United States
Environmental Protection
Agency

Office of Water
(4303)

EPA 821-B-00-004
July 2000

Method Guidance and Recommendations for Whole Effluent Toxicity (WET) Testing (40 CFR Part 136)

Table of Contents

Tables

Figures

Disclaimer

This document, *Method Guidance and Recommendations for Whole Effluent Toxicity (WET) Testing (40 CFR Part 136)*, is provided to help implement national water quality-based permitting under the National Pollutant Discharge Elimination System (NPDES) Program. This guidance document does not, however, substitute for the Clean Water Act (CWA) or EPA's regulations, nor is it a regulation itself. Thus, it cannot impose legally binding requirements on EPA, States, Tribes, or the regulated community and may not apply to a particular situation based upon case-specific circumstances. The material presented herein is intended solely for guidance and does not alter any statutory or regulatory requirements, or requirements in an NPDES permit. EPA, State, and Tribal decision makers retain the discretion to adopt approaches on a case-by-case basis that differ from this guidance where appropriate. EPA may change this guidance in the future.

Executive Summary

In 1995, the U.S. Environmental Protection Agency (EPA) published a final rule standardizing 17 whole effluent toxicity (WET) test methods for use in NPDES (National Pollutant Discharge Elimination System) monitoring [60 FR 53529; October 16, 1995]. These WET test methods measure the aggregate acute and chronic toxicity of an effluent using standardized freshwater, marine, and estuarine plants, invertebrates, and vertebrates. The inclusion of WET methods in the NPDES program completes an integrated strategy for water quality-based toxics control that fulfills the Clean Water Act's mandate to protect aquatic life and prohibit the discharge of toxic pollutants in toxic amounts.

This document provides guidance and recommendations on the conduct of the approved WET test methods and interpretation of WET test results reported under the NPDES program. This guidance partially fulfills the obligations of a legal settlement agreement that resolves a judicial challenge to the WET final rule. The document provides guidance on the following issues: nominal error rate adjustments, confidence intervals, concentration-response relationships, dilution series, and dilution waters. A summary of the guidance and recommendations for each issue is provided below.

- **Nominal error rate adjustments** - The WET method manuals (USEPA, 1993c; USEPA, 1994a; USEPA, 1994b) recommend a nominal error rate (or alpha level) of 0.05 when using hypothesis testing to determine test results. This guidance clarifies that alpha may be reduced to 0.01 when sublethal endpoints from *Ceriodaphnia* or fathead minnow tests are reported under NPDES permit requirements, or when WET permit limits are derived without allowing for receiving water dilution. In these situations, however, the alpha level should be reduced only in tests that meet a set criterion for test sensitivity, since reductions in alpha also reduce statistical power. Specifically, the percent minimum significant difference (%MSD) calculated for the test using an alpha of 0.01 should be less than or equal to a set criterion. Increased replication may be necessary to meet the %MSD criterion when using an alpha of 0.01. This document provides guidance on determining the need for additional test replication, as well as the entire decision process for reducing the alpha level in hypothesis testing.
- **Confidence intervals** - Point estimation techniques described in the WET method manuals are used to generate effect concentrations and associated 95% confidence intervals. Software used to conduct these statistical procedures occasionally does not provide the associated confidence intervals. This may arise when the test data are inappropriate for the assumptions or requirements of the statistical method chosen. In these cases, statistical flowcharts provided in the WET method manuals should guide the analyst to more appropriate techniques. Confidence intervals also may not be generated if the calculated point estimate is outside of the test concentration range. In this case, confidence intervals are not applicable because exact point estimates are not reported. For the inhibition concentration percentage (ICp) procedure, there are additional

anomalous circumstances when confidence intervals are not generated due to limitations of the software.

- **Concentration-response relationships** - The concentration-response relationship established between the concentration of a toxicant and magnitude of the response, is a fundamental principle of toxicology. EPA recommends the use of this concentration-response concept as a test review step to assist in determining the validity of WET test results. When unexpected concentration-response relationships are encountered, a thorough review of test performance, test conditions, and the particular concentration-response pattern exhibited should be conducted to determine whether the derived effect concentrations are reliable or anomalous. This document recommends review steps for 10 different concentration-response patterns that may be encountered in WET test data. Based on the review, it may be determined that calculated effect concentrations are reliable and should be reported, that calculated effect concentrations are anomalous and should be explained, or that the test was inconclusive and the sample should be retested.
- **Dilution series** - This guidance clarifies that the WET method manuals do not require the use of a specific dilution series for all WET tests. The dilution series for a specific test should be selected to optimize the precision of calculated effect concentrations and assist in establishing concentration-response relationships. Recommendations for selecting an appropriate dilution series include: considering historic WET testing information for the given effluent, using the receiving water concentration as a test concentration, bracketing the receiving water concentration with test concentrations, adding test concentrations within a given range of interest, and increasing the dilution factor used to space effluent concentrations.
- **Dilution waters** - This guidance clarifies that an acceptable dilution water for WET testing is appropriate for the objectives of the test; supports adequate performance of the test organisms with respect to survival, growth, reproduction, or other responses that may be measured in the test (i.e., consistently meets test acceptability criteria for control responses); is consistent in quality; and does not contain contaminants that could produce toxicity. If the objective of the test is to determine the absolute toxicity of an effluent, EPA recommends the use of a standard synthetic dilution water. A consistent, high purity natural water source (e.g., uncontaminated seawater or treated well water) also may be appropriate for determining the absolute toxicity of an effluent when specific criteria given in this guidance are met. If the objective of the test is to determine the toxicity of an effluent in the receiving system, a local receiving water is recommended for use as dilution water provided that the receiving water meets specific criteria. The receiving water should be collected as a grab sample from upstream or near the final point of effluent discharge, have adequate year-round flow, support adequate performance of the test organisms, be consistent in quality, be free of contaminants that would produce toxicity, and be free from pathogens and parasites that could affect WET test results. If the local receiving water fails to meet any of these criteria for use, a synthetic dilution water adjusted to approximate the chemical characteristics of the receiving water is recommended.

1 Introduction

This chapter provides a brief introduction to whole effluent toxicity (WET) testing and describes the regulatory background and context of WET testing. This chapter also describes the purpose of this document and outlines the issues addressed in each chapter.

What is whole effluent toxicity (WET) and how is it measured?

Whole effluent toxicity (WET) is defined as "the aggregate toxic effect of an effluent measured directly by an aquatic toxicity test" [54 FR 23868 at 23895; June 2, 1989]. Aquatic toxicity test methods designed specifically for measuring WET have been codified at 40 CFR part 136 [60 FR 53529; October 16, 1995]. These WET test methods employ a suite of standardized freshwater, marine, and estuarine plants, invertebrates, and vertebrates to estimate acute and short-term chronic toxicity of effluents and receiving waters. Specific test procedures for conducting the approved WET tests are included in the following three test method manuals:

- U.S. Environmental Protection Agency. 1993c. *Methods for Measuring the Acute Toxicity of Effluents and Receiving Waters to Freshwater and Marine Organisms,* 4th ed., EPA 600/4-90/027F. U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Cincinnati, OH.
- U.S. Environmental Protection Agency. 1994. *Short-term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Freshwater Organisms,* 3rd ed., EPA 600/4-91/002. U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Cincinnati, OH.
- U.S. Environmental Protection Agency. 1994. *Short-term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Marine and Estuarine Organisms*, 2nd ed., EPA 600/4-91/003. U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Cincinnati, OH.

These three method manuals (WET method manuals) were incorporated by reference into 40 CFR part 136 in the 1995 rule. As regulations, use of these methods and adherence to the specific test procedures outlined in the WET method manuals is required when monitoring WET under the National Pollutant Discharge Elimination System (NPDES). Of course, the extent that such procedures are "requirements" depends on the text of the WET method manuals themselves. Words of obligation, such as "must" or "shall" indicate a required procedure. When WET method manuals use discretionary terms such as "may" or "should" the manuals provide flexibility so that the laboratory analyst may optimize successful test completion (USEPA, 1996a).

What is the regulatory background of WET testing?

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." Along with other specific goals, CWA section 101(a)(3) states that "it is the national policy that the discharge of toxic pollutants in toxic amounts be prohibited." EPA has pursued this goal through the implementation of the water quality standards program and the NPDES permitting program. These programs have adopted an integrated strategy of water quality-based toxics control that includes the following approaches:

- Chemical-specific control approach
- Whole effluent toxicity (WET) control approach
- Biological criteria/bioassessment and biosurvey approach

To implement this strategy, States and Tribes are encouraged to define numeric or narrative water quality standards that include chemical-specific criteria, criteria for whole effluent toxicity, and biological criteria. Some states have included numeric criteria for WET, while others have relied on narrative criteria such as, "free from toxics in toxic amounts". These water quality standards and criteria are maintained by controlling the discharge of pollutants through the NPDES permitting program. When a discharge causes or has a reasonable potential to cause or contribute to the excursion of numeric or narrative water quality standards, a water quality-based effluent limit in the NPDES permit will be issued to control the discharge. This includes permit limits for WET if the discharge causes, has a reasonable potential to cause, or contributes to the excursion of water quality standards for WET, including narrative criteria for toxicity.

Further explanation of the regulatory role and background of WET can be found in the WET method manuals (USEPA, 1993c; USEPA, 1994a; USEPA, 1994b) and in EPA's *Technical Support Document for Water Quality-based Toxics Control* (USEPA, 1991b).

What is the purpose of this document?

This guidance is intended to clarify the published WET method manuals on selected issues regarding the conduct of WET tests and interpretation of WET test results. This document provides additional guidance and recommendations to EPA Regional, State, Tribal, and local regulatory authorities; regulated entities; and environmental laboratories on these selected issues. Proper implementation of the guidance provided in this document should enhance successful WET test completion, result interpretation, and the application of WET testing in the NPDES program.

EPA developed this guidance document as part of efforts to resolve litigation over the rulemaking that standardized and approved the WET test methods for use in NPDES monitoring [60 FR 53529; October 16, 1995]. In a settlement agreement, EPA agreed to

provide guidance and recommendations on five specific technical issues. Each of these issues is addressed in a separate chapter of this guidance document.

- **Nominal error rate adjustments** - Chapter 2 explains the concept of a nominal error rate (or alpha level) and the effect of alpha on false positive rates, false negative rates, and test sensitivity. This chapter clarifies the circumstances when the alpha level for WET hypothesis testing may be reduced from 0.05 to 0.01. This chapter also provides guidance and recommendations for assuring that test sensitivity is not adversely affected by reductions in alpha. This guidance includes procedures for measuring test sensitivity, determining the need for additional test replication, and comparing test sensitivity to recommended criteria.
- **Confidence intervals** - Chapter 3 clarifies the circumstances under which confidence intervals are not generated and/or not capable of generation when using point estimation techniques.
- **Concentration-response relationships** - Chapter 4 explains the concept of a concentration-response relationship and describes how this concept may be used as a WET test review step. This chapter identifies various forms of concentration-response relationships encountered in WET testing and provides guidance on evaluating and interpreting results from these concentration-response relationships.
- **Dilution series selection** - Chapter 5 provides guidance on selecting appropriate dilution series for WET tests. This guidance provides recommendations for modifying the dilution series to assist in determining the existence of a concentration-response relationship and improving point estimate precision.
- **Dilution water** - Chapter 6 clarifies what EPA considers to be acceptable dilution water for WET testing. This chapter provides guidance on selecting an appropriate dilution water based on the objectives of the WET test and the quality and consistency of available dilution water sources. Guidance is provided regarding when to use the following waters for dilution: receiving water, standard synthetic water, and synthetic water adjusted to approximate receiving water characteristics. This chapter also clarifies the use of dual controls when dilution water differs from the water used to culture test organisms.

What other clarification and guidance documents has EPA published on WET?

The final WET methods rule [60 FR 53529; October 16, 1995] incorporated the WET method manuals (USEPA, 1993c; USEPA, 1994a; USEPA, 1994b) by reference. EPA provided further guidance and clarifications regarding the use of the WET test methods in a memorandum dated April 10, 1996 from Tudor Davies, Director of the EPA Office of Water's Office of Science and Technology. This memorandum, titled "Clarifications Regarding Flexibility in 40 CFR Part 136 Whole Effluent Toxicity (WET) Test Methods" (USEPA, 1996a), provided clarification on the following WET test issues: pH and ammonia control, temperature, hardness, test dilution concentrations, and acceptance criteria for *Champia parvula*.

In January 1999, EPA published an errata sheet for the WET method manuals (USEPA, 1999). This errata sheet amended the approved WET test methods to correct typographical errors and omissions, provide technical clarification, and establish consistency among the 1995 WET rule language and the three WET method manuals.

EPA has recently published a guidance document titled, *Understanding and Accounting for Method Variability in Whole Effluent Toxicity Applications Under the National Pollutant Discharge Elimination System Program*, (USEPA, 2000). This guidance document is intended to provide regulatory authorities with an understanding of WET test variability and provide guidance on accounting for and minimizing WET test variability and its effects on the regulatory process.

2 Nominal Error Rate Adjustments

The WET method manuals (USEPA, 1993c; USEPA, 1994a; USEPA, 1994b) recommend a nominal error rate (or alpha) of 0.05 when using hypothesis testing to determine WET test results. Under certain circumstances, it may be appropriate to reduce alpha to 0.01. This chapter provides an explanation of the concept and use of a nominal error rate and provides guidance on when alpha may be reduced.

When is a nominal error rate used?

A nominal error rate is used in the statistical method of hypothesis testing. According to the WET method manuals, effect concentrations for effluent toxicity tests may be generated by point estimation techniques or hypothesis testing techniques (see Section 9 of USEPA, 1994a; USEPA, 1994b). Point estimation techniques are used to generate effect concentrations such as LC50 (median lethal concentration), EC50 (median effect concentration), or IC25 (25% inhibition concentration) values. Hypothesis testing techniques are used to generate NOEC (No-Observed-Effect-Concentration) and LOEC (Lowest-Observed-Effect-Concentration) values. Both statistical techniques have advantages and disadvantages (Grothe *et al.*, 1996), and regulatory authorities may choose to base WET permit limits on effect concentrations generated using either technique. The WET method manuals (see Section 9 of USEPA, 1994a; USEPA, 1994b) state that point estimation techniques are the preferred statistical methods for calculating effect concentrations in WET tests under the NPDES permit program.

What is a nominal error rate?

The concept of hypothesis testing relies on the ability to distinguish statistically significant differences between a control treatment and other test treatments (e.g., effluent concentrations). In terms of classical statistics, the hypothesis testing techniques (whether Dunnett's Test, t-Test with Bonferroni adjustment, Steel's Many-One Rank Test, or Wilcoxon Rank Sum Test with Bonferroni adjustment) test the null hypothesis ($H_\varnothing$) that there is no difference between the control treatment and other test treatments (the effluent is not toxic). This null hypothesis is rejected (the effluent is determined to be toxic) if the difference between the control treatment and any other test treatment is statistically significant. In order to determine when the difference between treatments is large enough to be statistically significant and to warrant rejection of the null hypothesis, the statistician or analyst selects a nominal error rate. This nominal error rate is an intended upper bound on the probability of incorrectly rejecting the null hypothesis (determining that the effluent is toxic) when it is in fact true (the effluent is not toxic). In selecting the nominal error rate, the analyst is deciding what level of uncertainty

he/she is comfortable with in making this type of error (determining that the effluent is toxic when it is not). The larger the nominal error rate, the greater the probability of incorrectly rejecting the null hypothesis (determining that the effluent is toxic when in fact it is not). In classical statistics, the error of incorrectly rejecting the null hypothesis is termed a Type I error, and the nominal error rate selected to place an intended upper bound on the probability of this error is termed alpha (α). To remain consistent with statistical terminology, the nominal error rate will be referred to as alpha in the remainder of this document. An alpha of 0.05 means a 5% probability of making a Type I error and is associated with a 95% level of significance (i.e., on average 1 test in 20 tests could produce a Type I error).

How is the alpha level related to specific types of errors?

Figure 2.1 describes the possible correct and erroneous decisions that can be made in hypothesis testing. In making the decision to reject or accept the null hypothesis, two types of error are possible. An incorrect decision can be made by determining that a sample is toxic when in fact it is not (Type I error), or determining that a sample is not toxic when in fact it is (Type II error). These errors also may be commonly referred to as false positive error and false negative error, respectively. The alpha level that is selected by the statistician or analyst in a hypothesis test represents the probability of making a Type I error (or the Type I error rate). The probability of a Type II error (or the Type II error rate) is represented by beta (β).

