three years but may be conducted at any time there is a need, the auditor may choose to send checksheets in time for them to be completed by the lab or take them to be filled in during the audit.

The actual audit, if for WET testing only, would involve one auditor and last one or two days depending on the scope of tests done in the lab. If the lab does other testing, the audit team may involve as many as five, and the audit may last as many as three days (or longer if required, but none have to date). The audit consists of an in-briefing, a thorough audit of personnel qualifications and equipment/supplies status (which were reported as part of the application), facility adequacy, sample management, records keeping/data management, performance evaluation study data (if applicable), the overall quality assurance program, status of quality control testing results (to see if the lab is meeting data quality objectives which were approved in the QA manual), and a check to see that current methods/SOPs are readily available and being followed. An out-briefing follows the audit during which the audit team informally summarizes major findings, both good and bad.

Following the audit, our program allows us 30 calendar days to prepare a written report. Depending on the scope of testing, this report, which addresses each of the factors discussed above, may be only 3 or 4 pages, or many more, and might include several attachments providing guidance or assistance to the lab. The secondary objective of our program as specified in the code is to assist labs in achieving the ability to meet required standards of performance, a perhaps novel but very effective approach to achieving desired capability in accredited labs. Historically, we have been deficient in meeting the 30-day report requirement, which has caused us to change our accreditation strategy. Using a fixed-price contract to encourage prompt reporting, we now contract out the audit task to a highly-qualified auditor whose last audit report was delivered within 10 days of the audit.

Performance audits (PE studies) are required in our program twice each year, and system audits are preferably conducted every three years with the code allowing four years for documented cause. At this time, we see no need to exceed three years for future audits of WET testing labs.

E.4.6 Describe Any Specific Implementation Guidance That Your State Has Developed to Assist Permit Writers. How Is the Guidance Available to the Public?

We have developed and kept updated suggested language for use in NPDES permits and fact sheets for POTWs and industries. The suggested language is a part of templates (“shells”) for permits and fact sheets that permit writers use as they draft a permit. We also have a “Permit Writer’s Manual” (USEPA 1996a) which addresses species choice, WET monitoring frequency, recommendations for number of test concentrations, etc. The “Permit Writer’s Manual” was developed with public input/review and is available to the public for the cost of printing.

E.4.7 Describe How Your State Provides or Utilizes Any Toxicity Testing Training

We had extensive training in all of our offices at the beginning of our use of WET testing in water quality-based permitting early in the 1990s. Because of budget constraints, because WET test review and technical assistance are centralized functions, and because of the availability of permit writing guidance in the “Permit Writer's Manual” and suggested permit language, we no longer hold WET training sessions.

E.5 RESPONSES FROM WISCONSIN DEPARTMENT OF NATURAL RESOURCES

E.5.1 Describe How Your State Evaluates Reference Toxicant and Effluent Test Results

Reference toxicant and effluent test data is sent directly to the Biomonitoring Coordinator in Madison (central office). Certified labs are required to perform reference toxicant tests (using NaCl, specified dilutions and dilution water) on a monthly basis. Acute and chronic reference toxicant results are evaluated using the point-estimation techniques described in the EPA manual (LC_{50}, IC_{25}). Control charts (graphical and tabular) representing the mean LC_{50} or IC_{25} and upper and lower control limits (mean $\pm$ 2 standard deviations) are established for each species, using data from the previous 20 months. Any exceedance of

either the upper or lower control limit after establishment of the control chart requires a review of the culture and test systems. Missing or out-of-range data must be explained (if possible) and may result in invalidation of Washington Pollution Discharge Elimination System (WPDES) test results conducted during the same period.

Each test report for all effluent tests is reviewed by the Biomonitoring Coordinator for completeness, adherence to QA and test acceptability requirements, and for compliance with the WPDES permit. Deviations from permit requirements, test acceptability criteria, or other factors may cause tests to be repeated.

E.5.2 Explain How Your State Reviews Reference Toxicant Data for Laboratory Performance

(See above.)

In addition to the regular review by the Biomonitoring Coordinator, reference toxicant data is reviewed by the Department's WET Laboratory Auditor prior to on-site laboratory inspections. The laboratory provides copies of bench sheets, water quality data, current control chart data, and any explanations of out-of-range test results for each test type and organism combination. The materials are reviewed for appropriate test frequency, proper test conditions, test result validity, and proper responses to out-of-range events.

E.5.3 Describe Any Additional QA/QC Criteria Your State Has Developed and Implemented Within Your State

Test acceptability requirements, based on current “State of Wisconsin Aquatic Life Toxicity Testing Methods Manual, Edition 1”:

Testing must be separated from culturing activities (separate rooms with separate ventilation systems; if closed incubators are used, culturing & testing may not be contained within the same incubator)

For Static Renewal Acute Tests:

Pretest Requirements (Requirements For Culture Acceptability)

— *C. dubia*:
  - Average Number Of Neonates In 3 Broods $\geq$ 15
  - Mean Survival $\geq$ 80 percent
  - Number Of Neonates In Each Brood $\geq$ 8
  - Age Of Organism $\leq$ 24-H

— Fathead Minnows:
  - Age Of Organism 1- 14 Days
  - Sample Requirements
  - Holding Time $\leq$ 36-H
  - Temperature During Collection & Prior To Shipping $\leq$ 4 °C
  - Temperature Upon Arrival At The Laboratory $\leq$ 10 °C

Test Requirements (Requirements For Test Acceptability)

— Temperature 20 ° $\pm$ 1 °C

— Dissolved Oxygen > 40 percent and < 100 percent saturation

— Effluent - pH $\geq$ 6.0 and $\leq$ 9.0.

— Control Survival $\geq$ 90 percent

For Static Renewal Chronic Tests:

Pretest Requirements (Requirements For Culture Acceptability)

— *C. dubia:*
  - Average Number Of Neonates $\geq$ 20
  - Mean Survival $\geq$ 80 percent
  - Neonates Used In Test Must Be From 3rd Or Subsequent Brood
  - Number Of Neonates In 3rd Or Subsequent Brood $\geq$ 8
  - Age Of Organism $\leq$ 24-H; Released Within Same 8-H Window

— Fathead Minnows:
  - Age Of Larvae $\leq$ 24-H
  - Sample Requirements
  - Holding Time $\leq$ 36-H
  - Temperature During Collection & Prior To Shipping $\leq$ 4 °C
  - Temperature Upon Arrival At The Laboratory $\leq$ 10 °C

Test Requirements (Requirements For Test Acceptability)

— Temperature 25 ° $\pm$ 1 °C
— Dissolved Oxygen > 40 percent and < 100 percent saturation
— Effluent - pH $\geq$ 6.0 and $\leq$ 9.0
— Control Survival $\geq$ 80 percent
— *C. dubia* Mean Control Reproduction $\geq$ 15 Neo./Adult; > 60 percent produce 3 broods
— Fathead Minnow Mean Control Biomass $\geq$ 0.25 mg/individual

The Wisconsin Department of Natural Resources (WDNR) is in the process of updating it's WET Methods Manual. Future methods (2nd Edition expected in 2001) will include *additional* or *revised* test acceptability criteria:

For Static Renewal Acute Tests:

Pretest Requirements (Requirements For Culture Acceptability)

— Fathead Minnows:
  - Age Of Organism 4 - 14 Days
  - Sample Requirements
  - Temperature Upon Arrival At The Laboratory $\leq$ 6 °C

Test Requirements (Requirements For Test Acceptability)

— Control Variability CV < 40 percent

For Static and Static Renewal Chronic Tests:

Sample Requirements

— Temperature Upon Arrival At The Laboratory $\leq$ 6 °C

Test Requirements (Requirements For Test Acceptability)

— Control Variability - Fathead Minnow & *C. dubia* CV < 40 percent
— Control Variability - *R. subcapitata* CV < 20 percent
— *C. dubia* Male Production < 20 percent in controls & < 20 percent all concentrations
— *C. dubia* Mean Control Reproduction >80 percent produce 3 broods
— *R. subcapitata* Control Performance Cell Density > 1 X 10^6 cells/ml at end of test

E.5.4 Describe Any Efforts Your State Has Made to Minimize Test Method Variability

1. Close review of each test result submitted with consistent adherence to test protocol test acceptability criteria.

2. Investigations of test results from disagreeing “split” effluent sample analyses.

3. State specific methods: In order to limit the variability that may occur when different procedures are used by different labs, WDNR requires strict adherence to clearly specified methods, regarding: (a) sampling procedures (volume, type, storage conditions, etc.); (b) holding times; (c) test duration; (d) deviations in feeding & environmental conditions (light, pH, temperature, DO, etc.); (e) dilution water; (f) number of concentrations and replicates tested; and (g) number of organisms per replicate.

   Each of these is addressed in EPA methods, but flexibility is allowed so labs can make tests fit in specific situations. The more flexibility allowed in test methods, the higher the chance that tests will be done differently between labs or between tests, resulting in increased WET variability. In order to control WET variability and improve the consistency of methods used by Wisconsin labs and permittees, WDNR created the "State of Wisconsin Aquatic Life Toxicity Testing Methods Manual," Edition 1 (PUBL-WW-033-96) (Methods Manual) and incorporated it by reference into NR 149.22 and NR 219.04, Wis. Adm. Code, in 1996. The Methods Manual contains specific procedures regarding testing and sampling procedures, types of tests, quality control/quality assurance procedures, test acceptability criteria (see above), etc., that labs must follow when performing WET tests for permit compliance.

4. Implementation of a WET Laboratory Certification program. In order to insure labs are of the highest quality and are able to demonstrate a serious commitment to a quality assurance/control program, WDNR, under State statutes, certifies labs to perform WET tests. In order for a lab to apply for certification for WET testing, the lab must submit a completed application and a quality assurance plan to the lab certification program and pass an on-site evaluation. WET labs must have an ongoing reference toxicant program, a review process for all test data and reporting, a good sample custody system, proper equipment maintenance, dilution water quality monitoring, facility maintenance, and attention to test organism health, and make other demonstrations of good lab practices in order to pass an audit.

5. The WDNR's WET Team strives to continually improve the WET program. The WET Team is now revising the Methods Manual to require that labs verify the training and qualifications of their staff, to include test acceptability criteria related to variability, and other changes to further improve WET test quality and reduce variability (see above).

E.5.5 Explain How Your State Reviews or Conducts Performance Lab Audits

Inspections may be announced or unannounced. Prior to the inspection, the auditor reviews laboratory SOPs and recent reference toxicant results for adherence to WDNR protocols. The actual inspection consists of an opening conference, a walk-through of the laboratory facilities, and a closing conference. During the opening conference, the auditor discusses the SOP review and general procedures in the laboratory. During the walk-through, the auditor examines equipment, written documentation, cultures, and laboratory procedures. He/she will also interview lab personnel to insure that they understand lab quality assurance and methods requirements. The closing conference serves as a review of observations and comments during the inspection. After the inspection, the auditor generates an inspection report, to which the laboratory will be given 60 days to respond. If there are significant deficiencies discovered during the inspection, and the laboratory fails to fix those deficiencies satisfactorily within the allotted time, the laboratory's certification may be revoked.

E.5.6 Describe Any Specific Implementation Guidance That Your State Has Developed to Assist Permit Writers. How Is the Guidance Available to the Public?

The WDNR created the "WET Program Guidance Document" in 1996, as a companion document to the "State of Wisconsin Aquatic Life Toxicity Testing Methods Manual," in order to provide guidance and

clarification of existing rules, for WDNR staff, permittees, labs, consultants, and others. The WET Guidance Document is updated as program needs dictate, at least once yearly, and can be obtained by contacting the Biomonitoring Coordinator at: WDNR, Bureau of Watershed Management, P.O. Box 7921, 101 S. Webster St., Madison, WI, 53707-7921; email: flemik@dnr.state.wi.us; or at http://www.dnr.state.wi.us/org/water/wm/ww/biomon/biomon.htm.

