CALIFORNIA ASSOCIATION of SANITATION AGENCIES

1225 8th Street, Suite 595• Sacramento, CA 95814 • TEL: (916) 446-0388 • www.CASAweb.org

December 21, 2018

Submitted via email to commentletters@waterboards.ca.gov

Felicia Marcus, Chair
State Water Resources Control Board (State Water Board)
1001 I Street, 25th Floor
Sacramento, CA 95814

Subject: CASA Comments on Proposed Toxicity Provisions

Dear Chair Marcus:

The California Association of Sanitation Agencies (CASA) appreciates the opportunity to comment on the proposed Toxicity Provisions within the Water Quality Control Plan for Inland Surface Waters, Enclosed Bays, and Estuaries of California (Toxicity Provisions), as well as the supporting appendices and the Draft Staff Report (Staff Report). CASA is an association dedicated to protecting public health and the environment through effective wastewater treatment. We promote sustainable practices such as water recycling, biosolids management, and renewable energy production. We represent over 100 public agencies in California and focus on advocacy, education, and leadership.

CASA has been working with members and staff of the State Water Resources Control Board (State Water Board) for several years on various approaches to addressing toxicity on a statewide level. There have been many positive changes to the Toxicity Provisions between the 2012 version and the current proposal, and we appreciate the recent workshops and materials designed to clarify the approach staff is recommending to address toxicity for inland surface waters. Unfortunately, many of our previously articulated concerns regarding fundamental elements of the Toxicity Provisions remain unresolved. The required use of a particular species (*Ceriodaphnia dubia*) is at the core of our concerns related to test variability, incorrect determinations of toxicity, and ultimately increased violations based on inaccurate measures of real toxicity. In addition, there are several areas we will discuss below where clarification is needed to ensure the overall regulatory approach is implementable and makes sense. Finally, to the extent that other wastewater association commenters (including the Bay Area Clean Water Agencies, Central Valley Clean Water Association, and Southern California Alliance of POTWs) address other implementation issues not discussed in detail here, we support those comments and incorporate them by reference.

Ongoing Concerns

1. Imposition of Numeric Limits for Whole Effluent Toxicity Testing is Inappropriate and Will Not Improve Environmental Outcomes

In response to the various iterations of the Toxicity Provisions that have been released, CASA has consistently submitted comments noting that numeric objectives and associated numeric limits for

chronic toxicity are both unnecessary and inappropriate. Numeric limits will not result in greater environmental protection than narrative limits with numeric triggers, which have been sufficiently protective of receiving water beneficial uses for more than a decade. This position has not changed.

From an overarching perspective, it is important to remember that whole effluent toxicity (WET) testing is a biological test, not a chemical test. Unlike chemical testing, the effects measured must be compared to effects on unexposed organisms. Further, "toxicity" is not a pollutant per se, but rather a response or condition that results if (presumably) chemicals are present in amounts or combinations deemed harmful to certain organisms.

Perhaps more importantly, no proactive or immediate reactive actions can be performed to prevent or control toxicity based on the violation of a numeric toxicity limit (aka a "test failure") until a contaminant cause has been identified. This is particularly true of POTW dischargers. Until the source of the toxicant has been identified through an appropriate Toxicity Reduction Evaluation (TRE) and/or Toxicity Identification Evaluation (TIE) process, it is impossible to proactively address toxicity because the cause is typically unknown. Thus, numeric limits do nothing more than impose liability on dischargers for circumstances generally outside of their control, based on the presence of unknown chemicals for which there may be no specific objectives, all while the discharger investigates for the cause of the apparent toxicity.

Instead, narrative limits with appropriate numeric triggers are far more suitable given the inherent differences between standard chemical testing and WET testing. Narrative limits are equally protective of the environment while avoiding the additional costs, compliance, and liability issues created for public agencies through imposition of numeric limits. A positive test result for toxicity should be used as a trigger to investigate what specific chemicals or classes of chemicals may have caused the test failure, not to impose fines and penalties. It is because of these features of WET testing, and the difficulties inherent in the implementation of a test that looks for impacts of unknown constituents on living organisms, that the use of numeric objectives and limitations based on WET testing in NPDES permits has been controversial for so long. It is also these underlying reasons why numeric objectives and limits for toxicity are unnecessary, inappropriate, and not well suited to the nature of the tests.

2. Mandating Use of the Test of Significant Toxicity (TST) in the Toxicity Provisions is Inappropriate, Particularly When Applied to the *Ceriodaphnia* Reproduction Endpoint

We continue to have significant concerns with incorporation of the Test of Significant Toxicity (TST) into the Toxicity Provisions for a variety of reasons. First, the TST has never been through a United States Environmental Protection Agency (USEPA) public review-and-comment rulemaking process, which is required when a new method is proposed for NPDES testing. A formal rulemaking must be conducted by the USEPA to incorporate the TST into the promulgated WET methods, before the TST can be required in California for purposes of measuring compliance in NPDES permits.

Other commenters have focused on the legality of using an unapproved method like the TST in this context, so our comments will not get into greater detail on that issue. To the extent that CASA's prior comment letters and the comments of other wastewater associations and entities address

components of the toxicity plan that relate to the imposition of numeric limits and use of the TST, we incorporate those comments by reference.

