Southern California Stormwater Monitoring Coalition

Potential Impacts of Toxicity Testing Variability on Waterbody Impairment Status

DRAFT White Paper

Prepared by:
J. Michael Trapp, PhD and Karin Peternel
Michael Baker International
Carlsbad, California

This page intentionally left blank

Abstract

A recent SMC funded intercalibration study for toxicity has revealed significant variability in *C. dubia* reproduction toxicity results. This variability has created concerns about the potential to affect toxicity results in real world environmental samples, and thus the possible listing or non-listing of waterbodies as impaired on the federal 303(d) list. This paper evaluates eight representative sites listed for toxicity within SMC member agency jurisdiction. Variability observed in runoff, Cu spiking, and lab control samples during the intercalibration were applied to individual toxicity tests to determine if their result status could change from pass to fail or fail to pass. Results show that under the lower treatment +/- 22% at least 50% of tests would change results and under the higher date treatment +/- 46% as many as 76-91% of samples could change results. When these potential changes were applied to 303(d) listing criteria, 7 of the 8 sites could change their listing status. These revelations present significant management implications. As a result it is suggested that the use of the *C. dubia* reproduction toxicity be discontinued or efforts undertaken to improve the consistency of its results and understand the roots of the observed variability.

Introduction

One of the primary goals of the Southern California Stormwater Monitoring Coalition (SMC) is to conduct the science based project to support stormwater management. SMC's research efforts include developing regional stormwater monitoring infrastructure that is expected to lead to increased comparability among the numerous monitoring programs and laboratories in Southern California. Stormwater monitoring samples collected from water bodies throughout California are analyzed at various laboratories using standardized procedures. Results from these analyses have been used for management activities and to drive regulatory actions. To help ensure full confidence in the data produced by these labs the SMC has funded intercalibration studies the most recent of these is a toxicity intercalibration. This study identified poor comparability in toxicity testing between laboratories, as well as within individual laboratories.

Stormwater management agencies in Southern California spend nearly $1 million per year on toxicity sampling alone. As a result of the variability observed in the intercalibration study significant concerns were raised among SMC member organizations that ramifications could result in potential false positive or negative toxicity results. The consequences of which could result in the incorrect status of waterbodies on the Clean Water Act (CWA) 303(d) listings or the initiation of unnecessary toxicity reduction evaluation (TRE) investigations. Both of these potential outcomes are costly and time consuming.

This paper aims to determine if the variability observed in the laboratory intercalibration results could impact changes to the passing or failing of real world samples in a non-biased investigation. Further the variability on individual tests are explored in regulatory samples to determine if changes in results could potentially cause changes to 303(d) impairment status.

Background

Regulations and Policy

The goal of the CWA, passed in 1972, is "to restore and maintain the chemical, physical, and biological integrity of the Nation's waters" (33 U.S.C. Sec. 1251(a)). CWA Section 303(d) requires each state to develop, update, and submit to the U.S. Environmental Protection Agency (EPA) for approval a list of water bodies not meeting water quality standards or "beneficial uses." This List of Impaired Waters, submitted biennially, is referred to as the 303(d) list. The guidance by which the California Water Boards comply with the listing requirements of Section 303(d) is the Water Quality Control Policy (2015), a standardized approach used to meet the overall goals of achieving water quality standards and maintaining beneficial uses.