E.5.7 Describe How Your State Provides or Utilizes Any Toxicity Testing Training

The Biomonitoring Coordinator provides one-on-one training, as needed, for WDNR staff and permittees (usually as permits are reissued with new WET requirements). The University of Wisconsin-Madison State Lab of Hygiene (who provides WET testing and research services to WDNR) can provide hands-on WET training to WDNR staff, permittees, and/or new staff at contract laboratories, at their request. WDNR staff, permittees, and contract lab staff have also attended The Society of Environmental Toxicology and Chemistry's two-day WET course and statistical analysis course.

This page intentionally left blank.

APPENDIX F

IMPROVEMENTS IN MINIMIZING WET TEST VARIABILITY BY THE STATE OF NORTH CAROLINA

This page intentionally left blank.

IMPROVEMENTS IN MINIMIZING WET TEST VARIABILITY BY THE STATE OF NORTH CAROLINA

F.1 Background

The North Carolina Division of Water Quality (NC DWQ) began in-house WET testing in the late 1970s. Data collected through the mid-1980s indicate that one in four NC NPDES facility effluents tested had the potential to cause acute toxicity instream during low stream flow/high effluent flow conditions (Eagleson et al. 1986). The Division began to require WET self-monitoring by individual facilities in 1985 through administrative letters. DWQ first implemented WET limits in NPDES permits in 1987. As of March 29, 2000, 554 facilities are required to perform some type of WET monitoring; 453 of these have limits. North Carolina permittees have demonstrated compliance rates consistently above 90 percent since the additional TAC were implemented. Chronic *Ceriodaphnia dubia*, acute *C. dubia*, and acute fathead minnow are the primary test types used.

The Division uses two primary strategies to enhance data quality: (1) individual report review and (2) laboratory certification.

Division personnel review each analysis report for the following test acceptability criteria:

- Sample type (specified by permit)
- Sample hold time
- Sample temperature upon receipt at lab
- Control treatment water pH and dissolved oxygen
- Control water hardness*
- Effluent treatment dissolved oxygen
- Test type (specified by permit)
- Replication
- Effluent dilution (specified by permit)
- Control survival and/or reproduction
- Percentage of control organisms producing three broods (*Ceriodaphnia* chronic)
- Control organism reproduction coefficient of variation (*Ceriodaphnia* chronic)*
- Test duration

*NC State criteria

The reviewer may also statistically analyze data sets when the result is unclear based on a cursory review of the data.

The Division's Water Quality Rules specify that WET analyses associated with NPDES permits must be performed by certified laboratories. The Division implemented the laboratory certification program in 1988. Key requirements of that program are specific qualifications for laboratory supervisors, a reference toxicant testing program, annual inspections and audits, and performance evaluation analyses.

Laboratory Supervisor Qualifications

Laboratory supervisors must have either a Bachelor of Science degree in biology or a closely related field and three years of experience in aquatic toxicity testing, or a Master of Science degree in biology or a closely related field and one year of experience in aquatic toxicity testing.

Reference Toxicant Testing Program

The laboratory must maintain a reference toxicant testing program for each organism and test type category (chronic and acute). A reference toxicant test should be performed every two weeks for each organism used in acute WET testing. Alternatively, acute reference toxicant tests may be performed such that NC NPDES acute tests are performed within one week of an acute reference toxicant test for the organism in question. Similarly, a reference toxicant test should be performed once per month for each organism used in chronic WET testing. Alternatively, tests may be performed such that NC NPDES chronic tests are performed within two weeks of a chronic reference toxicant test. To maintain certification for an organism, reference toxicant tests must be performed at least quarterly.

Annual Inspection and Audit

The Division conducts at least one inspection per year at each laboratory. Most inspections are announced, but may be performed without notice. Inspections include the following activities:

- Inspect facilities, equipment, and QA procedures according to the laboratory's standard operating procedures
- Examine living and preserved test organisms
- Review reference toxicant testing program documentation
- Inspect meters and meter calibration records
- Trace randomly selected test records

Performance Evaluation Analyses

The Division may distribute unknown samples to laboratories up to three times per year for analysis. The Division constructs acceptability criteria using the pooled results of the analyses. Laboratories generating results outside of the acceptable range must repeat the analysis. Two consecutive out-of-range results result in decertification. A decertified laboratory regains certification by generating acceptable results on two follow-up analyses.

F.2 Data Evaluation (1992-94) Summary

In January 1992, NC DWQ began recording reproduction data from *Ceriodaphnia* chronic pass/fail tests performed by NC DWQ-certified laboratories in association with NPDES permit requirements. The majority of NC facilities with WET limits use this test. NC pass/fail tests consist of two treatments: a control and a critical concentration, each with 12 replicates. The purposes of the data base were to evaluate the sensitivity of the analysis, assess performance characteristics of the analyses, and evaluate performance of individual laboratories. Analysis was limited to test results with normally-distributed reproduction data.

In 1994, NC DWQ investigators reviewed the PMSD and MSD as a percentage of the control mean for each test (Rosebrock et al. 1994). Evaluation of the data indicated a correlation between PMSD and timing of test termination. EPA methods allow the test to be terminated once 60 percent of the control organisms produce three broods. Therefore, the percentage of adults producing a third brood at test termination may

vary from 60 to 100 percent. Plotting PMSD versus percent of control organisms producing three broods clearly showed that higher percentages of control organisms producing three broods were associated with lower PMSDs (Figure F-1).

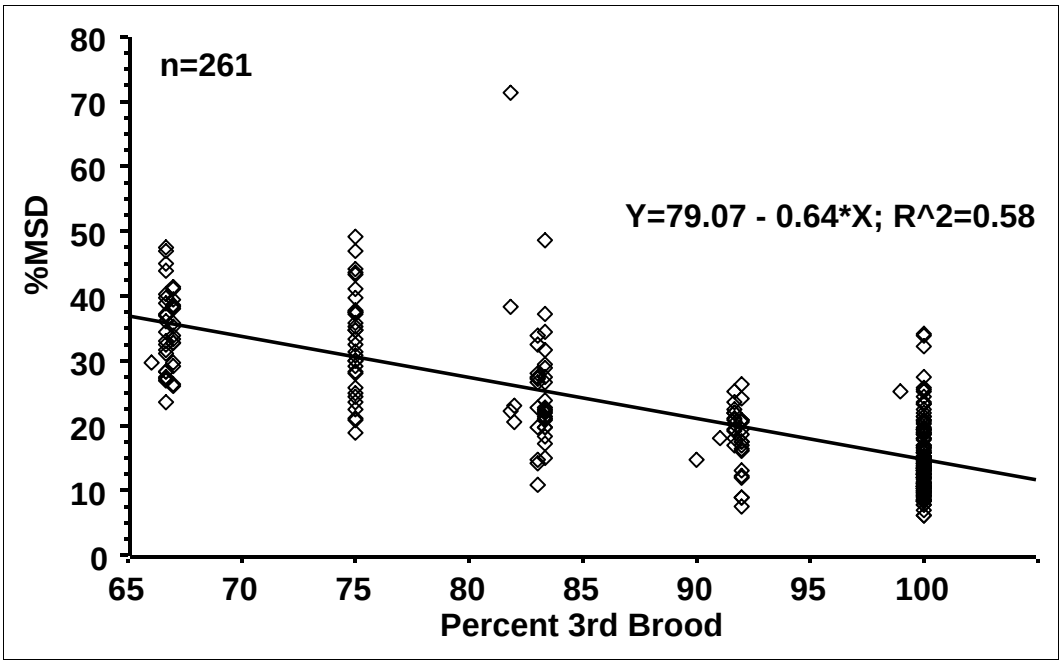

Figure F-1. PMSD versus percent control organisms producing three broods (1994).

Percentile analysis of the PMSD data produced a median PMSD of 20. This means that the "average" analysis, defined as the median, can statistically detect as small as a 20 percent difference between the treatment and control organism reproduction.

Percentile analysis of the CV data for control organism reproduction produced a median of 17 percent and a 95th percentile of 40 percent. This means that 95 percent of the control data sets produced CVs at or below 40 percent.

F.3 North Carolina Chronic Protocol Modifications

Using results from the data evaluations described above and empirical knowledge gained from experience with the test, NC DWQ made several changes to its chronic *Ceriodaphnia* protocol to improve sensitivity, precision, and practical application of test results in its compliance program. These changes were implemented in two stages in late 1994 and early 1996.

December 1994 Changes

- Exclusion of 4th brood and higher neonates from the reproduction analysis
- Addition of a TAC requiring that at least 80 percent of the control organisms produce three broods
- Addition of a TAC requiring that the test be terminated no later than seven days after initiation

January 1996

- Addition of a TAC requiring that the control organism reproduction CV be less than 40 percent
- Specification that for an effluent treatment to be considered as producing an effect, the reproduction mean must be statistically significantly lower than the control mean **and** represent at least a 20 percent reduction from the mean

Reducing the CV of the control reproduction can be shown mathematically to result in reductions in the MSD and PMSD, producing a more sensitive test. Placing an upper limit on the CV will eliminate less sensitive tests. Excluding 4th brood neonates from the reproduction analysis and requiring that at least 80 percent of the control organisms produce a 3rd brood will reduce the control organism reproduction CV.

The specification of at least a 20-percent reduction in reproduction from the control effectively sets a lower limit on test sensitivity. DWQ's experience has shown that high-quality laboratories can produce extremely sensitive tests that can detect very small differences between treatment and control reproduction. Unfortunately, this can be a disincentive for laboratories to produce high quality tests because some clients will gravitate toward laboratories that produce compliant test results. Less sensitive tests will more likely produce such results.

F.4 Evaluation of Program Modifications

The North Carolina data base affords the opportunity to evaluate the effectiveness of additional TAC and changes to the test protocol as they relate to the variability of WET test results. Effluent data for individual laboratories, and across all tests and laboratories, were examined to discern the impact of program changes on laboratory performance. Data were partitioned into two data bases, one for effluent tests completed before December 1994 (termed Pre-1995) and one for effluent tests completed after January 1996 (termed Post-1995). Pass/Fail tests were included in the evaluation. Only tests that did not have a significant mortality effect were considered. Two measures of laboratory performance were calculated using the reproductive data from the tests: PMSD and control CV. The PMSD data set contains all tests reported for compliance. The control CV data set contains all unique controls that were reported by the laboratories and used in compliance calculations. Conclusions reflect the cumulative impact of all changes made to the program from late 1994 to early 1996.

F.5 Overall Test Performance

Pre-1995 and Post-1995 percentile values were generated for the PMSD and the control CV combined across all tests and laboratories (Table F-1). For the PMSD, the median value decreased from 21 percent to 16 percent and the 90th percentile from 39 percent to 31 percent, indicating an overall increase in test sensitivity. The narrower interquartile range of Post-1995 PMSD values (IQR=12 percent), compared with the interquartile range of Pre-1995 PMSD (IQR=16 percent), implies an improvement in the ability of laboratories to achieve similar levels of test sensitivity. (The interquartile range is the difference between the 75th and 25th percentiles of the cumulative distribution function and is a measure of spread of the distribution.) For the control CV, the median value was reduced from 15 percent to 13 percent and the 90th percentile from 34 percent to 28 percent. The overall decrease in the control CV reflects the capacity of laboratories to improve their performance as measured by a decrease in control variability relative to the control mean. Changes in test acceptability criteria and in test protocols improved the consistency of control performance quantified by the reduction in the interquartile range of the control CV Pre-1995 (IQR=15 percent) and Post-1995 (IQR=10 percent).