However, the practical issues associated with the TST as applied to certain freshwater species, notably the *Ceriodaphnia dubia* reproduction endpoint, are of primary concern to CASA. When USEPA first proposed approval for use of the *Ceriodaphnia dubia* reproduction endpoint in NPDES testing, there was litigation over the rule, and the court in the *Edison Electric* case ordered USEPA to amend the test method to include safeguards to protect against identifying non-toxic samples as toxic. USEPA's safeguards included a requirement to run multiple concentrations and look at the response to see if the results made sense. The safeguards also included application of variability criteria. The rationale for this safeguard is that a clearer understanding is gained with more information from running multiple dilutions (e.g. at 20-40-60-80-100% effluent), to see if a valid pattern of increasing effects with increasing concentrations is obtained. The TST as required in the Toxicity Provisions strips away USEPA's safeguards by only looking at 100% effluent (or the Instream Waste Concentration (IWC)). Finally, we have other significant concern with use of the TST in combination with the *Ceriodaphnia dubia* reproduction endpoint, which is discussed in greater detail below.

Core Implementation Concern

Eliminate or Modify the "Reproduction" Endpoint for the *Ceriodaphnia dubia* Chronic Freshwater Method Until Fundamental Testing Issues Are Resolved

CASA's primary and overriding concern is the continued use of the reproduction endpoint for the *Ceriodaphnia dubia* (water flea) chronic freshwater method. This species is the primary source of unacceptable testing variability, and will inevitably lead to increased instances of incorrect determinations of toxicity, and attendant violations, particularly when combined with numeric pass/fail limits and the use of the TST.

This endpoint is particularly troublesome for toxicity testing because the result is derived from counting how many offspring each water flea produces. In the absence of any other contributing factors, this figure can range from 15 to 45 offspring in a non-toxic control, resulting in a range whose upper bound is 300% higher than its lower bound. With such a high inherent variability among non-toxic control treatments, it is exceptionally difficult to reliably identify a 25% percent effect in the reproduction endpoint, which is the management decision currently identified in this draft of the Toxicity Provisions, and to determine the effect is caused by toxicity instead of natural variation.

Compounding this concern, use of the TST exacerbates the problem presented by use of the water flea because the TST strips away essential safeguards found in the promulgated test procedures, such as analyzing the data from multiple dilution tests. Although the Toxicity Provisions require that the dilution series be run, the information obtained from that important step cannot be used.

In addition, research conducted in this area by USEPA, the Southern California Coastal Water Research Project (SCCWRP), and the State Water Board has shown consistently that the high within-test variability associated with this reproduction endpoint results in a higher frequency of toxicity

detections when evaluated using the TST compared to the no observed effect concentration (NOEC), particularly when compared to those observed for the other species and endpoints. In light of these findings and scientific consensus about the limitations of the *Ceriodaphnia* reproduction endpoint, and in conjunction with currently available information suggesting that the other species and endpoints contained in Table 1 **(Toxicity Provisions at p. 6)** may be robust enough for application of the TST in a regulatory context, it is clear that the *Ceriodaphnia dubia* reproduction endpoint is simply not amenable to the TST statistical method.

In the peer-reviewed publication of the State Water Board/USEPA "Test Drive" study,[1] USEPA concluded that although the TST exhibited a similar or lower frequency of toxicity detections than the NOEC approach for most of the test endpoints examined when the mean effect was less than the 25% standard in the regulatory management decision (RMD), "the Ceriodaphnia reproduction… endpoints exhibited a somewhat opposite pattern (Table 1)." The authors further identified that the "chronic *Ceriodaphnia* reproduction endpoint yielded the largest number of tests declared toxic using the TST when the mean effect in the effluent was less than the toxic RMD of 25% (13 of 29 tests or 45%; Table 2)…the proportion of *Ceriodaphnia* tests having this outcome is approximately twice the proportion observed in the entire study (45 vs 23%, respectively)." Thus, while the Staff Report supporting the Toxicity Provisions frequently cites the Test Drive as evidence that the TST works and is reliable overall, the data within the Test Drive demonstrates that the *Ceriodaphnia dubia* reproduction endpoint does not follow that trend, is not reliable, and is in fact highly variable.

This observation was subsequently affirmed and corroborated in a SCCWRP-conducted interlaboratory comparison funded by the Stormwater Monitoring Coalition.[2] In this study, the TST resulted in incorrect determinations of toxicity for half (50%) of the non-toxic blank samples (laboratory dilution water) tested with *Ceriodaphnia dubia*. While recognizing that the reason for this observed toxicity has not been identified, the report recommended that future studies should "conduct the experimental manipulations to identify the source of this inter-laboratory variability" to "confirm this anomalous result." Absent that additional research, and in the light of the scientific unreliability of the *Ceriodaphnia dubia* reproduction endpoint, we think it is inappropriate for the Toxicity Provisions to include numeric toxicity limits based on this measure of toxicity with its unacceptably low precision.

Beyond the scientific literature, recent ambient testing by the Delta Regional Monitoring Program (Delta RMP) also experienced challenges with its *Ceriodaphnia dubia* chronic toxicity testing and data interpretation. Testing included ambient samples with conductivity outside of the organisms' tolerance range; therefore, secondary controls with low conductivity were also tested, as recommended by the Surface Water Ambient Monitoring Program (SWAMP) guidance.[3] Reproduction in these secondary controls was significantly lower than in the standard laboratory control in 14 of the 23 tests. These data suggest that water quality differences between samples or controls can contribute to the observed effects, and recent laboratory testing improved reproduction in low-conductivity laboratory control water with the addition of standard nutrients

[1] Environmental Toxicology and Chemistry, Vol. 32, No. 5, pp. 1101–1108, 2013
[2] SCCWRP Technical Report 956. December 2016. Stormwater Monitoring Coalition Toxicity Testing Laboratory Guidance Document. Kenneth C. Schiff and Darrin Greenstein, Southern California Coastal Water Research Project.
[3] https://www.waterboards.ca.gov/water_issues/programs/swamp/swamp_iq/toxicity.html

(i.e., biotin, sodium selenate, and vitamin B12). Additional monitoring and testing by the Delta RMP will be done to better understand this issue, but it is clear that *Ceriodaphnia dubia* are a sensitive test organism and their reproduction can reflect effects from constituents other than contaminants.

