United States Environmental Agency
Office of Wastewater Management (4203M)
EPA 833-R-10-003 (June/2010)
http://www.epa.gov/npdes/permitbasics

National Pollutant Discharge Elimination System Test of Significant Toxicity Implementation Document

June 2010

NATIONAL POLLUTANT DISCHARGE ELIMINATION SYSTEM
TEST OF SIGNIFICANT TOXICITY
IMPLEMENTATION DOCUMENT

An Additional Whole Effluent Toxicity Statistical Approach for Analyzing Acute and Chronic Test Data

U.S. Environmental Protection Agency
Office of Wastewater Management
Water Permits Division
1200 Pennsylvania Avenue, NW
Mail Code 4203M
EPA East – Room 7135
Washington, DC 20460

June, 2010

NOTICE AND DISCLAIMER

This document provides the basis for implementing the Test of Significant Toxicity (TST) approach under the National Pollutant Discharge Elimination System (NPDES) for permitting authorities (states and Regions) and persons interested in analyzing whole effluent toxicity (WET) test data using the traditional hypothesis testing approach as part of the NPDES Program under the Clean Water Act (CWA). This document describes what the U.S. Environmental Protection Agency (EPA) believes is another statistical option to analyze valid WET test data for NPDES WET reasonable potential and permit compliance determinations. The document does not, however, substitute for the CWA, an NPDES permit, or EPA or state regulations applicable to permits or WET testing; nor is this document a permit or a regulation itself. The TST approach does not result in changes to EPA's WET test methods promulgated at Title 40 of the *Code of Federal Regulations* Part 136. The document does not and cannot impose any legally binding requirements on EPA, states, NPDES permittees, or laboratories conducting or using WET testing for permittees (or for states in evaluating ambient water quality). EPA could revise this document without public notice to reflect changes in EPA policy and guidance. Finally, mention of any trade names, products, or services is not and should not be interpreted as conveying official EPA approval, endorsement, or recommendation.

CONTENTS

TABLES

Table **Page**

EXECUTIVE SUMMARY

The U.S. Environmental Protection Agency (EPA or the Agency) has developed a new statistical approach that assesses the whole effluent toxicity (WET) measurement of wastewater effects on specific test organisms' ability to survive, grow, and reproduce. The new approach is called the Test of Significant Toxicity (TST) and is a statistical method that uses hypothesis testing techniques based on research and peer-reviewed publications. The TST approach examines whether an effluent, at the critical concentration (e.g., in-stream waste concentration or IWC, as recommended in EPA's Technical Support Document (TSD) (USEPA 1991) and implemented under EPA's WET National Pollutant Discharge Elimination System (NPDES) permits program and the control within a WET test differ by an unacceptable amount (the amount that would have a measured detrimental effect on the ability of aquatic organisms to thrive and survive). EPA Regions and their NPDES states can still use EPA's TSD approaches. The TST approach is another statistical option to analyze valid WET test data.

Since the inception of EPA's NPDES WET Program in the mid 1980s, the Agency has striven to advance and improve its application and implementation under the NPDES Program. The TST approach explicitly incorporates test power (the ability to correctly classify the effluent as non-toxic, also see reference in the glossary under power) and provides a positive incentive to generate valid, high quality WET data to make informed decisions regarding NPDES WET reasonable potential (RP) and permit compliance determinations. Once the WET test has been conducted (using multiple effluent concentrations and other requirements as specified in the EPA WET test methods), the TST approach can be used to analyze the WET test results to assess whether the effluent discharge is toxic at the critical concentration. Performing the EPA WET test where the minimum five required test concentrations (pursuant to the EPA WET test methods) can establish a concentration-response curve. The TST approach is designed to be used for a two concentration data analysis of the IWC or a receiving water concentration (RWC) compared to a control concentration. Using the TST approach, permitting authorities will have more confidence when making NPDES determinations as to whether a permittee's effluent discharge is toxic or non-toxic. Use of the TST approach does not result in any changes to EPA's WET test methods; however, a facility might desire to modify its future WET tests by increasing the number of replicates over the minimum required (USEPA 1995, 2002a, 2002b, 2002c) by the approved EPA WET test method to increase test power, which is the probability of declaring an effluent *non-toxic* if the organism response at the IWC is truly acceptable. If WET tests have already been performed, the WET data generated cannot be modified to increase the number of test replicates because the TST analysis is done on valid WET data generated within a WET test.

The TST approach was developed on the basis of extensive analyses and detailed research. EPA used valid WET data from *more than 2,000* WET tests to develop and evaluate the TST approach. The TST approach was tested using *nine different* WET test methods comprising *twelve biological endpoints* (e.g., reproduction, growth, survival) and representing most of the different types of WET test designs currently in use. *More than one million* computer simulations were also used to select error rates achieving EPA's regulatory management decisions for the TST approach.

Background

In the NPDES Program, an effluent sample is declared toxic relative to a permitted WET limit if the no observed effect concentration (NOEC) is less than the permitted IWC using a hypothesis statistical approach. In that traditional hypothesis approach, the question being answered is, "Is the mean response of the organisms the same in the control and at the IWC?" The hypothesis testing approach has four possible outcomes: (1) the IWC is truly toxic and is declared toxic, (2) the IWC is truly non-toxic and is declared non-toxic, (3) the IWC is truly toxic but is declared non-toxic, and (4) the IWC is truly non-toxic but is declared toxic. The latter two possible outcomes represent decision errors that can occur with any hypothesis testing approach. In the NPDES WET Program, those two types of errors can occur when test control replication is very good (i.e., test is very precise) so that a very small difference between IWC and control is declared toxic (outcome [4] above), and when test control replication is poor (i.e., the test is very imprecise) so that *even large differences in organism response between the IWC and control cannot be distinguished as statistically different, and the effluent is incorrectly classified as non-toxic (outcome [3] above).*

Organism responses to the IWC and control are unlikely to be exactly the same. The difference might be so small that even if statistically significant, it would be considered biologically negligible. Another approach for assessing an effluent's toxicity on the basis of collected WET data might be to rephrase the question, "Does the mean WET test response in the control and the IWC differ by a defined biological amount?" That approach is known as the *test of bioequivalence,* which the Food and Drug Administration has successfully used to evaluate drugs, as have many researchers in other biological fields. Using the TST approach, the question is, "Is the organism response at the IWC less than or equal to a fixed fraction of the control response (e.g., 75 percent of the control mean response)?" That fixed fraction, expressed as a decimal between 0.00 and 1.00, is termed "*b"* in the TST approach. Thus, the hypothesis being tested is written as follows: mean response [IWC] $\leq b \times$ mean response [control].

The TST approach requires defining what is considered toxic. For *chronic* testing (i.e., for both lethal and sublethal toxicity test endpoints) in EPA's NPDES WET Program, the *b* value in the TST analysis is set at 0.75, which means that a 25 percent effect (or more) is considered evidence of unacceptable *chronic* toxicity. IWC responses substantially less than a 25 percent effect would be interpreted to have a lower risk potential. The regulatory management decision (RMD) for *acute* WET methods is set at 0.80, which means that a 20 percent effect (or more) is considered evidence of unacceptable *acute* toxicity. The acute RMD toxicity threshold is higher than that for chronic WET methods because of the severe environmental implications of acute toxicity (lethality or organism death). For more discussion on the *b* values of 0.75 (chronic toxicity) and 0.80 (acute toxicity), see Section 2.1 of this document.

EPA's RMDs using the TST approach identify true toxicity in WET tests most of the time when it occurs, while also minimizing the probability that the IWC is declared toxic when in fact it is not. That objective requires additional RMDs regarding acceptable maximum false positive *(β or beta using a TST approach) and false negative rates (α or alpha using a TST approach***)**. In the TST approach, the RMDs are defined as (1) declare a sample toxic at least 75 percent of the time (alpha, $\alpha \leq 0.25$) when there is unacceptable toxicity (20 percent effect for acute and 25 percent effect for chronic test methods), and (2) declare an effluent non-toxic no more than 5 percent

(beta, $\beta \leq 0.05$) of the time when the mean effect at the critical effluent concentration is ≤ 10 percent for both acute and chronic WET tests (including for sublethal endpoints). For more discussion on the RMDs, see Section 2.1 of this document.

On the basis of EPA's analyses, the alpha levels shown in Table ES-1 are recommended for the nine WET test methods examined using the TST approach. An important feature of the TST approach is that the false negative error rate (rate of declaring a toxic effluent to be non-toxic) is established, which, under the traditional hypothesis testing approach, had not been established by EPA previously. For more discussion on the inclusion of the beta error rate in the TST approach, see Section 1.2 of this document and Section 1.1 on the current approach in EPA's 1991 TSD. A demonstrated benefit of the TST approach is that increasing within-test replication (the test power) results in a *lower* rate of WET tests being declared toxic using the TST approach when the IWC is truly non-toxic.

Results obtained from the TST analyses using the *nine* EPA test methods should be applicable to other EPA WET methods not examined. For example, results generated under this project for the fish *Pimephales promelas* survival and growth test is extrapolated to other EPA fish survival and growth tests (e.g., *Menidia* sp., *Cyprinus variegatus*, *Atherinops affinis*) because those test methods use a similar test design (e.g., number of replicates, number of organisms tested) and measure the same endpoints.

Summary

More than 2,000 WET test results and more than one million simulations were conducted to develop the technical basis for the TST approach. The approach builds on the strengths of the traditional hypothesis testing approach, including use of robust statistical analyses and published EPA documents regarding WET data analysis and interpretation. The TST approach yields a rigorous statistical interpretation of valid WET data by incorporating transparent RMDs and established alpha and beta error rates, which can provide incentives to generate test results having greater test power. Because the approach considers statistical test power, its use will result in greater confidence in WET regulatory decisions. In addition, the TST approach provides a positive incentive for the permittee to generate valid, high quality WET data by either increasing the number of test replicates for the IWC and the control within a test and/or achieving better precision within a test through improved WET test method performance (e.g., a high level of quality assurance and quality control).

Permitting authorities should consider the practical programmatic shift from the traditional hypothesis testing approach to the TST approach by opening a dialogue with their regulated community. In addition, they might want to begin to identify what changes might be needed to assimilate the TST approach into any regulations, policy, guidance, and training in their respective NPDES WET Programs. Again, the traditional hypothesis testing approach under EPA's TSD is still considered valid as applied; however, that approach can now be advanced through the TST approach by providing new incentives to permittees to provide valid, high quality WET data.

Table ES-1. Summary of alpha (α) levels or false negative rates recommended for different WET test methods using the TST approach

EPA WET test method	*b* value	Probability of declaring a toxic effluent non-toxic False negative (α) error[a]
Chronic Freshwater and East Coast Methods		
Ceriodaphnia dubia (water flea) survival and reproduction	0.75	0.20
Pimephales promelas (fathead minnow) survival and growth	0.75	0.25
Selenastrum capricornutum (green algae) growth	0.75	0.25
Americamysis bahia (mysid shrimp) survival and growth	0.75	0.15
Arbacia punctulata (Echinoderm) fertilization	0.75	0.05
Cyprinodon variegatus (Sheepshead minnow) and *Menidia beryllina* (inland silverside) survival and growth	0.75	0.25
Chronic West Coast Marine Methods		
Dendraster excentricus and *Strongylocentrotus purpuratus* (Echinoderm) fertilization	0.75	0.05
Atherinops affinis (topsmelt) survival and growth	0.75	0.25
Haliotis rufescens (red abalone), *Crassostrea gigas (oyster), Dendraster excentricus, Strongylocentrotus purpuratus* (Echinoderm) and *Mytilus sp* (mussel) larval development methods	0.75	0.05
Macrocystis pyrifera (giant kelp) germination and germ-tube length	0.75	0.05
Acute Methods		
Pimephales promelas (fathead minnow), *Cyprinodon variegatus* (Sheepshead minnow), *Atherinops affinis* (topsmelt), *Menidia beryllina* (inland silverside) acute survival[b]	0.80	0.10
Ceriodaphnia dubia, *Daphnia magna, Daphnia pulex*, *Americamysis bahia* acute survival[b]	0.80	0.10

Notes:

a. (1) declare a sample toxic at least 75 percent of the time (alpha ≤ 0.25) when there is unacceptable toxicity (20 percent effect for acute and 25 percent effect for chronic test methods) and (2) declare an effluent non-toxic no more than 5 percent of the time (beta ≤ 0.05) when the mean effect at the critical effluent concentration is 10 percent for both acute and chronic WET tests (including sublethal endpoints). For more discussion on the RMDs, see Section 2.1 of this document.

b. Based on four replicate test design

In addition, EPA recommends the following:

- Permitting authorities should decide up front which approach (the EPA's 1991 TSD approach, the TST approach, or another scientifically defensible approach that is sufficient to meet the statutory and regulatory requirements) they will follow (including for their RP procedures) and use the selected approach consistently in all their state NPDES permits. Permitting authorities should ensure that the most environmentally protective approach is consistently used across all permits when assessing valid WET data (e.g., WET RP) for

NPDES permit requirements (e.g., WET limits, monitoring frequencies, toxicity identification evaluation/toxicity reduction evaluation) and avoid selecting the approach that underestimates the true toxicity of the permitted effluent discharge.

- Where a *small data set* exists (fewer than four valid WET tests performed and reported in the previous 5 years), permitting authorities should use the TSD approach for determining RP. With small WET data sets, the TSD's RP multiplying factor is more conservative for environmental water quality protection purposes than the TST. The TST approach is intended for larger data sets (four or more) because it does not use an RP multiplying factor.
- If WET tests have already been performed, the WET data generated cannot be modified to increase the number of test replicates within a test. The decision to increase the number of within test replicates is a decision that needs to be made before conducting the WET tests.
- Where a permittee has concerns about WET data quality, EPA recommends increasing the number of replicates in tests, even if the permitting authority has not yet adopted the TST approach.