Table F-1. PMSD and Control Organism CV

	PMSD		CV	
	Pre 1995	Post 1995	Pre 1995	Post 1995
# Tests	4110	5471	2478	3401
Min	0.055	0.049	0.033	0.034
Max	0.839	0.676	0.835	0.400
Median	0.212	0.160	0.155	0.133
IQR	0.164	0.118	0.150	0.103
10th Percentile	0.105	0.095	0.078	0.077
25th Percentile	0.142	0.116	0.103	0.097
50th Percentile	0.212	0.160	0.155	0.133
75th Percentile	0.306	0.233	0.253	0.200
90th Percentile	0.391	0.307	0.343	0.285

F.6 Individual Laboratory Performance

Comparison of effluent data across multiple laboratories provides information about the influence of program changes on individual laboratory performance. Data for a laboratory (Lab 1) with low sensitivity were compared to data from a laboratory (Lab 2) with high sensitivity. Pre-1995 and Post-1995 percentile values were generated for the PMSD combined across all tests for each of the two laboratories (Table F-2). The performance of Lab 2, represented by the distribution of PMSD, was essentially the same Pre-1995 and Post-1995. However, the performance of Lab 1 improved, as evidenced by the changes in medians (33 percent to 18 percent), changes in the 90th percentile (46 percent to 32 percent), and the slight decrease in the width of the interquartile range (13 percent to 12 percent). Additionally, the Post-1995 medians for the two laboratories were relatively close (18 percent and 12 percent) percent for Lab 1 and Lab 2, respectively. A comparison of the cumulative distribution functions for each laboratory indicates that performance was more consistent across laboratories after implementing program changes (Figures F-2 and F-3).

Table F-2. Lab 1 versus Lab 2 PMSD

	Pre-1995		Post-1995	
	Lab 1	Lab 2	Lab 1	Lab 2
# Tests	921	545	1424	466
Min	8.8	5.5	6.8	5.5
Max	67.3	48.9	67.6	39.9
Median	33.5	11.7	18.2	12.5
IQR	13.3	5.5	11.9	4.4

The distribution of PMSD values within a laboratory compared to distributions in other laboratories was examined Pre-1995 and Post-1995 (Figures F-4 and F-5). The range in median values across all laboratories Pre-1995 was 12 percent to 36 percent. Post-1995, the range in median values was 10 percent to 27 percent, indicating a decrease in the overall spread among laboratories. The range in PMSD values within a laboratory was 22 percent to 78 percent Pre-1995. The Post-1995 range in PMSD values within a laboratory compared across laboratories was 17 percent to 61 percent, indicating a narrowing of the range of values within a laboratory (Table F-3). A similar comparison was made using the control CV as an indicator of laboratory ability (Figures F-6 and F-7). The median control CV varied across laboratories from 9 percent to 30 percent Pre-1995. Post-1995, the median control CV ranged across laboratories from 9 percent to 26 percent, a slight improvement in the comparability of control CV. The range in control CVs within a laboratory was 21 percent to 79 percent Pre-1995, while the range in control CVs within a laboratory Post-1995 was 17 percent to 36 percent. Overall, laboratories are generating data with more consistency across, as well as within, laboratories after implementing additional TAC and modifications to testing protocols.

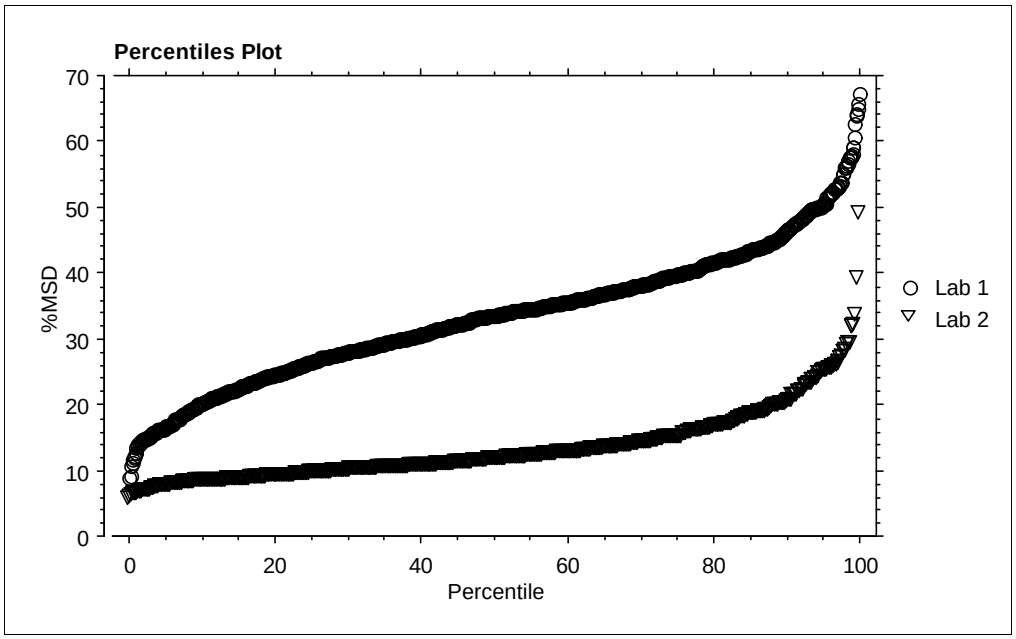

Figure F-2. Laboratory 1 versus Laboratory 2 Pre-1995 PMSD (species: *Ceriodaphnia dubia*).

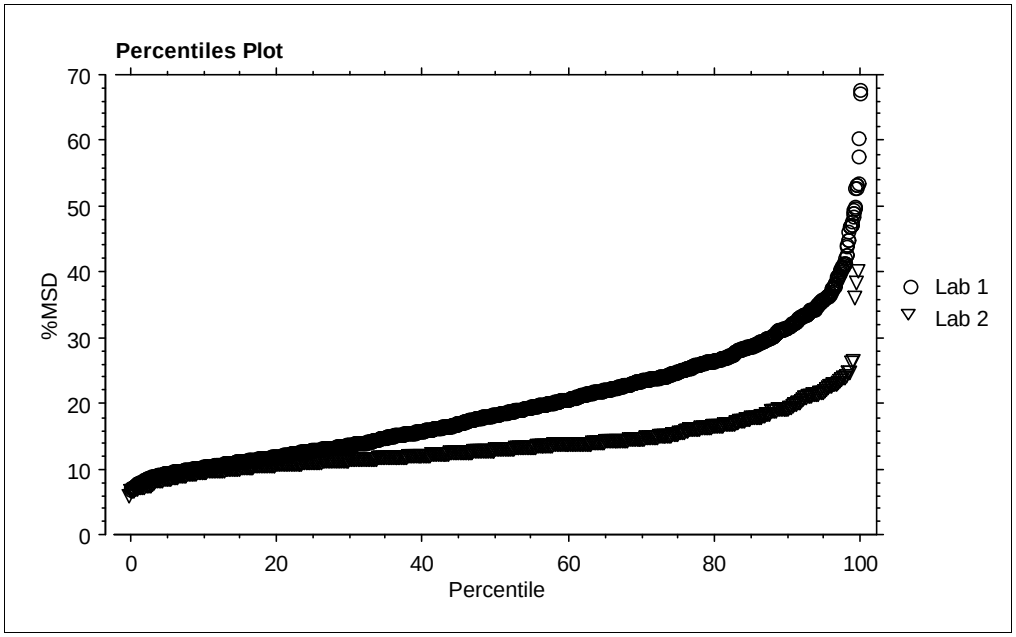

Figure F-3. Laboratory 1 versus Laboratory 2 Post-1995 PMSD (species: *Ceriodaphnia dubia*).

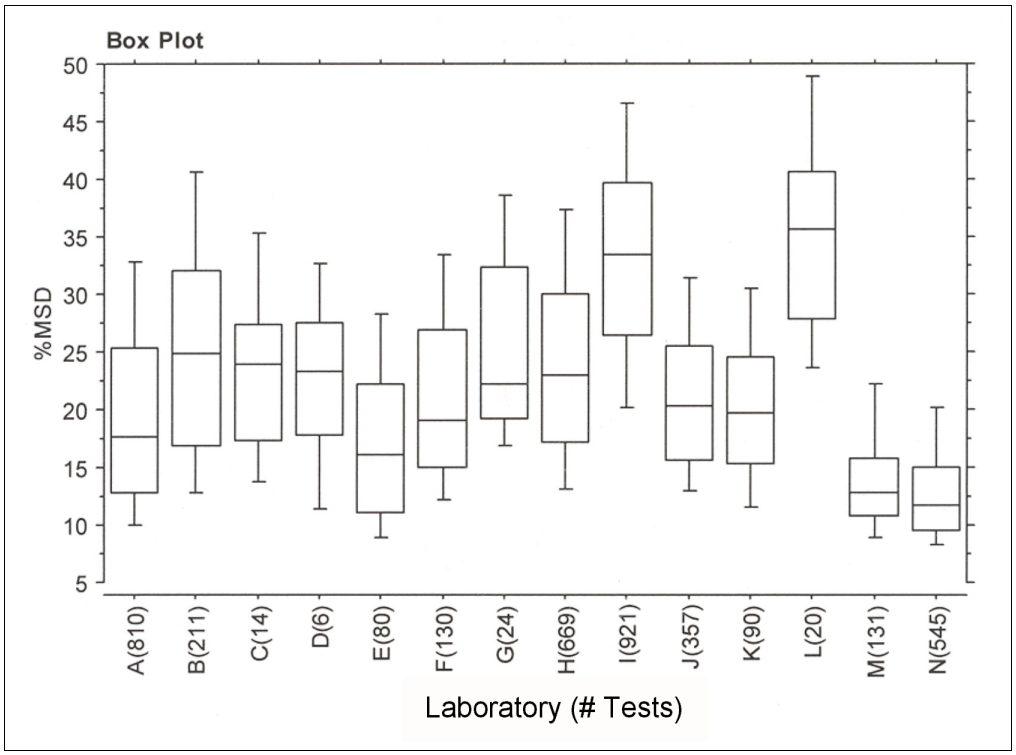

Figure F-4. Individual laboratory performance—Pre-1995 PMSD (species: *Ceriodaphnia dubia*).

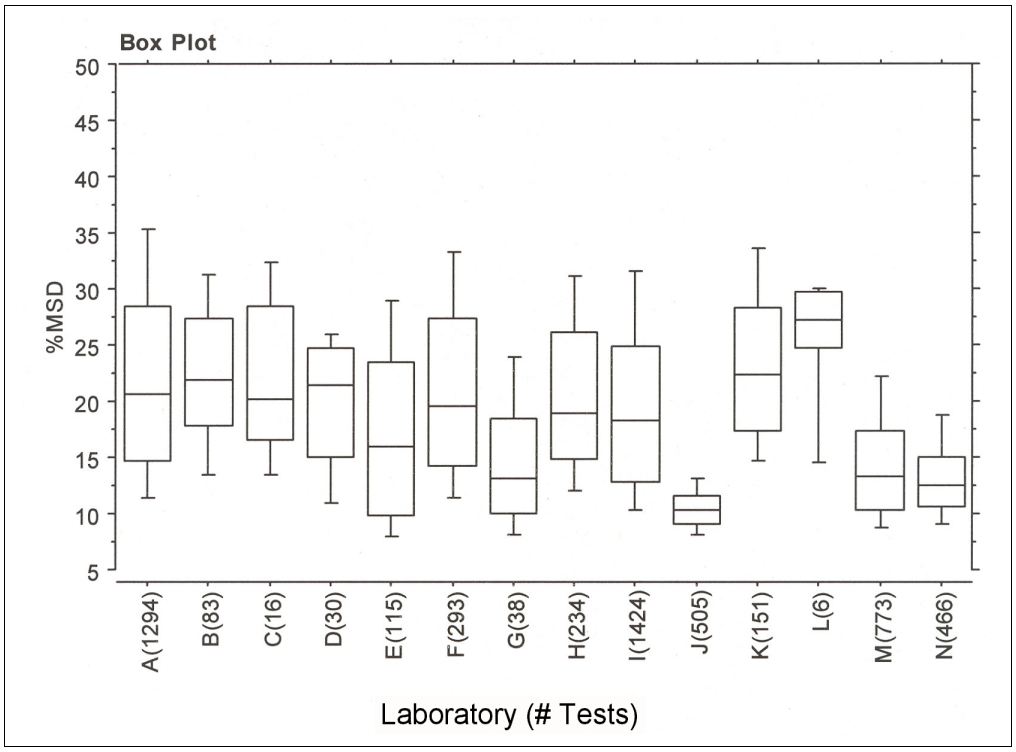

Figure F-5. Individual laboratory performance—Post-1995 PMSD (species: *Ceriodaphnia dubia*).