Attached to this letter is a comprehensive white paper that summarizes the findings above in greater detail and utilizes the data from the Test Drive and SCCWRP study to highlight the variability of the *Ceriodaphnia dubia* reproduction endpoint using the TST.[4] The purpose of the analysis in the white paper is to summarize the existing chronic toxicity *Ceriodaphnia dubia* reproduction test data from prior studies that were 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 necessary and advisable to develop recommendations for reducing observed variability. While these studies have been somewhat limited in size, together they indicate a lack of confidence in the accuracy of the test results for the *Ceriodaphnia dubia* reproduction endpoint when the TST is used. Because of this problem, we believe that the *Ceriodaphnia dubia* reproduction endpoint should not be included as the basis for numeric limits in the Toxicity Provisions at this time.

As always, CASA is willing to partner with the State Water Board and others to work on resolution of these real issues going forward, including working together collaboratively on Toxicity Provisions that solve any real toxicity issues. CASA is also interested in exploring a partnered study with industry experts, the State Water Board, and other agencies including dischargers, to resolve the issues related to the *Ceriodaphnia dubia* reproduction endpoint. This study could be used to inform future use of this species as an indicator of toxicity, and to reduce test interferences.

However, any application of a regulatory limit associated with this species should not be considered until the problems identified by USEPA, SCCWRP, and others are addressed, and the solutions can be appropriately implemented. CASA and other stakeholders are in the process of developing an alternative approach to address this issue, and we look forward to working with State Water Board members and staff.

Additional Implementation Issues

1. The Provisions Should Clarify That Routine Acute Toxicity Testing is Not Generally Expected to Occur When Chronic Testing is Already Occurring

We appreciate that the Toxicity Provisions specify that Regional Water Quality Control Boards (Regional Boards) are not required to conduct a Reasonable Potential Analysis (RPA) for acute toxicity. Specifically, the provisions state that a RPA is "not required" for both categories of POTW dischargers, but that the Regional Boards "may require POTW dischargers to conduct a REASONABLE POTENTIAL analysis for acute toxicity" and shall document that decision in the NPDES fact sheet or equivalent document. **(Provisions at p. 14 / Staff Report at p. 16)**

[4] Larry Walker Associates, Inc. 2018. *Ceriodaphnia dubia* Short-term Chronic Reproduction Test: Understanding the Probability of Incorrect Determinations of Toxicity in Non-toxic Samples. White Paper prepared for California Association of Sanitation Agencies. November.

For POTWs, chronic toxicity testing is generally as protective of beneficial uses as both acute and chronic toxicity testing. From previous discussions with staff, it is our understanding that circumstances would be exceedingly rare where a Regional Board would require a POTW already subject to routine monitoring for chronic toxicity to also conduct acute testing. The Staff Report supports this understanding and makes clear that POTWs should only be required to run acute toxicity under limited circumstances, such as when there are very high dilution rates or where an adequate chronic toxicity test does not exist. The economic analysis supports this proposition as well, as the chronic testing cost analyses assume no acute testing is taking place and the cost estimates do not account for the costs of acute testing in its "sample" facilities analysis. (**See Table 9-1, Staff Report at Page 245**). However, we believe additional language must be added into the Toxicity Provisions themselves to reflect the conclusions in the Staff Report, and to delineate the anticipated circumstances where testing for both acute and chronic toxicity might be ordered by a Regional Board. Thus, we request additional language in the Toxicity Provisions to clarify that, in general, when chronic testing is being performed, acute testing is not simultaneously required. CASA and other stakeholders are in the process of developing language that reflects this approach, and we look forward to working with State Water Board staff on this issue.

2. The Provisions Should Clarify That Regional Boards Should Generally Reduce Monitoring Frequency During a Toxicity Reduction Evaluation (TRE)

We appreciate that the Toxicity Provisions specify that the Regional Boards may approve a temporary reduction in the frequency of routine monitoring for dischargers conducting a TRE. (**Provisions at p. 18, Staff Report at p. 96-97).** This approach makes sense as the discharger typically would perform extensive testing during a TRE that would make chronic testing for compliance purposes redundant. In addition, if there is an ongoing toxicity issue during the TRE, it does not make sense for a discharger to continue to receive routine monitoring compliance "fails" that could result in violations while it is simultaneously conducting the TRE, which is the only remedial measure available to potentially address the toxicity. Finally, as the Staff Report acknowledges, reducing routine monitoring while a discharger is conducting a TRE "allow[s] the discharger to concentrate resources on finding and eliminating the source of toxicity." **(Staff Report at p. 98)** Accordingly, we request additional language in the draft Toxicity Provisions themselves to clarify that, in general, Regional Boards and their staff should grant a temporary reduction in the frequency of routine monitoring for dischargers conducting a TRE. CASA and other stakeholders are in the process of developing language that reflects this approach, and we look forward to working with State Water Board staff on this issue.