Figure 2.1. Possible decisions and outcomes in the hypothesis test.

		True State of Nature	
		H∅ is true (sample is not toxic)	H∅ is false (sample is toxic)
Decision	Accept H∅ (determine that sample is not toxic)	Correct decision	Type II error (false negative)
	Reject H∅ (determine that sample is toxic)	Type I error (false positive)	Correct decision

There are direct and indirect costs associated with both types of errors. False positives can create undue costs and effort involved in follow-up actions such as increased testing, Toxicity Identification Evaluation (TIE) and Toxicity Reduction Evaluation (TRE) procedures, possible fines for permit violations, and the potential for civil lawsuits. False negatives can cause the continuation of unchecked environmental degradation and the associated long-term cost of reclamation or restoration. Researchers have suggested that false negatives may be more costly than false positives because false positives may be quickly discovered by additional

testing, while false negatives may continue longer before being discovered (Thursby *et al.*, 1997). Since there are costs associated with each type of error, neither type of error should be ignored, and an effort should be made to minimize both types of error. However, because of the relationship between the Type I error rate (α) and the Type II error rate (β), reductions in one type of error generally cause an increase in the other. For instance, when test variability and test design are held constant, reducing the alpha level of a test increases the Type II error rate (β). This reduces the statistical power (defined as 1-β) of the test and limits the ability of the test to detect small effects as statistically significant. Because costs exist for both types of error, it is important to consider the impact of both types of error before reducing alpha.

What alpha level is recommended in the WET method manuals?

Traditionally, scientists have set alpha for biological studies at 0.01 to 0.1 (1 to 10%). The 0.01 level, at one extreme, provides a statistically conservative error rate that minimizes false positives. The 0.1 level, at the other extreme, provides a statistically more liberal error rate that results in increased statistical power. Zar (1984) states that a probability of 5% or less is commonly used as a criterion for rejection of the H_0, and that when the 5% chance of an incorrect rejection of the hypothesis is unacceptably high, then a 1% level of significance is sometimes used. The WET test method manuals recommend an alpha of 0.05 for hypothesis testing (see Section 9 of USEPA 1994a; USEPA 1994b). The experimental test designs of the WET test methods (e.g., replicates, treatments, number of organisms) have limits to the magnitude of toxic response that they are able to detect given a specific alpha level (Denton and Norberg-King, 1996; USEPA, 2000); smaller effects will generally not be detected. If the recommended test alpha level is reduced, the experimental test design may need modification (e.g. increased test replication) to maintain the same level of test sensitivity.

When can alpha be reduced?

The alpha level used for hypothesis testing in WET data analysis may be reduced from 0.05 to 0.01 when:

- sublethal endpoints (reproduction or growth) from *Ceriodaphnia dubia* or fathead minnow tests are reported under NPDES permit requirements, or
- the NPDES permit limit for WET was derived without allowing for receiving water dilution due to low dilution potential in the receiving system,

provided that the WET test is able to maintain adequate test sensitivity (as demonstrated by successfully meeting a set criterion for minimum significant differences [MSDs]) using an alpha of 0.01.

When should alpha not be reduced?

The alpha level of a test should not be reduced unless the regulatory authority allows or specifies an alpha of 0.01 in the NPDES permit (see "What is the recommended decision process for determining the appropriate alpha level?"). The alpha level of a test also should not be reduced if the test does not maintain adequate test sensitivity. This determination is made by comparing the test MSD (calculated using the reduced alpha of 0.01) to recommended maximum MSD levels (see "How can adequate test sensitivity be confirmed?"). If the test MSD (calculated using the reduced alpha of 0.01) is greater than the MSD criterion, alpha should not be reduced to 0.01, and results should be reported using the standard alpha level of 0.05.

How can adequate test sensitivity be confirmed?

As described above, alpha may be reduced only when the test maintains adequate test sensitivity. Adequate test sensitivity is determined by calculating the MSD for a given test and comparing this value to maximum MSD criteria. This procedure is described below.

- **Calculate test MSD** - To measure the sensitivity of the test, the minimum significant difference or MSD is calculated. The MSD is defined as the smallest difference between the control and another test treatment that can be determined as statistically significant in a given test. The MSD is a measure of statistical sensitivity that is dependent upon the within test variability, the alpha level selected for the test, and the test design (i.e., number of replicates and treatments). The MSD decreases (i.e., statistical sensitivity increases) with decreasing test variability, increased test replication, and increased alpha. According to the WET method manuals (USEPA, 1994a; USEPA, 1994b), the MSD may be calculated for Dunnett's multiple comparison test using the following equation:

$$MSD = d \times s_w \sqrt{\frac{1}{n_0} + \frac{1}{n_c}}$$

where:

d = Dunnett's t for the selected α and N - (k+1) degrees of freedom
s_w = square root of the error mean square from analysis of variance (ANOVA)
n_0 = number of replicates in the control
n_c = number of replicates for each effluent concentration
N = total number of replicates in the ANOVA
k = number of non-control treatments being compared to the control

The pooled variance estimate, s_w, is obtained from an analysis of variance (ANOVA). Test concentrations that exhibit 0% survival are excluded from the ANOVA for survival endpoints, and test concentrations greater than the NOEC for survival are excluded from the ANOVA for sublethal endpoints.

When the number of replicates is not the same for all test treatments, but variances are expected to be the same, the t-test with Bonferroni's adjustment is used for hypothesis tests (USEPA, 1994a; USEPA, 1994b). Under these circumstances, the MSD is calculated using the formula shown above, except that "*d*"is replaced by the standard t-statistic for a one-sided test at level 1-α/k, where *k* is the number of treatments being compared to the control. Further details and a table of critical values for t are provided in Appendix D of the WET method manuals (USEPA, 1994a; USEPA, 1994b).

The above equation (with the slight modification for unequal replicates, if needed) may be used to calculate the MSD for all tests in which results are derived from hypothesis testing, regardless of the hypothesis testing technique used (e.g., Dunnett's Test, t-test with Bonferroni adjustment, Steel's Many-One Rank Test, or Wilcoxon Rank Sum Test with Bonferroni adjustment). When a given data set does not meet the assumptions (e.g., normal distribution or homogeneous variance) necessary for the use of parametric hypothesis testing procedures (i.e., Dunnett's test or t-test with Bonferroni adjustment), the MSD still may be derived as described above for use as an approximate indicator of test sensitivity. However, when there are significant differences in variances among treatments, the best approach is to identify a variance-stabilizing transformation (preferably one which applies generally and not to just one test) and which leaves the treatment means approximately normal.

To facilitate the comparison of MSD values among tests and with established criteria, the MSD is generally expressed as a percentage of the mean control value for the given test. This transformation is conducted using the following equation:

$$\%\,MSD = \frac{MSD}{\text{Control mean}} \times 100\%$$

Other measures of test sensitivity, such as test power (1- β) also can be used to determine the statistical sensitivity of a test. However, the MSD is recommended in this guidance for determining the appropriateness of reducing alpha levels in hypothesis testing. The MSD is easily calculated and is generated by most statistical software packages used in WET test data analysis. In addition, the Pellston Workshop on Whole Effluent Toxicity (Chapman *et al.*, 1996; Denton and Norberg-King, 1996) and other researchers (Thursby *et al.*, 1997; Warren-Hicks *et al.*, 1999) recommend the use of MSDs to assure that acceptable statistical sensitivity is achieved. The MSD is currently used to access the acceptability of test sensitivity in the West Coast WET methods (USEPA, 1995), and criteria for acceptable MSD levels have been recommended for most of the approved WET test methods in a newly published EPA guidance document titled, *Understanding and Accounting for Method Variability in Whole Effluent Toxicity Applications Under the National Pollutant Discharge Elimination System Program* (USEPA, 2000).

- **Compare test MSD to maximum MSD criteria** - In EPA's recently published guidance document on WET method variability (USEPA, 2000), EPA recommends criteria for maximum MSD values in an effort to reduce method variability. EPA compiled a national database of WET reference toxicant test data from 75 laboratories and 23 test methods conducted over a 10-year period. EPA used these data to make inferences about WET test method variability and to evaluate recommendations for reducing variability. From an analysis of MSD values from these tests, it was determined that placing upper and lower bounds on MSDs improved test precision. Based on this finding, EPA recommended setting upper and lower limits for MSDs at the 10^{th} and 90^{th} percentiles of the MSD distribution compiled from this national database. Table 2.1 shows the recommended upper bounds on WET test MSDs for given test methods.

  EPA recommends that these maximum MSD criteria be met for all tests (USEPA, 2000), regardless of the alpha value used in hypothesis testing. Therefore, EPA recommends that alpha be decreased from 0.05 to 0.01 only when the test MSD (expressed as %MSD) calculated with the new, lower alpha (0.01) meets the criteria recommended in Table 2.1 (i.e., calculated test %MSD should be less than or equal to the value in Table 2.1 for the given method). If the calculated test %MSD is greater than the maximum criterion stated in Table 2.1, the test results should be reported using an alpha of 0.05. In order to meet these MSD criteria using an alpha of 0.01, additional test replication may be required (see Step 2 under "What is the recommended decision process for determining the appropriate alpha level?").

Table 2.1. Recommended maximum MSD (minimum significant difference) criteria for selected WET test methods and responses (adapted from Table 3-6 in USEPA, 2000).

WET test method	Biological Response	Maximum MSD Criterion (%MSD)
1000.0- Fathead Minnow, *Pimephales promelas*, Larval Survival and Growth Test	Growth	35
1002.0- Daphnid, *Ceriodaphnia dubia*, Survival and Reproduction Test	Reproduction	37
1003.0- Green Alga, *Selenastrum capricornutum*, Growth Test	Growth	23
1004.0- Sheepshead Minnow, *Cyprinodon variegatus*, Larval Survival and Growth Test	Growth	23
1006.0- Inland Silverside, *Menidia beryllina*, Larval Survival and Growth Test	Growth	35
1007.0- Mysid, *Mysidopsis bahia*, Survival, Growth, and Fecundity Test	Growth	32

What is the recommended decision process for determining the appropriate alpha level?

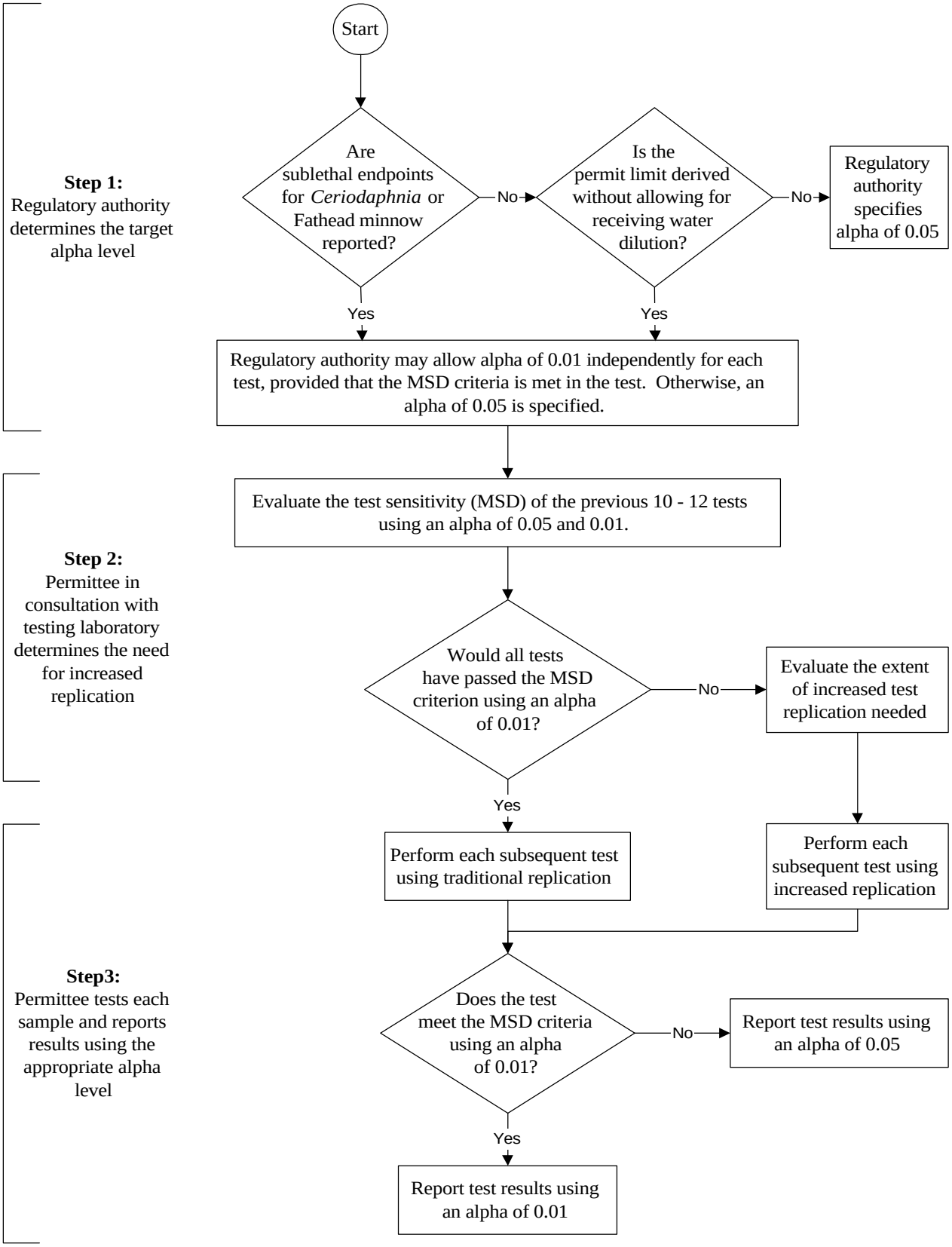
Figure 2.2 summarizes the recommended decision process for determining the appropriate alpha level for use in hypothesis testing. This figure is provided to assist regulatory authorities, permittees, and laboratories in this decision-making process. The recommended three-step decision process is described below.

- **Step 1** - In step one, the regulatory authority determines the target alpha level that will be specified in the permit. If either of the following circumstances apply, the regulatory authority may allow a target alpha of 0.01:
  - sublethal endpoints (reproduction or growth) from *Ceriodaphnia dubia* or fathead minnow tests are reported under NPDES permit requirements, or
  - the NPDES permit limit for WET was derived without allowing for receiving water dilution due to low dilution potential in the receiving system.

  The target alpha level is the alpha level that the analyst will attempt to use in the statistical analysis of test data for all samples of the given effluent. While a target alpha level may be specified for all tests, each test should be evaluated independently to determine if the target alpha level is appropriate (see Step 3). The regulatory authority should specify (as a permit condition) that when a target alpha level of 0.01 is allowed, the test MSD should not exceed the recommended MSD criterion for test sensitivity (Table 2.1). If the test fails to meet the MSD criterion using the target alpha level, results should be reported using the standard alpha of 0.05.

- **Step 2** - After the regulatory authority has determined that a target alpha level of 0.01 is allowable, the permittee should consult with the testing laboratory to determine if increased test replication is needed to meet the MSD criterion using the target alpha level. Since the MSD is a function of alpha, test variability, and test design (i.e., number of replicates and test treatments), an increase in the MSD caused by reducing alpha can be offset by an increase in test replication. Table 2.2 shows the increase in test replication needed to completely offset a reduction in alpha from 0.05 to 0.01. For instance, replication in the fathead minnow chronic test would need to be increased from four to seven replicates to maintain the same MSD level when alpha is decreased from 0.05 to 0.01 (assuming that variability remains constant).

Figure 2.2. Recommended decision process for determining the appropriate alpha level for WET hypothesis testing.



To determine the need for increased test replication, the permittee and testing laboratory should evaluate the laboratory's recent performance on tests with the given effluent. Laboratories that consistently conduct tests with low variability and high sensitivity (low MSDs) will require smaller increases in test replication than laboratories with high variability and low sensitivity (high MSDs). Laboratories should calculate MSDs for the previous 10 - 12 tests of the given effluent using an alpha of 0.05 and 0.01. While results from these tests already will have been reported using an alpha of 0.05, this exercise will provide the permittee with an idea of how often the laboratory might fail to meet the MSD criterion using the new, reduced alpha of 0.01. It is important that this evaluation is made using a single laboratory's performance (i.e., the laboratory that will perform testing with the new, reduced alpha) for the single effluent of interest. If all of the tests evaluated would have passed the MSD criterion using a reduced alpha of 0.01, then no increase in test replication will be necessary. If some of the tests evaluated would have failed the MSD criterion using a reduced alpha of 0.01, then increased test replication is needed.