Regulatory protection of beneficial uses is carried out through water quality objectives established in each of California's nine Regional Water Quality Control Board's (RWQCB) Basin Plans. These Basin Plans contain narrative toxicity objectives that require all water to be maintained free of toxic substances in concentrations that produce detrimental physiological responses in humans, plants, terrestrial animals, and aquatic organisms. In 2005, the Policy for Implementation of Toxics Standards for Inland Surface Waters, Enclosed Bays, and Estuaries of California (SIP) was adopted to provide a mechanism to implement the water quality criteria established in the California Toxics Rule. The SIP requires that the Regional Water Boards determine compliance with narrative chronic toxicity objectives using EPA methodology to estimate the potential effects on the survival, growth, and reproduction of species, where compliance is determined by conducting tests on at least one species of aquatic plant, one invertebrate, and one vertebrate (EPA 1994). According to the EPA testing methods and 303(d) policy for toxicity, if any of three tested species exhibit toxicity, the site will be placed on the 303(d) list. The 2012 EPA 303(d) list includes 260 impaired water bodies for toxicity in California, of which 85 fall within the watersheds of the SMC member organizations located in RWQCBs Los Angeles, Santa Ana, and San Diego.

Intercalibration

A successful chemistry intercalibration study led by SMC (Gossett et al. 2004) brought attention to the need for a similar study and steps to be taken toward comparability for aquatic toxicity testing. There is an expectation of repeatability between independent test results from the same laboratory, and reproducibility between test results obtained from different laboratories. However, in response to different outcomes from laboratories and known uncertainties in testing protocols, SMC conducted its laboratory intercalibration study to assess comparability among and between labs that conduct toxicity testing, resulting in the Stormwater Monitoring Coalition Toxicity Testing Laboratory Guidance Document (Schiff and Greenstein, 2016). The interlaboratory intercalibration study conducted two rounds of testing using eight toxicity laboratories where a significant level of variability was found in the results. Reasonable comparability was found with the two marine species, yet poor comparability was found for the two freshwater species, *Ceridaphnia dubia* and *Hyalella azteca*. Table 1 reveals that Round 1 results showed poor comparability in both *H. azteca* survival and *C. dubia* survival *and* reproduction between laboratories, while Round 2 showed better comparability for *H. Azteca* survival and *C. dubia* survival, but poor comparability for *C. dubia* reproduction. Additionally, a blind sample of laboratory dilution water, prepared in accordance with standardized guidance and expected to be nontoxic, was found to be toxic by some of the participating labs. This prompted a review of test methods and standardization of protocols prior to the second round of interlaboratory testing. The results of the second round of testing also identified toxicity in the dilution water sample and relatively low comparability of the laboratories between testing events. Intercalibration exercises have found that although specific guidance and procedures exist, there is still variability in results within and between laboratories. Additional intercalibrations have been recommended, specifically for the *C. dubia* reproduction test, to assess sources of variability in both stormwater and lab dilution water.

Table 1. Summary of laboratory comparability scoring from SMC Toxicity Testing Laboratory Guidance Document (Schiff and Greenstein, 2016), Table ES-1.

Lab	Ceriodaphnia Survival		Ceriodaphnia Reproduction		Hyalella Survival	
	Round 1	Round 2	Round 1	Round 2	Round 1	Round 2
A	Moderate	High	Very High	Low	Low	High
B	Very High	High	Moderate	High	Low	High
C	Low	High	Low	High	Low	Very High
E	Moderate	-	Moderate	-	-	
F	Moderate	High	Moderate	Low	Low	Very High
G	High	-	High	-	-	-
H	Low	-	Low	-	-	-
I	High	Moderate	High	Low	Moderate	Very High
J	Low	High	Low	Low	High	Very High

Variability

The EPA test methods (EPA 2004) detail the potential for increased variability in results due to potential interferences in sampling and analysis. This, in addition to the variability seen between laboratories in the toxicity intercalibration study, brings into question the laboratory results used for water bodies placed on the 303(d) list for toxicity as the pollutant, as a false positive or negative could affect the resulting regulatory action or inaction. Enough uncertainty has been shown between and among laboratories (where specific method options or interpretations are left to each laboratory) to increase concerns about toxicity testing's inherent variability. These uncertainties can have a profound effect on test results, and therefore, on the regulatory actions placed on a water body.