3. The Provisions Should Clarify That Compliance Data Prior to Adoption of the Toxicity Plan Can be Used in Requests for Reduced Monitoring Frequency

We appreciate that the provisions include potential reduced routine monitoring schedules for chronic toxicity testing in specified circumstances. Specifically, the Toxicity Provisions allow the Regional Boards to approve a reduction in the frequency of routine monitoring when during the "prior five consecutive years" the MDEL and MMEL have not been exceeded and the toxicity provisions in the applicable NPDES permit have been followed. **(Provisions at p. 17)**

Unfortunately, the current language is written in such a way as to effectively prohibit consideration of positive compliance data gathered at any time before adoption of the new Toxicity Provisions. As noted above, while a specific reference is made to exceedances of the MDEL and MMEL, the MDEL and MMEL do not currently exist (and have not existed in previous years) in most permits, and therefore it would be impossible for agencies with existing, long records of positive compliance data and no prior toxicity issues to be granted a reduced monitoring frequency in the first five years after the Toxicity Provisions are implemented. We understand that this may be a drafting error and not necessarily the intent on the part of the Board to prohibit consideration of prior years' data, and we look forward to working with staff to develop language that addresses this issue.

4. The Provisions Should Address Implementation Issues Relating to the Number of Routine Monitoring Tests Conducted Within a Calendar Month

We appreciate that the Water Board has attempted to address the practical issues related to conducting multiple toxicity tests in a limited window with some of the changes to the Toxicity Provisions (e.g. allowing start dates to be varied among the regulated community and cross over months). Under this draft of the Toxicity Provisions, it still will be logistically difficult to comply in circumstances where an entity is required to conduct three (3) full tests within a calendar month. As has been acknowledged by State Water Board staff, initiating three tests within a thirty-day period is theoretically possible, but very difficult. Comments submitted by the Bay Area Clean Water Agencies (BACWA) provide additional detail regarding the logistics and difficulties of these tests, including one example where it may be impossible for an agency (SFPUC) to conduct three tests in a calendar month when there are wet-weather events. Thus, we concur with and reiterate BACWA's proposed amendments that would provide for an alternative approach to initiating three tests in a specified period.

5. The Economic Analysis Understates Actual Testing Costs, Fails to Account for Increased Costs Associated with Potential Acute Testing, and Fails to Account for the Increased Likelihood of Incorrect Determinations of Toxicity Resulting in Violations

CASA is concerned that the economic analysis contained in the Staff Report supporting the proposed Toxicity Provisions is inaccurate in several respects and thus understates the true costs of implementing these provisions. Specifically, the cost estimates **(Staff Report at pp. 241 – 249 and Table 9-1)** do not reflect the real costs of toxicity tests at contract laboratories. As articulated by BACWA in their written comments, POTWs in the Bay Area pay approximately $3,000 per sample, far above estimates in the economic analysis. In addition, the cost estimating methods do not include the costs of collecting and shipping samples to contract laboratories.

Also problematic, as noted above, is that the economic analysis references, but does not adequately articulate, the potential cost to dischargers if a Regional Board were to impose monthly acute toxicity routine monitoring requirements. The analysis notes this amount could be as much $9,468 per year, yet the "Potential Incremental Costs for Sample Facilities" table excludes the costs of acute testing entirely. If the State Water Board were to articulate in the provisions the relative rarity of the need for routine acute testing where chronic already taking place, as we suggest above, this estimation may be more defensible. However, as written, the economic analysis should at

minimum identify more accurate examples of what costs would be if Regional Boards were to impose acute testing requirements.

Finally, as articulated in greater detail by others in their written comments, the economic analysis entirely fails to account for the potential cost of increased violations from imposition of numeric limits and the TST. Staff has acknowledged that imposition of the Toxicity Provisions likely will lead to an increase in toxicity violations at wastewater facilities, yet nowhere in the economic analysis is the concomitant financial impact of such violations acknowledged or quantified. Both Regional Board enforcement actions and third-party lawsuits impose significant costs on local agencies, and these need to be estimated and articulated in the economic analysis.

We appreciate the opportunity to comment and look forward to discussing these issues with you in early 2019. If you have any questions or concerns, please do not hesitate to reach out to me directly at (916) 446-0388 or alink@casaweb.org. Thank you.

Adam D. Link
Director of Operations

cc: Karen Mogus, SWRCB
Rebecca Fitzgerald, SWRCB
Zane Poulson, SWRCB

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 Valley[1]. 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 TUc[a]			
Pimephales promelas (survival)	832	4 (0.5%)	2 (3.4%)
Pimephales promelas (growth)	834	20 (2.4%)	15 (25.4%)
Ceriodaphnia dubia (survival)	818	9 (1.1%)	7 (11.8%)
Ceriodaphnia dubia (reproduction)	820	131 (16.0%)	38 (64.4%)
Selenastrum capricornutum (growth)	835	76 (9.1%)	29 (49.2%)
Central Valley POTWs with Chronic Toxicity Trigger > 1 TUc[b]			
Pimephales promelas (survival)	128	2 (1.6%)	1 (14.3%)
Pimephales promelas (growth)	128	2 (1.6%)	1 (14.3%)
Ceriodaphnia dubia (survival)	137	0 (0.0%)	0 (0.0%)
Ceriodaphnia dubia (reproduction)	138	17 (12.3%)	3 (42.9%)
Selenastrum capricornutum (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 TUc 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 TUc, 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 TUc and were POTWs with either advanced secondary or tertiary level treatment.