Table F-3. Descriptive Statistics—PMSD

	Pre-1995						Post-1995					
Lab	**N**	**Min**	**Max**	**Range**	**Median**	**IQR**	**N**	**Min**	**Max**	**Range**	**Median**	**IQR**
A	810	6.0	83.9	77.9	17.6	12.6	1294	6.4	58.9	52.5	20.6	13.7
B	211	8.6	59.7	51.1	24.8	15.0	83	10.2	39.9	29.7	21.9	9.6
C	14	13.7	35.6	21.9	23.9	10.0	16	12.5	34.5	22.1	20.1	11.9
D	6	10.6	33.2	22.6	23.3	9.7	30	9.6	33.9	24.3	21.5	9.6
E	80	6.5	43.5	37.0	16.1	11.1	115	5.6	43.8	38.3	15.9	13.6
F	130	6.9	69.4	62.5	19.1	11.8	293	6.8	55.0	48.2	19.5	13.0
G	24	13.9	45.0	31.1	22.2	13.2	38	6.6	33.1	26.5	13.1	8.4
H	669	6.2	71.5	65.3	23.0	12.8	234	8.4	38.9	30.5	19.0	11.4
I	921	8.8	67.3	58.4	33.5	13.3	1424	6.8	67.6	60.8	18.2	11.9
J	357	8.7	69.8	61.1	20.4	9.7	505	6.4	26.0	19.5	10.2	2.5
K	90	9.7	55.5	45.8	19.7	9.1	151	8.3	47.6	39.3	22.4	10.9
L	20	22.0	59.0	37.0	35.7	12.9	6	13.4	30.1	16.7	27.2	5.0
M	131	6.4	49.9	43.5	12.9	5.0	773	4.9	40.3	35.3	13.3	6.9
N	545	5.5	48.9	43.4	11.7	5.5	466	5.5	39.9	34.4	12.5	4.4

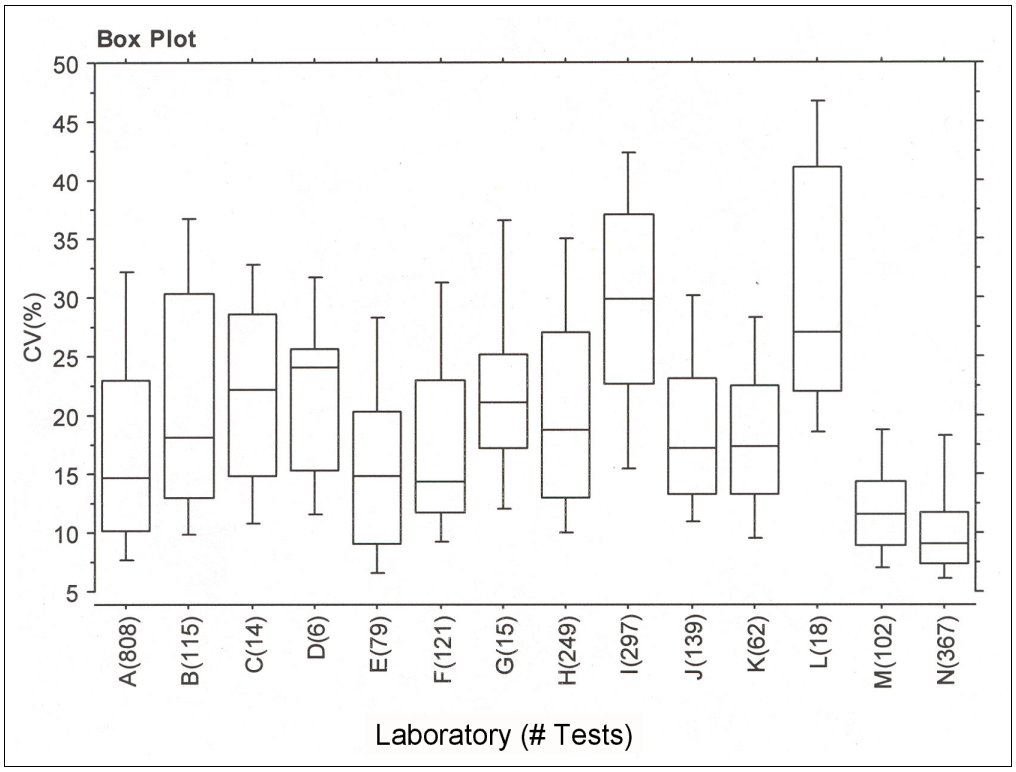

Figure F-6. Individual laboratory performance—Pre-1995 CV (species: *Ceriodaphnia duba*).

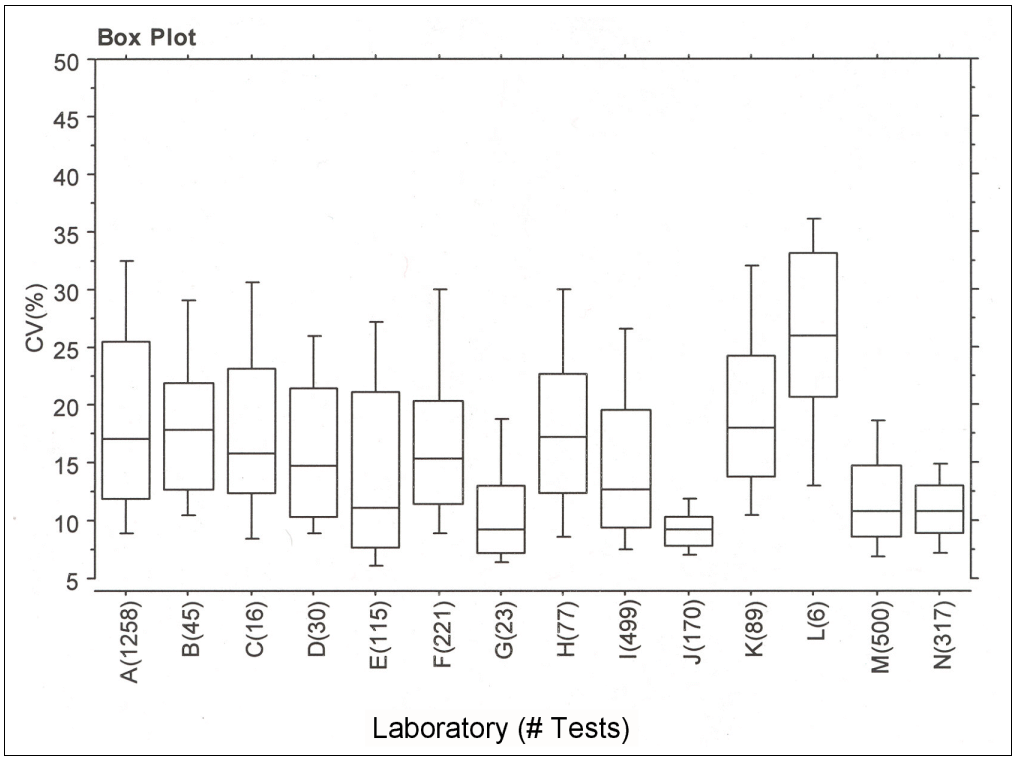

Figure F-7. Individual laboratory performance—Post-1995 CV (species: *Ceriodaphnia dubia*).

Table F-4. Descriptive Statistics—Coefficient of Variation (CV)

	Pre-1995						Post-1995					
Lab	**N**	**Min**	**Max**	**Range**	**Median**	**IQR**	**N**	**Min**	**Max**	**Range**	**Median**	**IQR**
A	808	0.041	0.835	0.794	0.146	0.129	1258	0.043	0.399	0.356	0.171	0.136
B	115	0.062	0.511	0.450	0.182	0.173	45	0.059	0.361	0.302	0.178	0.092
C	14	0.092	0.334	0.242	0.222	0.137	16	0.066	0.378	0.311	0.158	0.109
D	6	0.112	0.324	0.212	0.241	0.102	30	0.074	0.332	0.258	0.147	0.111
E	79	0.041	0.374	0.333	0.148	0.112	115	0.038	0.400	0.362	0.111	0.134
F	121	0.051	0.516	0.464	0.143	0.113	221	0.062	0.384	0.322	0.152	0.090
G	15	0.113	0.404	0.291	0.211	0.080	23	0.050	0.343	0.293	0.092	0.059
H	249	0.055	0.610	0.555	0.188	0.140	77	0.061	0.379	0.318	0.171	0.103
I	297	0.068	0.672	0.604	0.299	0.144	499	0.047	0.399	0.352	0.127	0.101
J	139	0.071	0.596	0.525	0.172	0.098	170	0.054	0.222	0.168	0.092	0.025
K	62	0.046	0.564	0.517	0.173	0.093	89	0.047	0.392	0.345	0.180	0.104
L	18	0.138	0.571	0.433	0.271	0.190	6	0.121	0.365	0.245	0.259	0.124
M	102	0.053	0.398	0.345	0.115	0.056	500	0.034	0.341	0.307	0.107	0.062
N	367	0.033	0.472	0.439	0.091	0.043	317	0.038	0.333	0.296	0.108	0.040

REFERENCES

Eagleson, K.W., S.W. Tedder, and L.W. Ausley. 1986. Strategy for whole effluent toxicity evaluations in North Carolina. In *Aquatic Toxicology and Environmental Fate: Ninth Volume, ASTM STP 921*. T.M. Poston, R Purdy, eds. American Society for Testing and Materials. Philadelphia, PA. 154-160.

Rosebrock, M.M., N.W. Bedwell, and L.W. Ausley. 1994. Indicators of *Ceriodaphnia dubia* chronic toxicity test performance and sensitivity. Poster presentation, Society of Environmental Toxicology and Chemistry 15th Annual Meeting, Denver, CO.

APPENDIX G

ANALYTICAL VARIABILITY IN REASONABLE POTENTIAL AND PERMIT LIMIT CALCULATIONS

This page intentionally left blank.

ANALYTICAL VARIABILITY IN REASONABLE POTENTIAL AND PERMIT LIMIT CALCULATIONS

Appendix G explains how analytical variability affects calculations used to determine reasonable potential and permit limits, and how such variability affects WET measurements. The appendix also considers suggested approaches to adjusting the reasonable potential and permit limit calculations to account for analytical variability. Only water quality-based effluent limitations are addressed because different considerations apply to technology-based limitations. While Appendix G addresses WET variability, its discussion and conclusions apply, with obvious modifications in terminology, to concentrations of chemical pollutants.