Definition of Evaluation

It is postulated that the variability seen in the interlaboratory intercalibration exercises may have extended to samples from real world sites that were tested and subsequently could have resulted in inaccurate results. If true, this could have an impact on the potential listing as a 303(d) impaired water body. The interlaboratory intercalibration exercises, as well as some of the standardizations to address variability, were conducted and reported in Schiff and Greenstein (2016). However, the majority of sites placed on the 303(d) list for toxicity resulted from laboratory testing that occurred significantly earlier than this study, and thus the inconsistencies were likely present.

To check the impact of the potential variability, the original datasets that were used as the basis for listing selected water bodies can be compared to the variability seen in the toxicity intercalibration study. This check will help to determine if similar conditions of variability occurred and if a potential false positives or negatives result could have affected the site's listing status.

A review of the data for sites in the San Diego region that were 303(d) listed for toxicity shows that most were listed based on a 2002-2003 ecological condition assessment of water bodies across the region. Results show that there were positive toxicity tests as a result of *C. dubia,* as well as a significant number of sites toxic to the algae species *Selenastrum capricornutum*. Thus it is possible that even if *C. dubia* tests resulted in false positives, algae toxicity could still result in 303(d) inclusion.

The need to minimize the variability in toxicity testing, and ensure sites placed on the 303(d) list be consistent as possible, is being recognized by the broader water quality community. This recognition extends to the State Board, where plans exist to revise its toxicity policy to that of 'provisions' within the next year. These revisions include statewide standards for sampling, a focus on quality of laboratories, and additionally the application of the test for significant toxicity (TST) statistical test (pers. comm., Z. Paulsen SWRCB). This study offers an opportunity to view variability in the context of real world regulatory outcomes and will help decision makers understand its ramifications.

To achieve these ends, this study will review the California 303(d) list to find a subset of representative sites to conduct the evaluation of potential impacts. Individual toxicity test results from these sites will subjected to variability based on thresholds typical to those observed during the intercalibration. Tests will then be rescored using the state toxicity criteria. The resulting results for each site will then be reevaluated to determine if any changes to 303(d) list impairments for toxicity could result.

Toxicity Listing Review

A full evaluation of every impaired water body in the state would not be realistic, so a criteria was developed to identify candidate sites for careful inspection. The 2012 CWA 303(d) list of 4,852 sites was reviewed for water bodies listed for toxicity and inclusion in this assessment using the following set of criteria:

- Included a toxic impairment (260 water bodies).
- Contained within SMC member regions of Los Angeles (Region 4), Santa Ana (Region 8), and San Diego (Region 9). The resulting list consisted of 19 water bodies listed for toxicity in the Los Angeles Region; two listed for unknown toxicity in Santa Ana; and 29 listed for toxicity in San Diego. (Sediment Toxicity was not considered)
- The SMC-associated reaches were then sorted by the total number of impairments to focus on sites with lower chances of complications resulting from multiple stressors. This resulted in toxicity alone at four sites, toxicity plus one other pollutant at seven sites, and toxicity with two to three other pollutants at four sites.

The initial list of water bodies selected for potential reassessment of data is shown in Table 2.

Table 2. Fifteen water bodies listed in 303(d) for toxicity and limited other constituents.

Region	Water Body	Listing 1	Listing 2	Listing 3	Listing 4
4	Santa Clara River Reach 1 (Estuary to Hwy 101 Bridge)	Toxicity			
9	Jamul Creek	Toxicity			
9	Poggi Canyon Creek	Toxicity			
9	Santa Ysabel Creek (above Sutherland Reservoir)	Toxicity			
9	Encinitas Creek	Toxicity	Selenium		
9	Loma Alta Creek	Toxicity	Selenium		
9	Moro Canyon Creek	Toxicity	Selenium		
9	Oso Creek (lower)	Toxicity	Selenium		
9	Poway Creek	Toxicity	Selenium		
9	Rose Creek	Toxicity	Selenium		
9	Santa Margarita River (Upper)	Toxicity	Phosphorus		
9	Dana Point Harbor	Toxicity	Copper	Zinc	
9	Segunda Deshecha Creek	Toxicity	Turbidity	Phosphorus	
8	San Diego Creek Reach 2	Unknown Toxicity	Sedimentation /Siltation	Nutrients	Indicator Bacteria
8	Lake Elsinore	Unknown Toxicity	Sediment Toxicity	Nutrients	Organic Enrichment