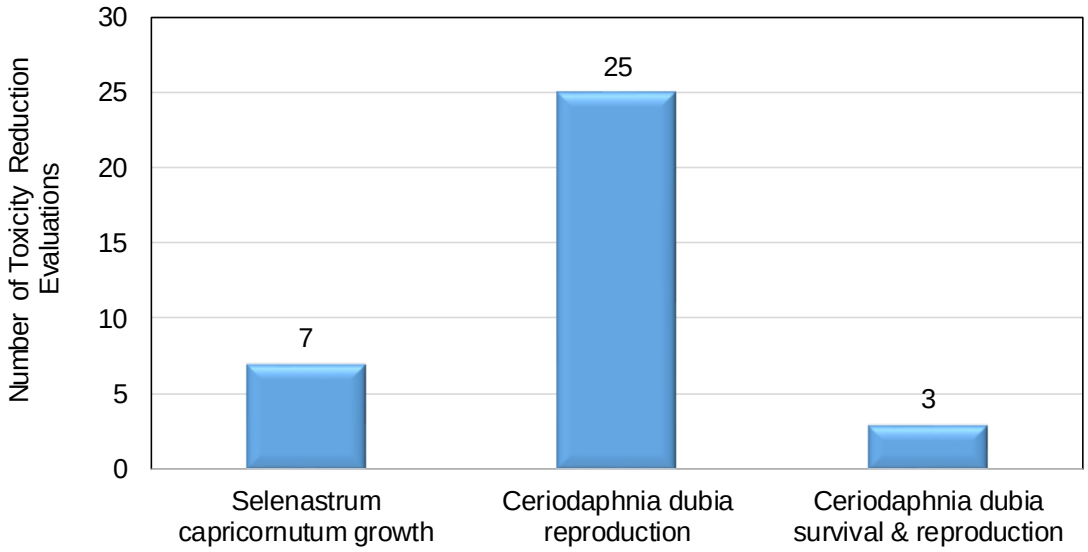

Figure 1. Central Valley POTW Toxicity Reduction Evaluation Test Organism and Endpoint

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 EDTA[2] (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 2011[3] 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 blind[4] analyses on blank[5] samples. The three studies described above provided the *Ceriodaphnia dubia* reproduction toxicity raw data that were analyzed in this white paper (EPA[6], SMC[7], WC[8]). 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
EPA-9337	100	Non-toxic	>100	Non-toxic	100	Non-toxic	9.95
EPA-9338	100	Non-toxic	>100	Non-toxic	100	Non-toxic	11.98
EPA-9340	100	Non-toxic	>100	Non-toxic	100	Non-toxic	0
EPA-9341	100	Non-toxic	>100	Non-toxic	100	Non-toxic	28.51, 16.60, 30.21, 27.23
EPA-9344	100	Non-toxic	>100	Non-toxic	100	Non-toxic	0
EPA-9349	100	Non-toxic	>100	Non-toxic	100	Non-toxic	5.52
EPA-9350	100	Non-toxic	>100	Non-toxic	50	Toxic	22.03
EPA-9356	100	Non-toxic	>100	Non-toxic	100	Non-toxic	7.47
EPA-9367	100	Non-toxic	>100	Non-toxic	50	Toxic	16.02
EPA-9371	100	Non-toxic	>100	Non-toxic	100	Non-toxic	1.4
EPA-9376	100	Non-toxic	>100	Non-toxic	100	Non-toxic	12.75
EPA-9379	100	Non-toxic	>100	Non-toxic	100	Non-toxic	11.24
EPA-9381	100	Non-toxic	>100	Non-toxic	100	Non-toxic	4.53
EPA-9382	100	Non-toxic	>100	Non-toxic	100	Non-toxic	6.55
EPA-9384	100	Non-toxic	>100	Non-toxic	100	Non-toxic	24.86, 39.88, 25.43
EPA-9402	100	Non-toxic	>100	Non-toxic	100	Non-toxic	5.73
EPA-9409	100	Non-toxic	>100	Non-toxic	100	Non-toxic	0
EPA-9410	100	Non-toxic	>100	Non-toxic	100	Non-toxic	2.82
EPA-9425	100	Non-toxic	>100	Non-toxic	100	Non-toxic	19.71, 20.00
EPA-9429	100	Non-toxic	>100	Non-toxic	100	Non-toxic	6.96
EPA-9432	100	Non-toxic	>100	Non-toxic	100	Non-toxic	9.04
EPA-9436	100	Non-toxic	>100	Non-toxic	100	Non-toxic	21.33
EPA-9439	100	Non-toxic	>100	Non-toxic	100	Non-toxic	14.81
EPA-9445	100	Non-toxic[a]	>100	Non-toxic	100	Non-toxic	5.08
EPA-9446	100	Non-toxic	>100	Non-toxic	100	Non-toxic	19.37, 10.81
EPA-9450	25	Toxic	15.88	Toxic	0	Toxic	13.92, 5.67, 65.98, 86.60, 78.87
SMC1-A	100	Non-toxic	>100	Non-toxic	50	Toxic	23.32

