

White Paper

***Ceriodaphnia dubia* Short-term Chronic Reproduction Test: Understanding the Probability of Incorrect Determinations of Toxicity in Non-toxic Samples**

Prepared for the California Association of Sanitation Agencies (CASA)

Prepared by Larry Walker Associates, Inc., Davis, California

November 28, 2018

EXECUTIVE SUMMARY

The results of *Ceriodaphnia dubia* short-term chronic tests are important in the NPDES permit compliance world given the frequency of occurrence of samples deemed to be toxic using this testing method. Questions regarding the variability of WET tests in general, and the *Ceriodaphnia dubia* short-term chronic reproduction test in particular, have existed for over twenty years. Various studies have been performed to address these questions, but questions and concerns remain, based on recent experience. The implementation of a new statistical approach to the interpretation of WET testing results in NPDES permits and the proposed adoption of numeric effluent limits based on WET testing results has amplified these concerns in California.

The purpose of this analysis is to summarize existing chronic toxicity *Ceriodaphnia dubia* reproduction test data from prior studies conducted on known non-toxic blank samples and to assess whether the results are sufficient to resolve concerns regarding the variability of interlaboratory *Ceriodaphnia dubia* test results or whether additional testing is advisable to develop recommendations to reduce observed variability.

The approach used in this white paper is to summarize the background of the short-term chronic WET testing program from its outset under the Clean Water Act, to identify what is known about the potential sources of WET testing variability, to document the significance of the *Ceriodaphnia dubia* chronic test as a driver of NPDES permit compliance actions, and, as a first step, to perform a statistical analysis using the raw data from three available “blank” studies, taking those data at face value, i.e., treating all samples meeting test acceptability criteria as valid, recognizing that concerns have been raised regarding the validity of certain data.

Following this approach, results were developed using the three common statistical approaches for interpretation of WET testing, including the Test of Significant Toxicity (TST), conventional hypothesis testing resulting in a determination of No Observable Effect Concentrations (NOEC), and point estimate techniques resulting in determination of Inhibition Concentrations at 25% effect level (IC25). Results from the statistical analysis were then considered in the overall context of concerns regarding the quality of “blank” samples, the methods used in the performance of the studies, and/or the quality of the laboratory work.

Prior Analysis Summary: Percent of Blank Samples Incorrectly Identified as Toxic

Study Name	NOEC	IC25	TST
EPA ^[a]	3.6%	7.1%	14%
SMC	36%	45%	55%
WC	33%	33%	38%

[a] The study originally included two test results that EPA subsequently wished to reject as being invalid. Their reason for rejecting the first result (EPA-9450) was due to a high pMSD, but the protocol dictates that a high pMSD has no impact on a toxicity test. The second result (EPA-9332) was considered valid in 2002 but 13 years later was stated not to have three broods. Both results were included in this analysis.

While the statistical analysis of available data indicates that the rate of incorrect determinations of toxicity was unacceptably high, the overall assessment found that current studies and data are insufficient to resolve outstanding questions regarding variability of *Ceriodaphnia dubia* chronic WET test results and incorrect determinations of toxicity. This study suggests that performance of a properly designed and implemented “blank” study with sufficient statistical power would have the potential to resolve concerns by better quantifying the occurrence of incorrect determinations of toxicity and, if appropriate, by examining measures that may reduce such results to an acceptable level while not impacting the ability to correctly detect toxic samples.

INTRODUCTION

Purpose

The purpose of the information provided in this white paper is to evaluate whether adequate data presently exist to determine the probability of falsely identifying non-toxic samples as toxic in assessing *Ceriodaphnia dubia* reproduction test results using various EPA statistical methods (i.e., Test of Significant Toxicity (TST), conventional hypothesis testing resulting in a determination of No Observable Effect Concentrations (NOEC), and point estimate techniques resulting in determination of Inhibition Concentrations at 25% effect level (IC25)). If available data are found to be insufficient, the additional purpose of this white paper is to offer recommendations regarding the need for additional research to improve understanding of this issue.

The drivers for the preparation of this white paper are threefold: (1) *Ceriodaphnia dubia* short-term chronic reproduction testing continues to provide an indication of low level toxicity in a significant number of highly treated POTW effluents, (2) inter-laboratory studies continue to suggest that there is significant variability in *Ceriodaphnia dubia* chronic reproduction test results among laboratories, including testing on blank samples, and (3) significant compliance costs associated with accelerated testing and TRE procedures to address *Ceriodaphnia dubia* chronic test results are currently occurring and are anticipated to continue or increase under the proposed State Water Resources Control Board (State Water Board) Toxicity Plan provisions of the Inland Surface Waters Plan. If identified, measures to reduce the occurrence of findings of toxicity in non-toxic samples while still finding toxicity where it exists would provide benefits to both the regulated and regulatory community in terms of conserved time and resources.