Table 2.2. Number of within-treatment replicates giving equivalent MSDs (minimum significant differences) at alpha = 0.05 and 0.01, for a test employing five concentrations and a control.

Number of replicates for alpha = 0.05	Number of replicates for alpha = 0.01
3	5
4	7
5	8
6	10
7	11
8	13
9	15
10	16

If increased test replication is needed, the extent of the increase should be determined by calculating the replication needed to pass the MSD criterion in the least sensitive of the 10 previous tests evaluated. This level of within-treatment replication will be sufficient to meet the MSD criterion in approximately 90% of tests conducted. The following steps and calculations should be followed to determine the needed increase in test replication across all treatments. A hypothetical example using *Ceriodaphnia dubia* 3-brood reproduction test data from 10 tests (Table 2.3) illustrates this determination. When unequal replication among treatments is desired (e.g., more replicates in the

control treatment than in other treatments), consult Dunnett (1964) for optimizing the allocation of replicates between the control and other treatments.

1. Determine the least sensitive of the previous 10 tests - Tabulate the results from the previous 10 tests conducted on the effluent of interest by a single laboratory (Table 2.3). For each test, include the mean control response, the error mean square (EMS) from the ANOVA, and MSDs calculated using an alpha of 0.05 and 0.01. The test with the highest MSD calculated using an alpha of 0.01 should be considered the least sensitive test of those evaluated. If replication varied among the tests evaluated, the least sensitive test should be identified as the test with the largest ratio of EMS to control mean. In the example given (Table 2.3), 2 of the 10 tests (tests 7 and 9) failed to meet the MSD criterion of 37% (Table 2.1) when using an alpha of 0.01. Test 9 should be determined to be the least sensitive test since the MSD of 43.81% is the largest observed in the previous 10 tests. The following calculations will determine the additional replication that would be needed for this test to pass the MSD criterion.

Table 2.3. Example results from 10 previous *Ceriodaphnia dubia* 3-brood reproduction tests.

Test	%MSD with alpha = 0.05	%MSD with alpha = 0.01	Error Mean Square (EMS)	Control mean
1	20.78	26.82	24.98	24.6
2	16.50	21.29	16.14	24.9
3	20.12	26.273	28.97	26.6
4	23.82	30.75	19.18	18.8
5	23.94	30.90	31.57	24.0
6	26.32	34.94	26.53	18.7
7	29.53	38.11	29.78	18.9
8	17.75	22.90	18.52	24.8
9	33.94	43.81	68.31	24.9
10	18.38	23.73	15.07	22.2

2. Transform %MSD criterion to MSD - The MSD criterion that should be met for all tests (Table 2.1) is expressed as a %MSD. This %MSD should be transformed to a MSD using the control mean performance in the least sensitive of the previous 10 tests that are being evaluated. Perform this transformation using the following equation:

$$MSD_{\max} = \frac{\%\,MSD \times \text{Control mean}}{100\%}$$

where:

MSD_{max} = the MSD that should have been met in the least sensitive of the previous 10 tests
%MSD = the %MSD criterion (Table 2.1)
Control mean = the mean control response in the least sensitive of the previous 10 tests

For the example given, the control mean for test nine should be used in conjunction with the MSD criterion for the *Ceriodaphnia dubia* chronic test method (Table 2.1) to calculate the MSD_{max} as:

$$MSD_{max} = \frac{37\% \times 24.9}{100\%}$$

$$MSD_{max} = 9.213$$

3. Calculate the square root of the error mean square (s_w) - The error mean square (EMS) is a measure of test variability that is obtained from an ANOVA of test data. To evaluate increased replication needs, use the EMS calculated in the least sensitive of the previous 10 tests. Calculate the square root of this EMS to obtain the variable s_w that is used in the calculation of test MSDs. In the example given, the EMS from test nine should be used to calculate s_w as:

$$s_w = \sqrt{EMS}$$

$$s_w = \sqrt{68.31}$$

$$s_w = 8.265$$

4. Calculate the MSD using an increase in test replication - Using the equation below and Table 2.4, calculate the MSD with an alpha of 0.01 and assuming one additional replicate per treatment.

$$MSD = d \times s_w \sqrt{\frac{1}{n_0} + \frac{1}{n_c}}$$

where:

d = Dunnett's t obtained from Table 2.4 using an alpha of 0.01 and the increased number of replicates
s_w = square root of the error mean square from the least sensitive of the previous 10 tests
n_0 = increased number of replicates in the control
n_c = increased number of replicates for each effluent concentration

For the example given, the MSD first should be calculated with one additional replicate (10 original replicates + 1 additional replicate = 11 replicates) to obtain:

$$MSD = 2.940 \times 8.265 \sqrt{\frac{1}{11} + \frac{1}{11}}$$

$$MSD = 10.36$$

Table 2.4. Comparison of critical Dunnett's values for five concentrations and a control using alpha = 0.05 and 0.01.[1]

Number of replicates	Degrees of freedom	alpha = 0.05	alpha = 0.01
3	12	2.502	3.420
4	18	2.407	3.206
5	24	2.362	3.107
6	30	2.335	3.049
7	36	2.318	3.012
8	42	2.305	2.986
9	48	2.296	2.967
10	54	2.289	2.952
11	60	2.284	2.940
12	66	2.279	2.931
13	72	2.275	2.923
14	78	2.272	2.916
15	84	2.269	2.910
16	90	2.267	2.905
17	96	2.265	2.901
18	102	2.263	2.897
19	108	2.261	2.894
20	114	2.260	2.891

[1] Critical values were calculated using the Dunnett's procedure in SAS (SAS Institute, 1990). Critical values were determined using equal replication in five test concentrations and a control. Degrees of freedom were determined as N - (k+1), where, N = total number of replicates in the experiment, and k = number of non-control treatments.

5. Determine if the increased replication meets the MSD criterion - If the MSD calculated in the above step is less than or equal to the MSD_{max} calculated in step 2, then the number of replicates used in this calculation is the appropriate replication that should be used in future testing. If the MSD calculated in the above step is greater

than the MSD_{max}, then repeat step 4 using one additional replicate. Continue to repeat step 4, each time with an additional replicate, until the MSD is less than or equal to the MSD_{max} calculated in step 2.

For the example given, the MSD calculated with 11replicates (10.36) was larger than the MSD_{max} (9.213) calculated in step 2, so additional replicates are needed. The above equation is repeated using one additional replicate until the calculated MSD meets the criterion. For this example, the criterion is first met at a level of 14 replicates:

$$MSD = 2.916 \times 8.265 \sqrt{\frac{1}{14} + \frac{1}{14}}$$

$$MSD = 9.109$$

Based on the above calculations for this example, the laboratory should use 14 test replicates per treatment in future testing using an alpha of 0.01.

- **Step 3 -** After a target alpha level of 0.01 has been specified (Step 1) and a decision has been made regarding the need for increased test replication (Step 2), testing may begin using the target alpha level (0.01) and the revised test design (i.e., replication). For each test that is performed, the MSD should be calculated and compared to the MSD criterion (Table 2.1). If the test meets the MSD criterion, the results may be reported using the target alpha level (0.01). If the test does not meet the MSD criterion, the results should be reported using the traditional alpha of 0.05. If more than 1 in 10 tests fail to meet the criterion, the permittee should reconsider the need and extent of increased replication.

3 Confidence Intervals

The WET method manuals (USEPA, 1993c; USEPA, 1994a; USEPA, 1994b) provide specific directions for the derivation of effect concentrations from WET tests. Effect concentrations recommended for reporting results from WET tests are either based on hypothesis testing (NOEC, LOEC) or point estimation (LC50, EC50, IC25). Multiple effect concentrations are possible for each WET method. For example, the potential endpoints reported for the fathead minnow larval survival and growth chronic test include an IC25 for growth, NOEC for growth, LC50 for survival, and a NOEC for survival. For each type of endpoint, flowcharts in the WET method manuals guide the analyst to the proper choice of statistical methods based on assumptions and determinations that can be made from the data. The proper statistical method can then be performed using EPA or commercially available software to derive the desired effect concentration. For point estimation techniques (LC50, EC50, IC25) the statistical methods generally produce an effect concentration with associated 95% confidence intervals. However, under certain circumstances confidence intervals are not produced or they are unreliable. This chapter provides clarification and guidance on the circumstances under which confidence intervals are not generated or are not suitable. Currently, confidence intervals are not reported in the permit compliance system but may be used in interpreting results of WET tests. Statements in this method guidance document regarding software refer to current versions of software available from USEPA at the following web site address: http://www.epa.gov/nerleerd/stat2.htm.

When are confidence intervals not generated by point estimation techniques?

Point estimation techniques may fail to generate confidence intervals if:

- **Test data do not meet specific assumptions required by the statistical methods -** Under these circumstances, an alternate statistical method should be used as indicated in the flowcharts for statistical analysis provided in the WET method manuals. These flowcharts guide the analyst to the proper statistical technique based on the appropriateness of data assumptions. In order to obtain reliable point estimates and confidence intervals from the Probit method, it is required that the data contain at least two partial mortalities (i.e., percent mortalities between 0 and 100%) and that the slope differ significantly from zero. If the assumption of two partial mortalities is not met, the software will provide a warning and neither point estimates nor confidence intervals will be generated. If the slope does not differ significantly from zero, point estimates will be generated without confidence intervals. In either of two situations (less than two partial mortalities or a significant Chi-square test indicating lack of fit to the model), the analyst should resort to use of the Spearman-Karber or Trimmed Spearman-Karber methods as indicated by the flowcharts in the WET method manuals. The

Spearman-Karber and Trimmed Spearman-Karber methods require at least one partial mortality to calculate an effect concentration and associated confidence intervals. If this assumption is not met by the data, EPA's Trimmed Spearman Karber software will automatically default to the use of the Graphical Method for determining point estimates. Since the Graphical Method does not estimate confidence intervals, EPA's Trimmed Spearman Karber software will produce a point estimate without confidence intervals and state that 95% confidence limits are not calculated. For sublethal effects, the inhibition concentration percentage (ICp) procedure is recommended for determining effect concentrations. Data assumptions for the ICp method are not tested by the ICp software. Thus, failure of test data to meet assumptions of the ICp method does not result in a failure to generate point estimates or confidence intervals.

- **Point estimates are outside of the test concentration range -** The Probit method may not produce confidence intervals if the generated point estimate is greater than the highest test concentration. In this case, the software will provide a warning that the slope is not significantly different from zero. The Spearman-Karber and Trimmed Spearman-Karber methods will produce neither point estimates nor confidence intervals if the point estimate is outside of the test concentration range. In this case, the software will produce an error message stating that the required trim is too large. The ICp method will not generate confidence intervals if a point estimate is above the test concentration range. The software will produce a warning that none of the group response means were less than 75% of the control mean. Whenever a point estimate lies above the test concentration range, the test result should be reported as greater than the highest test concentration (e.g., IC25 >100% or LC50 >100%). Whenever a point estimate lies below the test concentration range, the test result should be reported as less than the lowest test concentration (e.g., IC25 <6.25% or LC50 <6.25%). Under these circumstances, confidence intervals are not applicable since exact point estimates are not reported.
- **Specific limitations imposed by the software are encountered** - The ICp software may fail to generate confidence intervals if the number of random resamplings of the data used in the bootstrapping technique is not a multiple of 40. This may occur when the analyst selects a number of resamplings that is not a multiple of 40, or it may occur if one or more of the random resamples is automatically removed from the analysis. The ICp software will automatically remove random resamples that produce effect concentrations above the highest test concentration. If this occurs, the software will produce an error message that states that the number of resamplings was not a multiple of 40. The occurrence of this error increases with increasing test variability, increases as the point estimate approaches the highest test concentration, and increases with an increasing number of random resamples selected. This anomaly is due to a limitation of the ICp software and not necessarily an inherent limitation of statistical bootstrapping techniques upon which the software is based. For this reason, EPA recommends that confidence intervals for the ICp method not be reported or used in WET testing until the ICp software has been thoroughly reviewed by experts and possibly modified. This recommendation should not affect NPDES reporting in the interim since confidence intervals are not currently reported in the permit compliance system.

In summary, the choice of statistical methods, the choice of software for analysis, and the appropriateness of test data for those methods and software is important in generating reliable results. Computer programs for WET data analysis, modifications to those programs, data appropriateness for the programs, and user decision points within the programs should be evaluated by a statistician to verify that use of the programs is consistent with the WET method manuals and current statistical science. Laboratory analysts and regulatory authorities should also recognize that confidence intervals from statistical programs should always be considered approximate. Confidence intervals may not provide the exact coverage intended because of deviations from method assumptions. Lastly, investigators should keep informed of additional and improved techniques and software for WET data analysis that may become available.

4 Concentration-Response Relationships

This chapter is designed to explain the concept of a concentration-response relationship. This chapter also identifies common patterns of WET test data and provides guidance on using the concentration-response concept to review WET test results.

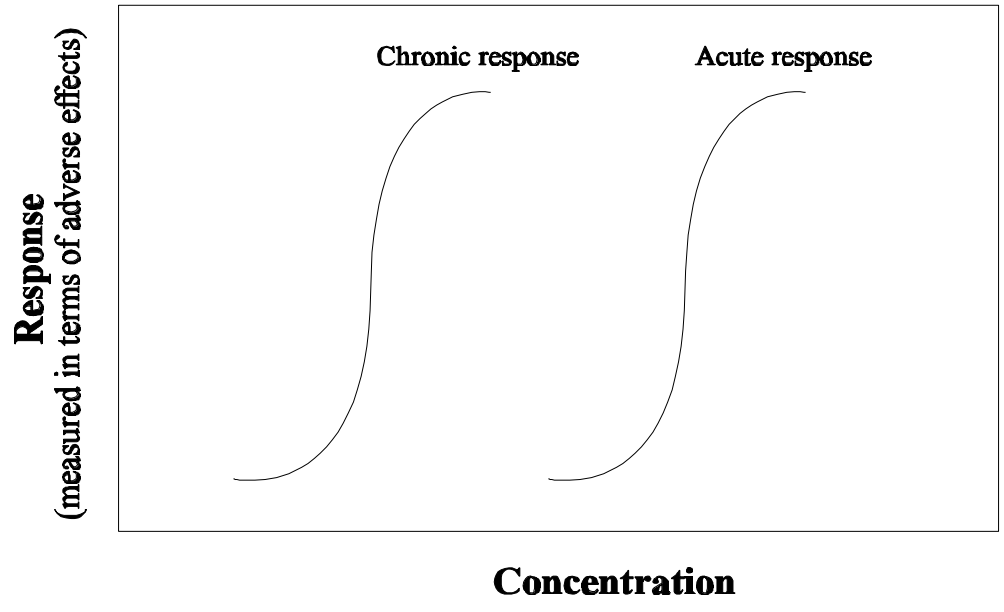
How will this guidance be incorporated into WET test methodology?

EPA plans to incorporate the guidance presented in this chapter into the WET method manuals (USEPA, 1993c; USEPA, 1994a; USEPA, 1994b). A proposal to amend the manuals is expected to appear in the *Federal Register* by March 2001.

What is the concentration-response relationship concept?

The concept of a concentration-response, or more classically, a dose-response relationship is "the most fundamental and pervasive one in toxicology" (Casarett and Doull, 1975). This concept assumes that there is a causal relationship between the dose of a toxicant (or concentration for toxicants in solution) and a measured response. A response may be any measurable biochemical or biological parameter that is correlated with exposure to the toxicant. The classical concentration-response relationship is depicted as a sigmoidal shaped curve (Figure 4.1), however, the particular shape of the concentration-response curve may differ for each coupled toxicant and response pair.

Figure 4.1. Classical concentration-response relationship.