EPA has evaluated methodologies to adjust for analytical variability in setting permit limits. These methodologies would allow permit limits to exceed acute and chronic wasteload allocations (WLAs), sometimes two-fold or more. EPA believes that such approaches contradict the intent and practice of current guidance and regulations directed at preventing toxicity. The TSD calculations were carefully designed to avoid setting limits that allow a discharge to routinely exceed WLAs. Attempts to use an "adjusted," smaller estimate of variability in the first step of the effluent limit calculation (calculating the long-term average from the WLA) while using the variability of measured toxicity in the second step (calculating limits from the LTA), as done in the "adjustment approaches," will risk setting limits that exceed WLAs because the second variability factor is larger than the first. EPA also believes that the TSD statistical approach is adequately protective. On average, it achieves the desired level of protectiveness that is described in the NPDES regulations (40 CFR 122.44(d)) and EPA guidance.

This review did not evaluate the "conservativeness" of other components of WET limits, such as the acute-to-chronic ratio (ACR) for WET, the suggested WET criterion values (TUa = 0.3 and TUc = 1.0), and the methods of calculating the WLA using models of effluent dilution. Instead, this review took the WLAa (or WLAa,c) and WLAc as given and considered the TSD statistical method per se.

G.1 TSD Statistical Approach to Reasonable Potential And Limit Calculations

This appendix provides a simplified description of the TSD approach. That approach is more completely described in the *Technical Support Document for Water Quality-Based Toxics Control* (USEPA 1991a). Reasonable potential calculations are described in Section 3.3 of that document. The calculation is only one component of a reasonable potential determination. Permit limit calculations are described in Section 5.4 and Appendix E of the TSD.

To evaluate reasonable potential or calculate permit limits, one needs a coefficient of variation (CV) representing the variability of toxicity or a pollutant in the effluent discharge. The TSD recommends that the CV of measured effluent data be used in all reasonable potential and effluent limit calculations without attempting to "factor out" analytical variability. The specification of this CV is at issue in the alternatives to the TSD statistical procedures discussed later in this appendix.

G.1.1 Reasonable Potential

The goal of the TSD reasonable potential calculation is to estimate the probable value of an upper bound (e.g., 99th percentile) of toxicity in an effluent discharge using limited data. For whole effluent toxicity (WET), data are expressed in toxic units (TU) before calculating the CV. TU = (100/effect concentration). For chronic toxicity, TUc = 100/NOEC or 100/IC25. For acute toxicity, TUa = 100/LC50. The TSD calculations assume that effluent toxicity values follow a lognormal distribution, at least approximately. There is abundant evidence supporting the lognormal distribution, but the TSD also

acknowledges that other distributions might be found more appropriate if sufficient data can support the finding.

The sample CV of effluent monitoring data is obtained in TU. If there are fewer than ten data points, the TSD recommends a default CV of 0.6. The TSD recommends basing a calculated CV on at least ten data points, collected at the same time intervals as intended for monitoring.

Even if there are fewer than ten data points, the maximum value for the data (e.g., TU_{max}) is used to calculate a projected maximum value. A nonparametric, upper tolerance bound is calculated to infer the population percentile represented by TU_{max} with probability P: $X_{P,n} = (1 - P)^{1/n}$. For example, with probability 0.99 the largest of five observations will exceed the 39.8th population percentile: $(1 - 0.99)^{1/5} = 0.398$. Next, the ratio between this percentile ($X_{P,n}$) and the population 99th percentile is estimated using moment estimators for a lognormal distribution:

Reasonable potential multiplier = $X_{0.99} / X_P = \exp(Z_{99}\,\sigma - 0.5\sigma^2) / \exp(Z_P\,\sigma - 0.5\sigma^2)$.

Here, σ^2 is estimated as $\log(1 + CV^2)$, using the default CV if necessary. The maximum projected value is the product of the observed TU_{max} and the reasonable potential multiplier. This value may be compared to the WLA, which is based upon the criteria continuous concentrations (CCC) or criteria maximum concentration (CMC) and the appropriate dilution factors (if applicable). The projected maximum value also may be multiplied by a dilution factor and compared directly to the CMC or CCC (TSD Section 3.3, Box 3-2). The TSD recommends using TUa = 0.3 and TUc = 1.0 either as numeric toxicity criteria or as a means of interpreting the narrative "no toxics in toxic amounts" criteria.

G.1.1.1 Permit limit calculation

The first step in determining the appropriate water quality-based effluent limits for an effluent discharge is to calculate wasteload allocations WLAa and WLAc that correspond to the water quality criteria for acute exposures and chronic exposures or the ambient values used in interpreting narrative criteria (e.g., no discharge of toxic pollutants in toxic amounts). This step is distinct and separate from the "statistical" steps for calculating permit limits or reasonable potential. The WLAs are "givens" in the statistical calculations.

WLAa and WLAc are found through either a direct steady-state calculation or a dynamic model simulation. In either case, any applicable mixing zone and critical stream flows are taken into account. For WET, WLAa is converted to WLAa,c using an ACR. WLAs must not be exceeded if the water quality standards of the receiving water are to be met.

The essential idea behind setting a permit limit using the TSD approach is to find the lognormal distribution (i.e., its mean value or LTA) that would allow no more than a specified percentage of single observations to exceed the WLAa and no more than a specified percentage of the 4-day averages of observations to exceed the WLAc. If this percentage is set at 1 percent, for example, then the 99th percentile of single observations must not exceed the WLAa, and the 99th percentile of 4-day averages must not exceed the WLAc. The 4-day averaging period comes from the typical definitions of chronic exposure and the CCC. The CV has already indirectly specified the distribution's standard deviation. Together, the CV and the LTA specify the appropriate distribution completely.

The calculations which lead to finding the LTAa,c and LTAc (corresponding to the WLAa and WLAc) work in the following manner. The ratio between the LTA and a percentile (X_P) is called a variability factor (VF_P). The VF is calculated from the CV, the percentile (95th or 99th), and the averaging period [1 day (no averaging) or 4 days].

Thus, $LTA = X_P / VF_P$

If we set X_P equal to the WLAa, we find:

$$LTAa,c = WLAa / VF_{99,\ 1\text{-day}}$$

and

$$LTAc = WLAc / VF_{99,\ 4\text{-day}}$$

The smaller of the two LTAs is selected as the LTA used to calculate a limit. This step assures that the limits will exceed neither the WLAa nor the WLAc.

Having selected the smaller LTA, the VF calculation is reversed. Following the TSD recommendations,

$$\text{“Maximum Daily Limit” (“MDL”)} = LTA * VF_{99,\ 1\text{-day}}$$

and

$$\text{“Average Monthly Limit” (“AML”)} = LTA * VF_{95,\ N\text{-day}}$$

(based on N observations)

Note that in calculating the average limit the TSD recommends using a 95th percentile (rather than a 99th percentile) and the number of observations N for averaging may be less than four (although the TSD recommends $N \geq 4$ for purposes of calculating average limits). Limits calculated using the TSD-recommended approach are always equal to or less than the WLAa and WLAc.

G.1.1.2 Analytical variability in the TSD procedures

Analytical variability is a part of the variability of measurements used to analyze reasonable potential and set water quality-based limits. All components of variability that will enter into the permit development process are included in the measurements and calculations used to evaluate reasonable potential and set limits. This insures that the WLA is not exceeded.

Some laboratories have suggested alternative statistical calculations to EPA. Sections G.3 and G.4 discuss these approaches. These alternative calculations, however, would allow limits to exceed the WLA. When a sample effluent toxicity equals the WLA exactly, analytical variability would be expected to cause tests to exceed the WLA about half the time. Limits set above the WLA could allow routine exceedances of the WLA. In contrast, limits set using the TSD approach will provide some margin of safety between the limit and the WLA, guarding to some extent against analytical variability. On average, the TSD approach, employing the CV of measurements, is expected to ensure that the WLA is not exceeded when measured toxicities remain within the limits.

G.2 Background on Analytical Variability and Variability of Measurements

This section describes how analytical variability may cause the variance (σ^2) of measured values to exceed the variance of toxicity. This discussion will assume that WET tests for one discharge are conducted by one laboratory. Thus, “analytical variability” here will refer to within-laboratory variability (repeatability) of WET test results.

G.2.1 Components of Measurement Variability

The variance of monthly or quarterly measurements of effluent toxicity depends on at least two components: the variance of the toxicity, which changes over time, and the variance owed to the analytical process (including calibration, if applicable). One could also distinguish a third component—sampling variance—if simultaneous samples differ in toxicity. Herein, this component will not be examined separately, but is combined with the variance in toxicity over time.

A direct way to estimate the analytical component of variability is to analyze the same sample of effluent on different occasions so that the analytical method is the only source of measurement variance. The sample must be measured on different days because real samples are measured at intervals of weeks to months and the analytical process can change subtly over time. Unfortunately, effluent samples may not retain the same toxicity for long. Therefore, saving a batch of sample and analyzing it once a month for several months may over-estimate analytical variability. Analyzing two or three subsamples on the same date may underestimate analytical variability because the measurement system changes between sampling dates. The organisms, laboratory technicians and procedures, and laboratory materials may all change subtly over time. It would be reasonable to design a study that measures analytical variability in both ways, using effluent subsamples on one occasion and using the same (stored) effluent sample on separate occasions, attempting to bracket the correct value of analytical variance. EPA is not aware of any such studies. Reference toxicant samples are expected to have the same potency on different occasions and are used routinely for laboratory quality assurance of WET test methods. This document summarizes the variability resulting from repeated (usually monthly) WET testing of reference toxicant samples in the same laboratory.

G.2.2 Effect of Analytical Variability on Measured Values

Because of analytical variability the probability distribution of measured values Y is "wider" than the distribution of true values X. Thus, the mean and high percentiles of measurements will exceed the percentiles of the true values.

One component of the variance of measurements is analytical variance. Simple but plausible assumptions lead to the equation $V_Y = V_X + V_A$. In other words, the variance of a measurement Y (toxicity) is the sum of the variances for toxicity (V_X) and the analytical variance (V_A). When this equation is approximately correct, then one suitable estimate of V_X is (V_Y - V_A), where the parameters V_Y and V_A are replaced by their sample estimates. This estimate may be biased (i.e., inaccurate) to some degree. Similar reasoning about the mean (EY) leads to EY = EX. Then $V_Y = V_X + V_A$ can be divided by EX^2 to give $CV_Y^2 = CV_X^2 + CV_A^2$. This reasoning requires two assumptions: variance is constant and unrelated to the mean, and there is little or no correlation between X and the magnitude of the analytical error. When X is distributed lognormally, these assumptions are not true, but may be suitable for transformed values like log(Y) and log(X).

G.2.3 Analytical Variability and Self-monitoring Data

EPA determines compliance with a limit on the basis of self-monitoring data. No special allowance is made for analytical variability. This is accounted for by the TSD statistical procedures used to determine the need for limits and calculate permit limits.

The permittee must ensure that the toxicity in the discharge is never great enough to result in a compliance measurement that exceeds the permit limit. The maximum discharge toxicity allowed by the treatment system must incorporate a margin of safety to account for the sampling and analytical variability that attends compliance measurements. In other words, to avoid exceedances of a limit, a treatment system will be designed so that the maximum discharge toxicity is somewhat lower than the permit limit. Most industrial and municipal treatment facilities should be able to implement such a design. When they are not, appeals based on fundamentally different factors and economic hardships may be feasible.