Site Descriptions

Monitoring data was obtained for the initial list of sites from the CEDEN database (www.ceden.org) and reviewed for the species type and the frequency at which each species or endpoint resulted in a positive test for toxicity. This list was narrowed down to 11 sites for which *C. dubia* was tested, and further reduced to 8 sites for which C. dubia reproduction resulted in a failed toxicity test. A brief description of these sites follows.

- Jamul Creek and Poggi Canyon Creek (toxicity) in the San Diego region underwent an ecological assessment in 2003 that indicated that both sites exceeded aquatic life thresholds for several water chemistry constituents, but there were more exceedances at Poggi Creek (4) than Jamul (3), and more pesticides were detected at Poggi (5) than at Jamul (2).
- The 3-mile segment of Encinitas Creek (toxicity, selenium) in Carlsbad, California was delisted for phosphorus due to flaws in the original listing (2004). Although analyzed for Benthic Community Effects, diazinon, nitrogen, turbidity, and total dissolved solids, it was not listed for any of those pollutants.
- Santa Ysabel Creek (toxicity) is located above the Sutherland Reservoir and is the largest tributary to the San Dieguito Creek, which drains to the Pacific Ocean. It is considered a reference site for monitoring the rest of the watershed, and generally has the fewest number of aquatic life threshold exceedances in the monitoring program.
- Loma Alta Creek (selenium, toxicity) in Oceanside, California is the natural drainage for about 6,400 acres of heavily urbanized land that runs along Oceanside Boulevard, including residential, commercial, and industrial uses. Its waters feed into the Loma Alta Slough, or estuary, which is the final 1,600 feet of the watershed before it reaches the ocean. Loma Alta Creek is known to have persistent algal blooms.
- Oso Creek (toxicity, selenium) is an approximately 13.5-mile (21.7 km) tributary of Trabuco Creek in southern Orange County. Draining about 20 square miles (52 km^2) in a region north of the San Joaquin Hills and south of the Santa Ana Mountains, the creek is Trabuco Creek's largest tributary and is part of the San Juan Creek drainage basin. The creek is channelized and polluted along much of its length.
- Santa Margarita River (toxicity, phosphorus) in San Diego and Riverside Counties underwent an ecological assessment using data collected from 1998 to 2005. Toxicity was moderate, although samples from all sites were toxic to the freshwater algae *Selenastrum capricornutum* on at least one sampling date. Fish tissues from the downstream Santa Margarita River site showed no evidence of impact. Bioassessment samples indicated that large areas of the watershed are in poor ecological condition, yet other areas of the watershed are in fair or good condition.
- Rose Creek (toxicity, selenium) in San Diego drains to Mission Bay. It was diverted and channelized in the first half of the 20th century and now enters Mission Bay through an artificial channel further east. Although diazinon was measured in the creek, water quality standards were not exceeded.

Data Analysis

In order to assess whether the variability seen in the intercalibration study could be quantified and compared to the results seen in the actual data from sites that have been placed on the 303(d) list, an estimate of the level of variability seen in the intercalibration study was conducted. This variability would then be applied to the datasets obtained from CEDEN.