ACRONYMS AND ABBREVIATIONS

CFR	Code of Federal Regulations
CV	coefficient of variation
CWA	Clean Water Act
DMR	discharge monitoring report
EC	effect concentration
EPA	U.S. Environmental Protection Agency
IC25	25 percent inhibition concentration
IWC	in-stream waste concentration
LC50	50 percent lethal concentration
LOEC	lowest observed effect concentration
MDL	maximum daily limit
NOEC	no observed effect concentration
NPDES	National Pollutant Discharge Elimination System
QA/QC	quality assurance/quality control
RMD	regulatory management decision
RP	reasonable potential
RPMF	reasonable potential multiplying factor
RWC	receiving water concentration
SWAMP	Surface Water Ambient Monitoring Program (California)
TAC	test acceptability criteria
TIE	toxicity identification evaluation
TRE	toxicity reduction evaluation
TSD	Technical Support Document for Water Quality-Based Toxics Control
TST	Test of Significant Toxicity
TU	toxicity unit
WET	whole effluent toxicity

GLOSSARY

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

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

Chronic Toxicity Test is a short-term test in which sublethal effects (e.g., reduced growth or reproduction) are usually measured in addition to lethality.

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

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

Effect Concentration (EC) is a point estimate of the toxicant concentration that would cause an observable adverse effect (e.g., mortality, fertilization). EC25 is a point estimate of the toxicant concentration that would cause an observable adverse effect in 25 percent of the test organisms.

False Negative is when the in-stream waste concentration is declared non-toxic but in fact is truly toxic. In the traditional hypothesis approach, false negative error rate is denoted by Beta (β). In the TST approach, false negative error rate is denoted as Alpha (α), which applies when the percent effect in the critical effluent concentration is $\geq$ 25% for a given test.

False Positive is when the in-stream waste concentration is declared toxic but in fact is truly non-toxic. In the traditional hypothesis approach, false positive error rate is denoted by Alpha (α). In the TST approach, false positive error rate is denoted as Beta (β), which applies when the percent effect in the critical effluent concentration is $\leq$ 10% for a given test.

Hypothesis Testing is a statistical approach (e.g., Dunnett's procedure) for determining whether a test concentration is statistically different from the control. Endpoints determined from hypothesis testing are no observed effect concentration and lowest observed effect concentration (LOEC). The two hypotheses commonly tested in WET are:

Null hypothesis (Ho): The effluent is non-toxic.

Alternative hypothesis (Ha): The effluent is toxic.

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

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

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

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

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

National Pollutant Discharge Elimination System (NPDES) is the national program for issuing, modifying, revoking and reissuing, terminating, monitoring and enforcing permits, and imposing and enforcing pretreatment requirements, under the Clean Water Act sections 307, 318, 402, and 405.

Power (or test power) in the context of the Test of Significant Toxicity approach, is the probability of correctly declaring an effluent non-toxic when, in fact, it has an acceptably low level of toxicity.

Precision is a measure of reproducibility (which is a statistical term about the ability to reproduce similar results across test replicates with in a test treatment) within a data set. Precision can be measured both within a laboratory (within-laboratory) and between laboratories (between-laboratory) using the same test method and toxicant.

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

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

Reasonable Potential (RP) is where an effluent is projected or calculated to cause an excursion above a water quality standard based on a number of factors including the four factors listed in Title 40 of the *Code of Federal Regulations* Part 122.44(d)(1)(ii).

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

Regulatory Management Decision (RMD) is the decision that represents the maximum allowable error rates and thresholds for toxicity and non-toxicity that would result in an acceptable risk to aquatic life.

Replicate is two or more independent organism exposures of the same treatment (i.e., effluent concentration) within a WET test. Replicates are typically separate test chambers with organisms, each having the same effluent concentration**.**

Sample is defined as a representative portion of a specific environmental matrix that is used in toxicity testing. For this document, environmental matrices could include effluents, surface waters, groundwater, stormwater, and sediment.

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

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

Test Acceptability Criteria (TAC) are test method-specific criteria for determining whether toxicity test results are acceptable. The effluent and reference toxicant must meet specific criteria as defined in the test method (e.g., for the *Ceriodaphnia dubia* survival and reproduction test, the criteria are as follows: the test must achieve at least 80 percent survival and an average of 15 young per surviving female in the control and at least 60% of surviving organisms must have three broods).

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

Type I Error (alpha α) is the error of rejecting the null hypothesis (H_o) that should have been accepted.

Type II Error (beta β) is the error of accepting the null hypothesis (H_o) that should have been rejected.

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

Welch's t-test is an adaptation of Student's t-test intended for use with two samples having unequal variances.

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

1.0 INTRODUCTION

Whole effluent toxicity (WET) test methods are laboratory procedures that measure biological effects (e.g., survival, growth, reproduction) on aquatic organisms exposed to effluents or storm water discharged to receiving waters in implementing the National Pollutant Discharge Elimination System (NPDES) Program under the Clean Water Act (CWA) section 402. Since the publication of EPA's *Technical Support Document for Water Quality-based Toxics Control* (TSD) (USEPA 1991), permitting authorities have requested alternative approaches for analyzing WET test data that would provide increased confidence in the data assessment and simplify the NPDES permit decision-making process with respect to WET. In response to those requests, EPA developed the TST approach as another statistical option to analyze valid WET test data. This document presents the NPDES programmatic features of the TST statistical approach for analyzing valid WET data and how it can be used to support permitting authorities and permittees when analyzing and interpreting WET test data. Use of the TST approach does not result in any changes to EPA's WET test methods, nor does it preclude the use of EPA's TSD approaches for analyzing valid WET data, or another scientifically defensible approach that is sufficient to meet the statutory and regulatory requirements.

1.1 Terminology and Concepts

This section briefly summarizes the major statistical concepts and terminology involved in WET analysis so as to give the reader a context with which to understand the TST approach and how it differs from current statistical approaches used to analyze valid WET data. This TST implementation document is not intended to provide a detailed discussion of WET test methods, data interpretation, or statistics, and it is assumed that the reader will consult EPA's TSD, WET test method documents, and other WET-related documents (e.g., *Understanding and Accounting for Method Variability in Whole Effluent Toxicity Applications*, USEPA 2000).

In the NPDES Program, WET tests examine organism responses to effluent, typically along a dilution series (USEPA 1995, 2002a, 2002b, 2002c). Acute WET methods measure the lethal response of test organisms exposed to effluent (USEPA 2002c). The principal response endpoints for those methods are the effluent concentration that is lethal to 50 percent of the test organisms (LC50) or the effluent concentration at which survival is significantly lower than the control. Chronic WET methods often measure both lethal and sublethal responses of test organisms. The statistical endpoints used in chronic WET testing are the no observed effect concentration (NOEC) and the 25 percent inhibition concentration (IC25). The NOEC endpoint is determined using a hypothesis testing approach that identifies the maximum effluent concentration at which the response of test organisms is not significantly different from the control. From a regulatory perspective, an effluent sample is declared toxic if the NOEC is less than the in-stream waste concentration (IWC) specified through the WET limitations in the permit. The IC25, by contrast, is a point estimation approach. It identifies the concentration at which the response of test organisms is 25 percent below that observed in the control concentration, and it interpolates the effluent concentration at which this magnitude of response is expected to occur. From a regulatory perspective, an effluent sample is declared toxic if the IC25 is less than the IWC specified through the WET limitations in the permit. This document focuses only on the hypothesis testing approach and not on point estimation approaches for analyzing and interpreting WET data.

In any hypothesis testing approach, two hypotheses are stated: the null hypothesis and the alternative hypothesis. The statistical concepts associated with the traditional hypothesis testing approach currently used in WET analysis are summarized in Table 1. Using that approach, the null hypothesis is that the IWC is non-toxic (i.e., the organism response at the IWC is equal to or better than the response in the test control). The alternative hypothesis is that the IWC is toxic (i.e., the organism response is worse in the IWC than in the control). With any hypothesis testing approach, two types of decision errors occur: (1) conclude that the null hypothesis is correct when in fact it is not or (2) conclude that the null hypothesis is incorrect (i.e., reject the null hypothesis) and thereby declare that the alternative hypothesis is correct, when in fact the null hypothesis is correct. In WET testing, the first type of error above is referred to as a *false negative*, meaning that the *IWC is declared non-toxic when in fact it is toxic*. The second type of error above is referred to as a *false positive* in WET testing, meaning that the *IWC is declared toxic when in fact it is not*.

In the traditional hypothesis testing approach summarized in Table 1, statisticians have assigned Greek letters to the two types of errors identified above. Alpha (or α) refers to the false positive error rate. Beta (or β) refers to the rate of false negatives. In the EPA WET test methods supporting the NPDES WET Program (USEPA 1995, 2002a, 2002b), α was established but β was not. Therefore, the application of α from the EPA test methods and implemented under EPA's TSD, recommended that the maximum rate of false positives that should be observed should be low (no more than 5 percent or $\alpha = 0.05$), but the rate of false negatives was not similarly controlled and is not currently evaluated in WET testing. As a result, the rate of false negatives in the NPDES WET Program has not been controlled. Put another way, the statistical power of these tests, the ability to correctly classify the IWC as toxic (where power is defined as $1-\beta$, Table 1) has not been controlled.

As noted previously in this section, a hypothesis testing approach determines whether the organism response at the IWC is significantly worse than that in the control. In practice, this statistical approach relies on two properties of the data: the average values in the control and the IWC (e.g., average fish weight in each test concentration), and the variability observed among replicates (i.e., organisms' responses from multiple replicates) within the IWC and the control. Whether the IWC is considered toxic depends on both of those data properties, which in many cases results in a well-established, statistically rigorous way to evaluate WET data. However, there are two types of situations in which the traditional hypothesis testing approach can yield equivocal results in WET testing: (1) in tests where within-test variability is high and (2) in tests where within-test variability is exceptionally low. In the first case, because within-test variability is high, it will be difficult to determine statistically whether the organism response to the IWC is worse than the control. That could result in more false negatives than would otherwise be the case. In the second case above, because within-test variability is very low, it will be relatively easy to show statistically significant differences in organism response between the IWC and the control. That could result in more false positives (as defined in the TST approach) than would otherwise be the case.

Table 1. Expression of null and alternative hypotheses used in traditional hypothesis testing and relationships between error rates and resulting decisions based on this approach. Entries correspond to the probability decision given in parentheses. The probability of a false positive (i.e., rejecting a null hypothesis that should not have been rejected) is represented by α and the probability a false negative (i.e., failing to reject the null hypothesis when it should have been rejected) is represented by β.

	True condition	
Decision	**Null hypothesis** Treatment mean ≥ Control mean **Sample is non-toxic**	**Alternative hypothesis** Treatment mean < Control mean **Sample is toxic**
Treatment mean ≥ Control mean **Sample is non-toxic**	Correct decision (1-α)	False negative (β)
Treatment mean < Control mean **Sample is toxic**	False positive (α)	Correct decision (1 – β) (power)

1.2 Background on the TST Approach

The TST is an alternative statistical approach for analyzing and interpreting valid WET data that also uses a hypothesis testing approach but in a different way, building on previous work conducted by EPA in the NPDES WET Program (USEPA 2000) and other researchers (Erickson and McDonald 1995; Shukla et al. 2000; Berger and Hsu 1996). The TST approach is based on a type of hypothesis testing referred to as *bioequivalence testing*. Bioequivalence is a statistical approach that has long been used in evaluating clinical trials of pharmaceutical products (Anderson and Hauck 1983) and by the Food and Drug Administration (Hatch 1996; Aras 2001; Streiner 2003). The approach has also been used to evaluate the attainment of soil cleanup standards for contaminated sites (USEPA 1989) and to evaluate effects of pesticides in experimental ponds (Stunkard 1990). In the context of the NPDES WET Program, the TST approach assesses whether the response of test organisms at the IWC (e.g., fish weight or number of neonates per female) is less than a predetermined proportion of the control response that is considered unacceptably toxic. *Once the WET test has been conducted (using multiple effluent concentrations and other requirements have been met as specified in the EPA methods), the TST approach is designed to be used for a two concentration data analysis of the in-stream waste concentration (IWC) or a receiving water concentration (RWC) compared to a control concentration.*

The null hypothesis using the TST approach is that the IWC is significantly more toxic (i.e., results in a worse organism response) compared to the control (see Table 2). The alternative hypothesis using the TST approach is that the IWC is non-toxic. Thus, the null and alternative hypotheses using the TST approach are opposite of what they are under the traditional hypothesis testing approach described in Section 1.1. In addition, the meaning of α and β are also opposite from what they represent in the traditional hypothesis approach. Under the TST approach, α is associated with false negatives, and β is associated with false positives. Statistical power using the TST approach is the ability to correctly classify the IWC as non-toxic (Table 2). The proportion or fraction of the control response that represents the toxicity threshold is denoted as *b* in the equations in Table 2 and is expressed as a decimal between 0.00 and 1.00. For example,

a *b* value set at 0.85 would mean that a response at the IWC that is at least 85 percent of the control response in the test (i.e., no more than a 15 percent effect) would be considered a lower risk for environmental impacts.

Using the TST hypothesis approach in the NPDES WET Program has several benefits. By incorporating *b* in the hypothesis equation, using the TST approach, there is explicit acknowledgement of the fact that the organism response at the IWC can be less than the control organism response by a certain amount and still be considered acceptable (i.e., non-toxic). In that way, truly non-toxic samples (as defined in the TST approach) can be addressed in a clearer manner than is possible with the traditional hypothesis testing approach as practiced in the NPDES WET Program. A low false positive rate in the TST approach is further addressed by having a low β ($\beta \leq 0.05$), which means more statistical power to identify an acceptable effluent (as defined by EPA's regulatory management decisions [RMDs]) as non-toxic in the NPDES WET Program. In addition, because the null hypothesis in the TST approach is opposite to what is used in the traditional hypothesis testing approach, false negatives are explicitly addressed (α in the TST approach addresses the false negative rate). As mentioned previously, the current NPDES WET Program does not control for false negatives. Thus, the TST approach allows permitting authorities to minimize the occurrence of false negatives (i.e., declaring the IWC non-toxic when it is actually exhibiting unacceptable toxicity), while also minimizing the occurrence of false positives (i.e., declaring the IWC toxic when it is actually acceptable). The TST approach has the added advantage of providing permittees with a clear incentive to improve the precision of test results (e.g., decrease within-test variability and/or use more replicates within a WET test than the minimum required in the EPA WET test method) to reach a definitive conclusion as to whether unacceptable toxicity is observed in a test. Thus, using the TST approach, a permittee can in fact *prove a negative*, i.e., that their effluent is acceptable (non-toxic).