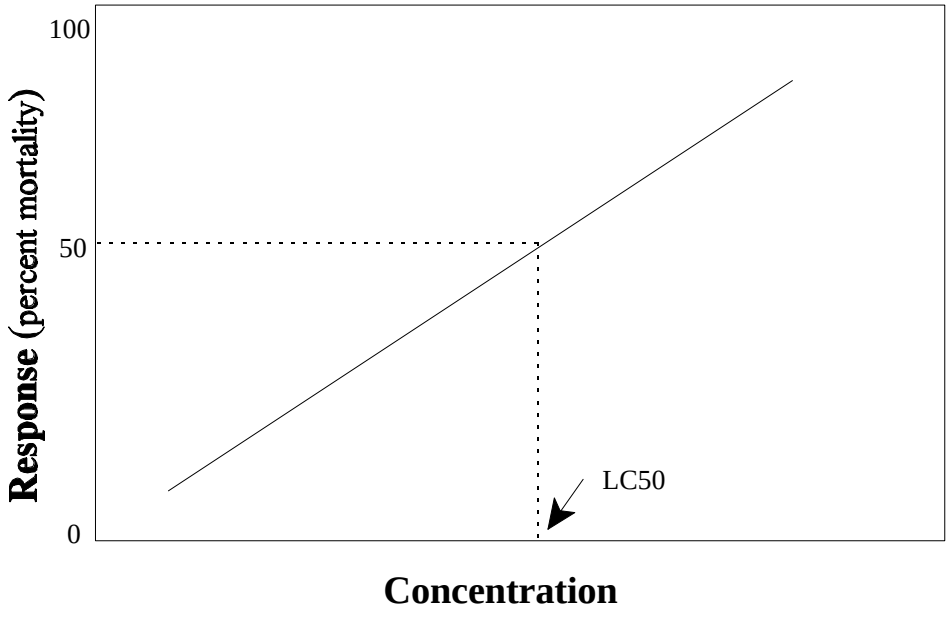
In general, more severe responses (such as acute effects) occur at higher concentrations of the toxicant, and less severe responses (such as chronic effects) occur at lower concentrations (Figure 4.1). A single toxicant also may produce multiple responses, each characterized by a concentration-response relationship.

In classical toxicology, concentration-response curves are generally displayed such that responses increase with increasing concentration (Figure 4.1). This is accomplished by defining responses in terms of adverse effects (e.g., mortality, reduction in growth, reduction in reproduction). The WET method manuals do not follow this convention; rather, responses are displayed in terms of survival, growth, and reproduction such that concentration-response curves for toxicants decrease with increasing concentration. This guidance will remain consistent with the convention established in the WET method manuals and will display concentration-response relationships for WET data such that responses decrease with increasing concentration.

How is the concentration-response concept used in WET testing?

The concentration-response concept is the basis for the determination of point estimates (LC50, EC50, IC25, etc.) in WET testing. A biological response (mortality, growth inhibition, reproductive inhibition, etc.) is measured at a range of effluent concentrations to develop a concentration-response curve. This curve, which is typically sigmoidal, is then linearized by various transformations of the data (e.g., probit transform) to assist in drawing conclusions from the relationship. From the resulting linearized concentration-response curve, a point estimate effect concentration can be calculated (Figure 4.2). The effect concentration is an estimate of the concentration of effluent that will produce a specific level of response (e.g., 50% mortality). In WET testing, effect concentrations such as the LC50, EC50, IC25 and IC50 are commonly used to report WET test results.

Figure 4.2. Example determination of point estimates from a concentration-response curve.



How can the concentration-response concept be used to review WET test results?

A corollary of the concentration-response concept is that every toxicant should exhibit a concentration-response relationship, given that the appropriate response is measured and given that the concentration range evaluated is appropriate. Use of this concept can be helpful in determining whether an effluent possesses toxicity and in identifying anomalous test results. An evaluation of the concentration-response relationship generated for each sample is an important part of the data review process that should not be overlooked. This chapter provides guidance on identifying valid concentration-response relationships and interpreting results from unexpected concentration-response patterns. This guidance on reviewing concentration-response relationships should be viewed as a component of a broader quality assurance and data review and reporting process that includes:

- **Review of test conditions** - The WET method manuals provide a summarized method-specific list of test conditions that should be followed in all WET test (e.g., test temperatures, number of replicates, test chamber sizes and volumes, lighting, feeding regimes, etc.). The conduct of each test should be reviewed to ensure that these conditions were met within the flexibility provided by the method manuals. The test conditions used in the test and any deviation from WET method manual requirements should be clearly reported. Daily measurements should be reviewed to ensure that values are within the acceptable ranges. Calibration of equipment should be verified and noted.
- **Review of test acceptability criteria** - The WET method manuals provide method-specific minimum criteria for the acceptability of tests (e.g., minimum control survival, reproduction, growth, or variability). These criteria are requirements of the methods, and any test not meeting the minimum test acceptability criteria should be considered invalid. All invalid tests should be repeated with a newly collected sample. While permit compliance should not be based on an invalid test, EPA's promulgation of the methods requires the results of all tests to be reported (valid or invalid).
- **Review of reference toxicant testing** - Reference toxicant testing is an important quality control practice that is required in the WET method manuals. Reference toxicant testing should be conducted on at least a monthly basis for each test method routinely conducted in a laboratory. WET test review should include evaluation of the most recent reference toxicant test and the reference toxicant cusum chart maintained by the laboratory. All reference toxicant tests should be conducted similarly (e.g., test duration, test conditions, test endpoint) to effluent tests being conducted. For instance, acute reference toxicant testing should be conducted to accompany acute testing of effluents, and short-term chronic reference toxicant testing should be conducted to accompany short-term chronic testing of effluents.
- **Review of organism culture health and performance** - EPA recommends that laboratories monitor and record the health and performance of organism cultures from which test organisms are obtained. For instance, the survival and reproduction of *Ceriodaphnia dubia* brood stock should be monitored and recorded during routine culture maintenance (i.e., water changes). This can be accomplished with a subset of 10 to 20 brood culture animals in individual culture vessels. This monitoring and documentation allows a laboratory to assess the current condition

of organism cultures prior to initiating a test and can allow the laboratory to postpone testing if organism cultures are unhealthy. This can potentially reduce the incidence of invalid tests and the cost associated with retesting. In the test review step, the documentation of culture health and performance can be useful in either identifying or eliminating poor culture health as a cause for marginal control performance in a test. Laboratories should maintain culture control charts (cusum charts) for survival, reproduction, growth, or other parameters for the appropriate species.

- **Review of test variability** - EPA recommends that the variability of each WET test, measured as a minimum significant difference (MSD) or percent MSD, be calculated and reported with all test results. EPA also recommends that laboratories maintain control charts for percent MSDs (USEPA, 2000). These control charts will allow laboratories to assess individual test variability in the context of typical variability within the laboratory. High test variability can result in insensitive tests or unexpected concentration-response relationships. Consult USEPA (2000) for additional guidance on WET test method variability.
- **Review of concentration-response relationships** - The guidance provided in this chapter may be used to assist in evaluating the concentration-response relationship as a part of the data review and reporting process. The succeeding section ("What are some patterns of concentration-response relationships typically seen in WET test data?") provides examples of common patterns in WET test data, discusses possible causes and solutions for unexpected patterns, and provides guidance on when to accept or reject test data based on the concentration-response concept. Some states have already developed similar guidance (Washington State Department of Ecology, 1997). It should be noted that the determination of a valid concentration-response relationship is not always clear cut. Data from some tests may suggest consultation with professional toxicologists and/or regulatory officials. Tests that exhibit unexpected concentration-response relationships also may indicate a need for further investigation and possible retesting. In general, when unexpected or apparently anomalous concentration-response relationships are encountered, EPA recommends the following:
  - **attempt to determine a cause for the response** - The above mentioned test review steps and specific guidance for individual concentration-response relationships (see "What are some patterns of concentration-response relationships typically seen in WET test data?") may assist in determining a cause for unexpected concentration-response relationships. Unexpected concentration-response relationships could be valid response patterns or anomalies resulting from Type I test error, high test variability, or other causes. If a given effluent consistently produces a specific, unexpected concentration-response relationship, there is likely a physical, chemical or biological cause. In situations where difficult-to-interpret concentration-response relationships are produced consistently by a given effluent, consultation with professional toxicologists is recommended. Toxicity identification evaluation (TIE) procedures (USEPA, 1991a; USEPA, 1992; USEPA, 1993a; USEPA, 1993b; USEPA, 1996b) also provide guidance that may be useful in determining a cause for such concentration-response relationships.
  - **follow guidance for specific concentration-response patterns** - The succeeding section ("What are some patterns of concentration-response relationships typically seen in WET

test data?") provides examples of 10 concentration-response patterns that may be exhibited by WET test data. This section provides guidance in interpreting each concentration-response pattern using a step-by-step review process. Based on this review, the guidance may recommend acceptance of the calculated results (e.g., NOEC or IC25) as valid and reliable, explanation of the calculated results as anomalous, or retesting.

- **increase testing frequency** - EPA recommends a testing frequency increase after any anomalous, questionable, or failing test result, with the number of tests and duration of testing to be determined by the regulatory authority.
- **coordinate with regulatory authorities, permittees, and testing laboratory** - EPA recommends that regulatory authorities, permittees, and testing laboratory personnel work together to resolve difficult-to-interpret WET test data. EPA also recommends that discussions be initiated as soon as possible when questions arise regarding WET test results.

This chapter provides additional guidance on reviewing test data; it is not the intent of this chapter to recommend the frequent disqualification and repetition of WET tests. Several warnings and safeguards should be considered when implementing the guidance in this chapter. First, unexpected concentration-response relationships should not occur with any regular frequency. Second, it is not recommended to screen only those tests in which toxicity is found at or below the receiving water concentration (RWC). If screening is to be done for unexpected concentration-response relationships, all tests should be screened in a similar manner. Third, all testing results should be reported to the regulatory authority, and the regulatory authorities should review all tests (including those disqualified and repeated). Regulatory authorities should be alert to patterns such as a high or increasing test rejection rate or a tendency for disqualified tests to show toxicity more often than tests accepted without qualification.

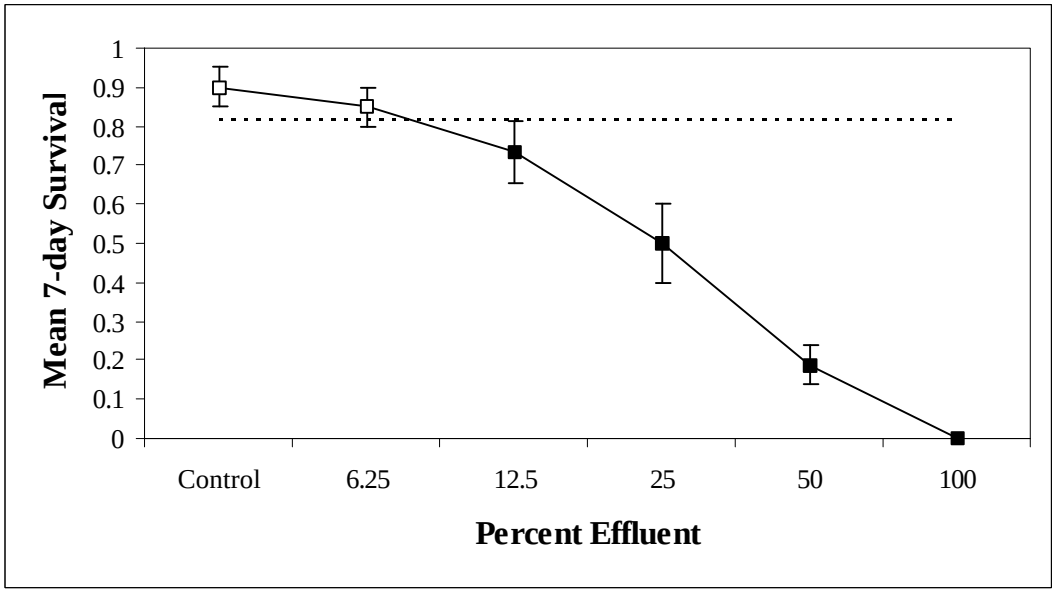
What are some patterns of concentration-response relationships typically seen in WET test data?

Ten concentration-response patterns that may appear in WET testing are individually described and illustrated below using hypothetical test data. This section provides guidance in interpreting each concentration-response pattern. The guidance focuses on determining a cause for unexpected concentration-response patterns by recommending a step-by-step review process. Based on this review, the guidance may recommend acceptance of the calculated results (e.g., NOEC or IC25) as valid and reliable, explanation of the calculated results as anomalous, or retesting. When retesting is recommended, this generally means beginning a new test on a newly collected sample since sample holding times are typically expired by the time results are obtained from the original test. Test results should be reported for all tests conducted, even if retesting is recommended.

1. Ideal concentration-response relationship

This response pattern (Figure 4.3) shows a clear concentration-response relationship, with multiple effluent concentrations identified as significantly different from the control. This pattern also shows a monotonic decrease in response, meaning that the response steadily decreases for each higher effluent concentration. This pattern is indicative of a well designed test with appropriately chosen concentrations that bracket the effluent's range of toxicity. Under these circumstances, the hypothesis testing and point estimation techniques recommended in the WET method manuals provide reliable results.

Figure 4.3. Ideal concentration-response relationship. [1]

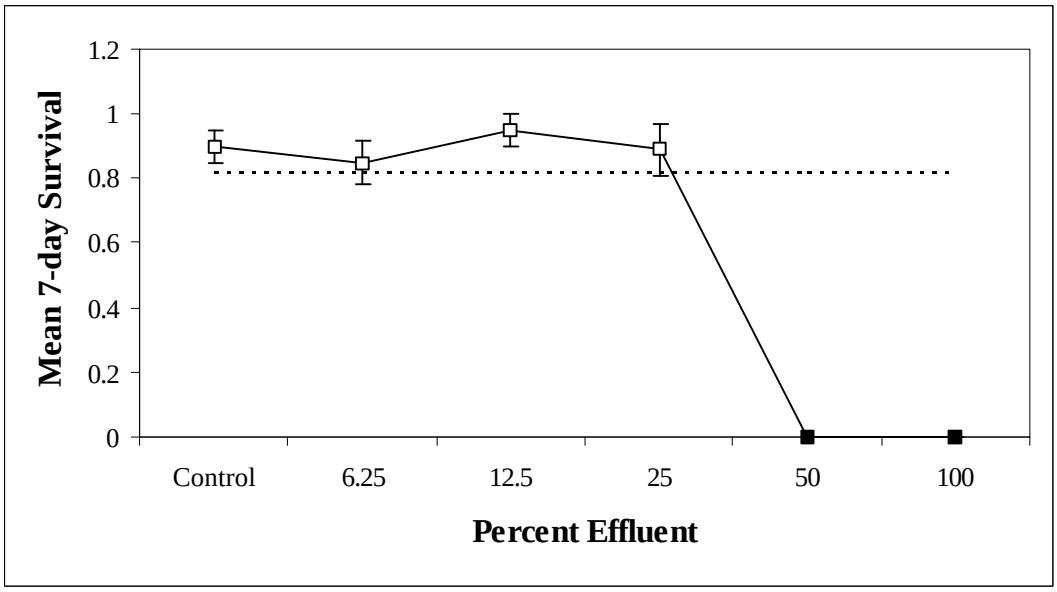


[1] Solid squares indicate data points that are statistically significantly different from the control, and hollow squares indicate data points that were not significantly different from the control. The dotted line shows the control mean minus the minimum significant difference (MSD); any test treatment response mean less than this value is considered to differ significantly from the control mean.

2. All or nothing response

The "all or nothing" response pattern is very common in WET test data. This response pattern (Figure 4.4) is characterized by a transition from no significant effect at one effluent concentration to a complete effect (100% mortality) at the next higher concentration. While not ideal, this pattern also represents a valid concentration-response relationship, and both hypothesis testing and point estimation techniques recommended in the WET method manuals will provide reliable results. This pattern of response is indicative of a steep concentration-response curve for the given effluent, and under these circumstances, the precision of the estimate may be improved by closer spacing of effluent concentrations (increased dilution factor) or the addition of intermediate effluent concentrations in future testing.