Laboratory ID	Result (percent)						Applicable %Effect
	NOEC		IC25		TST		
SMC1-B	100	Non-toxic	>100	Non-toxic	50	Toxic	13.46
SMC1-C	100	Non-toxic	44.44	Toxic	25	Toxic	24.40, 29.17, 35.12
SMC1-E	25	Toxic[b]	>100	Non-toxic	100	Non-toxic	7.97, 13.67, 13.44
SMC1-F	100	Non-toxic	>100	Non-toxic	100	Non-toxic	0
SMC1-G	100	Non-toxic	>100	Non-toxic	100	Non-toxic	0
SMC1-H	100	Non-toxic	>100	Non-toxic	100	Non-toxic	9.33
SMC1-I	50	Toxic	45.89	Toxic	25	Toxic	28.57, 59.52
SMC1-J	50	Toxic	72.08	Toxic	50	Toxic	39.62
SMC1-Referee	25	Toxic	41.67	Toxic	25	Toxic	36.00, 82.67
SMC2-A	100	Non-toxic	74.16	Toxic	50	Toxic	27.62, 19.52, 23.81, 30.48
SMC2-Ad	100	Non-toxic	>100	Non-toxic	100	Non-toxic	11.63
SMC2-B	100	Non-toxic	>100	Non-toxic	100	Non-toxic	10.22
SMC2-Bd	100	Non-toxic[c]	>100	Non-toxic	100	Non-toxic	11.35
SMC2-C	100	Non-toxic	>100	Non-toxic	100	Non-toxic	28.76, 19.74
SMC2-Cd	100	Non-toxic	>100	Non-toxic	100	Non-toxic	0
SMC2-G	50	Toxic	65.19	Toxic	25	Toxic	30.89, 14.66, 48.69
SMC2-Gd	100	Non-toxic	6.059	Toxic	12.5	Toxic	45.79, 41.58, 32.63, 23.16
SMC2-I	50	Toxic	79.33	Toxic	50	Toxic	18.61, 42.59
SMC2-Id	50	Toxic	63.78	Toxic	50	Toxic	64.79
SMC2-J	50	Toxic	83.33	Toxic	50	Toxic	29.22
SMC2-Jd	100	Non-toxic	>100	Non-toxic	100	Non-toxic	9.32
WC01	25	Toxic	31.85	Toxic	25	Toxic	46.11, 54.44
WC02	100	Non-toxic	>100	Non-toxic	50	Toxic	18.27, 17.77
WC03	100	Non-toxic	-	-	100	Non-toxic	0
	-	-	>100	Non-toxic	-	-	-
WC04	100	Non-toxic	>100	Non-toxic	100	Non-toxic	15.67
WC05	50	Toxic	69.85	Toxic	50	Toxic	61.14
WC06	50	Toxic	37.86	Toxic	12.5	Toxic	18.14, 12.56, 36.74, 65.58
WC07	100	Non-toxic	>100	Non-toxic	100	Non-toxic	18.9
WC08	100	Non-toxic	>100	Non-toxic	100	Non-toxic	5
WC09	100	Non-toxic	>100	Non-toxic	100	Non-toxic	11.39
WC10	100	Non-toxic	11.61	Toxic	<6.25	Toxic	17.74, 50, 2.42, 27.42, 26.61
WC11	50	Toxic	46.46	Toxic	25	Toxic	20.99, 56.91
WC12	100	Non-toxic	>100	Non-toxic	100	Non-toxic	0

Laboratory ID	Result (percent)						Applicable %Effect
	NOEC		IC25		TST		
WC13	100	Non-toxic	12.14	Toxic	<6.25	Toxic	7.75, 26.06, 42.25, 38.73, 23.24
WC14	100	Non-toxic	>100	Non-toxic	100	Non-toxic	1.54
WC15	100	Non-toxic	>100	Non-toxic	100	Non-toxic	7.37
WC16	100	Non-toxic	>100	Non-toxic	100	Non-toxic	0
WC17	100	Non-toxic	>100	Non-toxic	100	Non-toxic	14.56, 22.98
WC18	100	Non-toxic	>100	Non-toxic	100	Non-toxic	23.32
WC19	100	Non-toxic	>100	Non-toxic	100	Non-toxic	0
WC20	50	Toxic	>100	Non-toxic	100	Non-toxic	20.06
WC21	100	Non-toxic	94.01	Toxic	50	Toxic	26.61
WC22	100	Non-toxic	>100	Non-toxic	100	Non-toxic	2.03
WC23	25	Toxic	3.344	Toxic	<6.25	Toxic	46.72, 62.30, 45.90, 75.41, 61.48
WC24	12.5	Toxic	50.8	Toxic	25	Toxic	19.13, 24.5, 56.04
WC25	100	Non-toxic	43.91	Toxic	25	Toxic	33.04, 33.91
WC26	6.25	Toxic	36.66	Toxic	25	Toxic	20.87, 18.38, 31.15, 41.12

[a] The result was originally reported as 50%, but both the pMSD and %Effect were below the lower bound of 13. An individual 100% concentration test was performed, and the results demonstrated no toxicity.

[b] The TUc was originally reported as 8, but both the pMSD and %Effect were below lower bound of 13. An individual 25% concentration test was performed and shown to be non-toxic, so the toxic concentration was changed to 50% or 4 TUc.

[c] The TUc was originally reported as 2, but both the pMSD and %Effect were below lower bound of 13. An individual 100% concentration test was performed, and the results demonstrated no toxicity.

Gray shading indicates the test results which did not meet the control acceptability criteria. These results were not included in the analysis.

Toxicity Data Results

Chi-Square Comparison Test

The Chi-Square test statistically compares the differences between multiple aspects of multiple tests and shows whether one aspect does not fit the pattern of the others. A Chi-Square test was used to identify whether differences within one dataset could be statistically attributed to a particular factor, i.e., whether more frequent fail results are statistically attributable to the use of a particular test.

The results of a Chi-Square test on the number of passing results reported by each of the three studies (EPA, SMC, WC) showed that the differences in pass rates between studies were not significant at the current sample sizes (**Table 5**). Increasing the sample size is not possible as these studies have been completed. Because differences were not significant, the study results can be combined for analysis (as in **Table 9**), and do not need to be analyzed separately (as in **Table 10**).

Table 5. Results of a Chi-Square Test Comparing Each Study

Study		NOEC	IC25	TST	All
EPA	# Pass Results	27	26	24	77
	Expected #	27.5	26	23.5	
SMC	# Pass Results	14	12	10	36
	Expected #	13	12	11	
WC	# Pass Results	14	14	13	41
	Expected #	15	14	13	

	Chi-Square Value	P-Value	Significant?
Pearson	0.261	0.992	No
Likelihood Ratio	0.261	0.992	No

The results of a Chi-Square test on pass/fail results (binary, not quantitative) showed that the differences between the three analyses (NOEC, IC25, and TST) were not significant at the current sample size, as the p-value associated with the analysis was greater than 0.05. The results of the test are shown in **Table 6**. A larger sample size would be necessary to determine whether the results are statistically significant. When each existing dataset is used three times (213 results for each analysis, or 639 results total), the p-value drops to 0.03, which shows statistical significance. Therefore, not more than 213 sample results would be necessary to provide results with statistical significance.