Table 2. Expression of null and alternative hypotheses using the TST approach and relationships between error rates and resulting decisions based on this approach. Entries correspond to the probability decision given in parentheses. The probability of a false positive (i.e., rejecting a null hypothesis that should not have been rejected) is represented by α and the probability a false negative (i.e., failing to reject the null hypothesis when it should have been rejected) is represented by β.

	True condition	
Decision	**Null hypothesis** Treatment mean ≤ b × Control mean **Sample is toxic**	**Alternative hypothesis** Treatment mean > b × Control mean **Sample is non-toxic**
Treatment mean ≤ b × Control mean **Sample is toxic**	Correct decision (1-α)	False positive (β)
Treatment mean > b × Control mean **Sample is non-toxic**	False negative (α)	Correct decision (1-β) (power)

2.0 TST METHODOLOGY

2.1 Regulatory Management Decisions for the TST Approach

Toxicity is not an absolute quantity but rather an effect that is determined relative to a control or reference sample using a given WET test method. In the TST approach, what is considered unacceptable or acceptable toxicity are explicit RMDs. For *chronic* testing in EPA's NPDES WET Program, the *b* value in the TST null hypothesis is set at 0.75, which means that a 25 percent effect (or more) is considered a demonstration of unacceptable toxicity in a given WET test. Using a 25 percent effect threshold as the *b* coefficient is consistent with EPA's use of a 25 percent inhibition concentration (IC25) as an acceptable WET endpoint for examining chronic WET data. Responses substantially less than a 25 percent effect would be interpreted as a lower risk potential. The unacceptable toxicity RMD threshold for acute WET methods is set higher than that for chronic WET methods because of the severe environmental implications of acute toxicity (lethality or organism death). Therefore, for *acute* WET tests, the *b* value in the TST approach is set at 0.80 (i.e., ≥ 20 percent effect in the effluent in acute WET tests is considered unacceptable).

For both acute and chronic WET test methods, the low-risk RMD threshold is set at a 10 percent mean effect at the IWC within a WET test. Thus, one can *prove the negative* (i.e., an effluent is *acceptable or considered non-toxic under NPDES*) if that condition is met in a WET test. For mean effect levels greater than 10 percent but less than the unacceptable toxicity RMD threshold (20 percent for acute and 25 percent for chronic WET tests), the TST approach will still declare the IWC non-toxic depending on within-test variability: the lower the variability in the WET test, the more likely the sample will be declared non-toxic on the basis of the mean responses observed under these test conditions.

EPA's RMDs using the TST approach are used to specify unacceptable toxicity in WET tests most of the time when it occurs (i.e., a low false negative rate). As mentioned previously, under the traditional hypothesis testing approach currently used in the NPDES WET Program, the false negative rate was not controlled. Using the TST approach, the false negative rate RMD is $0.05 \leq \alpha \leq 0.25$, which translates to at least 75 percent probability that an effluent causing unacceptable toxicity will be declared toxic. As noted in the previous paragraph, the unacceptable toxicity RMD threshold is defined as ≥ 20 percent effect of the IWC in acute WET tests and ≥ 25 percent effect of the IWC in chronic WET tests.

EPA also desires to minimize the probability that the IWC is declared toxic when in fact it is acceptable (i.e., low false positive rate). Under the traditional hypothesis testing approach currently used in the NPDES WET Program, the false positive rate is set at 0.05 or 5 percent. Therefore, in the TST approach, the desired false positive rate is also set at 0.05 or 5 percent ($\beta \leq 0.05$). A $\beta = 0.05$ in the TST approach means that 95 percent of the time, a truly acceptable effluent (≤ 10 percent mean effect at the IWC) will be declared non-toxic in the NPDES WET Program. Depending on the minimum WET test design required in the EPA methods (e.g., number of replicates and number of organisms per test concentration) and achievable laboratory control precision for a WET test method, α will be set between 0.05 and 0.25 while still

maintaining a $\beta \leq 0.05$. Extensive analyses were used to identify the lowest α for a given WET test method for which $\beta = 0.05$ and all other RMDs are met.

The RMD thresholds above represent boundaries in terms of desired α and β rates. An $\alpha = 0.20$ for a chronic test method, for example, means that the Type I error rate will be approximately 20 percent at a mean effect of 25 percent. At higher levels of effect in the IWC, actual Type I error rates would be lower; at lower mean effect levels in the IWC, Type I error rate would be somewhat higher, depending on the test method. Therefore, at mean effect levels between the 10 percent non-toxic RMD boundary and the unacceptable toxicity RMD boundary (20 percent for acute and 25 percent for chronic WET test methods), there are differing probabilities of an effluent being declared toxic depending on within-test variability and the difference in mean responses observed between control and IWC. As a result, there will be some instances in which TST will declare a test toxic, whereas the traditional hypothesis approach would declare that test non-toxic (particularly when within-test variability is high or the mean effect at the IWC is near 25 percent, as explained in Section 1.1). Similarly, there will be some instances in which TST will declare an effluent non-toxic but the traditional hypothesis approach would declare that test toxic (when within-test variability is low and the mean effect at the IWC is less than the 20 percent toxicity RMD threshold for acute test methods or 25 percent for chronic toxicity test methods, as explained in Section 1.1).

WET test design and the types of WET endpoints measured influence test sensitivity (e.g., control coefficient of variation or CV). Therefore, TST α error rates are identified for different types of test designs. For example, all fish chronic WET test methods that use a similar test design and have the same type of test endpoints (e.g., growth and survival) would have the same α value. Varying α by WET test design is appropriate for the TST approach. Given the way that the hypotheses are formulated in the TST approach (see Table 2), α represents what is considered β in the traditional hypothesis testing approach, and an acceptable β error was not identified in the current EPA TSD's approach to the EPA NPDES WET Program. Setting α as well as β in the TST approach addresses both false positives and false negatives.

2.2 Setting the Test Method-Specific Alpha Level

Several types of analyses were conducted to determine the appropriate α level for each WET test method. First, representative effluent and reference toxicant data meeting EPA WET test method's test acceptability criteria (TAC) were obtained from several state databases, which included multiple laboratories and wastewater effluents. Valid effluent WET data that met the following data selection requirements were considered to be a representative sample.

- Cover a range of NPDES permitted facility types, including both industrial and municipal permittees
- Represent many facilities for a given EPA WET test method (i.e., no one facility dominates the data for a given WET test method)
- Cover a range of target (design) effluent dilutions on which WET reasonable potential (RP) and NPDES permit compliance are based, ranging from 10 percent to 100 percent effluent concentrations
- Generated by several laboratories for a given EPA WET test method

- Cover a range of observed effluent toxicity for each EPA WET test method (e.g., NOECs range from < 10 percent to 100 percent effluent)

For each of the nine EPA WET test methods examined, control precision was calculated on the basis of valid WET data compiled in this project. A similar analysis was performed for the control response for each of the nine test methods (e.g., mean number of offspring per female in the chronic *Ceriodaphnia dubia* test method) to characterize typical achievable test performance in terms of control response.

A Monte Carlo simulation analysis (a statistical method) was used to estimate the percentage of WET tests that would be declared toxic using the TST approach as a function of different α levels, within-test variability (control and effluent variability), and different effect levels. That analysis identified probable false positive error rates (i.e., declaring an effluent toxic when in fact it is not) under all WET test scenarios encountered. Using the RMDs defined above, an appropriate α level was then identified for each WET test design given a desired β error of ≤ 5 percent (0.05) when there is a 10 percent mean effect at the IWC. By simulating thousands of WET tests for a given scenario (mean percent effect and control CV), the percentage of tests declared toxic under a given effluent assessment scenario could be calculated and compared with other scenarios.

3.0 USING THE TST APPROACH IN WET DATA ANALYSES

3.1 Summary of Test Method-Specific Alpha Values

On the basis of all the analyses conducted in this project, EPA recommends the following alpha levels when using the TST approach in a two concentration (i.e., two treatments) data analysis comparison (e.g., IWC and control) (see Table 3).

Table 3. Summary of alpha (α) levels or false negative rates recommended for different WET test methods using the TST approach

EPA WET test method	*b* value	Probability of declaring a toxic effluent non-toxic: False negative (α) error[a]
Chronic Freshwater and East Coast Methods		
Ceriodaphnia dubia (water flea) survival and reproduction	0.75	0.20
Pimephales promelas (fathead minnow) survival and growth	0.75	0.25
Selenastrum capricornutum (green algae) growth	0.75	0.25
Americamysis bahia (mysid shrimp) survival and growth	0.75	0.15
Arbacia punctulata (Echinoderm) fertilization	0.75	0.05
Cyprinodon variegatus (Sheepshead minnow) and *Menidia beryllina* (inland silverside) survival and growth	0.75	0.25
Chronic West Coast Marine Methods		
Dendraster excentricus and *Strongylocentrotus purpuratus* (Echinoderm) fertilization	0.75	0.05
Atherinops affinis (topsmelt) survival and growth	0.75	0.25
Haliotis rufescens (red abalone), *Crassostrea gigas (oyster), Dendraster excentricus, Strongylocentrotus purpuratus* (Echinoderm) and *Mytilus sp* (mussel) larval development methods	0.75	0.05
Macrocystis pyrifera (giant kelp) germination and germ-tube length	0.75	0.05
Acute Methods		
Pimephales promelas (fathead minnow), *Cyprinodon variegatus* (Sheepshead minnow), *Atherinops affinis* (topsmelt), *Menidia beryllina* (inland silverside) acute survival[b]	0.80	0.10
Ceriodaphnia dubia, *Daphnia magna, Daphnia pulex*, *Americamysis bahia* acute survival[b]	0.80	0.10

Notes:

a. (1) declare a sample toxic at least 75 percent of the time (alpha $\leq$ 0.25) when there is unacceptable toxicity (20 percent effect for acute and 25 percent effect for chronic test methods) and (2) declare an effluent non-toxic no more than 5 percent of the time (beta $\leq$ 0.05) when the mean effect at the critical effluent concentration is 10 percent for both acute and chronic WET tests (including sublethal endpoints). For more discussion on the RMDs, see Section 2.1 of this document.

b. Based on four replicate test design

3.2 Calculating Statistics for Valid WET Data Using the TST Approach

Appendix A includes a step-by-step guide for using the TST approach to analyzing WET test data. The appendix also includes a statistical flowchart and several examples. Note that the WET test method should follow the test condition requirements as specified in EPA's approved WET methods (USEPA 1995, 2002a, 2002b, 2002c).

The TST approach is used to statistically compare organism responses from two concentrations (i.e., treatments) of the WET test, the IWC and the control. Percent data (quantal data), such as percent survival or percent germination from a WET test, is first transformed as required in the EPA WET test manuals. Other types of WET data (e.g., growth or reproduction data) are not transformed. Data are then analyzed using Welch's t-test, a well-known modification of the standard t-test (Zar 1996), which is appropriate for the TST approach (see Appendix A).

Appendix B lists the critical *t* values that apply to WET testing using the TST approach given the number of degrees of freedom and the α level that applies for a given WET test method from Table 3 of this document. If the calculated *t* value for the WET test is greater than the critical *t* value (see Table B-1), the null hypothesis is rejected, i.e., the test result is *Pass* and **the effluent is declared *non-toxic***. If the calculated *t* value is less than the critical *t* value in Appendix B, the null hypothesis is not rejected, i.e., the test result is *Fail* and **the effluent *is* declared toxic.** Appendix A contains examples that demonstrate the formulae used in the TST approach and are designed to illustrate how the outcome is influenced by within-test variability and the mean effect of the IWC using the TST approach. Four different case examples are presented, three of which have equal variances between control and IWC: (1) *Ceriodaphnia* reproduction data having relatively high within-test variability, (2) *Ceriodaphnia* reproduction data having relatively low within-test variability and the same effect as in Example 1, (3) growth data from two fathead minnow chronic WET tests, both with relatively high within-test variability but small mean effect at the IWC; one test was conducted with the minimum number of replicates required in the EPA WET test method (four replicates) and the other test was conducted a priori with six replicates per concentration; and (4) calculations using the TST approach for an acute fathead minnow WET test.

Case Example #1 in Appendix A: ***Demonstrates a benefit of the TST approach by addressing false negatives***. A WET test that has relatively high within-test variability for a given WET test method and has an effect at the IWC approaching the RMD threshold (25 percent in this case because it is a chronic WET test) *is declared toxic* using the TST approach. Using the traditional hypothesis testing approach as recommended in the TSD, such test data typically lead to a conclusion that the effluent is not toxic (i.e., a false negative).

Case Example #3 in Appendix A: ***Demonstrates the benefits of increased within-test replication using the TST approach***. Increasing the replication before conducting the test, which thereby improves the precision and power of the WET test, increases the chances of rejecting the null hypothesis and declaring a truly acceptable effluent as *non-toxic* using the TST approach. That increases the ability to *prove the negative*, i.e., that an effluent is declared not toxic.

The TST approach can also be used for ambient toxicity (i.e., receiving water) tests and stormwater toxicity testing programs because the TST approach compares two treatments (for application of the TST approach to ambient toxicity testing, see Appendix C).

4.0 IMPLEMENTING THE TST APPROACH IN WET NPDES PERMITS

The TST approach is an alternative approach for analyzing and interpreting valid WET data. Use of the TST approach does not result in any changes to EPA's WET test methods. WET limits are simpler to communicate and understand (for example permit language for acute and chronic WET monitoring using the TST statistical analysis approach, see Appendix D) than the TSD approach. EPA recommends that permitting authorities decide up front which approach (the 1991 TSD approach, the TST approach, or another scientifically defensible approach that is sufficient to meet the statutory and regulatory requirements) they will incorporate and consistently use in their state's NPDES implementation procedures, including their RP procedures. The permitting authority should use the selected WET statistical approach consistently in all of their state NPDES permits.