Background

The USEPA's focus on short-term chronic Whole Effluent Toxicity (WET) testing began in the 1980s as one of many actions taken to implement the Clean Water Act. Whole effluent toxicity is the aggregate toxic effect of an aqueous sample (e.g. effluent, receiving water) measured directly by an aquatic toxicity test. Aquatic toxicity tests are laboratory experiments that measure the biological effect (e.g. growth, survival, reproduction) of effluent or receiving waters on test organisms. In aquatic toxicity tests, organisms of a specific species are held in test chambers under controlled laboratory conditions and exposed to one or more concentrations of an aqueous sample, such as a reference toxicant, effluent, or receiving water, and observations are made at pre-determined exposure periods. At the end of the test, the responses of the test organisms are used to estimate the effects of the aqueous sample being tested. The primary advantage of the use of WET tests is the fact that the test integrates the effects of all chemicals in the tested sample.

Starting in 1985, USEPA published its first versions of short-term methods for estimating the chronic toxicity of effluents and receiving waters. In March 1991, USEPA issued the Technical Support Document for Water Quality-based Toxics Control (TSD) which, among other things, described the intended use of WET testing in NPDES permits. USEPA promulgated WET test methods in October 1995 for use in the NPDES program.

As a result of this direction from USEPA, since the mid-1990s, short-term chronic WET tests have been used in NPDES permits throughout California. In freshwater situations, the USEPA "three species" suite of short-term chronic tests has been commonly used in NPDES permits.

This suite of tests includes water flea (*Ceriodaphnia dubia*) survival and reproduction, algae (*Selanastrum capricornutum*) growth, and fathead minnow (*Pimephales promelas*) larval growth and survival.

In 2000, a statewide approach to the use of short-term chronic toxicity tests in NPDES permits was established by the State Water Board in the State Implementation Plan (SIP), which described the implementation of toxicity provisions contained in the Inland Surface Waters Plan. The SIP prescribed a narrative permitting approach, with toxicity test result “triggers” leading to action including accelerated monitoring and, where monitoring continued to show toxicity, performance of TRES. At the time of adoption of this statewide approach, the common statistical methods used to interpret toxicity results in California were the hypothesis testing (NOEC) approach and the point estimate (IC25) approach.

A new statewide Toxicity Plan is expected to be adopted by the State Water Board in 2019, replacing the current NPDES permitting approach in the SIP. That Plan is anticipated to change from the above-described narrative approach to the use of numeric toxicity effluent limitations in NPDES permits and the use of a new statistical approach, the Test of Significant Toxicity (TST), to interpret test results.

Since the issuance of the TSD and culminating in the promulgation of WET test methods in 1995, concerns were voiced by the regulated community regarding the use of short-term chronic tests for NPDES permit compliance. Issues of concern have included variability of biological test results, differences in methodologies and capabilities among laboratories, inability to dependably determine causation of chronic toxicity, and choice of statistical methods to interpret test results.

Those concerns led to a lawsuit (WestCAS and Edison Electric Institute et al. vs. USEPA) in the mid-1990’s. The litigation was settled in July 1998. The following points were contained in the settlement agreement:

- Agreement by USEPA to perform multi-laboratory evaluation of twelve WET tests,
- Agreement by USEPA to prepare guidance on how to take WET test variability into account in NPDES permits, and
- Agreement by USEPA to consider minor clarifications to WET test methods.

In response to the litigation, USEPA performed a study in the late 1990’s titled *Understanding and Accounting for Method Variability in Whole Effluent Toxicity Applications Under the National Pollutant Discharge Elimination System Program* (EPA 833-R000-003, June 2000). The document was produced to address questions raised regarding WET test method variability and to satisfy a requirement of the July 1998 settlement agreement. The document addresses potential sources of variability associated with WET testing, steps to minimize variability, and means to address variability in the NPDES permitting program. The document was produced by an EPA workgroup and was externally peer reviewed. Important findings of the document included:

- The TSD approach to reasonable potential determinations and effluent limit derivation is “appropriately protective.” It would be inappropriate to make adjustments to the TSD methodology to account for toxicity test variability in NPDES permitting.
- The hypothesis test procedures prescribed in USEPA’s WET methods will provide adequate protection against false conclusions that an effluent is toxic. The use of an alpha factor of 0.05 established the expected maximum rate of such errors (i.e., one in twenty

samples). The hypothesis test procedure is designed to provide an error rate no greater than the alpha value when default assumptions are met. If a test is properly conducted and correctly interpreted, incorrect determinations of toxicity should be impossible.