G.2.4 Imprecision in WET Estimates, Reasonable Potential Determinations and Limits

Although WET tests provide protection against false positives, the estimates (NOEC, EC25, LC50) from WET tests, like all estimates based on limited data, are imprecise. That is, the exact level of toxicity in a sample is estimated with "error" (imprecision). This imprecision can be reduced by providing a suitable number of organisms and replicates for each test. The numbers required for EPA WET method test

acceptability are *minimums*. Test precision will be approximately proportional to the square root of the number of replicates. Thus, a doubling of replication may increase the precision of a test endpoint response (survival, growth, reproduction) to roughly 70 percent of its former level. For example, consider these calculations for fathead minnow growth (USEPA 1994a, pp. 102-105): the standard error of the difference between a treatment and the control is $Sw\sqrt{(1/n_T + 1/n_C)}$, which in one test took the value $(0.0972)\sqrt{(1/4 + 1/4)} = (0.0972)(0.707) = 0.0687$. If the root mean squared error Sw had been the same but the number of replicates had been doubled, the standard error would have been 0.0486. Dunnett's critical value would have been 2.24 instead of 2.36, and the MSD 0.109 instead of 0.162. With a doubling of replication, the test would be able to detect a 16-percent reduction from the control rather than a 24-percent reduction.

For reasonable potential and limit calculations, WET data are accumulated over a year or more to characterize effluent variability over time. This sampling program may not fully characterize effluent variability if too few samples are taken, if the sampling times and dates are not representative, or if the duration of the sampling program is not long enough to represent the full range of effluent variability. For reasonable potential and limits, the key quantity being estimated is the variance (or CV). A large number of samples is required to estimate a variance or CV with much precision. Confidence intervals for the variance and CV can be calculated easily and carried through the calculations for reasonable potential and effluent limits (Section G.1). Even when assumptions are not strictly met, this information may provide a useful perspective on the uncertainty of the calculation.

G.2.5 Between-laboratory Variability in Reasonable Potential and Permit Limit Calculations

It is inappropriate to use estimates of between-laboratory variability in calculations of reasonable potential and permit limits. Such estimates do not represent the variability affecting measurements of effluent discharge toxicity. In most cases, only one laboratory will produce the data for one discharge. In some cases, there will be a change of laboratory over time, which needs to be handled case-by-case. Using estimates of between-laboratory variability to represent the analytical component of variance for one discharge is equivalent to assuming that each new sample is sent to a new laboratory selected at random from the population of laboratories conducting the test method. This approach does not occur in practice.

Between-laboratory differences in test sensitivity are important and need to be addressed. To some extent, apparent differences in sensitivity between laboratories (Warren-Hicks et al. 1999) may be owed to several factors, including use of unstable reference toxicants like SDS (Environment Canada 1990), errors in calculating and recording stock concentrations (Chapter 3 of the Variability Guidance, SCTAG 1996), differences in dissociation and bioavailability of metal ions, comparisons of non-comparable ionic forms (e.g., potassium chromate versus potassium dichromate, SCTAG 1996), use of different waters, health of organisms, and varying techniques.

Within-laboratory variability should be reflected in regulatory calculations. If the data being used for reasonable potential or permit limit calculations consist of effluent measurement data reported by two or more laboratories, there are ways to account for between-laboratory differences:

- If the same laboratories are used in the same proportion or frequency, and the measurements for different laboratories represent different sampling dates, the measurement data may be treated as if they come from one laboratory. This may increase the estimated variance and the average monthly limit, which is not in the interest of the permittee. It would be better to select one laboratory, based on the variance of its reported reference toxicant test results.

- If only one laboratory has reported data on each date, with the different laboratories either reporting over different time spans or over the same time span on alternate dates, EPA recommends a pooled

estimate of variance. Calculate the sample variance S^2 for log(TU) separately for each laboratory, and combine the data in the following formula:

$$\text{pooled variance of } \log(X) = [(N_1 - 1)S_1^2 + (N_2 - 1)S_2^2] / [(N_1 - 1) + (N_2 - 1)]$$

(i.e., the analogous formula for more than two laboratories). The same result can be obtained by conducting a one-way analysis of variance on log(X) and using the mean squared error. This approach would be undesirable if the different laboratories sampled times or time spans that were known or expected to differ in the average or variance of TU. In that case, one would pool the data, treating it as if it had come from one laboratory (see above).

A change of testing laboratory by a permittee may result in a change in analytical (within-laboratory) variability of measurements and a change in "sensitivity." The average effect concentration may change. There may be between-laboratory differences in sensitivity to some toxicants, such as metals (Warren-Hicks et al. 1999).

Ideally, a permittee will anticipate a change of the testing laboratory. Permittees should compare reference toxicant test data from current and candidate replacement laboratories, selecting a laboratory with acceptable variability and a similar average effect concentration. Regulatory authorities should compare reference toxicant data for old and new laboratories when interpreting a series of WET test results that involves a change of laboratory.

Some areas may help reduce laboratory differences in average effect concentration for the same reference toxicant test protocol. These include standardization and reporting of stock culture conditions (such as loading, age structure, age-specific weight, and other conditions), standardization of dilution water for reference toxicant tests, and reporting to verify such practices. Other areas for consideration include test protocols, test acceptability criteria, and dilution water. Another approach that could be evaluated further is conducting a reference toxicant test with each effluent test, and normalizing the effluent response using the toxicant response.

G.3 Adjustment Approaches To Account For Analytical Variability in Setting Permit Limits

G.3.1 Adjustment Approaches To Account for Analytical Variability

Methods have been proposed for determining reasonable potential and calculating permit limits by adjusting the calculations based on analytical variability. The more general principles are discussed here, details of these methods are outlined in Section G.4. The focus of these discussions is the limit calculation, although similar principles apply to the reasonable potential calculation.

The idea behind the proposed "adjustment methods" for calculating water quality-based effluent limits is to estimate the distribution of toxicity values using data on measured effects concentrations and analytical variability, and then to factor out analytical variability from some steps in the process of calculating limits. In proposed adjustment methods for calculating effluent limitations one would (1) estimate the variance of effluent concentrations (this entails subtracting an estimate of the analytical variance from the variance of effluent measurements, e.g., $V_X = V_Y - V_A$, or an equivalent calculation using CVs); (2) calculate the LTAa and LTAc using the TSD approach and the adjusted variance V_X; and (3) calculate the limit (from the lower of the two LTAs) using the variance of measurements V_Y. Because the V_Y necessarily exceeds V_X, these methods would result in limits that would exceed calculated WLAs, depending on other assumptions made in the limit calculations. As a result, the discharge may allow instream WET to routinely exceed the criterion limits, a condition that should not occur.

G.3.2 Adjustment Equations

As noted above, the adjustment approaches are based on the TSD statistical approach, modified to subtract analytical variability from the LTA calculation. These approaches refer to V_X as the "true" variance. In what follows, the sample estimate of V_X is S^2_{True}. Thus, $S^2_{True} = S^2_{Meas} - S^2_{Analy}$ (where S^2 is the sample estimate of variance) is used to calculate the LTAs and S^2_{Meas} is used to calculate the limits from the smallest of the two LTAs. The TSD equations as applied to WET would be adjusted as follows:

When the LTAa,c is the smallest LTA,

$$MDL = WLAa,c * (VF_{99,\ 1\text{-day},\ Meas} / VF_{99,\ 1\text{-day},\ True})$$
$$AML = WLAa,c * (VF_{95,\ N\text{-day},\ Meas} / VF_{99,\ 1\text{-day},\ True})$$

When LTAc is the smallest LTA (and assuming that the chronic criterion is a 4-day average)

$$MDL = WLAc * (VF_{99,\ 1\text{-day},\ Meas} / VF_{99,\ 4\text{-day},\ True})$$
$$AML = WLAc * (VF_{95,\ N\text{-day},\ Meas} / VF_{99,\ 4\text{-day},\ True})$$
$$\text{where } N = \text{samples/month (for purposes of AML calculation)}$$

The VF (variance factor) is the ratio of a percentile to a mean, in this case for the lognormal distribution.

$$VF_{99,\ 1\text{-day},\ Meas} = \exp(Z_{99}\, S_{Meas} - 0.5S^2_{Meas})$$
$$VF_{99,\ 1\text{-day},\ True} = \exp(Z_{99}\, S_{True} - 0.5S^2_{True})$$
$$VF_{95,\ n\text{-day},\ Meas} = \exp(Z_{95}\, S_{n\text{-day},\ Meas} - 0.5S^2_{n\text{-day},\ Meas})$$
$$VF_{99,\ 4\text{-day},\ True} = \exp(Z_{99}\, S_{4\text{-day},\ True} - 0.5S^2_{4\text{-day},\ True})$$

while
$$S^2_{Meas} = \log(1 + CV^2_{Meas})$$
$$S^2_{True} = \log(1 + CV^2_{True})$$
$$S^2_{N\text{-day},\ Meas} = \log(1 + CV^2_{Meas}/N)$$
or
$$S^2_{N\text{-day},\ Meas} = S^2_{Meas}/N = \log(1 + CV^2_{Meas})/N$$
$$S^2_{4\text{-day},\ True} = \log(1 + CV^2_{True}/4)$$
or
$$S^2_{4\text{-day},\ True} = S^2_{True}/4 = \log(1 + CV^2_{True})/4$$

G.3.3 Consequences of Adjustment Approaches

As an example of the consequences of applying an adjustment methodology to water quality-based effluent limit calculations, one may consider the following scenario. In this scenario, such a methodology would allow calculation of an average monthly limit (AML) exceeding the chronic WLA (a four-day average value) even when sampling frequency for the calculation is set at the recommended minimum of four samples per month. It is acceptable for the MDL (a single sample) to exceed the chronic WLA or for the AML to exceed the chronic WLA if the AML calculation is based on less than four samples per month. Note, however, that the TSD recommends always assuming at least four samples per month when calculating the AML.

Table G-1 below offers an example of MDLs and AMLs calculated using the TSD approach and an approach that adjusts the CV for analytic variability. This adjustment would allow effluent limits that exceed the WLA on the premise that analytical variability tends to make measured values larger than actual effluent values. Thus, this approach assumes that the "true" monthly average would be below the WLAc even though the limit and the measured monthly average may be above the WLAc.

EPA believes that these assumptions are invalid. Therefore, EPA cannot recommend an approach that makes such assumptions as part of national guidance to regulatory authorities. EPA is not recommending national application of an "adjustment approach" to either reasonable potential or effluent limit calculation

procedures. EPA continues to recommend the TSD approach, which ensures that effluent limits and, thereby, *measured* effluent toxicity, are consistent with calculated WLAs.

Table G-1. Sample Effluent Limit Calculations Using EPA's TSD Approach and an Adjustment Approach (USEPA 1991a)

WLA_c	Probability Basis	Approach	LTA_c	MDL	AML
10	MDL = 99th percentile AML = 95th percentile	TSD	4.4	17.6	7.7
10	MDL = 99th percentile AML = 95th percentile	Adjustment approach	6.43	25.8	***11.2*** *
10	MDL = 99th percentile AML = 99th percentile	TSD	4.4	17.6	9.99
10	MDL = 99th percentile AML = 99th percentile	Adjustment approach	6.43	25.8	***14.6*** *

Assumptions: Chronic LTA/WLA controls calculations, WLA = 99th percentile probability basis, n = 4 (sampling frequency for AML calculation), Total CV = 0.8 and Adjusted CV = 0.4 are used in calculations.
(*) These numbers exceed the WLAc.

G.3.4 Related Concerns

In addition to addressing the differences between measured and "true" values in the reasonable potential and effluent limit calculations, related concerns regarding WET testing and the water quality-based effluent permits process have been raised as reasons for adjusting the TSD statistical procedures.