4.1 Reasonable Potential (RP) WET Analysis

NPDES permitting authorities conducting an RP analysis must follow Title 40 of the *Code of Federal Regulations* (CFR) section 122.44(d)(1) to determine whether a discharge will "cause, have the [RP] to cause, or contribute to" an excursion of a numeric criterion or a narrative WET criterion. Some states have state-specific WET RP approaches in their water quality control plan or other NPDES policy or guidance.

For RP calculations using the TST approach, EPA recommends that permitting authorities use all valid WET test data generated during the current permit term and any additional valid data that are submitted as part of the permit renewal application. The TST RP approach necessitates having at least a minimum of four valid WET tests to address effluent representativeness (see EPA's TSD, Chapter 3, p. 57, under Step 2 in the section *Steps in Whole Effluent Characterization Process*). EPA also recommends that states request that their permittees provide the actual test endpoint responses for the control (i.e., control mean) and IWC concentration (i.e., IWC mean) for each WET test conducted to make it easier for permit writers to find the necessary WET test results when determining WET RP. WET test data are then analyzed according to the TST approach using the IWC and control test concentrations for all the valid WET test data available. For data sets with fewer than four valid WET data points, RP should be assessed using EPA's TSD RP approach because it addresses small WET data sets by incorporating an RP multiplying factor (see Section 3.3.2 of the TSD, p. 54) to account for effluent variability in small WET data sets. If WET test data are available and the TST statistical approach indicates that the IWC is toxic in any WET test, RP has been demonstrated (40 CFR 122.44(d)(1)(i)). Similar to the TSD approach, the TST approach can establish the existence of RP for WET even when no tests have been declared toxic using the TST to address concerns regarding the "potential to cause or contribute to toxicity." Appendix E presents the approach used to determine RP using the TST approach.

Note that using the TST approach might be to the permittee's advantage. If the permittee decides to incorporate additional replicates for the control and the IWC within a WET test, beyond the minimum required in the WET test method, the test power is increased. More test replicates increases test power, which means a higher probability of declaring a sample as non-toxic using the TST approach *if the effluent is truly non-toxic*. A demonstration is provided in Appendix A (Case Example #3), which illustrates that as an intended consequence of the TST approach

methodology. Thus, using the TST approach, a permittee has a greater ability to *prove the negative* (i.e., their effluent does not have RP).

In those cases where the WET RP outcome is *yes*, a WET limit is expressed in the permit. In those situations where the RP outcome is *no*, WET monitoring requirements should still be incorporated in the permit. Also in the permit, a test result of *Fail* (i.e., sample declared toxic) during monitoring, would trigger additional steps in the permit. In either of those situations—either a WET limit or a WET monitoring requirement, if toxicity is demonstrated—states should specify an approach to address toxicity in the permit. Doing so often includes increased frequency of WET testing and additional permit requirements to perform a toxicity reduction evaluation.

4.2 NPDES WET Permit Limits

Using the TST approach, WET NPDES permit limits would be expressed as *no significant toxicity of the effluent at the IWC using the TST analysis approach*. A test result of *Pass* is when the calculated *t* value is greater than *the critical t value*. A test result of *Fail* is when the calculated *t* value is less than *the critical t value*.

Beyond assessing WET data for the NPDES Program, WET tests are used to assess toxicity of receiving water (watershed assessment for CWA section 303(d) determinations) and stormwater samples. Often as a first assessment of receiving or stormwater toxicity, researchers test a control and a single concentration (e.g., 100 percent receiving water or stormwater). In such cases, the TST approach can be used in the same way a t-test is used. Such analysis is used to determine whether organism response in a specified ambient concentration is significantly different than the control organism response (for further information, see Appendix C).

5.0 RECOMMENDATIONS FOR NPDES IMPLEMENTATION OF THE TST APPROACH

5.1 EPA Regions and NPDES States (Permitting Authorities)

Permitting authorities should consider adding the TST approach to their implementation procedures for analyzing valid WET data for their current NPDES WET Program. Permitting authorities should consider the practical programmatic shift from the traditional hypothesis testing approach to the TST approach by opening a dialogue with their regulated community. In addition, they might want to begin to identify what changes might be needed to assimilate the TST approach into any regulations, policy, guidance, and training within their respective NPDES WET Programs. EPA also recommends that permitting authorities decide up front which RP approach (the 1991 TSD approach, the TST approach, or another scientifically defensible approach that is sufficient to meet the statutory and regulatory requirements) the permitting authority will incorporate and consistently use in their state's NPDES implementation procedures. The permitting authority should then use the WET statistical approach (either the TSD approaches or the TST data analysis approach) selected throughout all its state NPDES permits. Again, the traditional hypothesis testing approach recommended in EPA's TSD is still considered valid as applied; however, that approach can now be advanced through the TST approach by providing new incentives to permittees to generate valid, high quality WET data.

The RMDs incorporated into the TST approach were selected on the basis of considerable research and analysis involving several of the EPA WET test methods. Lower *b* values (i.e., for chronic test methods using a 0.70 instead of 0.75 *b* is unacceptable) are not recommended because it would mean that a lower fraction of test control response (i.e., greater effect at the IWC) is considered acceptable. EPA chose the acute and chronic *b* values to minimize effects on aquatic ecosystems. Likewise, the alpha values identified by EPA using the TST approach were determined on the basis of the predetermined *b* values and therefore should *not* be altered.

The permitting authority should consider carefully how the TST approach will be implemented in NPDES permits. Example permit language is shown in Appendix D. In consideration of maintaining NPDES WET Program implementation consistency, the TST approach should be used in place of, *and not in addition to,* the traditional hypothesis testing (NOEC) approach for WET analysis.

5.2 NPDES Permittees

One of the intended benefits of the TST approach is that increasing the precision and power of the WET test increases the chances of declaring a truly *acceptable effluent as non-toxic*. The permittee has greater control over the interpretation of WET test results using the TST approach because the RMDs are transparent, and the level of WET data quality needed to obtain unequivocal results can be determined beforehand. For example, conducting tests with more test replicates improves the power of the WET test, which can then support and provide a defensible basis for a permittee's demonstration that its effluent is *acceptable (i.e., in compliance with the permit*) if the mean effect is truly within the RMDs as defined in the TST approach. Using the TST approach, there is a *lower* rate of WET tests declared toxic for tests that are truly acceptable because of the increased power of the WET test when the permittee increases its number of

replicates in a WET test or achieves better replication within a test through improved test method performance. Thus, the TST approach increases the ability of the permittee to *prove the negative*, that the effluent is *non-toxic* if it is truly *acceptable*. Where a permittee has concerns about WET data quality, EPA recommends increasing the number of replicates in tests, even if the permitting authority has not yet adopted the TST approach.

6.0 SUMMARY OF THE TST APPROACH

EPA's TSD approaches are valid and can still be used by EPA Regions and their NPDES states. The TST approach is another statistical option for analyzing valid WET test data. The TST approach can be applied to acute (survival) and chronic (sublethal) endpoints and is appropriate to use for both freshwater and marine EPA WET test methods. The TST approach requires no more time or expertise than is presently expended when using the TSD hypothesis testing statistical approach and can be used with a well-recognized statistical test. Below is a brief outline of both the TST and TSD hypothesis testing approaches relevant to the information in this document and a short list of the benefits derived when using the TST approach.

TST Approach

- Considered additional guidance only—TST is a statistical approach for analyzing WET test data as an alternative option to the traditional hypothesis testing approach provided in EPA's TSD
- Expresses NPDES WET permit limit "as no significant toxicity of the effluent at the in-stream waste concentration" using the TST analysis approach
- Provides a positive incentive to NPDES permittees to generate valid, high quality WET data to the permitting authority by improving test performance or increasing the number of replicates within a WET test (which increases statistical power of WET test)
- Addresses both false negative (declared non-toxic when actually toxic) and false positive (declared toxic when actually non-toxic) error rates in a WET test

Traditional Hypothesis Test (EPA TSD)

- Existing approaches remain valid and can still be used by NPDES permitting authorities
- In existing guidance, WET permit limits are expressed as *no observed effect concentration (NOEC)* at the IWC
- Provides relatively less incentive to permittees to generate high quality valid, WET data or to increase the number of replicates within a WET test to increase statistical power of a WET test
- False negative error rate in a WET test is not addressed

Benefits When Using the TST Approach in WET Data Analysis

- ***The TST approach*** is similar to statistical concepts used in other EPA programs and at other federal agencies
- ***Transparent RMDs.*** RMDs are transparent because they are incorporated into the WET data analysis process, e.g., what effect level is considered toxic and what effect level is considered acceptable.
- ***WET test method-specific alpha and beta error rates.*** Both error rates are directly incorporated into the TST statistical approach, thereby increasing confidence in WET test interpretation.
- ***High quality WET test data incentive.*** Provides a positive incentive for the permittee to generate valid, high quality WET data; better test performance (lower within-test

variability) helps ensure appropriate WET decisions using the TST approach (e.g., a truly acceptable effluent will be declared non-toxic).

- ***Streamlined, simpler statistical analysis.*** Flowchart for analyzing valid WET data under the TST approach is much simpler because fewer statistical tests are needed.
- ***RP analysis is simpler.*** Because the calculation of the individual test result, using the TST statistical approach, incorporates both error rates in the analysis, the RP determinations can rely on a direct calculation of the percent effect at the IWC. Thus, the RP procedures are much simpler to use than the RP statistical procedures recommended in the TSD.

7.0 LITERATURE CITED

Anderson, S., and W. Hauck. 1983. A new procedure for testing equivalence in comparative bioavailability and other clinical trials. *Communications in Statistics–Theory and Methods* 12:2663–2692.

Aras, G. 2001. Superiority, non-inferiority, equivalence, and bioequivalence–revisited. *Drug Information Journal* 35:1157–1164.

Berger, R., and J. Hsu. 1996. Bioequivalence trials, intersection–union tests and equivalence confidence sets. *Statistical Science* 11:283–319.

Erickson, W., and L. McDonald. 1995. Tests for non-inferiority of control media and test media in studies of toxicity. *Environmental Toxicology and Chemistry* 14:1247–1256.

Hatch, J. 1996. Using statistical equivalence testing in clinical biofeedback research. *Biofeedback and Self-Regulation* 21:105–119.

Shukla, R., Q. Wang, F. Fulk, C. Deng, and D. Denton. 2000. Bioequivalence approach for whole effluent toxicity testing. *Environmental Toxicology and Chemistry* 19:169–174.

Streiner, D. 2003. Unicorns *do* exist: A tutorial on "proving" the null hypothesis. *Canadian Journal of Psychiatry* 48(11):756–761.

Stunkard, C. 1990. *Tests of proportional means for mesocosms studies*. Technical Report. University of Maryland, Department of Measurement, Statistics, and Evaluation, College Park, MD.

USEPA (U.S. Environmental Protection Agency). 1989. *Methods for Evaluating the Attainment of Cleanup Standards, Vol. I – Soils and Solid Media*. EPA 230/02-89-042. U.S. Environmental Protection Agency, Office of Policy, Planning and Evaluation, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 1991. *Technical Support Document for Water Quality-based Toxics Control*. EPA/505/2-90-001. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (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. EPA/600/R-95-136. U.S. Environmental Protection Agency, National Exposure Research Laboratory, Cincinnati, OH, and Office of Research and Development, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2000. *Understanding and Accounting for Method Variability in Whole Effluent Toxicity Applications Under the NPDES Program*. EPA/833-R-00-003. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2002a. *Short-Term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Freshwater Organisms*. 4th edition. EPA/821/R-02-013. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2002b. *Short-Term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Marine and Estuarine Organisms*. 3rd ed. EPA/821/R-02-14. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2002c. *Methods for Measuring the Acute Toxicity of Effluents and Receiving Waters to Freshwater and Marine Organisms*. 5th ed. EPA/821/R-02-012. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

Zar, J. 1996. *Biostatistical Analysis* 3rd ed. Prentice Hall Publishers, New Jersey.

APPENDIX A

STEP-BY-STEP PROCEDURES FOR ANALYZING VALID WHOLE EFFLUENT TOXICITY DATA USING THE TEST OF SIGNIFICANT TOXICITY APPROACH

APPENDIX A: STEP-BY-STEP PROCEDURES FOR ANALYZING VALID WET DATA USING THE TST APPROACH

The following is a step-by-step guide for using the TST approach to analyze valid WET data for the NPDES Program. This guide is applicable for a two-concentration valid WET data analysis of an in-stream waste concentration (IWC) or a receiving water concentration (RWC) compared to a control concentration. For further information regarding conducting WET tests and proper quality assurance/quality control needed, see the EPA WET test method manuals. Refer to the flowchart shown in Figure A-1 in this appendix as you proceed through this guide.

Step 1: Conduct WET test following procedures in the appropriate EPA WET test method manual. That includes following all test requirements specified in the method (USEPA 1995 for chronic west coast marine methods, USEPA 2002a for chronic freshwater test methods, USEPA 2002b for chronic east coast marine test methods, and USEPA 2002c for acute freshwater and marine WET test methods).

Step 2: For each test endpoint specified in the WET test method manual (e.g., survival and reproduction for the *Ceriodaphnia* chronic WET test method), follow Steps 3–7 below. Note that the guide refers to an effluent concentration tested, which is assumed to be the IWC as specified in the permit or a receiving water concentration for ambient testing. For example, if no mixing zone is allocated, the IWC is 100 percent effluent.