Figure 4.4. All or nothing concentration-response relationship. [1]

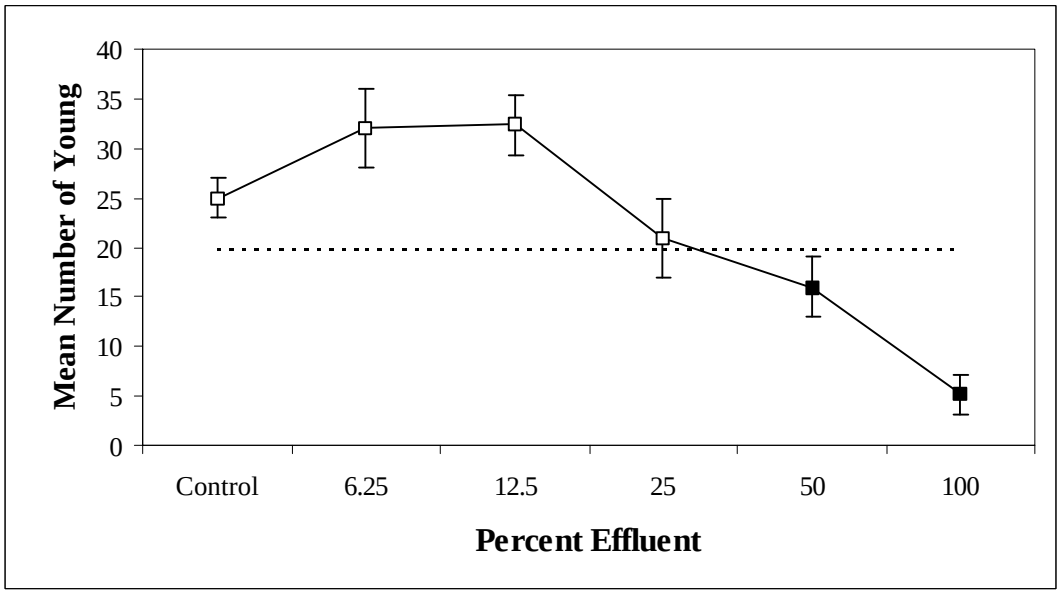


[1] Solid squares indicate data points that are statistically significantly different from the control, and hollow squares indicate data points that were not significantly different from the control. The dotted line shows the control mean minus the minimum significant difference (MSD); any test treatment response mean less than this value is considered to differ significantly from the control mean.

3. Stimulatory response at low concentrations and detrimental effects at higher concentrations

A stimulatory response is a nonmonotonic concentration-response relationship characterized by a measured increase in the response (stimulation) at low concentrations. This stimulation at low concentrations can be followed by a detrimental effect at higher concentrations (Figure 4.5) or by no effect at higher concentrations (see Section 4 following). Davis and Svendsgaard (1993) found that such nonmonotonic concentration-response relationships occurred in 12-24% of the toxicological studies surveyed. The stimulatory response pattern characterized in Figure 4.5 is typically found with sublethal endpoints such as reproduction, growth, fertilization, or larval development. For instance, test organism reproduction may increase (relative to the control) at low concentrations of an effluent and decrease relative to the control at higher concentrations. This concentration-response pattern, while nonmonotonic, is still a valid concentration-response relationship, and both hypothesis testing and point estimation techniques recommended in the WET method manuals will provide reliable results.

Figure 4.5. Stimulation at low concentrations and significant effects at high concentrations. [1]

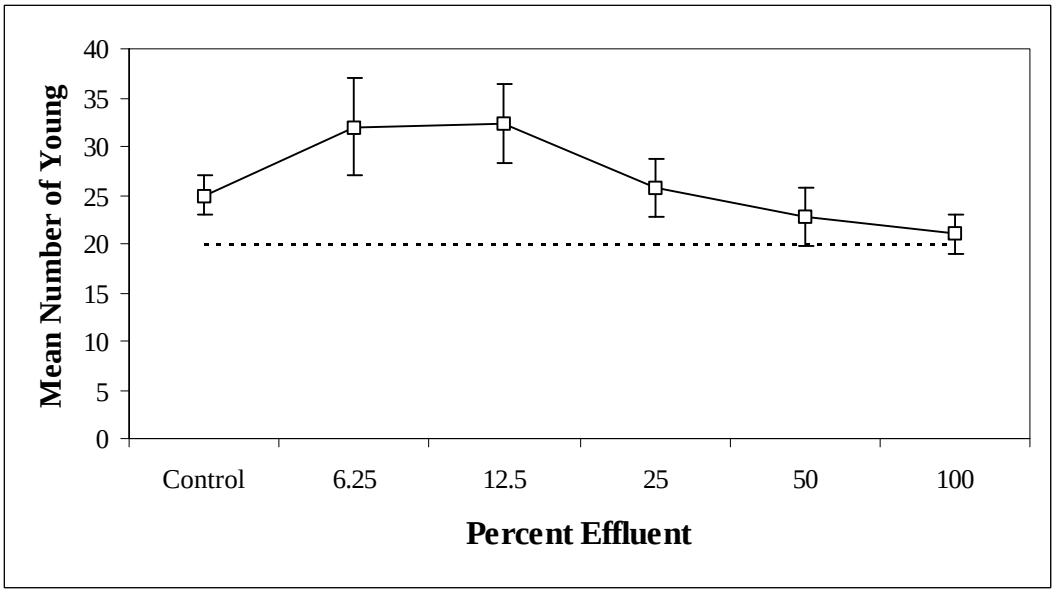


[1] Solid squares indicate data points that are statistically significantly different from the control, and hollow squares indicate data points that were not significantly different from the control. The dotted line shows the control mean minus the minimum significant difference (MSD); any test treatment response mean less than this value is considered to differ significantly from the control mean.

4. Stimulation at low concentrations but no significant effect at higher concentrations

This concentration-response relationship is similar to the previous example in that stimulation is observed at lower concentrations, but in this case, higher concentrations do not produce significant effects (Figure 4.6). In this situation, hypothesis testing techniques should produce reliable results, assuming that adequate test sensitivity is achieved. Results from point estimation techniques should be interpreted carefully when this response pattern is encountered, because the inhibition concentration percentage (ICp) procedure may produce effect concentrations (particularly IC25s) that indicate toxicity at effluent concentrations where the response is comparable to the control response. The ICp procedure assumes that responses: (1) are from a random, independent, and representative sample of test data; (2) follow a piecewise linear response function; and (3) are monotonically non-increasing, meaning that the mean response for each higher concentration is less than or equal to the mean response for the previous concentration. If the data are not monotonically non-increasing, the ICp procedure adjusts the response means using a "smoothing" technique that averages adjacent means (see Appendix M of USEPA, 1994a). This technique averages response means (including that of the control) with those of the next highest test concentration until responses are monotonically non-increasing. In cases where the responses at the low effluent concentrations are much higher than in the control, the smoothing process may result in a large upward adjustment in the control mean. This can lead to an IC25 result that is less than the highest test concentration, even though the highest test concentration was not statistically different from the control treatment and even if a percent difference of less than 25% was observed between the control response and the response at the highest test concentration.

Figure 4.6. Stimulation at low concentrations but no significant effect at higher concentrations. [1]



[1] Solid squares indicate data points that are statistically significantly different from the control, and hollow squares indicate data points that were not significantly different from the control. The dotted line shows the control mean minus the minimum significant difference (MSD); any test treatment response mean less than this value is considered to differ significantly from the control mean.

If the response pattern depicted in Figure 4.6 (stimulation at low concentrations but no significant effect at higher concentrations) is encountered, the following review steps should be taken in addition to standard test review procedures:

- **Evaluate the concentration range** - If the highest concentration used in the test was less than 100% effluent (or the highest achievable effluent concentration for marine tests), the effluent should be retested using higher test concentrations to establish if a valid concentration-response relationship exists. This may not be necessary if the permit limit is set at much lower than 100% effluent and test results indicate no toxicity at the permit limit level and at least one concentration above the permit limit.
- **Compare hypothesis testing results and point estimates** - If there is agreement between the NOEC and the IC25 for tests producing the concentration-response pattern depicted in Figure 4.6 (i.e., neither value indicates toxicity at or below the permitted RWC, or both values indicate toxicity at or below the RWC) the test results should be reported and considered valid. If, however, the NOEC indicates no toxicity at the RWC (i.e., NOEC greater than or equal to RWC) but the IC25 is calculated as less than the RWC, the remaining recommended actions should be taken.
- **Evaluate control response** - It is possible that the response pattern depicted in Figure 4.6 could result from poor performance in the controls rather than stimulation at the lower test concentrations. This poor control performance could cause a toxic effect at higher test concentrations not to be detected. To evaluate this possibility, compare the control response

to the normal control performance for the laboratory. If (1) a particular test exhibits the response pattern depicted in Figure 4.6, (2) there is disagreement between NOEC and IC25 estimates, and (3) the mean control response is well below the laboratory's normal range of control performance; retesting of the effluent is recommended even if the minimum test acceptability criteria have been met. For example, if a laboratory consistently achieves a control mean of 25-30 neonates for the *Ceriodaphnia dubia* 3-brood chronic test, a control mean of 15-18 neonates (in conjunction with a non-ideal concentration-response curve and disagreement between the NOEC and IC25) would warrant retesting. In this situation, suppressed control performance could be considered as the cause for this response pattern rather than stimulation. A review of control performance should also investigate the possibility of poor performance in a single replicate substantially reducing the mean control response. In this case, retesting is also recommended.

- **Evaluate the test sensitivity** - Discrepancies between IC25 and NOEC values could be due to low test sensitivity. To determine if this is the case, evaluate the sensitivity of the test by comparing the test MSD to MSD criteria for the given test method (see Chapter 2 of this guidance and USEPA, 2000) and to the laboratory's historical test sensitivity performance. Laboratories are encouraged to track test sensitivity (as %MSDs) for tests conducted over time. If a test exhibits the response pattern depicted in Figure 4.6 and the test MSD is above maximum recommended criteria for the method or above the laboratory's typical range, the sample should be retested.
- **Evaluate the ICp calculation** - If a test exhibits the response pattern depicted in Figure 4.6 and it has been determined from the above actions that the pattern is not due to poor control performance or low test sensitivity, then discrepancies between the NOEC and IC25 may be due to bias from the ICp smoothing technique. To determine if this is the case, calculate the observed percent difference between the response at the RWC and the control as:

$$\frac{\left(m_c - m_{RWC}\right)}{m_c} \times 100\%$$

where:

m_c = mean control response

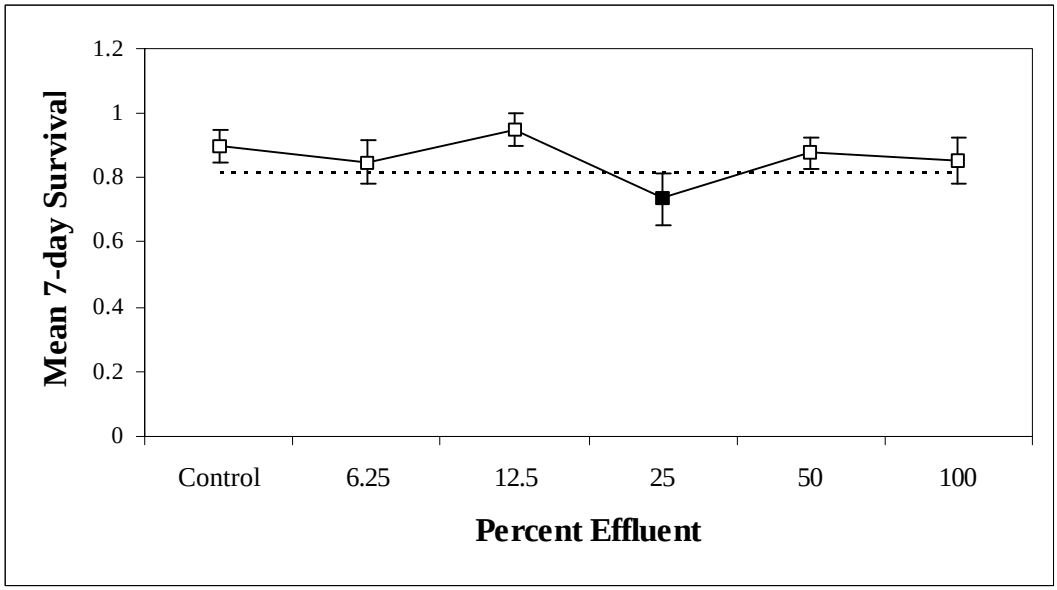
m_{RWC} = mean response at the receiving water concentration (RWC)

If the observed percent difference between the response at the RWC and the control is less than 25% and the response at the RWC is not statistically significantly different from the control response, then a calculated IC25 of less than the RWC should be noted as anomalous and the effluent determined to be non-toxic at the RWC. If the observed percent difference is equal to or greater than 25%, then the calculated IC25 should be considered valid.

5. Interrupted concentration-response: significant effect bracketed by non-significant effects

This response pattern is characterized by a single test concentration showing a significant difference from the control while adjacent higher and lower test concentrations do not differ significantly from the control (Figure 4.7). When this response pattern is encountered, point estimation techniques generally will yield reliable results, but hypothesis testing results should be interpreted carefully. The method manual definitions of NOEC (the highest concentration of toxicant in which the values for the observed responses are not statistically significantly different from the controls) and LOEC (the lowest concentration of toxicant in which the values for the observed responses are statistically significantly different from the controls) were intended for situations where the concentration-response relationship is monotonically non-increasing. Under these circumstance, the NOEC and LOEC are always adjacent values with the NOEC being the test concentration just below the LOEC. In circumstances where the concentration-response relationship is non-monotonic (as in Figure 4.7), the identification of NOEC and LOEC values is severely compromised (Chapman *et al.*, 1996). For this response pattern, the following review actions should be taken in addition to standard test review procedures to determine the validity of results obtained by hypothesis testing:

Figure 4.7. Interrupted concentration-response: significant effect bracketed by non-significant effects. [1]



[1] Solid squares indicate data points that are statistically significantly different from the control, and hollow squares indicate data points that were not significantly different from the control. The dotted line shows the control mean minus the minimum significant difference (MSD); any test treatment response mean less than this value is considered to differ significantly from the control mean.

- **Check for test condition or procedural errors** - The concentration-response relationship depicted in Figure 4.7 could result from test conditions errors (such as pH, DO, salinity, or temperature excursions) occurring in isolated test replicates. This concentration-response

pattern also could be due to procedural errors such as failure to properly randomize test organisms or test chamber placement. The laboratory should verify that all test conditions were within ranges required by the WET method manuals for the given test method. The laboratory should verify that the assignment of test organisms to individual treatments was properly randomized (Davis *et al*., 1998). This can be complete randomization or block randomization (as with the *Ceriodaphnia dubia* 3-brood reproduction test). The laboratory also should verify that the positions of test chambers within the experiment were properly randomized. If test condition or procedural errors are identified, the sample should be retested.

- **Evaluate within-treatment variability** - It is possible for poor performance in a single replicate to bias the mean response for a given test concentration and cause that concentration to differ significantly from the control. For this reason, the within-treatment variability should be evaluated for the significantly different treatment. If the variability (standard deviation or CV) for that treatment is considerably greater than for other treatments, then responses of individual replicates should be investigated. This investigation may show that a single outlier replicate has biased the treatment mean. If this is the case and the responses from all but the single outlier replicate are consistent with the control response, then the sample should be retested.
- **Evaluate test sensitivity** - When the response pattern depicted in Figure 4.7 is encountered, it is important to evaluate test sensitivity. If test sensitivity is low (e.g. high MSD values), large effects at higher test concentrations may not be detected as statistically significant. To evaluate test sensitivity, compare the MSD for the test to benchmark criteria for the given test method (see Chapter 2 of this guidance and USEPA, 2000) and to the laboratory's historical test sensitivity performance. As previously mentioned, laboratories are encouraged to track test sensitivity (as %MSDs) for tests conducted over time. If test sensitivity is low (i.e., MSDs are above maximum recommended criteria or typical laboratory performance), then the sample should be retested. Consult Section 6.4 in USEPA (2000) for additional guidance on implementing upper and lower bounds on test sensitivity.