G.3.4.1 Compounding protective assumptions

Approaches to "account for analytical variability" by adjusting the calculations for reasonable potential and limits usually state that several conservative assumptions are employed. In the TSD approach, a water quality-based effluent limit is the result of three key components: (1) a criterion concentration; (2) a calculated dilution or mixing-zone factor; and (3) a statistical calculation procedure that employs a CV based on effluent data. The conservative assumptions cited may involve deriving the criterion concentration, and assuming dilution and low-flow conditions, in addition to the probability levels used in the TSD statistical calculations. Even if these assumptions were considered conservative, the TSD statistical procedure remains valid. As explained above, the TSD statistical approach is *appropriately* protective, provided that the WLA is accepted as given. It is inappropriate for regulatory authorities to modify the TSD's correctly conceived statistical approach in order to compensate for assumptions intrinsic to derivation of the WLA that are perceived as over protective. Therefore, EPA does not believe that it is appropriate to adjust the TSD statistical methodology for conducting reasonable potential and calculating permit limits to address concerns about how WLAs are calculated.

G.3.4.2 Test sensitivity and method detection limit

EPA does not employ method detection limits (MDLs: 40 CFR part 136 Appendix B) for WET methods. For effect concentrations derived by a hypothesis test (LOEC and NOEC), the alpha level of the test provides one means of providing a functional equivalent of an MDL. The hypothesis test prescribed in the method provides a high level of protection from "false positives." For point estimates (ECp, ICp, LCp), a valid confidence region provides the equivalent of a hypothesis test. EPA will provide clarification regarding when confidence intervals are not or cannot be generated for point estimation procedures, including the ICp procedure. This variability guidance cites recommendations (Chapman et al. 1996a, Baird et al. 1996, Bailer et al. 2000) regarding alternative point estimation methodologies.

While protecting against false positives, hypothesis tests and confidence intervals, will provide little protection from toxicity unless the test method is designed to detect a suitable effect size. The two most commonly used chronic tests are incapable of routinely detecting effects of 20 percent to 30 percent (Denton and Norberg-King 1996) when employed by many laboratories using the minimum recommended number of replicates and treatments. To provide suitable test sensitivity, regulatory authorities should consider requiring more replication, a suitable minimum significant difference (MSD), or suitable effect sizes and power, particularly for the control and IWC test concentrations (e.g., Denton and Norberg-King 1996; Washington State Department of Ecology 1997, Ch. 173-205 WAC). It may be desirable to specify that a statistically significant effect at the IWC must exceed some percentage difference from the control before it is deemed to have regulatory significance. Combining these approaches, an effective strategy would require that a test consistently be able to detect the smallest effect size (percent difference between the control and the IWC) that would compromise aquatic life protection, and to disregard very small, statistically significant effects. To further these ends, this guidance document sets an upper limit to the value of MSD/(Control Mean), defining the maximum acceptable value. This document also sets a lower limit to the effect size, defined by 100×(Control Mean - Treatment Mean)/(Control Mean), which can be regarded as "toxic" in a practical sense (see Section 6.4).

The alpha level of a hypothesis test or confidence interval cannot be decreased from that level (α = 0.05) recommended for WET methods without sacrificing test power and sensitivity of the method. Alpha should not be decreased without a corresponding increase in sample size that would preserve the power to detect biologically significant effects. EPA will issue guidance on when the nominal error rate (alpha level) may be adjusted in the hypothesis test for some promulgated WET methods (USEPA 2000a).

G.4 Technical Notes on Methods of Adjusting For Analytical Variability

This section describes and comments on several adjustment methodologies suggested to EPA as alternatives to the TSD statistical calculations.

G.4.1 Notation

Explanations may help clarify the notations in this section. The symbols VX, V[X], and σ^2_X all mean: the variance of X. Standard deviation (σ_X) is the square root of the variance. The mean (average) is symbolized as EX and also as μ_X.

When X is lognormally distributed, there is a potential for confusing the mean and variance of log(X) with the mean and variance of X. Typically (and in the TSD), when X is lognormally distributed, the parameters will be given for log(X) as follows: X ~ lnorm(μ, σ). This is read as "X is distributed lognormally with the mean of log X equal to μ (mu) and the standard deviation of log X equal to σ (sigma)." Better notation would be X ~ lnorm(μ_{logX}, σ_{logX}); recommended terms for the parameters are "mu-logX" and "sigma-logX." The mean and variance of X for this distribution are

$$\mu_X = EX = \exp(\mu_{logX} + 0.5*\sigma^2_{logX})$$
$$\sigma^2_X = VX = \exp(2*\mu_{logX} + \sigma^2_{logX}) * [\exp(\sigma^2_{logX}) - 1]$$

To avoid confusion, the symbols EX and VX are used in preference to μ_X and σ^2_X to signify the mean and variance of X. Usually, mu and sigma are used only as symbols for the mean and standard deviation of log(X), that is, μ_{logX} and σ_{logX}, in the context of lognormal distributions. Below, μ_{logX} and σ_{logX} are abbreviated to μ and σ, with the addition of subscripts like "Effl" and "Meas" to further distinguish the intended quantity.

CV may be used to symbolize parametric values or their sample estimates, with the meaning indicated in the text. Symbols S^2_{Effl} , S^2_{Meas} , and S^2_{Analy} will represent sample estimates of variances $\sigma^2_{logX, Effl}$, $\sigma^2_{logX, Meas}$, and $\sigma^2_{logX, Analy}$.

G.4.2 General Comments on Analytical Variance as a Component of the Variance of Measurements

Two simple models lead to the same equation. The first model assumes that each measurement Y is the sum of a concentration X and an analytical error ϵ, that is $Y = X + \epsilon$. The analytical error ϵ may be positive or negative and has mean zero and variance V_A. X and ϵ are uncorrelated. (This is a strong assumption; it may be approximately correct only for some transformation of the data.) Then $V_Y = V_X + V_A$. The second, hierarchical, model assumes that X follows a distribution P_X with mean and variance E_X and V_X. Each measurement Yt (t indexes the time of measurement) follows another distribution having mean Xt and variance V_A. V_A is assumed to be constant, independent of Xt. (This is a strong assumption which may be approximately correct only for some transformation of the data.) Then, it can be shown that $V_Y = V_X + V_A$. The same models and assumptions lead to EY = EX. These models and assumptions are not correct when X is lognormally distributed. In that case, the models might provide reasonable approximations to the behavior of log(X) and log(Y). If EY = EX and $V_Y = V_X + V_A$ are both correct, then $V_Y = V_X + V_A$ can be divided by EX^2 to give $CV_Y^2 = CV_X^2 + CV_A^2$. In this case, the parameters V_X and CV_X^2 might be estimated by using sample estimates in the expressions $(V_Y - V_A)$ and $(CV_X^2 - CV_A^2)$, respectively. Such estimates will be somewhat biased.

G.4.3 Commonwealth of Virginia Approach

The Commonwealth of Virginia Toxics Management Program Implementation Guidance (1993) (revised on August 25, 1994) prescribes a method of accounting for analytical variability of WET data. A synopsis of the method follows. Symbolic notation has been changed; the numbered "steps" below were created for this synopsis.

1. Obtain the CV of WET monitoring data. This will be 0.6 (default value) if fewer than ten data are available. If there are at least ten data, a computer program (*described in Guidance Memo 93-015*) is used. *"Only acute test data are considered here because the LC_{50} is a statistically derived point estimate from a continuous data set. Also, the LC_{50}s must be real numbers. Values reported as '> 100%' should not be used in the calculation. .... Enter either LC_{50}s or TU_as for <u>the most sensitive species</u> into the program."* [Comments on Step 1: LC50 and TU values are not equivalent; they will not have the same CV values. The exclusion of ">100%" values will tend to bias the CV of TUs toward larger values.]

2. Calculate $S^2_{logX,\ Effl} = S^2_{logX,\ Meas} + S^2_{logX,\ Analy}$, using $S^2_{logX,\ Analy} = 0.20$. If $CV_{X,\ Meas} < 0.47$ (implying that $S^2_{logX,\ Meas} < 0.20 = S^2_{logX,\ Analy}$), instead use $S^2_{logX,\ Effl} = S^2_{logX,\ Meas}$. (These subscripts are not used in the Guide.) The value for $S^2_{logX,\ Analy}$ is based on data provided by several laboratories conducting tests for Virginia permits for the five most common species, using cadmium chloride as the reference toxicant. The Guide states that these data yielded a geometric mean CV_X of 0.47, and $0.20 = \ln(1 + 0.47^2)$; the last formula is the relation between the parametic variance and CV of a lognormal variate. [Comments on Step 2: The calculations should employ sample variances of log(TU), not sample CVs, in the interest of accuracy and precision. The estimate $S^2_{logX,Effl}$ is a discontinuous function, decreasing toward zero as $S^2_{logX,\ Meas}$ decreases toward 0.2, then jumping to 0.2 and decreasing again toward zero as $S^2_{logX,\ Meas}$ decreases further. The default value of $S^2_{logX,\ Effl}$ becomes $\ln(1 + 0.60^2) - \ln(1 + 0.47^2) = 0.11$.]

3. Calculate LTAa,c and LTAc as in the TSD, using $S^2_{logX,\ Effl}$ instead of $S^2_{logX,\ Meas}$, and using Z_{97}, the 97th percentile Z-statistic, instead of Z_{99}. WLA and LTA values are in units of TUc. The smaller of LTAa,c and LTAc is selected as LTA_{min}.

4. Calculate the "MDL" limit from LTA_{min} as in the TSD, now using $S^2_{logX,\ Meas}$ rather than $S^2_{logX,\ Effl}$ and still using the 97th percentile Z-statistic. No procedure is described for a limit of averages ("AML").

   By using this procedure, the WLAa,c may be exceeded when the CV of measurements exceeds 0.47 (because then the estimate $S^2_{logX,\ Effl} < S^2_{logX,\ Meas}$). The maximum ratio of Limit to WLA occurs when the CV of observations is just over 0.47, when the ratio of Limit to WLA is just over 2. Numerical evaluations (Table G-2) show that the daily limit can exceed the WLAa,c. The daily limit (DL or MDL) should be compared to the WLAa,c. It is not unusual for the daily limit to exceed the WLAc when LTAc is smaller than LTAa,c. This outcome does not necessarily indicate a problem. Instead, the regulatory authority should compare the average limit to WLAc in this case (see "Modified TSD Approach" below).

Table G-2. Numerical Effect of State of Virginia WET Limit Calculation on Ratio of Daily Limit to WLA

CV_{Meas}	S^2_{Effl}	$S^2_{Effl,\ 4\text{-day average}}$	Ratio of Daily Limit to WLAa,c	Ratio of Daily Limit to WLAc
0.10	0.01	0.00	1.00	1.09
0.20	0.04	0.01	1.00	1.19
0.30	0.09	0.02	1.00	1.27
0.40	0.15	0.04	1.00	1.35
0.45	0.18	0.05	1.00	1.38
0.470	0.1996	0.0538	2.097	1.393
0.471	0.0004	0.0002	2.026	2.042
0.50	0.02	0.01	1.65	1.87
0.60	0.11	0.03	1.39	1.76
0.70	0.20	0.06	1.28	1.74
0.80	0.29	0.09	1.22	1.72
0.90	0.39	0.13	1.18	1.71
1.00	0.49	0.17	1.16	1.70

The State of Virginia Guide, Appendix D, also states: "*Because the statistical approach evaluates both acute and chronic toxicity of the effluent, only one limit is necessary to protect from both acute and chronic toxicity. The limit is expressed only as a maximum daily limit (MDL) because the frequency of monitoring will typically be less than once per month. If the testing is to be monthly, then the MDL can also be expressed as an average monthly limit (AML).*" [Comment: a single MDL limit is not as protective as the combination of limits, one for single observations (MDL) and another for averages (for example, the quarterly or annual average). Refer to the TSD (USEPA 1991a, Section 5.3).]

G.4.4 Rice Approach

James K. Rice's unpublished draft, "Laboratory QC and the Regulatory Environment: Relation Between Method Performance and Compliance" prescribes a method of accounting for analytical variability of WET data. The document was provided with a notation that the typescript was originally submitted to EPA as a comment on the draft "TSD," presumably in the period 1989 to 1991. A synopsis of the method follows. The numbered "steps" below were created for this synopsis. Calculations and symbols have been

simplified. This synopsis omits many detailed observations that provide context and guidelines for readers intending to apply Rice's method.