Note: If there is no variance (i.e., zero variance) in the endpoint in both concentrations being compared (i.e., all replicates in each concentration have the same exact response), then skip the remaining steps in the flowchart and do the following. Compute the percent difference between the control and the other concentration (e.g., IWC) and compare the percent difference against the RMD values of 25% for chronic and 20% for acute endpoints. Percent mean effect is calculated as:

$$\%\ \text{Effect at IWC} = \frac{\text{Mean Control Response} - \text{Mean Response at IWC}}{\text{Mean Control Response}} \times 100$$

If the percent mean response is $\geq$ the RMD, the sample is declared toxic and the test is "*Fail*". If the percent mean response is < the RMD, the sample is declared non-toxic and the test is "*Pass*".

Step 3: For data consisting of proportions from a binomial (response/no response; live/dead) response variable, the variance within the i^{th} treatment is proportional to $P_i (1 - P_i)$, where P_i is the expected proportion for the treatment. That clearly violates the homogeneity of variance assumption required by parametric procedures such as the TST procedure because the existence of a treatment effect implies different values of P_i for different treatments, *i*. Also, when the observed proportions are based on small samples, or when P_i is close to zero or one, the normality assumption might be invalid. The arcsine square root (arcsine $\sqrt{P}$) transformation is used for such data to stabilize the variance and satisfy the normality requirement. The square root of percent data (e.g., percent survival, percent fertilization), expressed as a decimal fraction (where 1.00 = 100 percent) for each treatment, is first calculated. The square root value is then

arcsine transformed before analysis in Step 4. Note: Excel and most statistical software packages can calculate arcsine values.

Step 4: Conduct Welch's t-test (Zar 1996) using Equation 1:

Equation 1

$$t = \frac{\overline{Y_t} - b \times \overline{Y_c}}{\sqrt{\frac{S_t^{\,2}}{n_t} + \frac{b^2 S_c^{\,2}}{n_c}}}$$

where

$\overline{Y_c}$ = Mean for the control

$\overline{Y_t}$ = Mean for the IWC

$s_c^{\,2}$ = Estimate of the variance for the control

$s_t^{\,2}$ = Estimate of the variance for the IWC

n_c = Number of replicates for the control

n_t = Number of replicates for the IWC

b = 0.75 for chronic test methods; 0.80 for acute test methods

Note on the use of Welch's t-test: Welch's t-test is appropriate to use when there are an unequal number of replicates between control and the IWC. When sample sizes of the control and treatment are the same (i.e., $n_t = n_{c)}$, Welch's t-test is equivalent to the usual Student's t-test (Zar 1996).

Step 5: Adjust the degrees of freedom (df) using Equation 2:

Equation 2

$$\upsilon = \frac{\left(\frac{S_t^{\,2}}{n_t} + \frac{b^2 S_c^{\,2}}{n_c}\right)^2}{\frac{\left(\frac{S_t^{\,2}}{n_t}\right)^2}{n_t - 1} + \frac{\left(\frac{b^2 S_c^{\,2}}{n_c}\right)^2}{n_c - 1}}$$

For tests using Welch's t-test, df is the value obtained for *v* in Equation 2 above. Because *v is* most likely a non-integer, round *v* to the next smallest integer, and that number is the df.

Step 6: Using the calculated *t* value from Step 4, compare that *t* value with the critical *t* value table in Appendix B using the test method-specific alpha values shown in Table A-1. To obtain the correct critical *t* value, look across the table for the alpha value that corresponds to the WET test method (for the alpha value, see Appendix A, Table A-1) and then look down the table for the appropriate df.

Step 7: If the calculated *t* value is less than the critical *t* value, the IWC is declared toxic and the test result is *Fail*. If the calculated *t* value is greater than the critical *t* value, the IWC is not declared toxic and the test result is *Pass*.

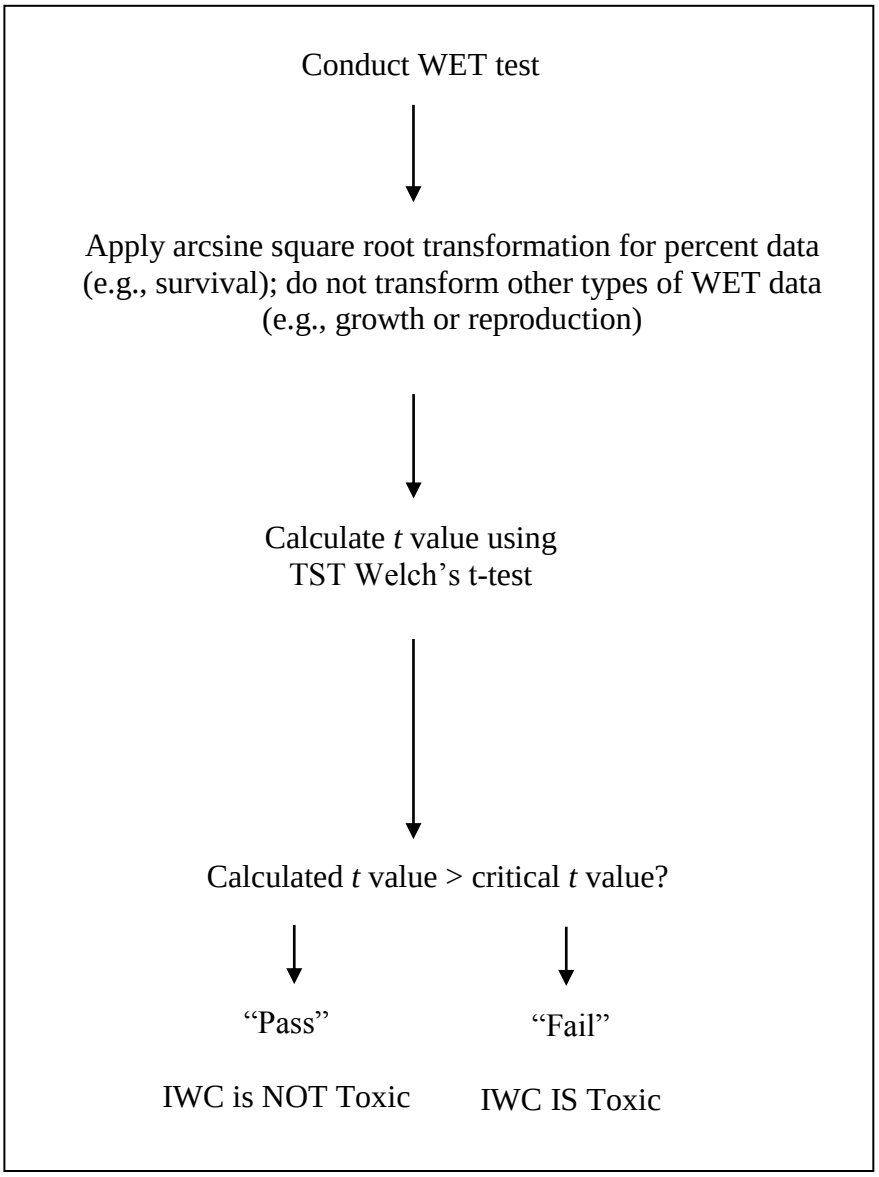

Figure A-1. Statistical flowchart for analyzing valid WET data using the TST approach for control and the IWC, receiving water, or stormwater.

Table A-1. Summary of alpha (α) levels or false negative rates recommended for different WET test methods using the TST approach

EPA WET test method	*b* value	Probability of declaring a toxic effluent non-toxic: False negative (α) error[a]
Chronic Freshwater and East Coast Methods		
Ceriodaphnia dubia (water flea) survival and reproduction	0.75	0.20
Pimephales promelas (fathead minnow) survival and growth	0.75	0.25
Selenastrum capricornutum (green algae) growth	0.75	0.25
Americamysis bahia (mysid shrimp) survival and growth	0.75	0.15
Arbacia punctulata (Echinoderm) fertilization	0.75	0.05
Cyprinodon variegatus (Sheepshead minnow) and *Menidia beryllina* (inland silverside) survival and growth	0.75	0.25
Chronic West Coast Marine Methods		
Dendraster excentricus and *Strongylocentrotus purpuratus* (Echinoderm) fertilization	0.75	0.05
Atherinops affinis (topsmelt) survival and growth	0.75	0.25
Haliotis rufescens (red abalone), *Crassostrea gigas (oyster), Dendraster excentricus, Strongylocentrotus purpuratus* (Echinoderm) and *Mytilus sp* (mussel) larval development methods	0.75	0.05
Macrocystis pyrifera (giant kelp) germination and germ-tube length	0.75	0.05
Acute Methods		
Pimephales promelas (fathead minnow), *Cyprinodon variegatus* (Sheepshead minnow), *Atherinops affinis* (topsmelt), *Menidia beryllina* (inland silverside) acute survival[b]	0.80	0.10
Ceriodaphnia dubia, *Daphnia magna, Daphnia pulex*, *Americamysis bahia* acute survival[b]	0.80	0.10

Notes:

a. (1) declare a sample toxic at least 75 percent of the time (alpha $\leq$ 0.25) when there is unacceptable toxicity (20 percent effect for acute and 25 percent effect for chronic test methods) and (2) declare an effluent non-toxic no more than 5 percent of the time (beta $\leq$ 0.05) when the mean effect at the critical effluent concentration is 10 percent for both acute and chronic WET tests (including sublethal endpoints). For more discussion on the RMDs, see Section 2.1 of this document.

b. Based on four replicate test design

Case Example 1: Chronic *Ceriodaphnia* Reproduction Test with High Within-Test Variability

Step 1: Conduct WET test

Replicate/statistic	Control	Treatment
1	27	32
2	38	28
3	27	25
4	34	28
5	37	20
6	35	15
7	30	27
8	31	31
9	36	31
10	39	30
Mean	33.4	26.7
Std. deviation	4.402	5.417
N (# of replicates)	**10**	**10**

Step 2: Follow Steps 3–7 for each endpoint required in the test method

The following example is for chronic *Ceriodaphnia dubia* reproduction endpoint only.

Step 3: Transform data using an arcsine square root transformation, if necessary

Not necessary because reproduction is not percent data.

Step 4: Conduct Welch's t-test

$$t = \frac{\overline{Y}_t - b * \overline{Y}_c}{\sqrt{\frac{S_t^{\,2}}{n_t} + \frac{b^2 S_c^{\,2}}{n_c}}} = \frac{26.7 - (0.75 \times 33.4)}{\sqrt{\frac{29.34}{10} + \frac{(0.75)^2\,(19.38)}{10}}} = 0.82$$

Step 5: Adjust the df

$$v = \frac{\left(\frac{S_t^{\,2}}{n_t} + \frac{b^2 S_c^{\,2}}{n_c}\right)^2}{\frac{\left(\frac{S_t^{\,2}}{n_t}\right)^2}{n_t - 1} + \frac{\left(\frac{b^2 S_c^{\,2}}{n_c}\right)^2}{n_c - 1}} = \frac{\left(\frac{29.34}{10} + \frac{(0.75)^2 (19.38)}{10}\right)^2}{\frac{\left(\frac{29.34}{10}\right)^2}{10 - 1} + \frac{\left(\frac{(0.75)^2\,(19.38)}{10}\right)^2}{10 - 1}} = 15$$

Step 6: Calculated *t* value > critical *t* value? 15 df and test method alpha = 0.20 (Table A-1)

Critical *t* value = 0.87

0.82 < 0.87

Step 7: Declare effluent toxic or not

Calculated *t* < critical *t* value. Therefore, ***effluent is declared toxic; test result is FAIL***.

Case Example 2: Chronic *Ceriodaphnia* Reproduction Test with Low Within-Test Variability

Step 1: Conduct WET test

Replicate/statistic	Control	Treatment
1	29	31
2	38	28
3	31	25
4	34	28
5	36	22
6	35	21
7	30	27
8	31	26
9	36	29
10	34	30
Mean	33.4	26.7
Std. deviation	2.989	3.268
N (# of replicates)	10	10

Step 2: Follow Steps 3–7 for each endpoint required in the test method

The following example is for chronic *Ceriodaphnia dubia* reproduction endpoint only.

Step 3: Transform data using an arcsine square root transformation, if necessary

Not necessary because reproduction is not percent data.

Step 4: Conduct Welch's t-test

$$t = \frac{\overline{Y}_t - b \times \overline{Y}_c}{\sqrt{\frac{S_t^{\,2}}{n_t} + \frac{b^2 S_c^{\,2}}{n_c}}} = \frac{26.7 - (0.75 \times 33.4)}{\sqrt{\frac{10.68}{10} + \frac{(0.75)^2\,(8.93)}{10}}} = 1.32$$

Step 5: Adjust the df

$$v = \frac{\left(\frac{S_t^{\,2}}{n_t} + \frac{b^2 S_c^{\,2}}{n_c}\right)^2}{\frac{\left(\frac{S_t^{\,2}}{n_t}\right)^2}{n_t - 1} + \frac{\left(\frac{b^2 S_c^{\,2}}{n_c}\right)^2}{n_c - 1}} = \frac{\left(\frac{10.68}{10} + \frac{(0.75)^2 (8.93)}{10}\right)^2}{\frac{\left(\frac{10.68}{10}\right)^2}{10 - 1} + \frac{\left(\frac{(0.75)^2\,(8.93)}{10}\right)^2}{10 - 1}} = 16$$

Step 6: Calculated *t* value > critical *t* value? 16 df and test method alpha = 0.20 (Table A-1)

Critical *t* value = 0.86

1.32 > 0.86

Step 7: Declare effluent toxic or not

Calculated *t* > critical *t* value. Therefore, ***effluent is declared Non-Toxic; test result is PASS***.

Case Example 3: Benefit of Increased Replication in Chronic Fish Growth Test with Low Mean Effect and High Within-Test Variability

Step 1: Conduct WET test

Replicate/statistic	Control	Treatment
1	0.366	0.303
2	0.399	0.379
3	0.354	0.311
4	0.422	0.236
Mean	0.385	0.307
Std. deviation	0.031	0.058
N (# of replicates)	**4**	**4**

Step 2: Follow Steps 3–7 for each endpoint required in the test method

Step 3: Transform data using an arcsine square root transformation, if necessary

Not necessary because growth is not percent data.