Table 6. Results of a Chi-Square Test Comparing Each Analysis

Analysis		Fail	Pass	All
NOEC	Count	16	55	71
	Expected count	20	51	
IC25	Count	19	52	71
	Expected count	20	51	
TST	Count	24	47	71
	Expected count	20	51	
All	Count	59	154	213

	Chi-Square Value	P-Value	Significant?
Pearson	2.297	0.317	No
Likelihood Ratio	2.286	0.319	No

Toxicity Data Fail Rate

The "fail rate" is the number of blank samples that showed toxicity out of the total number of samples. As blank samples are expected to show no toxicity, the fail rate is expected to be zero. However, when working with living organisms, there are uncertainties involved (e.g., organisms which fail to reproduce regardless of water toxicity) and these are accounted for by a maximum acceptable fail rate of 1 in 20 (5%)[10]. Therefore, even when using blank water, one out of 20 toxicity tests may acceptably fail. The actual number of failed tests reported during the three studies were compared with this acceptable rate.

There were 71 data points (TUc results) available from the combined three studies of blank water. The study fail rates were compared with the maximum acceptable fail rate for each of the three analyses. It was found that for each analysis type, the fail rate of the combined data set exceeded the acceptable fail rate (**Table 7**).

Table 7. Fail Rates of Blank Toxicity Test Analyses

Analysis	Sample size	Fail rate	Acceptable fail rate
NOEC	71	23%	5%
IC25	71	27%	5%
TST	71	34%	5%

When the three analyses were analyzed separately by study, the results of one analysis (the NOEC analysis of the EPA study) had a fail rate below the acceptable fail rate, but all other individual studies' tests had fail rates above the acceptable rate (**Table 8**).

Table 8. Fail Rate of Toxicity Test Analyses by Study

Study	Analysis	Sample size	Fail rate
EPA	NOEC	28	3.6%
	IC25	28	7.1%
	TST	28	14%
SMC	NOEC	22	36%
	IC25	22	45%
	TST	22	55%
WC	NOEC	21	33%
	IC25	21	33%
	TST	21	38%

[10] USEPA, 2000. *Understanding and Accounting for Method Variability in Whole Effluent Toxicity Applications Under the National Pollutant Discharge Elimination System Program* (EPA 833-R000-003), June.

T-Tests and Power Analyses

Statistical differences between two datasets can be determined using a paired t-test when each dataset contains the same number of samples. The statistical significance of the differences determined (mean paired difference) can be estimated using the p-value. A p-value below 0.05 is considered to represent statistical significance. The confidence level, or number of samples required to achieve a higher confidence level, can be determined using a power analysis. A confidence level or power of 80% has been considered to be "sufficient" by the statistician Jacob Cohen (1988)[11].

The statistical differences between the results of each blank analysis and the results of an "ideal" analysis (containing all passing results, or TUc set equal to 1) were found to be significant (p-values below 0.05). However, only the TST analysis showed a confidence level over 80% (at 93%). As shown in **Table 9**, the number of samples necessary for a confidence level of 80% was found to be 88 for the NOEC analysis and 126 for the IC25 analysis[12].

Table 9. Results of Comparison Between Blank Analyses and "Ideal" Analyses

Analysis	Mean Paired Difference	Standard Deviation	p-value	Significant difference?	Confidence Level	Sample size	Samples Required for 80% Confidence
NOEC	0.620	2.045	0.013	Yes	71%	71	88
IC25	0.495	1.963	0.037	Yes	55%	71	126
TST	0.930	2.225	0.001	Yes	93%	71	47

When the results of the three studies were analyzed separately (i.e., each one compared with an "ideal" all passing dataset), only the results from the SMC study TST analysis showed a difference from the "ideal" results with both statistical significance and confidence over 80% (shown in **Table 10**). There were insufficient data for confidence in the results of the other tests and study results. However, the Chi-Square test results indicate that separating the analyses by study is not necessary, as the study results are not significantly different from each other.

[11] Cohen, Jacob. *Statistical Power Analysis for the Behavioral Sciences*, Second Edition. 1988.
http://www.utstat.toronto.edu/~brunner/oldclass/378f16/readings/CohenPower.pdf
https://effectsizefaq.com/category/type-ii-error/

[12] As noted above, the maximum acceptable fail rate for toxicity tests is 5% (1 in 20 samples). To show 80% confidence in the differences between a theoretical dataset with this fail rate and an "ideal" all-passing dataset, when the failures are set to the smallest possible values (2 TUc for NOEC and TST and 1.01 TUc for IC25), 159 samples are required.

Table 10. Results of Comparison Between Blank Analyses and "Ideal" Analyses, by Study

Study	Analysis	Mean paired difference	Standard Deviation	p-value	Significant difference?	Confidence Level	Sample size	Samples Required for 80% Confidence
EPA	NOEC	0.107	0.567	0.326	No	16%	28	223
	IC25	0.192	1.000	0.319	No	16%	28	215
	TST	0.643	2.831	0.240	No	21%	28	155
SMC	NOEC	0.545	0.912	0.011	Yes	76%	22	24
	IC25	0.984	3.275	0.174	No	27%	22	89
	TST	1.182	1.708	0.004	Yes	87%	22	19
WC	NOEC	1.381	3.514	0.087	No	40%	21	53
	IC25	0.387	0.692	0.019	Yes	68%	21	28
	TST	1.048	1.802	0.015	Yes	72%	21	26

When the results of each analysis were compared with each other using the combined data set, the IC25 and TST test results showed differences that were statistically significant. However, as shown in **Table 11**, over 130 data points would be necessary to show this with at least 80% confidence. The differences between NOEC and IC25 and between NOEC and the TST were not statistically significant and would require hundreds of samples to achieve 80% confidence.