1. Obtain the CV of WET monitoring data (measured values), and the CV of the analytical method, in symbols $CV_{X, Meas}$ and $CV_{X, Analy}$. Sample size is not addressed, but the text indicates that "a large number" of measurements are needed to characterize variability and bias.

2. Solve for $CV_{X, Effl}^2$ in $CV_{X, Meas}^2 = CV_{X, Analy}^2 + CV_{X, Ttue}^2 + (CV_{X, Analy}^2 * CV_{X, Effl}^2)$, after substituting the sample estimates of $CV_{X, Meas}^2$ and $CV_{X, Analy}^2$. Thus, solve

   $CV_{X, Effl}^2 = (CV_{X, Meas}^2 - CV_{X, Analy}^2) / (1 + CV_{X, Analy}^2)$.

   [Comment: This formula assumes a model such as Measurement = (Concentration * Recovery), with multiplicative errors for Concentration and Recovery. This is one plausible model, especially for data that are distributed lognormally. Another plausible model would lead to the formula $CV_{X, Meas}^2 = CV_{X, Analy}^2 + CV_{X, Ttue}^2$.]

3. Calculate LTA values as in the TSD, using $CV_{X, Effl}$ instead of $CV_{X, Meas}$, and use Z_{99}, the 99th percentile Z-statistic. First calculate $\sigma^2_{logX, Effl} = \ln(1 + CV_{X, Effl}^2)$ for the variance of log(TU), and $\sigma^2_{logX, Effl, n} = \ln(1 + (CV_{X, Effl}^2)/n)$ for an n-day average. Then $LTA_{Effl} = WLA * \exp(0.5\sigma^2_{logX, Effl, n} - Z_P\, \sigma_{logX, Effl, n})$. Rice then calculates $LTA_{meas} = (R/100) * LTA_{Effl}$, where R is the percent recovery of the analytical method. [Comments: Many chemical methods are now calibrated instrumentally so that E[R] = 100 percent. It will be assumed herein that R = 100 percent for WET methods. There is no discussion of, or accounting for, the sampling error (the uncertainty) that attends the estimates of R or σ^2, of the sample sizes required to estimate these well. The example does not encompass the derivation and comparison of acute versus chronic LTAs using estimates of the variance of single observations and averages and selection of the smaller one, as in the 1991 TSD. Rice's method could easily be modified for the current TSD approach (see for example, the State of Virginia method, above).

4. Calculate the MDL and AML limits from the LTA as in the TSD, now using $\sigma^2_{logX, Meas}$ rather than $\sigma^2_{logX, Effl}$, and using the 99th percentile Z-statistic. Thus,

   $$MDL = LTA_{meas} * \exp(-0.5\sigma^2_{logX, Meas, 1} + Z_P\, \sigma_{logX, Meas, 1})$$
   $$AML_n = LTA_{meas} * \exp(-0.5\sigma^2_{logX, Meas, n} + Z_P\, \sigma_{logX, Meas, n})$$

   Using this procedure, the limits exceed the WLAc.

   $$MDL = WLAc * (VF_{.99, 1, Meas} / VF_{.99, 4, Effl}) > WLAc$$
   $$AML_n = WLAc * (VF_{.99, n, Meas} / VF_{.99, 4, Effl}) > WLAc \text{ if } n \le 4$$

   The AML can exceed WLAc even if n >4, depending upon the variance values. Because the current TSD approach of comparing LTAa,c and the LTAc had not been developed by the time of Rice's report, he did not apply his procedure to the WLAa,c.

G.4.5 Amelia River Report

The Amelia River Report (USEPA 1987, Appendix G) describes a similar approach, estimating $S^2_{logX, Effl} = S^2_{logX, Meas} + S^2_{logX, Analy}$ (without any provision for the case $S^2_{logX, Meas} \le S^2_{logX, Analy}$), calculating LTA from WLA using $S^2_{logX, Effl}$, and calculating the limits using $S^2_{logX, Meas}$.

G.4.5.1 Modified TSD approach

The methods described above predate the current TSD statistical approach and differ from it. As noted in the previous section, one could consider how the current TSD statistical approach could be modified to account for analytical variability using the same principles. The LTAs would be calculated using a variance estimate $S^2_{Effl} = S^2_{Meas} - S^2_{Analy}$, the smallest would be selected, and limits would be calculated from the smaller LTA using S^2_{Meas}. Table G-3 compares the current and modified calculations for whole effluent toxicity. Numerical calculations appear in Tables G-4 and G-5.

Table G-3. A Comparison of the Current TSD Calculation of Limits with a Modification That Takes into Account the Analytical Variability

Method	Smallest LTA	Limits
TSD statistical approach	LTAa,c	MDL = WLAa,c ($VF_{.99, 1, Meas}$ / $VF_{.99, 1, Meas}$) = WLAa,c AML = WLAa,c ($VF_{.95, N, Meas}$ / $VF_{.99, 1, Meas}$) < WLAa,c
	LTAc	MDL = WLAc ($VF_{.99, 1, Meas}$ / $VF_{.99, 4, Meas}$) < or > WLAa,c AML = WLAc ($VF_{.95, N, Meas}$ / $VF_{.99, 4, Meas}$) < WLAc
TSD modified to use S^2_{Effl} to calculate LTA	LTAa,c	MDL = WLAa,c ($VF_{.99, 1, Meas}$ / $VF_{.99, 1, Effl}$) > WLAa,c AML = WLAa,c ($VF_{.95, N, Meas}$ / $VF_{.99, 1, Effl}$) < or > WLAa,c
	LTAc	MDL = WLAc ($VF_{.99, 1, Meas}$ / $VF_{.99, 4, Effl}$) < WLAc AML = WLAc ($VF_{.95, N, Meas}$ / $VF_{.99, 4, Effl}$) < or > WLAc

Symbols for estimates based on data (sample estimates):

S^2_{Meas} sample variance of natural logs of measured TUs
S^2_{Analy} sample variance of natural logs of measurements on the same or TU
S^2_{Effl} estimate of variance of natural logs of TUs
$S^2_{Effl} = S^2_{Meas} - S^2_{Analy}$

$VF_{P, N, xxxx} = \exp(Z_P S_{xxx, N} - 0.5 S^2_{xxx, N})$ estimates the ratio of the P-th percentile to the mean for a lognormal variate: the P-th percentile is $\exp(\mu + Z_P \sigma)$ and the mean is $\exp(\mu + 0.5\sigma^2)$. The mean of a 4-day average of lognormal observations is assumed to be lognormal (Kahn, H.D., and M.B. Rubin. 1989. Use of statistical methods in industrial water pollution control regulations in the United States. *Environmental Monitoring and Assessment* 12:129-148).

The variance estimates may change with and be a function of the TU.
"N" is the number of samples (measurements) intended for use in determining compliance with the average limit, not the number of data used to calculate the sample variances used in setting limits.
It can be shown that LTAc < LTAa,c implies that WLAc < WLAa,c

For WET, WLAa,c = WLAa * ACR. It is assumed that the variance of observations (S^2_{Meas}) equals or exceeds the analytical variance (S^2_{Analy}). Numerical comparisons appear in Tables G-2 to G-4.

Calculations in Tables G-4 and G-5 show the numerical effect of adjustment on permit limits in relation to the WLA. These tables show the ratio of the limit to the WLA. For these calculations, S^2_{Meas} was calculated as $\log(1 + CV^2_{Meas})$, while $S^2_{Meas, 4\text{-}day} = \log(1 + CV^2_{Meas}/4)$, giving slightly different numerical results than if $S^2_{Meas, 4\text{-}day} = S^2_{Meas}/4 = \log(1 + CV^2_{Meas})/4$. The first formula is prescribed in the TSD, Box 5-2 and Table 5-1. The tables show the combinations of CV values used for CV_{Meas} and CV_{Analy}. The variance of TUs was calculated as $S^2_{Effl} = S^2_{Meas} - S^2_{Analy}$ using $S^2_{Meas} = \log(1 + CV^2_{Meas})$ and $S^2_{Analy} = \log(1 + CV^2_{Analy})$.

Table G-4. Ratio of MDL to WLA-LTA from WLA and CV_{Effl} and Limit from LTA and CV_{meas}

	LTAac is Smallest Ratio is MDL:WLAa,c					**LTAc is Smallest Ratio is MDL:WLAc**				
	CV_{Analy}					**CV_{Analy}**				
CV_{Meas}	0.1	0.2	0.3	0.4	0.5	0.1	0.2	0.3	0.4	0.5
0.1	1.25	0.00	0.00	0.00	0.00	1.25	0.00	0.00	0.00	0.00
0.2	1.06	1.55	0.00	0.00	0.00	1.28	1.55	0.00	0.00	0.00
0.3	1.04	1.17	1.90	0.00	0.00	1.38	1.47	1.90	0.00	0.00
0.4	1.03	1.11	1.31	2.28	0.00	1.48	1.55	1.69	2.28	0.00
0.5	1.02	1.09	1.22	1.48	2.68	1.58	1.63	1.73	1.93	2.68
0.6	1.02	1.07	1.16	1.33	1.65	1.66	1.70	1.79	1.93	2.18
0.7	1.01	1.06	1.13	1.26	1.47	1.72	1.76	1.83	1.94	2.12
0.8	1.01	1.05	1.11	1.21	1.37	1.77	1.81	1.87	1.96	2.10
0.9	1.01	1.04	1.10	1.18	1.30	1.81	1.84	1.90	1.98	2.09
1.0	1.01	1.04	1.08	1.16	1.26	1.84	1.86	1.91	1.98	2.08

[a] The LTA was calculated using the WLA and $_{Cveffl}$. The limit was calculated using the LTA and $CV_{meas.}$

Table G-5. Ratio of AML to WLA

	LTAa,c is smallest ratio is AML:WLAa,c					**LTAc is smallest ratio is AML:WLAc**				
	CV_{Analy}					CV_{Analy}				
CV_{Meas}	0.1	0.2	0.3	0.4	0.5	0.1	0.2	0.3	0.4	0.5
0.1	1.08	0.00	0.00	0.00	0.00	1.08	0.00	0.00	0.00	0.00
0.2	0.80	1.17	0.00	0.00	0.00	0.96	1.17	0.00	0.00	0.00
0.3	0.69	0.78	1.26	0.00	0.00	0.92	0.98	1.26	0.00	0.00
0.4	0.61	0.66	0.78	1.36	0.00	0.89	0.93	1.01	1.36	0.00
0.5	0.55	0.59	0.66	0.80	1.45	0.85	0.88	0.94	1.05	1.45
0.6	0.51	0.53	0.58	0.66	0.82	0.83	0.85	0.89	0.96	1.08
0.7	0.47	0.49	0.53	0.58	0.68	0.80	0.82	0.85	0.90	0.98
0.8	0.44	0.46	0.49	0.53	0.60	0.77	0.79	0.82	0.86	0.92
0.9	0.42	0.43	0.45	0.49	0.54	0.75	0.76	0.79	0.82	0.87
1.0	0.40	0.41	0.43	0.46	0.50	0.73	0.74	0.76	0.79	0.83

NOTE: If the AML were set at a 99th percentile value, all ratios would exceed 1.00. It is not surprising that the ratio in the table for AML is less than 1, should not come close to one, because the 95th percentile was used in the second part of the equation. The ratio should be constantly less than one in order to protect water quality criteria.

[a] The LTA was calculated using the WLA and $_{Cveffl}$. The limit was calculated using the LTA and $CV_{meas.}$