Step 4: Conduct Welch's t-test

$$t = \frac{\overline{Y}_t - b \times \overline{Y}_c}{\sqrt{\frac{S_t^2}{n_t} + \frac{b^2 S_c^2}{n_c}}}$$

$$= \frac{0.307 - (0.75 \times 0.385)}{\sqrt{\left(\frac{0.00342}{4} + \frac{(0.75)^2\,(0.00096)}{4}\right)}} = 0.58$$

Step 5: Adjust the df

$$v = \frac{\left(\frac{S_t^2}{n_t} + \frac{b^2 S_c^2}{n_c}\right)^2}{\frac{\left(\frac{S_t^2}{n_t}\right)^2}{n_t - 1} + \frac{\left(\frac{b^2 S_c^2}{n_c}\right)^2}{n_c - 1}}$$

$$= \frac{\left(\frac{0.00342}{4} + \frac{(0.75)^2\,(0.00096)}{4}\right)^2}{\frac{\left(\frac{0.00342}{4}\right)^2}{4-1} + \frac{\left(\frac{(0.75)^2\,(0.00096)}{4}\right)^2}{4-1}} = 4$$

Step 6: Calculated *t* value > critical *t* value? 4 df, alpha = 0.25 (Table A-1); Critical *t* value = 0.74

0.58 < 0.74

Step 7: *Effluent is declared toxic, test result is FAIL.*

Step 1: Conduct WET test

Replicate/statistic	Control	Treatment
1	0.366	0.303
2	0.399	0.379
3	0.354	0.311
4	0.422	0.236
5	0.343	0.364
6	0.407	0.247
Mean	0.382	0.307
Std. deviation	0.032	0.058
N (# of replicates)	**6**	**6**

Step 2: Follow Steps 3–7 for each endpoint required in the test method

Step 3: Transform data using an arcsine square root transformation, if necessary

Not necessary because growth is not percent data.

Step 4: Conduct Welch's t-test

$$t = \frac{\overline{Y}_t - b \times \overline{Y}_c}{\sqrt{\frac{S_t^2}{n_t} + \frac{b^2 S_c^2}{n_c}}}$$

$$= \frac{0.307 - (0.75 \times 0.382)}{\sqrt{\frac{0.00342}{6} + \frac{(0.75)^2 (0.00101)}{6}}} = 0.79$$

Step 5: Adjust the df

$$v = \frac{\left(\frac{S_t^2}{n_t} + \frac{b^2 S_c^2}{n_c}\right)^2}{\frac{\left(\frac{S_t^2}{n_t}\right)^2}{n_t - 1} + \frac{\left(\frac{b^2 S_c^2}{n_c}\right)^2}{n_c - 1}}$$

$$= \frac{\left(\frac{0.00342}{6} + \frac{(0.75)^2\,(0.00101)}{6}\right)^2}{\frac{\left(\frac{0.00342}{6}\right)^2}{6-1} + \frac{\left(\frac{(0.75)^2\,(0.00101)}{6}\right)^2}{6-1}} = 7$$

Step 6: Calculated *t* value > critical *t* value? 7 df, alpha = 0.25 (Table A-1); Critical *t* value = 0.71

0.79 > 0.71

Step 7: *Effluent is declared Non-Toxic; test result is PASS.*

Case Example 4: Fish Acute Toxicity Test Example

Step 1: Conduct WET test

Replicate/statistic	Control	Treatment
1	10	10
2	10	8
3	10	9
4	10	8
Mean	10	8.75
Variance	0.000	0.917
N (# of replicates)	**4**	**4**

Step 2: Follow Steps 3–7 for each endpoint required in the test method

The following example is for acute *Pimephales promelas* survival endpoint only.

Step 3: Transform data using an arcsine square root transformation

Replicate/statistic	Control	Treatment
1	1.412	1.412
2	1.412	1.107
3	1.412	1.249
4	1.571	1.107
Mean	1.412	1.218
Variance	0.000	0.021
N (# of replicates)	**4**	**4**

Step 4: Conduct Welch's t-test

$$t = \frac{\overline{Y}_t - b \times \overline{Y}_c}{\sqrt{\frac{S_t^{\,2}}{n_t} + \frac{b^2 S_c^{\,2}}{n_c}}} = \frac{1.218 - (0.80 \times 1.412)}{\sqrt{\frac{0.021}{4} + \frac{(0.80)^2\ (0.000)}{4}}} = 1.2297$$

Step 5: Adjust the df

$$v = \frac{\left(\frac{S_t^{\,2}}{n_t} + \frac{b^2 S_c^{\,2}}{n_c}\right)^2}{\frac{\left(\frac{S_t^{\,2}}{n_t}\right)^2}{n_t - 1} + \frac{\left(\frac{b^2 S_c^{\,2}}{n_c}\right)^2}{n_c - 1}} = \frac{\left(\frac{0.021}{4} + \frac{(0.80)^2\ (0.000)}{4}\right)^2}{\frac{\left(\frac{0.021}{4}\right)^2}{4-1} + \frac{\left(\frac{(0.80)^2\ (0.000)}{4}\right)^2}{4-1}} = 3$$

Step 6: Calculated *t* value > critical *t* value? 3 df, alpha = 0.10 (Table A-1)

Critical *t* value = 1.64

1.229 < 1.64

Step 7: Declare effluent toxic or not

Therefore, ***effluent is declared toxic; test result is FAIL***.

Literature Cited

USEPA (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. EPA/600/R-95-136. U.S. Environmental Protection Agency, National Exposure Research Laboratory, Cincinnati, OH, and Office of Research and Development, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2002a. *Short-Term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Freshwater Organisms*. 4th edition. EPA/821/R-02-013. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2002b. *Short-Term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Marine and Estuarine Organisms.* 3rd ed. EPA/821/R-02-14. U.S. Environmental Protection Agency, Office of Science and Technology, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2002c. *Methods for Measuring the Acute Toxicity of Effluents and Receiving Waters to Freshwater and Marine Organisms*. 5th ed. EPA/821/R-02-012. United States Environmental Protection Agency, Office of Water, Washington, DC.

Zar, J. 1996. *Biostatistical Analysis*. 3rd ed. Prentice Hall Publishers, New Jersey.

APPENDIX B

CRITICAL *t* VALUES FOR THE TEST OF SIGNIFICANT TOXICITY APPROACH

Table B-1. Critical values of the *t* distribution. One tail probability is assumed.

Degrees of freedom			Alpha		
	0.25	**0.20**	**0.15**	**0.10**	**0.05**
1	1	1.3764	1.9626	3.0777	6.3138
2	0.8165	1.0607	1.3862	1.8856	2.92
3	0.7649	0.9785	1.2498	1.6377	2.3534
4	0.7407	0.941	1.1896	1.5332	2.1318
5	0.7267	0.9195	1.1558	1.4759	2.015
6	0.7176	0.9057	1.1342	1.4398	1.9432
7	0.7111	0.896	1.1192	1.4149	1.8946
8	0.7064	0.8889	1.1081	1.3968	1.8595
9	0.7027	0.8834	1.0997	1.383	1.8331
10	0.6998	0.8791	1.0931	1.3722	1.8125
11	0.6974	0.8755	1.0877	1.3634	1.7959
12	0.6955	0.8726	1.0832	1.3562	1.7823
13	0.6938	0.8702	1.0795	1.3502	1.7709
14	0.6924	0.8681	1.0763	1.345	1.7613
15	0.6912	0.8662	1.0735	1.3406	1.7531
16	0.6901	0.8647	1.0711	1.3368	1.7459
17	0.6892	0.8633	1.069	1.3334	1.7396
18	0.6884	0.862	1.0672	1.3304	1.7341
19	0.6876	0.861	1.0655	1.3277	1.7291
20	0.687	0.86	1.064	1.3253	1.7247
21	0.6864	0.8591	1.0627	1.3232	1.7207
22	0.6858	0.8583	1.0614	1.3212	1.7171
23	0.6853	0.8575	1.0603	1.3195	1.7139
24	0.6849	0.8569	1.0593	1.3178	1.7109
25	0.6844	0.8562	1.0584	1.3163	1.7081
26	0.684	0.8557	1.0575	1.315	1.7056
27	0.6837	0.8551	1.0567	1.3137	1.7033
28	0.6834	0.8546	1.056	1.3125	1.7011
29	0.683	0.8542	1.0553	1.3114	1.6991
30	0.6828	0.8538	1.0547	1.3104	1.6973
inf	0.6745	0.8416	1.0364	1.2816	1.6449

APPENDIX C

APPLICATION OF THE TEST OF SIGNIFICANT TOXICITY APPROACH TO AMBIENT TOXICITY PROGRAMS

APPENDIX C: APPLICATION OF THE TST APPROACH TO AMBIENT TOXICITY PROGRAMS

In ambient and stormwater toxicity testing, a laboratory control and a single concentration (i.e., 100 percent ambient water or stormwater) are often tested. In these two-concentration WET tests, the objective is to determine if a given sample or site water is toxic, as indicated by a significantly different organism response compared to the control. In the WET testing design, the determination of Pass or Fail (i.e., non-toxic or toxic) is ascertained using a traditional t-test (USEPA 2002c). EPA test methods recommend (USEPA 1995, 2002a, 2002b, 2002c) that the statistical significance (i.e., Pass/Fail) of a two-sample test design for ambient and stormwater toxicity testing be determined only using either a modified t-test (if homogeneity of variance is not achieved) or a traditional t-test (if homogeneity of variance is achieved).

To demonstrate the value of the TST approach in ambient toxicity programs, ambient toxicity test data from California's Surface Water Ambient Monitoring Program (SWAMP) was used for 409 chronic tests for *Ceriodaphnia dubia* and 256 chronic tests for *Pimephales promelas* using EPA's 2002 WET test methods (USEPA 2002a). Valid WET data for each EPA WET test method were subjected to the same statistical analyses as described in Section 2 of this document.

Chronic *Ceriodaphnia dubia* Ambient Toxicity Tests

Table C-1 summarizes results of the 409 *Ceriodaphnia dubia* ambient toxicity tests analyzed and an $\alpha = 0.20$ for this test method. Although the majority of the tests examined resulted in the same decision using either the TST or the traditional t-test approach, approximately 6 percent of the tests (24 tests) would have been declared non-toxic using the traditional t-test approach with mean effect levels $\geq$ 25 percent. In addition, 2 percent of the tests (7 tests) would have been declared toxic using the traditional t-test approach at mean effect levels < 15 percent and as low as 7 percent.

Table C-1. Comparison of results of chronic *Ceriodaphnia* ambient toxicity tests using the TST approach and the traditional t-test analysis. $\alpha = 0.20$ and *b* value = 0.75 for the TST approach. $\alpha = 0.05$ for the traditional hypothesis testing approach

Both approaches declare toxic	Only TST declares toxic	Only traditional approach declares toxic	Both approaches declare non-toxic
19.8%	5.9%	1.7%	72.6%

Figure C-1 shows ranges of CV values observed in *Ceriodaphnia dubia* ambient toxicity tests for those samples declared toxic using either the TST approach or the traditional t-test, but not both approaches. As expected, within-test variability was relatively high (higher CVs) for those tests found non-toxic using a t-test but toxic using the TST approach. The results demonstrate the lack of control of false negative rates using the traditional hypothesis testing approach when control variability is relatively high. Under those conditions, the traditional t-test did not have the power to detect toxicity when it was present. Figure C-1 also demonstrates that the TST approach recognizes a negligible effect as non-toxic when within-test variability is relatively low and the

mean percent effect is well below the risk management level of 25 percent. Under such conditions, the traditional t-test declared some samples toxic using this WET test method, even when the mean effect was as little as 7 percent. The TST approach, however, declared all such samples non-toxic using the recommended $\alpha = 0.20$. Thus, the TST approach reduces the number of tests declared as toxic when effects are actually well below the risk management decision.

Chronic *Ceriodaphnia* ambient WET tests that are identified as non-toxic (Pass) using the traditional hypothesis approach (NOEC) generally have high within-test variability (high control CVs) as compared to using the TST approach.

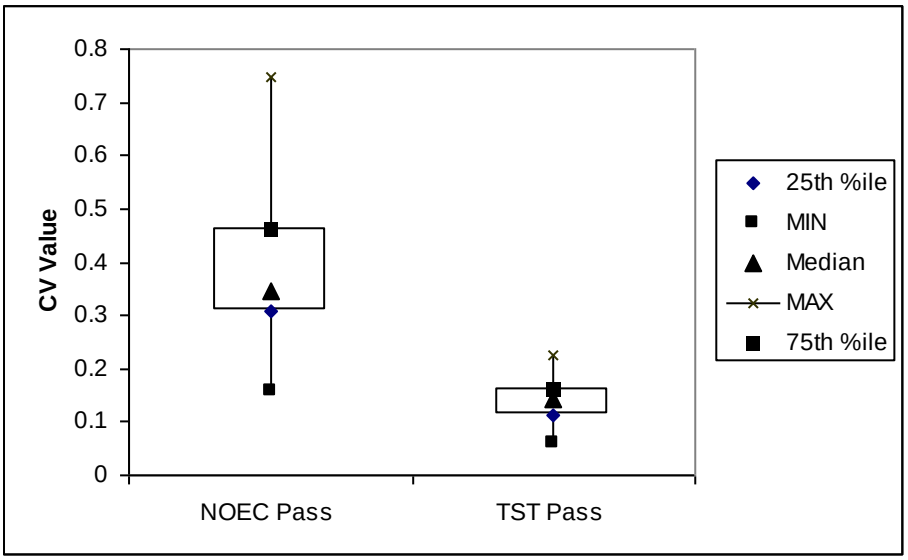

Figure C-1. Range of CV values observed in chronic *C. dubia* ambient toxicity tests for samples that were found to be non-toxic using the standard t-test but toxic using the TST approach (NOEC Pass) and for those samples declared toxic using t-test but not the TST approach (TST Pass). California's SWAMP WET test data.