- The variability of promulgated WET methods is within the range of variability experienced in other types of analysis (e.g. chemical analyses used in NPDES permit compliance determinations).
- Standardizing the choice of reference toxicants, the concentrations to be tested, and the range of acceptable effect concentrations for each test method would help resolve quality assurance problems.
- The data analysis performed for this study did not reveal the potential sources and causes of variability, such as using different sources of test organisms, dilution water, and food. Assessment of these sources of variability would require well-designed studies.
- Interim coefficients of variation (CVs) were developed for promulgated WET methods, pending completion of additional USEPA inter-laboratory studies (see the USEPA 2001 study described below).

Of particular interest for this white paper are (1) the findings regarding incorrect determinations of toxicity under the hypothesis testing approach and (2) the conclusion that evaluation of sources of variability in WET test results would require well-designed special studies.

PROBLEM DEFINITION

Why is a blank study potentially important?

The proper evaluation of any test method requires the assessment of “known” samples. In toxicity testing, a blank sample is the only known sample available. If, in fact, the occurrence of determinations of toxicity in non-toxic samples is unacceptably high for a given method, this would cause a significant waste of resources. A study to quantify and thereby reduce such results would be valuable.

The 2000 USEPA study identified a number of categories of short-term chronic WET test variability, which can be described as follows:

- Variations within the standard methods, e.g., test organisms, dilution water, feeding, sample handling, renewals, control tests, and dilution series;
- Variations within laboratories, e.g., all of the above plus analysts and application of standard methods;
- Variations among laboratories, e.g., all of the above plus analysts and application of standard methods;
- Different methods of statistical analysis, e.g., hypothesis testing (NOEC), point estimates (IC25).

The 2000 USEPA study also identified the following sources of inter-laboratory variability:

- Different concentrations in a dilution series,
- Different dilution waters,
- Different foods and feeding regimes,
- Differences in cultures of test organisms (genotypic and phenotypic differences in sensitivity), and

- Different reference toxicant methods or reporting.

Depending on the specific design elements, an inter-laboratory blank study could help answer questions regarding some of the sources of variability both within and between laboratories and could identify potential solutions to reduce that variability.

Why focus on the *Ceriodaphnia dubia* reproduction test?

As one example, recent work has been performed to assess the results of short-term chronic toxicity testing in the Central Valley for the period 2011 to 2017. The study has been performed as a special study funded by members of the Central Valley Clean Water Association (CVCWA), an association of publicly-owned treatment works (POTWs) located in the Central Valley¹. The results of that study indicate that *Ceriodaphnia dubia* short-term chronic reproduction tests are a primary source of findings of apparent toxicity in effluent samples from POTWs in the Central Valley, including many POTWs with advanced treatment systems. These results are consistent with results reported by others (e.g., LACSD, City of San Jose) and are believed to provide a representative sampling of California POTWs discharging to effluent-dependent water bodies.

As shown in **Table 1**, most (73%) of the POTWs with NPDES permits (those discharging to surface waters) in the Central Valley are tertiary facilities. Essentially all of these POTWs discharge to effluent-dominated streams and receive no dilution credit in their NPDES permits.

Table 1. Comparison of POTWs with NPDES Permits in Toxicity Special Study to All Central Valley POTWs.

	Total Central Valley POTWs	Number of POTWs in Study	Total Number of Chronic Toxicity Test Reports ^[a]
Treatment Type			
Secondary	15 (19%)	14 (21%)	620 (21%)
Advanced Secondary	6 (8%)	5 (8%)	145 (5%)
Tertiary	56 (73%)	47 (71%)	2,187 (74%)
Disinfection Process			
Chlorination ^[b]	39 (51%)	31 (47%)	1,144 (39%)
Ultraviolet ^[b]	38 (49%)	35 (53%)	1,808 (61%)

[a] Chronic toxicity test reports include routine chronic toxicity testing, accelerated testing, and TRE/TIE-related testing.

[b] Nevada County Sanitation District No. 1 Lake Wildwood Wastewater Treatment Plant converted from chlorination to ultraviolet light disinfection in 2013. The data analysis considers the disinfection method utilized at the time of the chronic toxicity test and groups information appropriately. For this table, this facility is considered an ultraviolet light disinfection facility.

As shown in **Table 2**, for the period from January 2011 to March 2017, *Ceriodaphnia dubia* chronic test reproduction results have yielded the greatest frequency of toxicity in effluent samples in the Central Valley for POTWs which have no dilution credit and must meet chronic toxicity triggers of 1 TUc. For Central Valley POTWs in this category, *Ceriodaphnia dubia* reproduction results have accounted for 55% of all toxicity trigger exceedances for the period in question. The frequency of toxicity trigger exceedances for the routine tests performed is shown to be 16% in **Table 2**.

¹ CVCWA, 2018. *Toxicity Special Study: Phase I Study Report*. Preliminary Draft. October.