C. dubia reproduction results are evaluated for toxicity on the basis of percent effect. The percent effect (closer to zero = nontoxic, closer to 100 = toxic) is as follows:

Percent Effect = (Mean control – mean sample) / mean control

Therefore as sample means get closer to the control means, the closer to zero the Percent Effect becomes and the sample is considered nontoxic, as control and treatment reproduction rates do not significantly differ. When smaller sample means are compared to the control means, the percent effect becomes closer to 100, and the sample is considered toxic. In the CEDEN and SWAMP datasets used to calculate toxicity for 303(d) listings, a two-tiered approach is applied to evaluate toxicity. The first is a hypothesis-testing approach comparing the organisms in the samples to the responses from the controls using a t-test statistical comparison. The second tier is a comparison to a threshold value that is 20% less than the control response. A site is considered toxic if the t-test is significant above an alpha value (0.05), and if the percent effect value is larger than the evaluation threshold. An analysis of the CEDEN data shows that the percent effect value used is 100 minus the calculated percent effect value (referred to as Percent Control) with a threshold value of 80. Therefore, any Percent Control less than 80 does not pass the second tier.

The magnitude of variability in sample results and laboratory dilution water based on the intercalibration study were used as the basis for determining the ranges of effects in this study. The values chosen for variability were based on the percent effect concentrations in the Laboratory water to apply to the individual test controls. The percent effect concentrations for Copper Spiking and Runoff were chosen to represent potential variability in environmental samples. The percent effect concentrations summarized in Figure 1 of the SMC Toxicity Testing Laboratory Guidance Document were used to calculate the standard deviation from the mean in Laboratory, Copper Spiking, and Runoff treatments. The three levels of variability two for the sample and one to the control are described below:

1) Adjust sample values to reach a percent effect of +/– 23% based on Runoff treatment

2) Adjust sample values to reach a percent effect of +/- 46% based on Copper Spiking treatment

3) Adjust control values to reach a percent effect of +/- 22% based on Laboratory control

These levels of variability were applied to the individual toxicity test data sets for the 8 sites where *C. dubia* reproduction toxicity experienced failure(s). Results from the application of the variability in the control and samples were then evaluated to determine if the treatment impacted the positive or negative result of the test and thus the potential status on the 303(d) list.

Findings

A comprehensive review of the data obtained from CEDEN shows the number of pass/fails for each species and endpoint for the 8 prioritized water bodies where *C. dubia* was tested. The analysis reported here primarily focuses on the *C. dubia* reproduction endpoint as it showed the greatest variability in the inter-laboratory study and thus likely the largest potential for impacting impairments. The eight sites containing *C. dubia* reproduction toxicity tests reported as passes and failures will be evaluated as a demonstration of the potential impact of the observed variability (Table 3). These 8 sites represent 12 failing and 21 passing individual sets of the individual *C. dubia* reproduction toxicity tests.

Table 3. Species and endpoints used as basis for toxicity testing for 303(d) listing of water bodies. All samples were taken between 2001 and 2003.

Species		*H. azteca*		*C. dubia*		*C. dubia*		*S. capricornutum*	
Endpoint		Survival		Reproduction		Survival		Count	
	WATER BODY	*Pass*	*Fail*	*Pass*	*Fail*	*Pass*	*Fail*	*Pass*	*Fail*
1	Poggi Canyon			2	1	3	0	0	3
2	Santa Ysabel			4	2	6	0	1	5
3	Moro Canyon	2	0	1	1	2	0	0	4
4	Poway Creek			3	1	4	0	0	4
5	Loma Alta Creek			1	3	4	0	0	4
6	Encinitas Creek			2	2	3	1	3	1
7	Oso Creek	1	0	2	1	2	1	0	4
8	Rose Canyon			2	1	3	1	1	3

Ceriodaphnia dubia Reproduction

Data treatments based on the three variabilities described above were applied to each of the individual *C. dubia* reproduction toxicity tests where one of the results (sample or control) are altered and the other remains the original value to determine if there is a resulting status change.

Assuming these variabilities could be positive or negative two potential sets of scenarios result: Fail to Pass (Scenario 1), and Pass to Fail (Scenario 2). The first set of scenarios would result in failures becoming passing results by bringing the two results into greater agreement by either increasing the sample result or decreasing the control. Table 4 shows the resulting number of passing and failing tests for each of the demonstration sites for the data treatments.