Similar to the *Ceriodaphnia* ambient test data, within-test variability was higher in those chronic fathead minnow ambient tests found *non-toxic* using a t-test but toxic using the TST approach (Figure C-2). Similarly, those tests declared *non-toxic* by the TST approach but toxic using t-test had lower within-test variability and mean effect levels < 25 percent (Figure C-2). Thus, similar to the chronic *Ceriodaphnia* ambient tests, data from chronic fathead minnow ambient tests demonstrate that the TST approach can provide as much protection as the traditional t-test approach while also identifying those samples that are truly acceptable from a regulatory management decision.

Fish ambient WET tests that are identified as non-toxic using the traditional hypothesis approach (NOEC) generally have high within-test variability (high control CVs) as compared to using the TST approach.

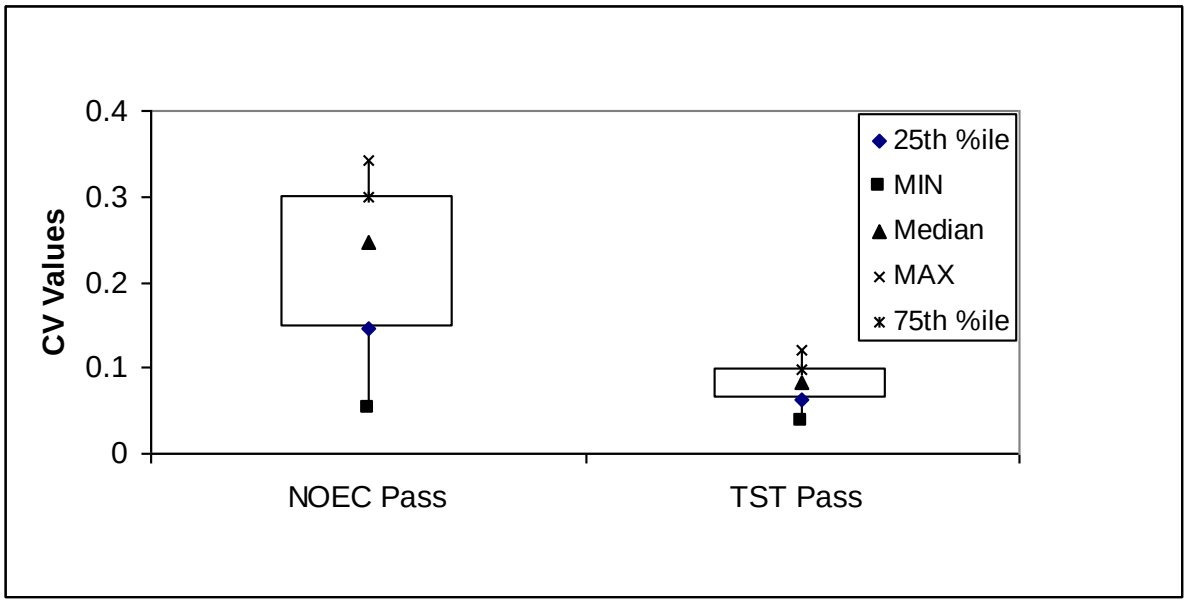

Figure C-2. Range of CV values observed in chronic *P. promelas* ambient toxicity tests for samples that were declared to be non-toxic using the standard t-test but toxic using the TST approach (NOEC Pass) and for those samples declared toxic using t-test but not the TST approach (TST Pass). California's SWAMP WET test data.

Literature Cited

USEPA (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*. EPA/600/R-95-136. U.S. Environmental Protection Agency, National Exposure Research Laboratory, Cincinnati, OH, and Office of Research and Development, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2002a. *Short-Term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Freshwater Organisms*. 4th edition. EPA/821/R-02-013. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2002b. *Short-Term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Marine and Estuarine Organisms.* 3rd ed. EPA/821/R-02-14. U.S. Environmental Protection Agency, Office of Science and Technology, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2002c. *Methods for Measuring the Acute Toxicity of Effluents and Receiving Waters to Freshwater and Marine Organisms*. 5th ed. EPA/821/R-02-012. United States Environmental Protection Agency, Office of Water, Washington, DC.

APPENDIX D

EXAMPLE NATIONAL POLLUTANT DISCHARGE ELIMINATION SYSTEM PERMIT LANGUAGE USING THE TEST OF SIGNIFICANT TOXICITY APPROACH

APPENDIX D: EXAMPLE NPDES PERMIT LANGUAGE USING THE TST APPROACH

ACUTE WHOLE EFFLUENT TOXICITY (WET) NPDES PERMIT LANGUAGE

xx. <u>Acute Whole Effluent Toxicity (WET) Requirements</u>

1. Monitoring Frequency

The permittee must conduct ***monthly/quarterly/semiannual*** acute toxicity tests on 24-hour composite effluent samples. Once each calendar year, at a different time of year from the previous years, the permittee must split a 24-hour composite effluent sample and concurrently conduct two toxicity tests using a fish and an invertebrate species; the permittee must then continue to conduct routine ***monthly/quarterly/semiannual*** toxicity testing using the single, most sensitive species.

Acute toxicity test samples must be collected for each point of discharge at the designated NPDES sampling station for the effluent (i.e., downstream from the last treatment process and any in-plant return flows where a representative effluent sample can be obtained). During years ***1, 2, 3, 4, and 5*** of the permit, a split of each sample must be analyzed for all other monitored parameters at the minimum frequency of analysis specified by the effluent monitoring program.

2. Freshwater Species and WET Test Methods

Species and short-term WET test methods for estimating the acute toxicity of NPDES effluents are in the fifth edition of *Methods for Measuring the Acute Toxicity of Effluents and Receiving Waters to Freshwater and Marine Organisms* (EPA/821/R-02/012, 2002; Table IA, 40 CFR Part 136). The permittee must conduct 96-hour static renewal toxicity tests with the following vertebrate and invertebrate species, respectively:

- ***Vertebrate:*** The fathead minnow, *Pimephales promelas* (Acute Toxicity Test Method 2000.0)
- ***Invertebrate:*** The daphnid, *Ceriodaphnia dubia* (Acute Toxicity Test Method 2002.0)

3. Acute WET Permit Triggers

a. There are no acute toxicity effluent limits for this discharge. For this permit, the determination of Pass or Fail from a multiple-effluent concentration acute toxicity test at the IWC is determined using the Test of Significant Toxicity (TST) approach that is described in the *National Pollutant Discharge Elimination System Test of Significant Toxicity Implementation Document* (EPA/833/R-10-003). The acute WET permit trigger is any one WET test where a test result is *Fail* (during the monthly reporting period) at the acute in-stream waste concentration (IWC). For this discharge, the IWC is **XXX** percent (e.g., either is 100 percent or an effluent at the mixing zone to be determined at the time of permit issuance) effluent. To calculate

either a Pass or Fail of a multiple-effluent concentration acute toxicity test at the IWC, follow the instructions in Appendix A in the *National Pollutant Discharge Elimination System Test of Significant Toxicity Implementation Document*. A *Pass* result indicates no toxicity of the multiple-effluent concentration test at the IWC, and a *Fail* result indicates toxicity of the multiple-effluent concentration test at the IWC. The permittee must report either a *Pass* or a *Fail* on the Discharge Monitoring Report (DMR) form. If a result is reported as Fail, the permittee must follow Section 6 (Accelerated Toxicity Testing and TRE/TIE Process) of this permit.

- OR -

3. Acute WET Permit Limit

b. There is an acute toxicity effluent limit for this discharge. For this permit, the determination of Pass or Fail from a multiple-effluent concentration acute toxicity test at the IWC is determined using the Test of Significant Toxicity (TST) approach which is described in the *National Pollutant Discharge Elimination System Test of Significant Toxicity Implementation Document* (EPA/833-R-10-003). The acute WET permit trigger is any one WET test where a test result is *Fail* (during the monthly reporting period) at the chronic in-stream waste concentration (IWC). For this discharge, the IWC is **XXX** percent (e.g., either is 100 percent or an effluent at the mixing zone to be determined at time of permit issuance) effluent. To calculate either a Pass or Fail of the multiple-effluent concentration acute toxicity test at the IWC, follow the instructions in Appendix A in the *National Pollutant Discharge Elimination System Test of Significant Toxicity Implementation Document*. A Pass result indicates no toxicity of the multiple-effluent concentration at the IWC and a Fail result indicates toxicity of the multiple-effluent concentration test at the IWC. The permittee must report either a *Pass* or a *Fail* on the DMR form. If a result is reported as Fail, the permittee must follow Section 6 (Accelerated Toxicity Testing and TRE/TIE Process) of this permit.

4. Quality Assurance – EPA WET Test Methods

a. Quality assurance measures, instructions, and other recommendations and requirements are in the EPA 2002 WET test methods manual previously referenced.

b. This permit is subject to a determination of Pass or Fail from a multiple-effluent concentration acute toxicity test at the IWC (for statistical flowchart and procedures, see Appendix A, Figure A-1 of the *National Pollutant Discharge Elimination System Test of Significant Toxicity Implementation Document*). The acute in-stream waste concentration (IWC) for this discharge is **XXX** percent effluent.

c. Effluent dilution water and control water should be prepared and used as specified in the EPA WET test methods manual *Methods for Measuring the Acute Toxicity of Effluents and Receiving Waters to Freshwater and Marine Organisms* (EPA/821/R-02/012, 2002).

d. If organisms are not cultured in-house, concurrent testing with a reference toxicant must be conducted. If organisms are cultured in-house, monthly reference toxicant testing is sufficient. Reference toxicant tests and effluent toxicity tests must be conducted using the same test conditions (e.g., same test duration).

e. If either the reference toxicant or effluent toxicity tests do not meet all test acceptability criteria in the EPA WET test methods manual, the permittee must resample and retest within 14 days.

f. If the discharged effluent is chlorinated, chlorine must not be removed from the effluent sample before toxicity testing without written approval by the permitting authority.

5. Initial Investigation TRE Work Plan

Within 90 days of the permit effective date, the permittee must prepare and submit to the permitting authority a copy of its Initial Investigation Toxicity Reduction Evaluation (TRE) Work Plan (1–2 pages) for review. That plan must include steps the permittee intends to follow if toxicity is measured above an acute WET permit limit or trigger and should include the following, at minimum:

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

b. A description of methods for maximizing in-house treatment system efficiency, good housekeeping practices, and a list of all chemicals used in operations at the facility.

c. If a Toxicity Identification Evaluation (TIE) is necessary, an indication of who would conduct the TIEs (i.e., an in-house expert or outside contractor).

6. Accelerated Toxicity Testing and TRE/TIE Process

a. If an acute WET permit limit or trigger is exceeded and the source of toxicity is known (e.g., a temporary plant upset), the permittee must conduct one additional toxicity test using the same species and EPA WET test method. This WET test must begin within 14 days of receipt of WET test results exceeding an acute WET permit limit or trigger. If the additional toxicity test does not exceed an acute WET permit limit or trigger, the permittee may return to the regular testing frequency.

b. If an acute WET permit limit or trigger is exceeded and the source of toxicity is not known, the permittee must conduct six additional toxicity tests using the same species and EPA WET test method, approximately every two weeks, over a 12-week period. This testing must begin within 14 days of receipt of WET test results exceeding an acute WET permit limit or trigger. If none of the additional toxicity tests exceed an acute WET permit limit or trigger, the permittee may return to the regular testing frequency.

c. If one of the additional toxicity tests (in paragraphs 6.a or 6.b) exceeds an acute WET permit limit or trigger, within 14 days of receipt of this WET test result, the permittee must initiate a TRE using, according to the type of treatment facility, EPA WET TRE manual, *Toxicity Reduction Evaluation Guidance for Municipal Wastewater Treatment Plants* (EPA/833/B-99/002, 1999) or EPA WET TRE manual, *Generalized Methodology for Conducting Industrial Toxicity Reduction Evaluations* (EPA/600/2-88/070, 1989). In conjunction, the permittee must develop and implement a Detailed TRE Work Plan that must consist of the following: further actions undertaken by the permittee to investigate, identify, and correct the causes of toxicity; actions the permittee will take to mitigate the effects of the discharge and prevent the recurrence of toxicity; and a schedule for such actions.

d. The permittee may initiate a TIE as part of a TRE to identify the causes of toxicity using the same species and EPA WET test method and, as guidance, EPA WET TIE/TRE method manuals: *Methods for Aquatic Toxicity Identification Evaluations: Phase I Toxicity Characterization Procedures* (EPA/600/6-91/003, 1991); *Methods for Aquatic Toxicity Identification Evaluations, Phase II Toxicity Identification Procedures for Samples Exhibiting Acute and Chronic Toxicity* (EPA/600/R-92/080, 1993); *Methods for Aquatic Toxicity Identification Evaluations, Phase III Toxicity Confirmation Procedures for Samples Exhibiting Acute and Chronic Toxicity* (EPA/600/R-92/081, 1993).

7. Reporting of Acute Toxicity Monitoring Results

a. The permittee must submit a full laboratory report for all toxicity testing as an attachment to the Discharge Monitoring Report (DMR) for the month in which the toxicity test was conducted; the laboratory report must contain the following: the toxicity test results, the dates of sample collection and initiation of each toxicity test; all results for effluent parameters monitored concurrently with the toxicity test(s); and progress reports on TRE/TIE investigations.

b. The permittee must provide the actual test endpoint responses for the control (i.e., control mean) and IWC concentration (i.e., IWC mean) for each WET test conducted to make it easier for permit writers to find the necessary WET test results when determining WET RP.

c. The permittee must notify the permitting authority in writing within 14 days of exceedance of an acute WET permit limit or trigger. Such notification must describe actions the permittee has taken or will take to investigate, identify, and correct the causes of toxicity; the status of actions required by this permit; and schedule for actions not yet completed; or reason(s) that no action has been taken.

8. Permit Reopener for Acute Toxicity

In accordance with 40 CFR Parts 122 and 124, this permit may be modified to include effluent limitations or permit conditions to address acute toxicity in the effluent or receiving waterbody, as a result of the discharge; or to implement new, revised, or newly interpreted water quality standards applicable to acute toxicity.