Table 2. Central Valley POTWs Chronic Toxicity Trigger Exceedances, January 2011 to March 2017.

Test Organism/Endpoint	Total Number of Chronic Toxicity Tests	Number of Chronic Toxicity Trigger Exceedances (%)	Number and Percent of Central Valley POTWs Impacted
Central Valley POTWs with Chronic Toxicity Trigger = 1 TU_c^[a]			
<i>Pimephales promelas</i> (survival)	832	4 (0.5%)	2 (3.4%)
<i>Pimephales promelas</i> (growth)	834	20 (2.4%)	15 (25.4%)
<i>Ceriodaphnia dubia</i> (survival)	818	9 (1.1%)	7 (11.8%)
<i>Ceriodaphnia dubia</i> (reproduction)	820	131 (16.0%)	38 (64.4%)
<i>Selenastrum capricornutum</i> (growth)	835	76 (9.1%)	29 (49.2%)
Central Valley POTWs with Chronic Toxicity Trigger > 1 TU_c^[b]			
<i>Pimephales promelas</i> (survival)	128	2 (1.6%)	1 (14.3%)
<i>Pimephales promelas</i> (growth)	128	2 (1.6%)	1 (14.3%)
<i>Ceriodaphnia dubia</i> (survival)	137	0 (0.0%)	0 (0.0%)
<i>Ceriodaphnia dubia</i> (reproduction)	138	17 (12.3%)	3 (42.9%)
<i>Selenastrum capricornutum</i> (growth)	152	2 (1.3%)	1 (12.5%)

[a] There are 59 Central Valley POTWs that have a chronic toxicity trigger of 1 TU_c for all test organisms. Prior to December 1, 2014, the City of Woodland had a chronic toxicity trigger of 1 TU_c for *S. capricornutum*. Chronic toxicity data collected for the City of Woodland prior to this date were included in this subset of data.

[b] There are 7 (seven) Central Valley POTWs that have a chronic toxicity trigger of greater than 1 TU_c for all test organisms. After December 1, 2014, the City of Woodland had a chronic toxicity trigger of 2 TU_c for *S. capricornutum*. Chronic toxicity data collected for the City of Woodland after this date were included in this subset of data.

Approximately two-thirds of the POTWs with a chronic toxicity trigger of 1 TU_c experienced at least one chronic toxicity trigger exceedance for *Ceriodaphnia dubia* reproduction between 2011 and 2017. For Central Valley POTWs that have a chronic toxicity trigger greater than 1 TU_c, the *Ceriodaphnia dubia* reproduction test endpoint was exceeded more frequently than the other chronic toxicity endpoints and affected almost half of the POTWs in this category. When chronic toxicity was indicated for *Ceriodaphnia dubia* reproduction (i.e., the chronic toxicity trigger was exceeded), approximately a quarter of those tests showed effects less than 25%, and approximately three-quarters of those tests showed effects less than 50%.

Information on 35 TREs completed in the Central Valley in the period from 2011 to 2017 was compiled. The test organisms and endpoints which led to the performance of the 35 TREs are presented in **Figure 1**. As shown, 28 of the 35 TREs that were examined were triggered by short-term chronic *Ceriodaphnia dubia* test results. Thirty-one of the 35 TREs evaluated were associated with POTWs that have a chronic toxicity trigger of 1 TU_c and were POTWs with either advanced secondary or tertiary level treatment.

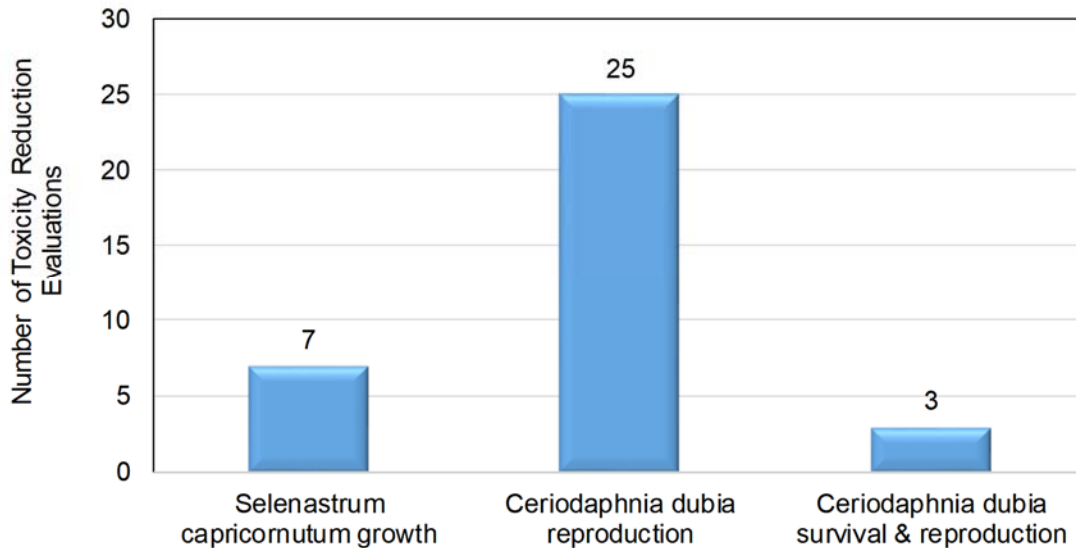


Figure 1. Central Valley POTW Toxicity Reduction Evaluation Test Organism and Endpoint