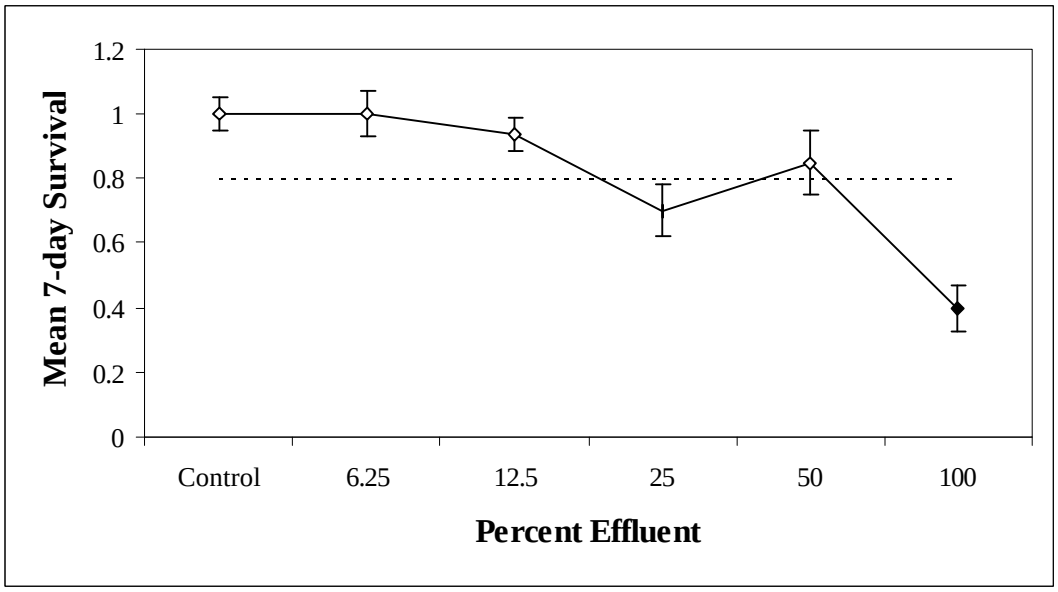
  If test sensitivity is moderate to high (i.e., MSDs below the maximum recommended criteria and within the laboratory's typical performance range) and none of the preceding evaluations have determined a cause for this response pattern, it is likely that the significantly different treatment is the result of a Type I error. A Type I error is the error of incorrectly rejecting the null hypothesis (assuming that the treatment is significantly different from the control) when in fact the null hypothesis is true (the treatment is not significantly different from the control). In this situation, due to the absence of a valid concentration-response relationship, the intermediate concentration that was determined by hypothesis testing to be statistically different from the control should be considered anomalous, and the NOEC should be determined as the highest concentration that was not significantly different from the control. Using Figure 4.7 to illustrate, the 25% concentration would be considered anomalous, the reported NOEC would be 100%, and the reported LOEC would be >100%. Under these circumstances, test results should still note that the 25% concentration was statistically

different from the control but was considered anomalous due to analysis of the concentration-response curve and the above review steps.

6. Interrupted concentration-response: non-significant effects bracketed by significant effects

This response pattern is similar to the previous response pattern in that the concentration-response curve is nonmonotonic (or interrupted), however, this response pattern is characterized by two or more test concentrations showing a significant difference from the control while an intermediate test concentration does not differ significantly from the control (Figure 4.8). When this response pattern is encountered, point estimation techniques will generally yield reliable results, but hypothesis testing results should be interpreted carefully. As mentioned for the previous concentration-response pattern, the identification of NOEC and LOEC values is severely compromised (Chapman *et al.*, 1996) when the concentration-response relationship is non-monotonic (as in Figure 4.8). For this response pattern, the test sensitivity should be evaluated as described below in addition to standard test review procedures to determine the validity of results determined by hypothesis testing.

Figure 4.8. Interrupted concentration-response: non-significant effects bracketed by significant effects. [1]



[1] Solid squares indicate data points that are statistically significantly different from the control, and hollow squares indicate data points that were not significantly different from the control. The dotted line shows the control mean minus the minimum significant difference (MSD); any test treatment response mean less than this value is considered to differ significantly from the control mean.

- **Evaluate test sensitivity** - When the response pattern depicted in Figure 4.8 is encountered, it is important to evaluate test sensitivity by comparing test MSDs to minimum and maximum MSD criteria recommended by EPA (USEPA, 2000). If the test MSD is lower than the minimum MSD criterion, only effects larger than the minimum MSD criterion should be

considered significant. For example, if the minimum MSD criterion for a method is 15% and the calculated test MSD is 10%, only effects greater than 15% difference compared to the control should be considered significant. If test sensitivity is low (i.e., test MSD is above maximum MSD criterion), the sample should be retested. If test sensitivity is moderate (i.e., test MSD is within minimum and maximum MSD criterion), the test results should be considered valid and the NOEC should be reported as the concentration below the LOEC. For the case depicted in Figure 4.8, a NOEC of 12.5% should be reported. Consult Section 6.4 in USEPA (2000) for additional guidance on implementing upper and lower bounds on test sensitivity.

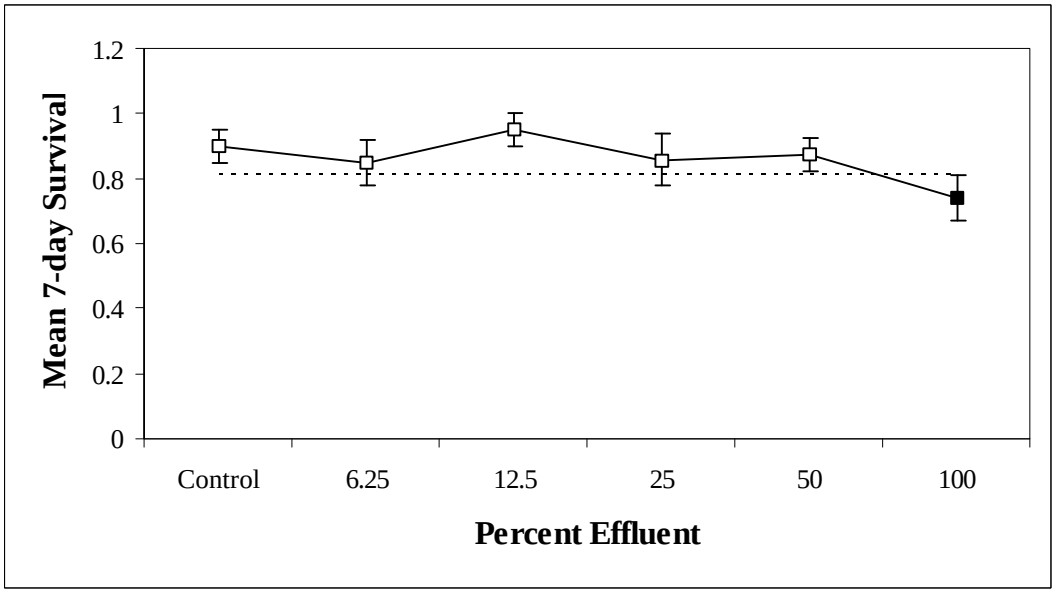
7. Significant effects only at highest concentration

This response pattern is characterized by only the highest test concentration producing a significantly different response from the control (Figure 4.9). This response pattern should be considered to be a valid concentration-response relationship and results determined by point estimation should be assumed to be reliable. Hypothesis testing results are also assumed to be reliable following the evaluation of test sensitivity as described below. If the response pattern depicted in Figure 4.9 (significant effects only at highest concentration) is encountered, the following review steps should be taken in addition to standard test review procedures:

- **Evaluate the concentration range** - When this response pattern occurs, the concentrations used for testing should be evaluated in future tests using this effluent. If the highest effluent concentration used in the test was less than 100% (or the highest achievable effluent concentration for marine tests), future testing using this sample should include at least one higher test concentration to confirm the presence of a concentration-response relationship. If the test used a 100% effluent concentration treatment, it is difficult to confirm a concentration-response relationship through retesting because concentrations are constrained to less than or equal to 100% in whole effluent testing. If this response pattern occurs commonly with a given effluent, future testing of the effluent should use a dilution factor of >0.5 such that test concentrations closer to the 100% effluent concentration are used (i.e., a dilution factor of 0.65 would provide a test concentration series of 18%, 27%, 42%, 65%, and 100%). This would provide a better opportunity to confirm a concentration-response relationship that may exist at the upper end of the concentration range. This approach should be used only if historical testing of the effluent indicates consistency and the effect concentration is not likely to fall below the adjusted test concentration series.
- **Evaluate test sensitivity** - Evaluate test sensitivity by comparing test MSDs to minimum and maximum MSD criteria recommended by EPA (USEPA, 2000). If the test MSD is lower than the minimum MSD criterion, only effects larger than the minimum MSD criterion should be considered significant. For example, if the minimum MSD criterion for a method is 15% and the calculated test MSD is 10%, only effects greater than 15% difference compared to the control should be considered significant. If test sensitivity is low (i.e., test MSD is above maximum MSD criterion), the sample should be retested. If test sensitivity is moderate (i.e., test MSD is within minimum and maximum MSD criterion), the test results should be

considered valid and the NOEC should be reported as the concentration below the LOEC. For the example given in Figure 4.9, a NOEC of 50% effluent should be reported. Consult Section 6.4 in USEPA (2000) for additional guidance on implementing upper and lower bounds on test sensitivity.

Figure 4.9. Significant effects only at highest concentration. [1]



[1] Solid squares indicate data points that are statistically significantly different from the control, and hollow squares indicate data points that were not significantly different from the control. The dotted line shows the control mean minus the minimum significant difference (MSD); any test treatment response mean less than this value is considered to differ significantly from the control mean.

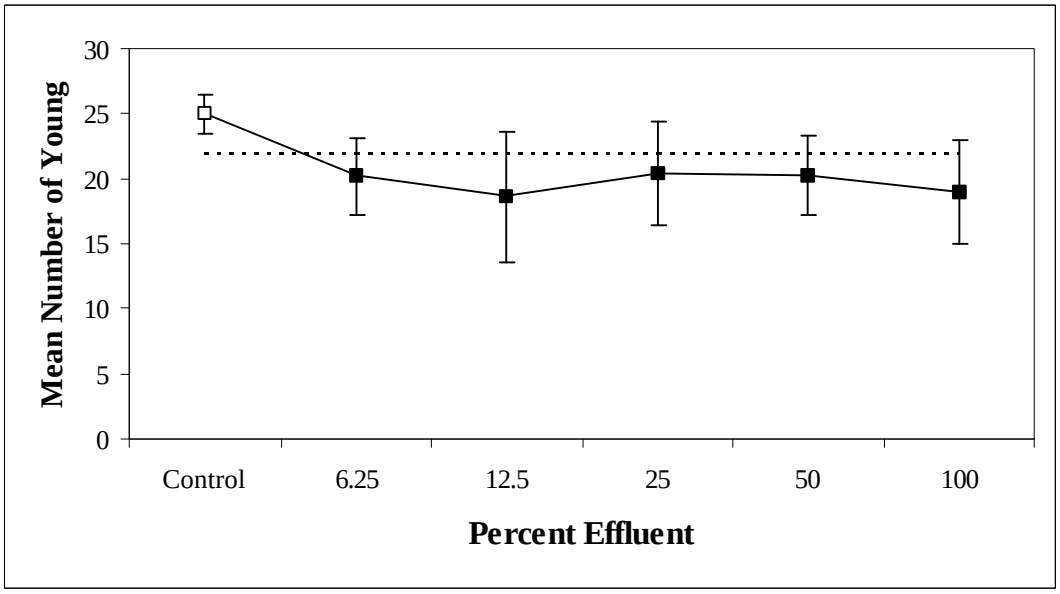
8. Significant effects at all test concentrations but flat concentration-response curve

This response pattern is demonstrated in Figure 4.10. All of the test concentrations produce a response that is significantly different from the control response, but a clear concentration-response relationship cannot be determined. This response pattern could be due to: (1) extremely low variability in the control, (2) an unusually high control response, (3) an inappropriate dilution water and improper use of dilution water controls, (4) inappropriate test dilution series, (5) potential pathogen effects in the effluent, (6) an unusual effluent-dilution water interaction. The following review actions should be taken to determine a cause for this concentration-response pattern and to subsequently determine the validity of calculated results.

- **Evaluate test sensitivity** - The response pattern depicted in Figure 4.10 may be an artifact of the data resulting from extremely precise control results and extremely high test sensitivity. Investigate this possibility by comparing test MSDs to minimum MSD criteria recommended by EPA (USEPA, 2000). If the test MSD is lower than the minimum MSD criterion, only effects larger than the minimum MSD criterion should be considered significant. For example, if the minimum MSD criterion for a method is 15% and the calculated test MSD is 10%, only effects greater than 15% difference compared to the control should be considered significant.

If test sensitivity is low (i.e., test MSD is above maximum MSD criterion), the sample should be retested. Consult Section 6.4 in USEPA (2000) for additional guidance on implementing upper and lower bounds on test sensitivity.

Figure 4.10. Significant effects at all test concentrations but flat concentration-response curve. [1]



[1] Solid squares indicate data points that are statistically significantly different from the control, and hollow squares indicate data points that were not significantly different from the control. The dotted line shows the control mean minus the minimum significant difference (MSD); any test treatment response mean less than this value is considered to differ significantly from the control mean.

- **Evaluate control response** - The concentration-response pattern depicted in Figure 4.10 could result from an unusually high response in the control treatment. Laboratories are encouraged to track the performance of controls in tests conducted over time. When the response pattern depicted in Figure 4.10 is exhibited, the control response for the test should be compared to historic control performance in the laboratory using the given dilution water. If the mean control response is above the normal range for that laboratory and dilution water, the sample should be retested.
- **Evaluate dilution water** - The improper use of dilution waters and dilution water controls could cause the concentration-response pattern depicted in Figure 4.10. It should be confirmed that test treatment concentrations were compared to the dilution water control and not a culture water control. A statistical comparison of the dilution water control and the culture water control should also be made if they are from different sources. If the dilution water control shows a statistically significant difference from the culture water control, alternate dilution waters should be considered and the sample retested (see Chapter 6 of this guidance).
- **Evaluate test concentrations** - If all test concentrations produce a complete effect (e.g., 100% mortality, zero reproduction, etc.), a flat concentration-response relationship will result. This concentration-response relationship should be considered valid, and it indicates high toxicity in

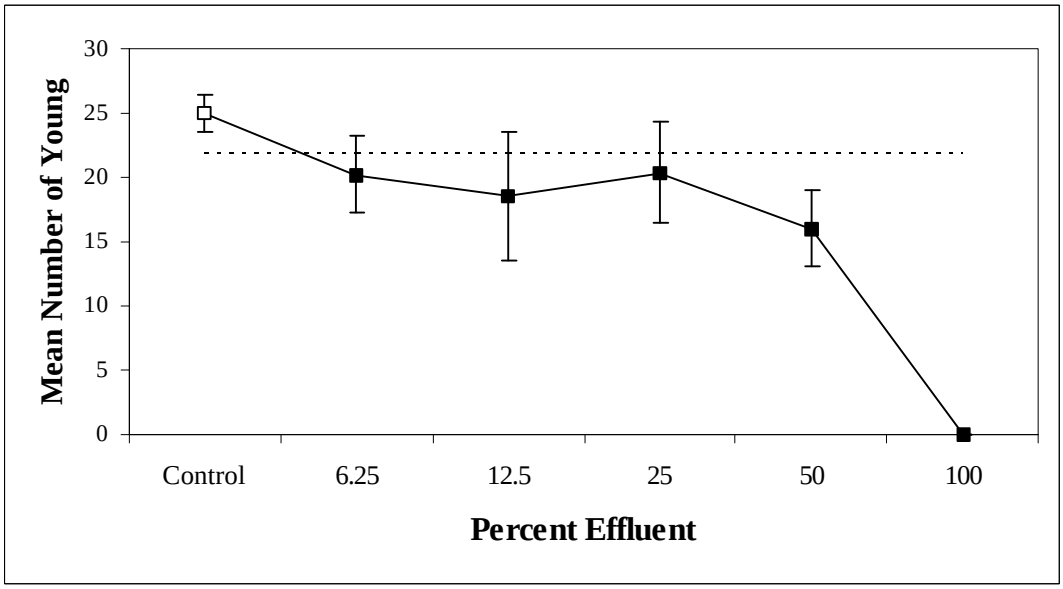
the sample. Assuming that the concentration range used in the test brackets the permitted RWC, it is not necessary to retest the sample, since the test results clearly indicate toxicity. If all test concentrations were significantly different from the control but did not produce complete effects (as in Figure 4.10), the dilution series should be investigated. It is possible that the test concentration range used for the test was too narrow to distinguish a shallow sloped concentration-response curve. Test concentrations may not have been low enough to produce no significant effect and may not have been high enough to produce severe effects. If this situation is suspected, the sample should be retested using an expanded dilution series range. Effluent concentrations that are lower than those used in the previous test should be added. Effluent concentrations that are higher than those used in the previous test also should be added (if possible) to assist in determining a concentration-response relationship.