Table 4. Resulting number of toxicity test results for C. dubia reproduction under different Scenario 1 variability conditions.

Scenario 1 Data Variability Treatments (*C. dubia* Reproduction)	Original Sample		Sample				Control	
			22%		46%		23%	
	Pass	*Fail*	*Pass*	*Fail*	*Pass*	*Fail*	*Pass*	*Fail*
Poggi Canyon	2	1	2	1	2	1	2	1
Santa Ysabel	4	2	4	2	6	0	4	2
Moro Canyon Creek	1	1	2	0	2	0	0	2
Poway Creek	3	1	4	0	4	0	3	1
Loma Alta Creek	1	3	2	2	4	0	1	3
Encinitas Creek	2	2	2	1	3	0	1	2
Oso Creek	2	1	3	0	3	0	1	2
Santa Margarita	4	0	4	0	4	0	4	0
Rose Canyon	2	1	3	0	3	0	2	1

Table 5 summarizes the number of individual test results that would change status and shows that in the first scenario when the lower sample variability treatment (22%) is applied, 7 of the 12 *C. dubia* reproduction toxicity test failures met the pass requirements and thus would have changes the test result. 11 of the 12 test failures could have passed based on the higher variability treatment (46%) and thus would have changed the results. Similarly, when the variability is applied to the control water (23%) 7 of the 12 failures could have passed changing the test result.

The second set of scenarios would increase the relative difference of the control and sample and thus could result in a passing test failing by lowering the sample count results or raising the control results. Table 6 shows the resulting test outcomes for these treatments. When the sample variability is applied, 11 of the 21 results changed from pass to fail with the lower degree of variability applied (22%) and 16 would change based on the higher degree of variability (46%). The 23% treatment for the control values resulted in 11 shifts from pass to fail.

Table 5. Resulting number of test result toxicity status changes for C. dubia reproduction under different variability conditions. Based on 8 demonstration sites with 12 failed tests and 21 passed tests.

Summary of Test Status Changes Based on Data Variability Treatments	**Samples**		**Control Water**
	22%	**46%**	**23%**
Scenario 1 - Fail to Pass (n=12)	7	11	7
Scenario 2 - Pass to Fail (n=21)	11	16	11

Table 6. Resulting number of toxicity test results for C. dubia reproduction under different Scenario 2 variability conditions.

Scenario 2 Data Variability Treatments (*C. dubia* Reproduction)	Original Sample		Sample 22%		Sample 46%		Control 23%	
	Pass	***Fail***	***Pass***	***Fail***	***Pass***	***Fail***	***Pass***	***Fail***
Poggi Canyon	2	1	1	2	0	3	1	2
Santa Ysabel	4	2	1	5	0	6	1	5
Moro Canyon Creek	1	1	0	2	0	2	0	2
Poway Creek	3	1	1	3	1	3	1	3
Loma Alta Creek	1	3	1	3	1	3	1	3
Encinitas Creek	2	2	1	2	1	2	1	2
Oso Creek	2	1	1	2	1	2	1	2
Santa Margarita	4	0	3	1	0	4	3	1
Rose Canyon	2	1	1	2	1	2	1	2

These results show the potential impact to the listing status of water bodies, particularly as even under the lower variability scenarios more than 50% of the test results could be overturned in either direction, and under the higher scenario 76 – 91% of the samples could be over turned (Figure 1). These demonstration sites can be considered analogous to potential impacts on all water bodies listed and non-listed for impairments as the pass/fail results could potentially be effected in either direction.