This information indicates that the *Ceriodaphnia dubia* short-term chronic reproduction test is a significant driver of current NPDES permit compliance activity for POTWs with advanced treatment facilities in California. Given the switch to the TST statistical method, the anticipated provisions of the State Water Board’s proposed Toxicity Plan, and the expected continuation of the use of the *Ceriodaphnia dubia* reproduction test in NPDES permits, there is no reason to anticipate that this situation will change in the future.

Haven’t prior studies already answered the question regarding incorrect determinations of toxicity?

To date, three studies have been performed which provide data to address the variability of blank sample WET test results. In response to the July 1998 litigation settlement agreement described above, USEPA performed an interlaboratory study that was finalized in 2001 titled *Final Report: Interlaboratory variability study of EPA short-term chronic and acute whole effluent toxicity test methods*. That study found that the rates of incorrect determinations of toxicity were within an acceptable range for almost all test methods, including the *Ceriodaphnia dubia* reproduction test. The results of that study were contradicted by the results of a second interlaboratory study performed by Moore et al. in 2000 which addressed specific questions regarding the *Ceriodaphnia dubia* reproduction test. The Moore et al. study addressed the issue of incorrect determinations of toxicity by sending blank samples to 16 certified laboratories. Of the 14 valid chronic tests received, toxicity was determined to exist in 6 of the blank samples. A more recent study performed by the Southern California Coastal Water Research Project for the Stormwater Coalition in Southern California in 2016 produced interlaboratory results for blank samples for a variety of short-term chronic toxicity tests, including the *Ceriodaphnia dubia* reproduction test. This study also found a high incidence of toxicity in blank samples in the *Ceriodaphnia dubia* reproduction test, over two rounds of analyses.

These contradictory results have caused continued concern over this issue within the regulated community. Additionally, none of the three studies evaluated the test results using the TST

approach, which is anticipated to be the approach used for future NPDES permit compliance in California. For these reasons, it is concluded that the question of incorrect determinations of toxicity using the *Ceriodaphnia dubia* short-term chronic test protocol has not been resolved to date.

METHODS

For this study, a two-step approach has been chosen. In the first step, available raw data from the three blank studies were analyzed at face value, to determine whether statistically significant determinations can be made based on NOEC, IC25 and TST results. Power analyses were used to determine whether adequate data exist to reach various conclusions with sufficient confidence. In the second step, a qualitative assessment was made regarding the validity of the raw data, taking into consideration feedback that has been received on the three studies in question. Results from these two steps were used in the development of findings.

As mentioned above, data from blank sample analyses for use in the statistical assessment performed for this white paper are available from the following three studies.

A. U.S. EPA, 2001. *Final report: Interlaboratory variability study of EPA short-term chronic and acute whole effluent toxicity test methods.* United States Environmental Protection Agency, Office of Water. Washington, DC. EPA-821-B-01-04. (Called the EPA study in this analysis.)

USEPA performed an interlaboratory study of 12 EPA short-term chronic and acute whole effluent toxicity test methods in 1999-2000. The study was performed under the terms of a settlement agreement to resolve a judicial challenge to the EPA 1995 promulgation of 17 whole effluent toxicity test methods.

The purpose of the study was to characterize (1) interlaboratory variability, (2) the rate of successful test completion and (3) the rate of incorrect determinations of toxicity for 12 WET test methods, which included the *Ceriodaphnia dubia* chronic test of survival and reproduction.

Successful test completion rates were higher than 90% for all tests except *Ceriodaphnia dubia* chronic (82%) and *Selanastrum* chronic (64-66%). Rates of incorrect determinations of toxicity were less than 5% for all tests except the *Selanastrum* chronic without EDTA² (33.3%). The USEPA authors found that the rate of incorrect determinations of toxicity for the *Ceriodaphnia dubia* chronic test was 3.7% (one out of 27 blanks showed toxicity). Interlaboratory variability was described by the coefficient of variation (CV) calculated for point estimates. Interlaboratory CVs of IC25s ranged from 10% to 58% for chronic test methods. Interlaboratory variability (CV) for *Ceriodaphnia dubia* chronic reproduction endpoints (IC25) was 35%. The within-laboratory CV for *Ceriodaphnia dubia* chronic reproduction results (IC25) was 17.4%.