- **Consider pathogen effect** - The concentration-response pattern depicted in Figure 4.10 could also be due to the presence of pathogens in the effluent. The most common identifier of pathogen effects are sporadic mortalities and extremely high variability between replicates. The pathogen effect is more common in tests using fish species than in invertebrate testing. This pathogen effect also may be evident only in chronic tests and not in acute tests. Pathogen effects also may be seasonal in occurrence. If within-treatment CVs for survival are >40% for effluent concentrations and relatively small for control replicates in standard synthetic water, pathogen effect should be considered. If pathogen effects are suspected in the effluent, this may be confirmed in subsequent side-by-side testing using the effluent and the effluent treated by brief exposure to UV light or the addition of antibiotics, or increasing the number of replicates and using less test organisms in each replicate. If pathogen effects in the effluent are confirmed, the sample should be retested and the regulatory authority should be consulted prior to changing testing procedures.
- **Continued testing** - If all of the above scenarios have been investigated and have not revealed the cause of the response pattern, the results should be considered valid; however, continued testing should be initiated in an effort to identify the cause of the response pattern. If an effluent consistently exhibits this response pattern, additional investigations could include chemical analysis or initiation of TIE procedures.

9. Significant effects at all test concentrations with a sloped concentration-response curve

This concentration-response pattern is similar to the pattern identified in item #8 above except a concentration-response curve can be identified at the higher effluent concentrations (Figure 4.11). This pattern is considered to be a valid concentration-response relationship, and point estimation techniques will generally yield reliable results. Results determined by hypothesis testing techniques should be interpreted carefully, and the cause for significantly different effects at low concentrations should be investigated as described for the response pattern described in item #8.

Figure 4.11. Significant effects at all test concentrations with a sloped concentration-response curve. [1]



[1] Solid squares indicate data points that are statistically significantly different from the control, and hollow squares indicate data points that were not significantly different from the control. The dotted line shows the control mean minus the minimum significant difference (MSD); any test treatment response mean less than this value is considered to differ significantly from the control mean.

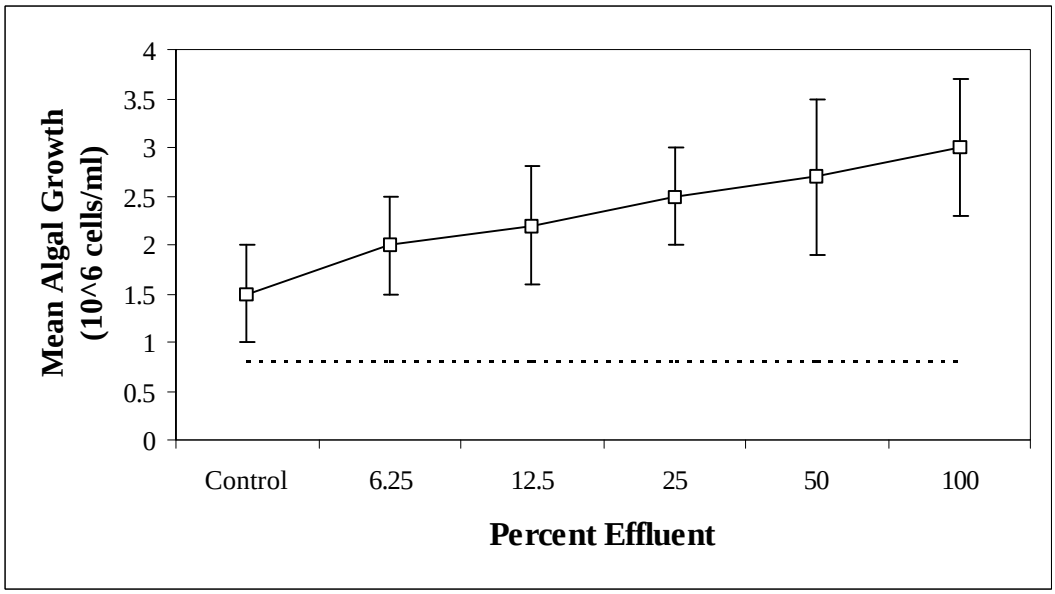
10. Inverse concentration-response relationship

This response pattern is characterized by a relationship in which adverse effects decrease with increasing effluent concentration (Figure 4.12). This situation is most often encountered in algal growth tests, and is typically caused by excess nutrients in the effluent. While a valid concentration-response relationship is demonstrated in this circumstance, the effluent should be considered nontoxic since the direction of the concentration-response relationship indicates decreasing adverse effects. It should be noted that while the effluent is considered non-toxic, the presence of excess nutrients still may pose a potential risk to the environment due to nutrient enrichment and oxygen depletion.

An inverse concentration-response pattern also may occur in tests other than algal growth assays when the dilution water used is a receiving water or synthetic water adjusted to approximate the receiving water characteristics. In such situations, the inverse concentration-response pattern can result from toxicity in the receiving water or the limitation of necessary components (i.e., hardness) in the receiving water or adjusted synthetic water. Under such circumstances, the objective of the toxicity test should be evaluated (see Chapter 6 of this guidance). If the objective of the test is to determine the toxicity of the effluent in the natural receiving water, then the results indicate no toxicity in the sample. If the objective of the toxicity test is to determine the absolute presence of toxicity in the effluent, the sample should be retested using a standard synthetic dilution water. Toxicity or limiting components in the receiving water or adjusted synthetic water may mask the

presence of low level toxicity in the effluent, making the absolute determination of toxicity in the effluent difficult.

Figure 4.12. Inverse concentration-response relationship. [1]



[1] Solid squares indicate data points that are statistically significantly different from the control, and hollow squares indicate data points that were not significantly different from the control. The dotted line shows the control mean minus the minimum significant difference (MSD); any test treatment response mean less than this value is considered to differ significantly from the control mean.

5 Dilution Series Selection

This chapter provides guidance on the selection of an appropriate dilution series for a WET test.

Do the WET method manuals specify a certain dilution series?

The WET method manuals (USEPA, 1993c; USEPA, 1994a; USEPA, 1994b) suggest, but do not require, a dilution series of 6.25%, 12.5%, 25%, 50%, and 100% effluent for most effluents. This dilution series should be used as a default when little information is known about the effluent being tested and when initial range finding indicates that the effect concentration of interest is within the 6.25% to 100% effluent range. In many situations, a more appropriate dilution series can be selected based on experience from repeated testing of a given effluent. The WET method manuals do recommend a dilution factor of 0.5 for preparing test concentrations. This recommendation does not fix the dilution factor, but is provided to establish a lower limit on the dilution factor. The use of dilution factors greater than 0.5 is encouraged when historical testing indicates that an effluent is relatively consistent and effect concentrations generally fall within a given range.

Why is selecting an appropriate dilution series important?

The selection of a dilution series (number and spacing of test concentrations) for WET tests is extremely important in producing reliable and precise results. This is most obvious for effect concentrations such as NOEC and LOEC values generated by hypothesis testing. These values are by definition limited to one of the effluent concentrations selected for the test. The precision of these values also is determined by the distance from the NOEC or LOEC to the next highest or lowest effluent concentration. For instance, using a standard dilution series of 6.25%, 12.5%, 25%, 50%, and 100% effluent, a measured NOEC value of 50% indicates that the transition from no observable effects to observable effects occurs somewhere between 50% and 100% effluent concentration (the NOEC-LOEC interval). If an alternative dilution series of 12.5%, 25%, 50%, 75%, and 100% were used for this test, then a NOEC of 50% would be a more precise estimate. In this test, the point of transition from no observable effect to observable effects is now known to lie between 50% and 75%.

The appropriate selection of a dilution series also is important for accurately identifying concentration-response relationships and increasing the precision of effect concentrations estimated from those relationships. For example, toxicants or effluents with steep concentration-response curves, often produce “all or nothing” results when using a standard dilution series of 6.25%,

12.5%, 25%, 50%, and 100% effluent. An "all or nothing" response means that one effluent concentration produces no effect and the next highest concentration produces a complete (e.g., 100% mortality) effect. Under these circumstances, the effect concentration is graphically determined between the no effect and complete effect concentrations. The effect concentration derived in this situation is less precise than when multiple concentrations with partial effects occur. The proper selection and spacing of dilutions can increase the opportunity of obtaining an ideal concentration-response relationship (see Chapter 4 of this guidance) that exhibits smooth transitions from no effect to partial effect to complete effect.

How might the dilution series or dilution sequence be modified to assist in determining a concentration-response relationship and improving the precision of calculated effect concentrations?

The preceding chapter identified and discussed 10 concentration-response patterns typically observed in WET testing. When applicable, recommendations for modifying the dilution series or dilution sequence were provided in the discussion of individual response patterns. In general, the following considerations and recommendations should improve the identification of concentration-response relationships and the precision of calculated effect concentrations.

- **Consider historic WET testing information for the given effluent** - Due to the importance of dilution series selection, this decision should be based on knowledge of the effluent from historical testing and permit information rather than simply on standard laboratory practice. Historic testing information on a given effluent will provide a typical range of effects that can characterize the consistency of the effluent's toxicity. This information is valuable and should not be overlooked. If historical testing shows toxicity consistently within a specified range of concentrations, the test dilution series for future tests can be selected to focus on that range. For example, if the LC50 for a given effluent is consistently between 50% and 100% effluent, it may be needless to continue testing concentrations as low as 6.25% effluent. A larger dilution factor, such as 0.75 could be used to provide a dilution series of 31.6%, 42.2%, 56.3%, 75%, and 100%. The analyst should be cautious not to narrow the range of concentrations too much, to avoid causing the effect concentration to fall outside the test concentration range when an unusually toxic sample is encountered.
- **Use the receiving water concentration as a test concentration** - As previously mentioned, a limitation of hypothesis testing is that NOEC and LOEC values are constrained only to effluent concentrations used in a test. Due to this limitation, hypothesis testing should be used only in situations where the toxicity of a particular effluent concentration of interest is to be evaluated (i.e., the receiving water concentration or RWC). In addition, the effluent concentration of interest, usually the RWC, should be included as one of the concentrations in the dilution series. Even if point estimation techniques are to be used for calculating effect concentrations, it is good practice to include the RWC as a test concentration in the dilution series.
- **Bracket the receiving water concentration with test concentrations** - Test concentrations selected should not only include the RWC, but also should bracket the RWC (unless the RWC

is 100%). This will allow the most precise determination of effect concentrations around the RWC and will aid in the determination of a valid concentration-response relationship.

- **Consider adding test concentrations within a given range of interest** - For better test resolution and more precise effect concentration estimates, additional test concentrations can be added within a given range of interest. This may be most beneficial when testing an effluent or toxicant that possesses a steep concentration-response relationship. Additional test concentrations placed between concentrations of no effect and complete effect may allow for partial effects to be measured and improve the precision of calculated effect concentrations. For instance, if no effect was observed at 100% effluent concentration and a complete effect was observed at 50% effluent concentration, an additional test concentration of 75% could be added to improve the precision of calculated effect concentrations. If historical testing information for this effluent indicates that effect concentrations are consistently between 50% and 100%, it may be possible to add the 75% concentration in place of the 6.25% concentration (i.e., 12.5%, 25%, 50%, 75%, and 100%). The addition of test concentrations also may be beneficial when very shallow concentration-response relationships are encountered. In this case, additional test concentrations should be added to extend the concentration range tested (e.g., 3.125%, 6.25%, 12.5%, 25%, 50%, and 100%).
- **Consider increasing the dilution factor used to space effluent concentrations** - Increasing the dilution factor for a test (i.e., reducing the space between concentrations) is encouraged if historic testing of the given effluent indicates relative consistency, and the given effect concentration is not expected to lie outside of the concentration range. Similar to adding test concentrations, increasing the dilution factor has the effect of narrowing the test focus on a concentration range of interest. This effect is accomplished while maintaining a logarithmic spacing of test concentrations, which is standard practice in toxicity testing. A possible disadvantage of increasing the dilution factor is that all of the test concentrations are typically changed when the dilution factor is altered; this may limit the comparability of results with previous testing, if test results are determined exclusively by hypothesis testing techniques. The comparability of point estimates should not be affected by alterations in the dilution factor.

6 Dilution Waters

This chapter provides guidance for selecting a dilution water that is appropriate for the objective of the WET test.

What does EPA consider to be an acceptable dilution water?

An acceptable dilution water for WET testing:

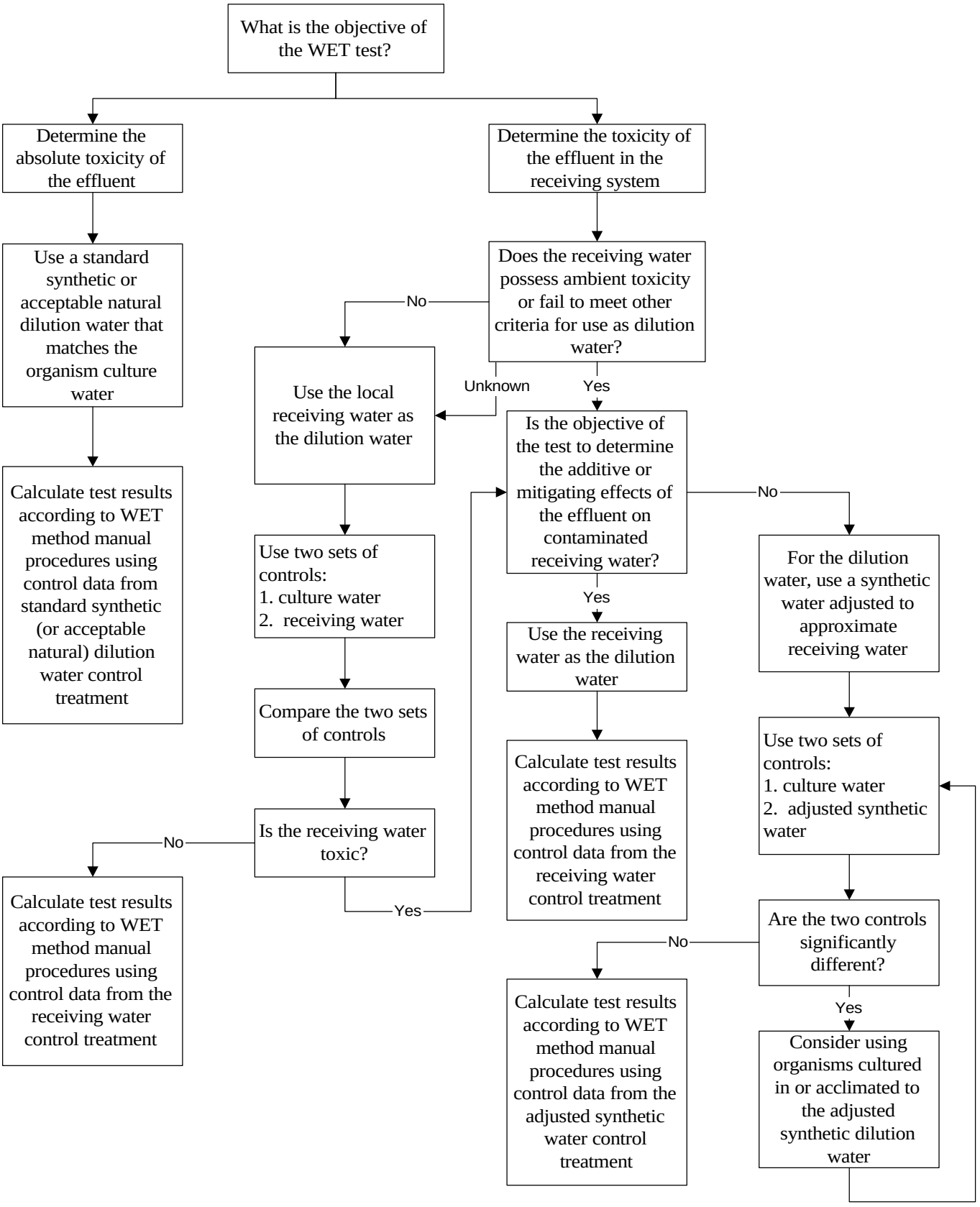
- is appropriate for the objectives of the test;
- supports adequate performance of the test organisms with respect to survival, growth, reproduction, or other responses that may be measured in the test (i.e., consistently meets test acceptability criteria for control responses);
- is consistent in quality; and
- does not contain contaminants that could produce toxicity.