CHRONIC WET NPDES PERMIT LANGUAGE

xx. Chronic Whole Effluent Toxicity (WET) Requirements

1. Monitoring Frequency

The permittee must conduct ***monthly/quarterly/semiannual*** chronic toxicity tests on 24-hour composite effluent samples. Once each calendar year, at a different time of year from the previous years, the permittee must split a 24-hour composite effluent sample and concurrently conduct three toxicity tests using a fish, an invertebrate, and an alga species; the permittee must continue to conduct routine ***monthly/quarterly/semiannual*** toxicity testing using the single, most sensitive species.

Chronic toxicity test samples must be collected for each point of discharge at the designated NPDES sampling station for the effluent (i.e., downstream from the last treatment process and any in-plant return flows where a representative effluent sample can be obtained). During years ***1, 2, 3, 4, and 5*** of the permit, a split of each sample must be analyzed for all other monitored parameters at the minimum frequency of analysis specified by the effluent monitoring program.

2. Freshwater Species and EPA WET Test Methods

Species and short-term EPA WET test methods for estimating the chronic toxicity of NPDES effluents are in the fourth edition of *Short-term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Freshwater Organisms* (EPA/821/R-02/013, 2002; Table IA, 40 CFR Part 136). The permittee must conduct static renewal toxicity tests with the following:

- Fathead minnow, *Pimephales promelas* (Larval Survival and Growth Test Method 1000.0)
- Daphnid, *Ceriodaphnia dubia* (Survival and Reproduction Test Method 1002.0);
- Green alga, *Selenastrum capricornutum* (also named *Raphidocelis subcapitata*) (Growth Test Method 1003.0).

3. Chronic WET Permit Triggers

a. There are no chronic toxicity effluent limits for this discharge. The chronic WET permit trigger is any one WET test (either biological endpoint of survival or sublethal) where a test result is *Fail* (during the monthly reporting period) at the chronic in-stream waste concentration (IWC). For this discharge, the IWC is **XXX** percent (e.g., either is 100 percent or an effluent at the mixing zone to be determined at time of permit issuance) effluent. To calculate either a Pass or Fail of the multiple-effluent concentration chronic toxicity test at the IWC, follow the instructions in Appendix A in the *National Pollutant Discharge Elimination System Test of Significant Toxicity Implementation Document* (EPA/833-R-10-003). A Pass result indicates no toxicity at the IWC, and a Fail result indicates toxicity at the IWC. The

permittee must report either a Pass or a Fail on the DMR form. If a result is reported as Fail, the permittee must follow Section 7 (Reporting of Chronic Toxicity Monitoring Results) of this permit.

- OR -

3. Chronic WET Permit Limits

b. There is a chronic toxicity effluent limit for this discharge. The chronic WET permit trigger is any one WET test (either biological endpoint of survival or sublethal) where a test result is *Fail* (during the monthly reporting period) at the chronic in-stream waste concentration (IWC). For this discharge, the IWC is **XXX** percent (e.g., either is 100 percent or an effluent at the mixing zone to be determined at time of permit issuance) effluent. To calculate either a Pass or Fail of the multiple-effluent concentration chronic toxicity test at the IWC, follow the instructions in Appendix A in the *National Pollutant Discharge Elimination System Test of Significant Toxicity Implementation Document* (EPA/833-R-10-003). A Pass result indicates no toxicity at the IWC, and a Fail result indicates toxicity at the IWC. The permittee must report either a Pass or a Fail on the DMR form. If a result is reported as Fail, the permittee must follow Section 7 (Reporting of Chronic Toxicity Monitoring Results) of this permit.

4. Quality Assurance – EPA WET Test Methods

a. Quality assurance measures, instructions, and other recommendations and requirements are in the EPA WET test methods manual previously referenced in this permit.

b. This permit is subject to a determination of Pass or Fail from a multiple-effluent concentration chronic toxicity test at the IWC (for statistical flowchart and procedures, see *National Pollutant Discharge Elimination System Test of Significant Toxicity Implementation Document*, Appendix A, Figure A-1). The chronic in-stream waste concentration (IWC) for this discharge is **XXX** percent (e.g., either is 100 percent or an effluent at the mixing zone to be determined) effluent.

c. Effluent dilution water and control water should be standard synthetic dilution water as described in the EPA WET test methods manual, *Short-term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Waters to Freshwater Organisms* (EPA/821/R-02/013, 2002). If the dilution water is different from test organism culture water, a second control using culture water must also be used.

d. If organisms are not cultured in-house, concurrent testing with a reference toxicant must be conducted. If organisms are cultured in-house, monthly reference toxicant testing is sufficient. Reference toxicant tests and effluent toxicity tests must be conducted using the same test conditions (e.g., same test duration).

e. If either the reference toxicant or effluent toxicity tests do not meet all test acceptability criteria in the EPA WET test methods manual, the permittee must resample and retest within 14 days.

f. Following Paragraph 10.2.6.2 of the freshwater EPA WET test methods manual, all chronic toxicity test results from the multi-concentration tests required by this permit must be reviewed and reported according to EPA guidance on the evaluation of concentration-response relationships in *Method Guidance and Recommendations for Whole Effluent Toxicity (WET) Testing (40 CFR Part 136)* (EPA/821/B-00-004, 2000).

g. If the discharged effluent is chlorinated, chlorine must not be removed from the effluent sample before toxicity testing without written approval by the permitting authority.

5. Initial Investigation TRE Work Plan

Within 90 days of the permit effective date, the permittee must prepare and submit to the permitting authority a copy of its Initial Investigation Toxicity Reduction Evaluation (TRE) Work Plan (1–2 pages) for review. That plan must contain steps the permittee intends to follow if toxicity is measured above a chronic WET permit limit or trigger and should include the following, at minimum:

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

b. A description of methods for maximizing in-house treatment system efficiency, good housekeeping practices, and a list of all chemicals used in operations at the facility.

c. If a Toxicity Identification Evaluation (TIE) is necessary, an indication of who would conduct the TIEs (i.e., an in-house expert or outside contractor).

6. Accelerated Toxicity Testing and TRE/TIE Process

a. If a chronic WET permit limit or trigger is exceeded and the source of toxicity is known (e.g., a temporary plant upset), the permittee must conduct one additional toxicity test using the same species and EPA WET test method. This WET test must begin within 14 days of receipt of WET test results exceeding a chronic WET permit limit or trigger. If the additional toxicity test does not exceed a chronic WET permit limit or trigger, the permittee may return to their regular testing frequency.

b. If a chronic WET permit limit or trigger is exceeded and the source of toxicity is not known, the permittee must conduct six additional toxicity tests using the same species and EPA WET test method, approximately every two weeks, over a 12 week period. This testing must begin within 14 days of receipt of WET test results exceeding a chronic WET permit limit or trigger. If none of the additional toxicity tests exceed a chronic WET permit limit or trigger, the permittee may return to their regular testing frequency.

c. If one of the additional toxicity tests (in paragraphs 6.a or 6.b) exceeds a chronic WET permit limit or trigger, within 14 days of receipt of this WET test result, the permittee must initiate a TRE using as guidance, according to the type of treatment facility, the EPA TRE manual, *Toxicity Reduction Evaluation Guidance for Municipal Wastewater Treatment Plants* (EPA/ 833/B-99/002, 1999) or EPA TRE manual, *Generalized Methodology for Conducting Industrial Toxicity Reduction Evaluations* (EPA/600/2-88/070, 1989). In conjunction, the permittee must develop and implement a Detailed TRE Work Plan that must contain the following: further actions undertaken by the permittee to investigate, identify, and correct the causes of toxicity; actions the permittee will take to mitigate the effects of the discharge and prevent the recurrence of toxicity; and a schedule for such actions.

d. The permittee may initiate a TIE as part of a TRE to identify the causes of toxicity using the same species and EPA WET test method and, as guidance, EPA WET TIE/TRE method manuals: *Toxicity Identification Evaluation: Characterization of Chronically Toxic Effluents, Phase I* (EPA/600/6-91/005F, 1992); *Methods for Aquatic Toxicity Identification Evaluations, Phase II Toxicity Identification Procedures for Samples Exhibiting Acute and Chronic Toxicity* (EPA/600/R-92/080, 1993); *Methods for Aquatic Toxicity Identification Evaluations, Phase III Toxicity Confirmation Procedures for Samples Exhibiting Acute and Chronic Toxicity* (EPA/600/R-92/081, 1993).

7. Reporting of Chronic Toxicity Monitoring Results

a. The permittee must submit a full laboratory report as an attachment to the DMR for all toxicity testing for the month in which the toxicity test was conducted; the laboratory report must contain the following: the toxicity test results, the dates of sample collection and initiation of each toxicity test; all results for effluent parameters monitored concurrently with the toxicity test(s); and progress reports on TIE/TRE investigations.

b. The permittee must provide the actual test endpoint responses for the control (i.e., control mean) and IWC concentration (i.e., IWC mean) for each WET test conducted to make it easier for permit writers to find the necessary WET test results when determining WET RP.

c. The permittee must notify the permitting authority in writing within 14 days of exceedance of a chronic WET permit limit or trigger. The notification must describe actions the permittee has taken or will take to investigate, identify, and correct the causes of toxicity; the status of actions required by this permit; and schedule for actions not yet completed; or reason(s) that no action has been taken.

8. Permit Reopener for Chronic Toxicity

In accordance with 40 CFR Parts 122 and 124, this permit may be modified to include effluent limitations or permit conditions to address chronic toxicity in the effluent or receiving waterbody, as a result of the discharge; or to implement new, revised, or newly interpreted water quality standards applicable to chronic toxicity.

APPENDIX E

WHOLE EFFLUENT TOXICITY REASONABLE POTENTIAL ANALYSIS USING THE TEST OF SIGNIFICANT TOXICITY APPROACH

APPENDIX E: WET RP ANALYSIS USING THE TST APPROACH

For reasonable potential (RP) calculations using the TST approach, EPA recommends that permitting authorities use all the valid WET test data generated during the current permit term and any additional valid data that are submitted as part of the permit renewal application. The permitting authority should be using at least a minimum of four valid WET tests to address effluent representativeness using the TST RP approach. WET test data are then analyzed according to the TST approach using the IWC and control test concentrations for all valid WET test data available. For the RP approach, data sets with fewer than four valid WET data points should be assessed using EPA's Technical Support Document (TSD) RP approach because it addresses small WET data sets by incorporating an RP multiplying factor (see Section 3.2.2 of the TSD, p. 54) to account for effluent variability in small WET data sets.

EPA also recommends that states request that their permittees provide the actual test endpoint responses for the control (i.e., mean of control) and IWC concentration (i.e., mean of IWC) for each WET test conducted to make it easier for permit writers to find the necessary data with which to calculate WET RP with this approach. EPA recommends that permitting authorities decide up front which approach (the 1991 TSD approach, the TST approach, or another scientifically defensible approach that is sufficient to meet the statutory and regulatory requirements) they will incorporate and consistently use in their state's NPDES implementation procedures, including for their RP procedures. Permitting authorities should consistently use the selected WET statistical approach in all the state NPDES permits.

All valid WET test data are then analyzed according to the TST approach using the IWC and control test concentrations. If WET test data are available and the TST statistical approach indicates that the IWC is toxic in any WET test ("effluent **cause(s)** toxicity"), RP has been demonstrated (40 CFR 122.44(d)(1)(i)). For example, if results of five WET tests are available using the TST approach and the results are Pass, Pass, Fail, Pass, Pass, because at least one test was a Fail (i.e., TST declared the effluent toxic in at least one test), RP has been demonstrated.

To address concerns regarding the "**potential** to cause or contribute to toxicity," a second assessment is applied to determine whether the effluent has RP even if all test results are *Pass* using the TST approach.

The current TST approach results in four outcomes with respect to RP at the IWC:

1. **Caused (effluent is toxic):** RP is demonstrated if any one test using the TST approach indicates a test result is *Fail* (i.e., using the statistical test (Appendix A) and *t* table (Appendix B), the test result is *Fail*; see Example A below in Table E-1);
2. **Potential to Cause:** Effluent has reasonable potential to cause (RP is demonstrated) if any test exhibits a mean effect at the IWC > 10 percent as compared to the mean control response, even if the test result is *Pass* using TST (see examples B-D, Table E-1); and
3. **No RP (effluent is non-toxic at the IWC):** Effluent does not cause or have reasonable potential to cause if the tests are each a *Pass* using the TST approach and the mean effect at the IWC is always $\leq$ 10 percent.

4. **Insufficient valid WET data (fewer than 4 tests or no data):** If fewer than four valid WET data are available, follow the TSD RP procedure for WET.

The second outcome is where the determination of RP is critical to demonstrate that the discharge has the reasonable *potential* to cause an excursion above the state toxicity water quality standards. In the TST approach, the regulatory management decision threshold for non-toxicity in WET tests under the NPDES WET Program is 10 percent mean effect at the IWC. At or below that mean effect level, the TST approach is designed to declare a WET test as non-toxic (i.e., *Pass*) most (at least 95 percent) of the time to help control for false positives. For purposes of RP assessment then, a 10 percent mean effect level at the IWC is used as a threshold, above which potential to cause is indicated, and the effluent has demonstrated RP. Any test with a mean effect at the IWC > 10 percent would demonstrate a potential for RP even if the TST test result is *Pass*. Equation E-1 below demonstrates how the effluent effect is calculated at the IWC.

$$\%\ \text{Effect at IWC} = \frac{\text{Mean Control Response} - \text{Mean Response at IWC}}{\text{Mean Control Response}} \times 100 \qquad \textbf{Equation E-1}$$

Table E-1. Examples illustrating the reasonable potential approach using TST and data from *Ceriodaphnia* chronic survival and reproduction WET tests

Example	Pass/Fail based on TST analysis	Mean control response	Mean response @ IWC	% effect at IWC	Reasonable potential?
A	Fail	26.3	17.0	35.4%	Yes
B	Pass	26.3	23.4	11.0%	Yes
C	Pass	28.6	22.0	23.1%	Yes
D	Pass	22.4	20.9	6.7%	No