Results from a total of 34 participant laboratories were used in the *Ceriodaphnia* chronic test method for the study, completing a total of 122 chronic tests. Of these 122 tests, 22 failed to meet test acceptability criteria and were deemed invalid; 3 tests yielded inconclusive results. Of the 34 laboratories, 24 produced valid *Ceriodaphnia dubia* chronic results in all tests. Of

² Ethylenediamine tetraacetic acid.

the 10 laboratories that had invalid tests, two had invalid results in all samples tested and 8 performed invalid tests in more than 50% of the samples tested. The single blank sample result with an incorrect determination of toxicity was produced by one of the labs that failed to perform valid *Ceriodaphnia dubia* chronic tests on any of the samples it tested.

- B. Moore, T.F., Canton, S.P., Grimes, M., 2000.** *Investigating the incidence of type I errors for chronic whole effluent toxicity testing using Ceriodaphnia dubia.* Environmental toxicology and chemistry 19.1 (2000): 118-122. (Called the WC (Western Coalition of Arid States) study in this analysis.)

A method blank-type study was performed in 1997 using the standard *Ceriodaphnia dubia* chronic WET test procedure [40 CFR 136 Method 1002.0] to investigate Type I errors (i.e., determinations of toxicity in non-toxic samples). Municipal wastewater entities contracted with 17 laboratories to perform tests on blank samples which were prepared using the standard dilution water formula for moderately hard water. Laboratories were not aware that the samples were blanks. Valid chronic test results were received for 16 of 17 samples based on control survival acceptance criteria and 14 of 17 samples were valid based on achievement of minimum reproduction acceptance criteria. Of those valid tests, no incorrect determinations of toxicity were found for the survival endpoint and 6 incorrect determinations of toxicity results were obtained for the reproductive endpoint. No plausible causes for the high incidence of incorrect determinations of toxicity were identified. It was concluded by the authors that contamination of the blank samples did not occur. It was also concluded by the authors that laboratory quality and/or poor laboratory performance did not cause the observed results. Changes in test acceptance criteria, combined use of NOEC and IC25 estimates to establish toxicity, use of monthly method blank tests and addition to reference toxicant tests to chart and assess laboratory performance, and reliance on mortality rather than sublethal endpoints were offered as recommendations to reduce Type I errors in the *Ceriodaphnia dubia* reproduction test.

- C. Schiff, K.C. and Greenstein, D., 2016.** *Stormwater Monitoring Coalition Toxicity Testing Laboratory Guidance Document.* Southern California Coastal Water Research Project, Technical Report 0956. (Called the SMC study in this analysis.)

The Southern California Stormwater Monitoring Coalition (SMC) includes fifteen (15) regulated and regulatory agencies ranging from Ventura in the north to San Diego in the south. The SMC conducted a laboratory intercalibration study to assess the comparability of toxicity test results for four test species, one of which was the *Ceriodaphnia dubia* chronic survival and reproduction test. The goal was to quantify intra- and interlaboratory variability for each test, and to make recommendations regarding steps to minimize that variability, where applicable.

Nine laboratories reported results for *Ceriodaphnia dubia* reproduction in an initial round of blind testing. Six laboratories provided results in a second round of *Ceriodaphnia dubia* reproduction testing. Each round of testing included four samples: laboratory dilution water, laboratory dilution water spiked with copper, runoff sample created with artificial rainfall, and a duplicate.

For most of the tests, laboratories produced data consistent with non-toxic samples when exposed to laboratory dilution water. The results from the *Ceriodaphnia dubia* reproduction tests were an exception. Laboratories exhibited a wide range of *Ceriodaphnia dubia*

reproduction test results for both “blank” samples and copper spiked samples. The authors of this report attributed this to variability in the test method and the need for a revised study design, or both. The authors also noted that the amount of *Ceriodaphnia dubia* chronic testing variability observed in their study was not uncharacteristic of the variability observed by others examining wastewater effluents, reference toxicants, or ambient media, citing results from the above described 2001 study by Moore et al. and a 2008 study by Diamond et al.

It should be noted that the findings of toxicity in blank samples from each of the three studies have been questioned for a variety of reasons. The primary criticism is that the percent effect observed for some of the blank samples is not believable, and that “something must have gone wrong” in the performance of the studies. The possible causes were suspected to be poor laboratory performance, contamination of the blank samples, or other unknown causes. The Moore et al. study (also known as the WC study) was criticized because its results were quite different from the results obtained in the 2001 USEPA study. For instance, the blank water used in the Moore et al. study was believed to be “slightly toxic,” as its make-up deviated slightly from the protocol. This criticism was made despite the fact that special steps were taken in that study in the preparation and analysis of the blank samples that were distributed for testing by multiple laboratories. The results from the SMC study have been questioned because the study was not specifically designed to address the issue of incorrect determinations of toxicity.