In the WET method manuals (USEPA, 1993c; USEPA, 1994a; USEPA 1994b), Section 7 describes the types of dilution water that may be used for WET testing depending upon the objectives of the test. This section provides procedures for preparing synthetic dilution waters and procedures for the collection and handling of receiving waters or natural dilution waters. The selection of the appropriate dilution water type should be made independently for each effluent based upon the objectives of the test, the condition and quality of ambient receiving water, in-stream dilution potential, and recommendations or requirements from local regulatory authorities.

How do I choose an appropriate dilution water?

Figure 6.1 is provided to assist in selecting an appropriate dilution water for WET testing. First, the choice of dilution waters should be consistent with the objectives of the WET test, thus the objective of testing should be clearly defined by the regulatory authority. Tests can be conducted in the standard reconstituted dilution water to assess the absolute toxicity of the effluent. The WET method manuals (USEPA, 1993c; USEPA, 1994a; USEPA 1994b) describe this as the primary objective of NPDES permit-related toxicity testing. To determine the toxicity of the effluent in the receiving system, tests can be conducted using receiving water for dilution or synthetic dilution water adjusted to approximate receiving water characteristics (USEPA, 1993c; USEPA, 1994a; USEPA 1994b; USEPA, 1996a). EPA's Technical Support Document discusses this objective in context of EPA's water quality based toxics control program (USEPA, 1991b).

Figure 6.1. Flowchart for appropriate selection and use of dilution water in WET testing.



What dilution water should I use when determining absolute toxicity of an effluent?

If the objective of the WET test is to determine the absolute toxicity of the effluent, then a standardized synthetic water is recommended for use as dilution water. A standardized synthetic dilution water has the following advantages: proven success in maintaining organism health, known chemical composition, reduced potential for effluent/dilution water interactions that may affect toxicity, and better test reproducibility and repeatability. Under some circumstances, a consistent, high purity natural water source (e.g., uncontaminated seawater or treated well water) may be used in lieu of a synthetic water to determine the absolute toxicity of an effluent. Such waters may be used if:

- the water is similar in physical and chemical composition to the standardized synthetic water (i.e., hardness, alkalinity, pH, salinity);
- the water is used consistently and successfully by the testing laboratory for culturing the test organisms; and
- survival and reproduction records demonstrating the successful use of the water for culturing are provided and approved by the local regulatory authority.

What dilution water should I use when determining the toxicity of an effluent in the receiving system?

If the objective of the WET test is to determine the toxicity of the effluent in the receiving system, the local receiving water may be the most appropriate choice of dilution water. The use of receiving water increases the environmental relevance of WET testing by simulating effluent/receiving water interactions in the test. This also improves the capacity of the WET test to predict in-stream effects. Despite these benefits, the local receiving water should first be evaluated to determine its appropriateness for use as dilution water. To be acceptable for use as dilution water, a receiving water should meet all of the following requirements:

- **The receiving water should be collected as a grab sample from upstream or near the final point of discharge for the effluent of interest.** The receiving water sample should be collected from as close to the point of discharge as possible while remaining outside of the influence of the discharge. This determination may be made by physical or chemical measurements or by preliminary testing. Once an appropriate collection site has been located, the location should be fully described and established as the standardized receiving water collection location for the effluent discharge of interest.
- **The receiving system should have adequate flow year round at the established receiving water collection location.** For instance, where the receiving water is classified as an intermittent stream or where zero flow conditions exist, the use of receiving water for dilution is inappropriate. Under these circumstances, a synthetic water adjusted to approximate the characteristics (pH, hardness, alkalinity) of the closest downstream perennial water should be used.

- **The receiving water should support adequate performance of the test organisms with respect to survival, growth, reproduction, or other responses that may be measured in the test.** This is a primary requirement for all dilution waters (see question, "What does EPA consider to be an acceptable dilution water?"). This means that the 100% receiving water concentration used as a dilution water control should consistently meet test acceptability criteria for control responses.
- **The receiving water should be consistent in quality and not contain contaminants that could produce toxicity.** This is a primary requirement for all dilution waters (see question, "What does EPA consider to be an acceptable dilution water?"). In the case of receiving waters, this requirement is evaluated by the use of dual controls. For each test using receiving water for dilution, a 100% receiving water control and a 100% culture water control should be run concurrently in the test and compared to determine the presence of toxicity in the receiving water (for more information on the use of dual controls, see the following question, "When and how do I use dual controls?"). If and when toxicity is identified in the receiving water, the use of receiving water for dilution should be discontinued. While it is recognized that receiving water characteristics are dynamic, the receiving water should consistently display no ambient toxicity. The presence of ambient toxicity may cause many receiving systems to be inappropriate for use as a dilution water source. In many circumstances the receiving system may be impacted by many other point and non-point sources of pollution. Use of receiving water that possesses consistent or intermittent ambient toxicity is discouraged in most cases. Test results are difficult to interpret, and low to moderate toxicity in the effluent is difficult to detect in the presence of contaminated dilution water. Receiving water that possesses ambient toxicity is recommended for use as dilution water only if the objective of the test is specifically to determine the additive or mitigating effects of the effluent on the contaminated receiving water.
- **The receiving water should be free from pathogens and parasites that could affect WET test results.** The presence of pathogens or parasites in the dilution water can cause sporadic mortalities in the test that are unrelated to effluent toxicity. Due to these sporadic mortalities, tests may fail to meet test acceptability criteria or anomalous concentration-response patterns may be produced. Receiving water that is confirmed or suspected to contain pathogens or parasites should not be used as dilution water.

If the local receiving water is inappropriate for use as dilution water due to failure to meet one of the above requirements, a synthetic dilution water adjusted to approximate the chemical characteristics (pH, hardness, alkalinity, salinity) of the receiving water should be used. The adjustment of synthetic dilution waters should be within the bounds of the test method and organism tolerances and should be conducted only for the purpose of matching dilution water to receiving water conditions. For most freshwaters in the U.S., a reasonable match can be obtained by adjusting the amounts of standard synthetic freshwater reagents (as described in Table 6 of Section 7 in the WET method manuals) to produce the desired hardness (from very soft to very hard). Mineral water also may be diluted appropriately (as described in Table 7 of Section 7 in the WET method manuals) to achieve the desired hardness. These standard preparations span the

range of hardness, pH, and alkalinity that is commonly found in U.S. waters. When the receiving water possesses an ionic balance that is atypical, the amounts of individual ion constituents in the synthetic freshwater preparation may be further adjusted to approximate the ionic balance of the receiving water. This may occur in coastal or arid regions, where the ionic composition may be more dominated by sodium and chloride ions than calcium and bicarbonate ions. For marine and estuarine testing, receiving water composition generally can be matched by preparing synthetic seawater at the appropriate salinity or adjusting the salinity of a natural seawater using deionized water, artificial sea salts, or hypersaline brine.

In the case of freshwater and marine testing, the preparation of synthetic dilution water can be adjusted to approximate the chemical characteristics of the receiving water; however, the dilution water should not be adjusted to match the properties of the effluent. High concentrations of common ions and ion imbalance in the effluent can be a source of toxicity (McCulloch *et al.*, 1993; Goodfellow *et al.*, 2000), and therefore should be included in the analysis of toxicity and not adjusted for in the test.

If an adjusted synthetic water is used for dilution and this water differs from the water used for culturing the organisms, dual controls are required by the WET method manuals as described below.

When and how do I use dual controls?

When the dilution water used in a test differs from the water used to culture, hold, and maintain the test organisms, an additional set of dilution water controls should be evaluated in the WET test. This is generally the case when a natural receiving water or an adjusted synthetic water is used for dilution, but additional controls also may be necessary for standard synthetic dilution waters if organisms are cultured in an alternative water. A culture water control should consist of 100% culture water, and a dilution water control should consist of 100% of the dilution water used in the test. These two controls should be run concurrently in the test and undergo the same test conditions.

Prior to the analysis of test treatment data, the two controls (dilution water control and culture water control) should be compared to determine if statistically significant differences exist. This comparison should be made using a t-test as described in Appendix H of the freshwater method manual (USEPA, 1994a) and Appendix G of the marine method manual (USEPA, 1994b). If there is no statistically significant difference between the two controls, the dilution water control should be used for further analysis and comparisons with the treatment groups. If a receiving water control is significantly different from the culture control, this may indicate ambient toxicity in the receiving water. In this case, the use of a synthetic dilution water adjusted to approximate the receiving water may be more appropriate. If an adjusted synthetic dilution water shows a significant difference from the culture control, this generally indicates that either the chemical adjustments of the dilution water were outside of the tolerance range of the test organism or

acclimation of the test organisms to the dilution water is necessary. In this situation, the analyst should consider using organisms cultured in water more similar to the dilution water or consider acclimating the test organisms to the adjusted dilution water prior to the test. These options, however, may increase test cost and may be impractical for laboratories that test effluents from numerous dischargers, each with specific dilution water requirements. For this reason, local regulatory authorities may wish to reevaluate test objectives for this effluent and consider the use of a standardized synthetic water.

How might the choice of dilution waters affect WET test results?

The selection of dilution waters can have significant impact on the results of a WET test. The physical and chemical properties of the dilution water can interact with contaminants in the sample to increase or reduce toxic effect. The presence of acid volatile sulfides (Di Toro *et al.*, 1992), hardness (Belanger *et al.*, 1989), and acidity (Schubauer-Berigan *et al.*, 1993) are all known to significantly affect the bioavailability (and hence the toxicity) of metals. Organic and other hydrophobic contaminants may bind or adsorb to colloids or organic matter in natural waters (Larson and Weber, 1994). These reactions could potentially decrease toxicity by reducing the free concentration of the contaminant, or increase toxicity for filter feeding, sediment dwelling, or sediment ingesting organisms through increased exposure and uptake of the contaminant from food sources. For these reasons, the selection of dilution water for WET testing should be carefully considered.

7 References

Belanger, S.E., J.L. Farris, and D.S. Cherry. 1989. Effects of diet, water hardness, and population source on acute and chronic copper toxicity to *Ceriodaphnia dubia*. *Arch. Environ. Contam. Toxicol.* 18: 601-611.

Casarett, L.J. and J. Doull. 1975. *Toxicology: The Basic Science of Poisons*. Macmillan Publishing Co., New York.

Chapman, G.A., B.S. Anderson, A.J. Bailer, R.B. Baird, R. Berger, D.T. Burton, D.L. Denton, W.L. Goodfellow, Jr., M.A. Heber, L.L. McDonald, T.J. Norberg-King, and P.J. Ruffier. 1996. Discussion synopsis, methods and appropriate endpoints. Chapter 3 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., SETAC Press, Pensacola, FL, pp. 51-82.

Davis, J.M. and D.J. Svendsgaard. 1993. Nonmonotonic dose-response relationships in toxicological studies. In *Biological Effects of Low Level Exposures: Dose-Response Relationship*. E.J. Calabrese, ed., Lewis Publishers, Boca Raton, FL, pp. 67-86.

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.

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., SETAC Press, Pensacola, FL, pp. 83-102.

Di Toro, D.M., J.D. Mahony, D.J. Hansen, K.J. Scott, A.R. Carlson, and G.T. Ankley. 1992. Acid volatile sulfide predicts the acute toxicity of cadmium and nickel in sediments. *Environ. Sci. Tech.* 26(1): 96-101.

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

Goodfellow, W.L., P.B. Dorn, L.W. Ausley, D.T. Burton, D.L. Denton, D.R. Grothe, M.A. Heber, T.J. Norberg-King, and J.H. Rodgers. 2000. Major Ion Toxicity in Effluents: A Review with Permitting Recommendations. *Environ. Toxicol. Chem.* 19(1): 175-182.

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

Larson, R.A. and E.J. Weber. 1994. *Reaction Mechanisms in Environmental Organic Chemistry*. Lewis Publishers, Boca Raton, FL.

McCulloch, W.L., W.L. Goodfellow and J.A. Black. 1993. Characterization, identification, and confirmation of total dissolved solids as effluent toxicants. In Environmental Toxicology and Risk Assessment: 2nd Volume. STP 1216. J.W. Gorsuch, F.J. Dwyer, C.J. Ingersoll and T.W. LaPoint, eds., American Society for Testing and Materials, Philadelphia, PA, pp. 213-227.

SAS Institute. 1990. SAS/STAT User's Guide, 4th Ed. Version 6, Cary, NC.

Schubauer-Berigan, M.K., J.R. Dierkes, P.D. Monson, and G.T. Ankley. 1993. pH-dependent toxicity of Cd, Cu, Ni, Pb, and Zn to *Ceriodaphnia dubia*, *Pimephales promelas*, *Hyalella azteca*, and *Lumbriculus variegatus*. *Environ. Toxicol. Chem.* 12(12): 1261-1266.

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(6): 1322-1329.

U.S. Environmental Protection Agency. 1991a. *Methods for Aquatic Toxicity Identification Evaluations: Phase I Toxicity Characterization Procedures*, 2nd ed., EPA/600/6-91/003. U.S. Environmental Protection Agency, Office of Research and Development, Environmental Research Laboratory, Duluth, MN.

U.S. Environmental Protection Agency. 1991b. *Technical Support Document for Water Quality-Based Toxics Control*. EPA/505/2-90/001. U.S. Environmental Protection Agency, Office of Water Enforcement and Permits and Office of Water Regulations and Standards, Washington, DC.

U.S. Environmental Protection Agency. 1992. *Toxicity Identification Evaluation: Characterization of Chronically Toxic Effluents, Phase I*, EPA/600/6-91/005F. U.S. Environmental Protection Agency, Office of Research and Development, Environmental Research Laboratory, Duluth, MN.

U.S. Environmental Protection Agency. 1993a. *Methods for Aquatic Toxicity Identification Evaluation: Phase II Toxicity Identification Procedures for Acutely and Chronically Toxic Samples*. EPA/600/R-92/080. U.S. Environmental Protection Agency, Office of Research and Development, Duluth, MN.

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

U.S. Environmental Protection Agency. 1993c. *Methods for Measuring the Acute Toxicity of Effluents and Receiving Waters to Freshwater and Marine Organisms*, 4th ed., EPA/600/4-90/027F. U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory (currently, National Exposure Research Laboratory), Cincinnati, OH.

U.S. Environmental Protection Agency. 1994a. *Short-term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Freshwater Organisms*, 3rd ed., EPA/600/4-91/002. U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Cincinnati, OH.

U.S. Environmental Protection Agency. 1994b. *Short-term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Marine and Estuarine Organisms*, 2nd ed., EPA/600/4-91/003. U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory (currently, National Exposure Research Laboratory), Cincinnati, OH.

U.S. Environmental Protection Agency. 1995. *Short-term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to West Coast Marine and Estuarine Organisms*, 1st ed., EPA/600/R-95/136. U.S. Environmental Protection Agency, Office of Research and Development, Cincinnati, OH.

U.S. Environmental Protection Agency. 1996a. Clarifications Regarding Flexibility in 40 CFR Part 136 Whole Effluent Toxicity (WET) Test Methods, April 10, 1996 memorandum from Tudor Davies, U.S. Environmental Protection Agency, Office of Science and Technology, Washington D.C.

U.S. Environmental Protection Agency. 1996b. Marine Toxicity Identification Evaluation (TIE): Phase I Guidance Document. EPA/600/R-95/054. U.S. Environmental Protection Agency, Environmental Effects Research Laboratory, Narragansett, RI.

U.S. Environmental Protection Agency. 1999. *Errata for Effluent and Receiving Water Toxicity Test 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 Estuarine Organisms.* January 1999. EPA/600/R-98/182. U.S. Environmental Protection Agency, Office of Research and Development, Duluth, MN.

U.S. Environmental Protection Agency. 2000. *Understanding and Accounting for Method Variability in Whole Effluent Toxicity Applications Under the National Pollutant Discharge Elimination System Program*. EPA/833/R-00/003. U.S. Environmental Protection Agency, Office of Wastewater Management, Washington, D.C.

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

Washington State Department of Ecology. 1997. *Laboratory Guidance and Whole Effluent Toxicity Test Review Criteria*. Washington State Department of Ecology Water Quality Program, Olympia, WA.

Zar, J.H. 1984. *Biostatistical Analysis*, 2nd ed. Prentice-Hall Engineering, Prentice-Hall Inc., Englewood Cliffs, N.J.