Table 11. Results of Comparison Between Blank Analyses

Comparison	Mean paired difference	Standard Deviation	p-value	Significant difference?	Confidence Level	Sample size	Samples Required for 80% Confidence
NOEC versus IC25	0.125	2.605	0.687	No	7%	71	3411
NOEC versus TST	-0.310	2.453	0.291	No	18%	71	494
IC25 versus TST	-0.435	1.757	0.041	Yes	54%	71	130

Conclusions based on statistical analysis

None of the individual studies had a sufficient number of samples to determine the rate of incorrect determinations of toxicity for the *Ceriodaphnia dubia* short-term chronic reproduction test with adequate statistical confidence (a minimum confidence level of 80% is necessary).

The results of a Chi-Square test showed no significant difference between the results from the three studies; therefore, the study results can be combined for this analysis. Out of 71 samples from all three studies, the TST analysis had a 34% fail rate (24 non-passing results), the IC25 analysis had a 27% fail rate (19 non-passing results), and the NOEC analysis had a 23% fail rate (16 non-passing results). These all exceed the acceptable fail rate of 1 out of 20, or 5%, which demonstrates the issue of variability of *Ceriodaphnia dubia* short-term chronic reproduction test results.

The statistical difference between the results of the TST analysis and the results of an "ideal" analysis was significant with a confidence level of 93%. A comparison of the IC25 versus TST

test results also showed a statistically significant difference, although at a confidence level of only 54%. This suggests that the TST analysis may have a fail rate statistically above that of the two other types of analysis and above the acceptable fail rate.

QUALITATIVE ANALYSIS

As noted previously, a number of questions have been raised about the data generated by the three studies in question.

The quality of the "blank" samples used in the studies that showed toxicity was questioned. It was suggested that they were incorrectly formulated or contaminated, particularly those samples that yielded large effect levels. The performance of the laboratories producing toxic results from the analysis of blank samples was also questioned, e.g. perhaps they did not properly perform the tests in accordance with promulgated standard methods.

If, in fact, these concerns were valid, this would call into question the raw data used in the statistical analysis presented above and impact the conclusions of that analysis. Unfortunately, the questions regarding the validity of the data produced by these studies are unanswerable. This leaves the overall issue of variability of *Ceriodaphnia dubia* short-term chronic reproduction test results unresolved.

One approach to attempt to resolve this issue would be to perform a tightly controlled study to quantify the rate of incorrect determinations of toxicity which addresses the questions and criticisms raised about the prior studies. Such a study would include testing on an array of synthetic dilution water blanks which would allow an examination of various dilution water compositions used by analytical laboratories in accordance with the flexibilities allowed under standard methods. Testing would be performed by twenty or more certified laboratories, using their standard dilution waters as controls. Dilution series would be run which would allow statistical analysis using the TST, NOEC and point estimate procedures. A referee laboratory would oversee the preparation of the samples to be tested and would certify the quality of the samples. An independent expert panel would review the methods, data analysis and reports prepared to describe study results and findings.

FINDINGS

The findings from the overall assessment are that available studies and data are insufficient to conclusively resolve outstanding questions regarding the variability of *Ceriodaphnia dubia* chronic WET test results and the evaluation of incorrect determinations of toxicity. While the statistical analysis of available data indicates that the incidence of incorrect determinations of toxicity was unacceptably high, questions regarding the validity of the raw data impact our ability to rely on those results. It is concluded that performance of a properly designed and implemented "blank" study with sufficient statistical power and oversight would have the potential to resolve concerns by identifying measures to reduce the occurrence of incorrect determinations of toxicity to an acceptable level while not impacting the ability to detect toxic samples.

References

Central Valley Clean Water Association (CVCWA), 2018. *Toxicity Special Study: Phase I Study Report.* Preliminary Draft. October.

Denton, D. L., Diamond, J. and Zheng, L. 2011. *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.

Moore, Timothy F., Canton, S. P., and 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.

Schiff, K.C., Greenstein, D., 2016. *Stormwater Monitoring Coalition Toxicity Testing Laboratory Guidance Document*. Southern California Coastal Water Research Project, Technical Report 0956.

United States Environmental Protection Agency (US EPA), 1991. *Technical Support Document for Water Quality-based Toxics Control*. United States Environmental Protection Agency, Office of Water. Washington, D.C. EPA 505/2-90-001. March.

US EPA, 2000. *Understanding and Accounting for Method Variability in Whole Effluent Toxicity Applications Under the National Pollutant Discharge Elimination System Program*. United States Environmental Protection Agency, Office of Wastewater Management. EPA-833-R-00-003. June.

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.

US EPA, 2011. *Whole Effluent Toxicity Test Drive Analysis of the Test of Significant Toxicity (TST)*. United State Environmental Protection Agency, EPA Region 9, Sacramento, CA. July.

Warren-Hicks, W.J., Parkhurst, B.R., 1995. *Issues in Whole Effluent Toxicity Test Uncertainty*. pp. 180-189. In: Grothe, D.R., Dickson, K.L., Reed-Judkins, D.K. (eds.). Whole Effluent Toxicity Testing: An Evaluation of Methods and Prediction of Receiving System Impacts. SETAC Press, Pensacola, FL.