Will the change to a Test of Significant Toxicity (TST) statistical approach resolve the potential need for a blank study?

None of the three studies described above relied upon the TST approach in the evaluation of blank samples. Therefore, the effect of changing to the TST approach on the evaluation of blank sample results is unknown.

A study was performed for the State Water Board by USEPA, Region 9 in 2011³ to assess the differences between results obtained by the NOEC and TST approaches to data interpretation. The State Water Board requested USEPA, Region 9 to perform this “test drive” which compared WET data results using the TST and the NOEC statistical approaches as a means to address stakeholder comments received at a November 2010 workshop. The primary objective of the analysis was to assess whether the two statistical approaches yielded similar or different determinations regarding the toxicity of whole effluent samples. The study did not assess the toxicity of “blank” samples.

The approach included effluent data gathered from over 25 stormwater and wastewater dischargers. The data were screened to ensure that test method acceptability criteria were met. A total of 837 test results were obtained and 775 results were deemed to be valid and usable by the study. As part of this study, a total of 84 chronic *Ceriodaphnia dubia* reproduction tests were evaluated. The study found that the frequency of toxicity detection based on the 84 *Ceriodaphnia dubia* short-term chronic reproduction test results was similar whether using either the NOEC or the TST statistical approach.

³ USEPA, 2011. *Whole Effluent Toxicity Test Drive Analysis of the Test of Significant Toxicity (TST)*. USEPA, Region 9. July.

Because the 2011 “test drive” did not yield blank sample data and did not evaluate the toxicity of blank samples, it did not provide definitive information regarding incorrect determinations of toxicity and did not resolve the potential need for a blank study.

STATISTICAL ANALYSIS OF AVAILABLE “BLANK” DATA

The error rate for a test method can be determined empirically by performing a statistically sufficient number of blind⁴ analyses on blank⁵ samples. The three studies described above provided the *Ceriodaphnia dubia* reproduction toxicity raw data that were analyzed in this white paper (EPA⁶, SMC⁷, WC⁸). The *Ceriodaphnia dubia* reproduction tests in these studies were all performed blind on blank water samples, which were prepared by different accredited test laboratories. A blank sample is expected to show no toxicity. The toxicity data results from the three studies were analyzed using three statistical methods: NOEC, IC25 and TST. The results of the three studies were compiled and reviewed for significance using paired T-tests. A power analysis was used to determine whether a statistically sufficient number of samples had been analyzed to determine an error rate for the *Ceriodaphnia dubia* reproduction test with appropriate statistical confidence (a confidence level of 80% is deemed necessary).

Description of Analysis

Toxicity Studies

During each of the three toxicity studies, ten replicates were analyzed by each laboratory at five concentrations (blank water was provided to each laboratory in place of analysis water for creating these concentrations) and a control. There were two exceptions during the EPA study, when only eight replicates were analyzed, and during the SMC study, when two labs failed to analyze all replicates for the 12.5% concentration. These exceptions are not expected to affect the results.

During the WC study, some samples were labeled “Reference toxicant,” some “Effluent,” and some “Process control sample,” although all samples were blank. During the SMC and EPA studies, the samples were not labeled in any distinguishing way. Unlike the other studies, the SMC study was not specifically designed to determine the rate of incorrect determinations of toxicity during toxicity testing.

The five sample treatment concentrations were 6.25, 12.5, 25, 50, and 100%, and each was compared with a control of laboratory water. As the provided water was blank, all concentrations were theoretically identical to each other and for each lab. Questions have been raised in the review of these studies as to the validity of this assumption. Differences could have been caused by different types of dilution water used by each lab and other unique methods employed within the parameters of the USEPA standard method for performance of *Ceriodaphnia dubia* reproduction testing.

⁴ The term “blind” refers to the fact that the laboratories were not aware that the sample water was pure, clean water, thereby removing the possibility of laboratory staff unconsciously or unintentionally performing the analysis in a manner different from any other sample.

⁵ The term “blank” refers to water that is supposedly clean and pure, without chemicals that could cause toxicity. A blank sample is expected to show no toxicity.

⁶ U.S. EPA, 2001.

⁷ Schiff, K.C., Greenstein, D., 2016.

⁸ Moore, T.F., Canton, S.P., and Grimes, M., 2000.

Each laboratory in the three studies was represented by an identification number for anonymity. Thirty-four laboratories provided *Ceriodaphnia dubia* reproduction data in the EPA study, nine in the SMC study (which included two separate analyses, duplicates, and a reference lab, for a total of 22 discrete results) and results from 17 laboratories were obtained in the WC study (some laboratories were issued multiple identification numbers). The laboratory identification numbers for each study are shown in **Table 3**.

Table 3. Laboratory Identification Numbers for Each Study

EPA ^[a]	SMC ^[b]		WC ^[c]
	Round 1	Round 2 ^[d]	
EPA-9332	SMC1-A	SMC2-A	WC1
EPA-9338		SMC2-Ad	WC2
EPA-9340	SMC1-B	SMC2-B	WC3
EPA-9349		SMC2-Bd	WC4
EPA-9350	SMC1-C	SMC2-C	WC5
EPA-9367		SMC2-Cd	WC6
EPA-9376	SMC1-E		WC7
EPA-9381	SMC1-F		WC8
EPA-9382	SMC1-G	SMC2-G	WC9
EPA-9384		SMC2-Gd	WC10
EPA-9425	SMC1-H		WC11
EPA-9429	SMC1-I	SMC2-I	WC12
EPA-9436		SMC2-Id	WC13
EPA-9450	SMC1-J	SMC2-J	WC14
EPA-9330		SMC2-Jd	WC15
EPA-9337	SMC1-Reference		WC16
EPA-9341			WC17
EPA-9344			WC18
EPA-9356			WC19
EPA-9371			WC20
EPA-9379			WC21
EPA-9402			WC22
EPA-9409			WC23
EPA-9410			WC24
EPA-9432			WC25
EPA-9439			WC26
EPA-9445			
EPA-9446			

[a] U.S. EPA, 2001.

[b] Schiff, K.C., Greenstein, D., 2016.

[c] Moore, T.F., Canton, S.P, and Grimes, M., 2000.

[d] A "d" following the Lab ID indicates a duplicate analysis. These were used in the study as equivalent samples, as all sample water was blank.

Toxicity Data Analyses

The data provided by the labs for each of the three studies was analyzed with CETIS software using three types of toxicity analysis:

- No Observable Effect Concentration (NOEC) hypothesis test,
- Inhibition Concentration 25 (IC25) point estimate test, and
- Test of Significant Toxicity (TST) hypothesis test.

1. NOEC Analysis

The NOEC hypothesis tests included the following, selected as appropriate for the dataset (normal or non-normal distributions, equal or unequal variances):

- Bonferroni Adjusted T-Test
- Dunnett Multiple Comparison Test
- Steel Many-One Rank Sum Test
- Wilcoxon/Bonferroni Adjusted Test

In hypothesis testing, an alpha value is selected which defines the false negative rate. The NOEC test used an alpha value of 5%.

2. TST Analysis

The TST is a statistical approach which does not require normality testing, and which uses bioequivalence hypothesis testing which examines whether there is a non-toxic effect at a single concentration of concern compared with a control.⁹ The TST Welch's t-Test was used and the alpha value was set to 20%.

3. IC25 Analysis

The IC25 point estimate test used was Linear Interpolation. The Linear Interpolation method is a procedure to calculate a point estimate of the test concentration that causes a given percent reduction in the reproduction of the organism. It assumes that the responses for each concentration proceed linearly (that each one is lower than the next).

The three analyses were performed on the laboratory results for all five dilutions and the control. The significant effect dilution (if any) and corresponding percent effect was determined for each result, as was the NOEC or IC25, TU and Percent Minimum Significant Difference (pMSD) (for hypothesis tests only).

In three cases of the NOEC test, the pMSD and percent effect were both below the lower pMSD bound when a significant effect had been identified. In these cases, the sample was deemed non-toxic if an individual analysis run on the next-highest dilution confirmed the result. In all three cases, the individual analysis confirmed no toxicity at the next-highest dilution, and the NOEC results were adjusted accordingly.

The control acceptability criteria were not met for five labs during the WC study (WC03, WC10, WC13, WC23, and WC25). All but one of these labs in this study produced results which

⁹ Denton, Debra L., Jerry Diamon, and Lei Zheng. "Test of significant toxicity: A statistical application for assessing whether an effluent or site water is truly toxic." *Environmental Toxicology and Chemistry* 30.5 (2011): 1117-1126

showed toxicity using the IC25 test and failed the TST test, although only one lab produced results that showed toxicity using the NOEC test. All results failing the control acceptability criteria were removed from this analysis.

A summary of the toxicity test analyses is shown in **Table 4**. Gray shading indicates the test results which did not meet the control acceptability criteria. These results were not included in the analysis. The raw toxicity test analysis results are shown in Appendix A.

Table 4. Summary of Toxicity Test Analyses

Laboratory ID	Result (percent)						Applicable %Effect
	NOEC		IC25		TST		
EPA-9330	100	Non-toxic	>100	Non-toxic	100	Non-toxic	10.34
EPA-9332	100	Non-toxic	92.98	Toxic	50	Toxic	13.3