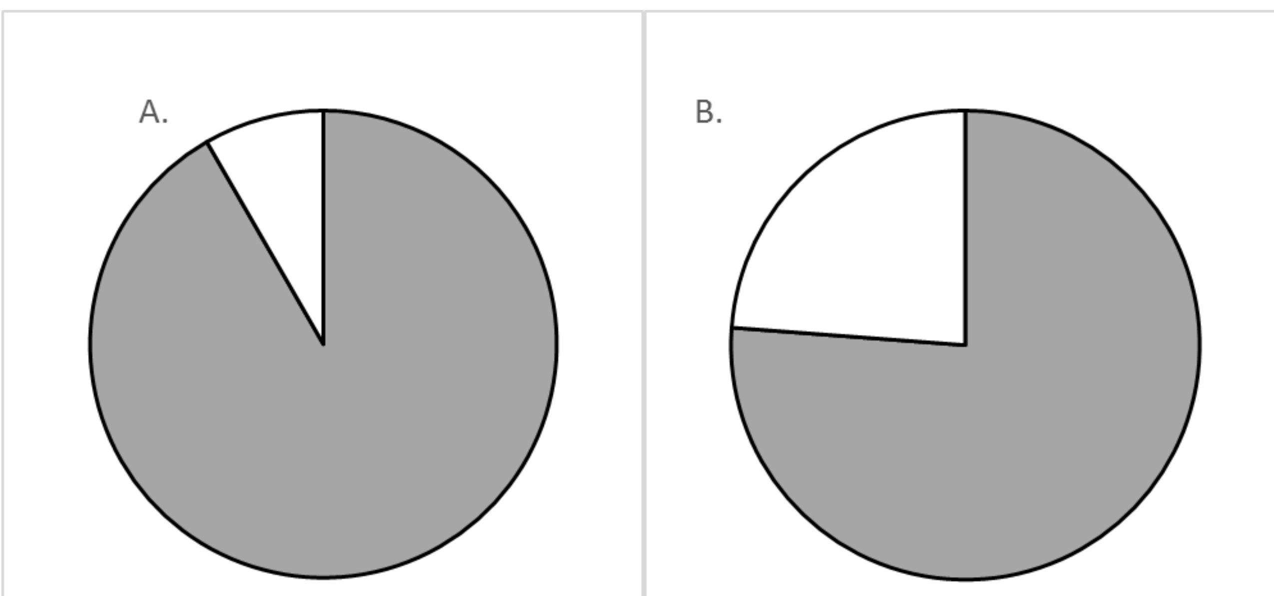

Figure 1. Pie charts showing the number of individual toxicity tests that could change from (a) fail to pass and (b) pass to fail based on the intercalibration Cu spiking variability applied to representative environmental samples where white remain the same and the shaded portions change test results.

Applying the results of these variations to the 303(d) listing criteria reveal that not only the individual test results can change but also the 303(d) listing status. A demonstration of this is Encinitas Creek, in that with all three of the Scenario 1 treatments the number of failing events would be reduced to 1. The reduction in *C. dubia* reproduction failures leaves one event failure (for *C. dubia* survival and *S. capricornutum* toxicity) and thus the listing criteria of 2 failing events is not met. If *S. capricornutum* toxicity, which was detected at all eight sites is not considered (discussed below), 7 of the 8 sites (all but Loma Alta Creek) would not meet the two failing event criteria to be listed as impaired under the Scenario 1 treatments. Thus, the potential variability observed in *C. dubia* reproduction presents a significant threat to compromise real world outcomes of individual toxicity tests as well as impairment listings.

Other Potential Variability in Species/Endpoints

Hyalella Azteca

The first round of intercalibration tested found low comparability in *H. azteca* samples. There was only one failure at three sites in this dataset (Jamul Creek), whereas it should be noted that data was collected and analyzed between 2001 and 2003 when this species was used infrequently. A review of these results only calls this one water body into question. However, the fact that Jamul Creek passed the two *C. dubia* tests, and the small margin by which it failed the *H. azteca* test, do bring potential variability into question. The one failure had a survival percent mean of 70 percent and a p=0.001. The passing sample had a mean of 78 percent and a p=0.051 (tests were run with alpha of less than 5%). The samples were taken approximately 3 weeks apart in spring 2003, with the failing sample taken during a higher base flow. Jamul Creek is noted as potentially impacted based on *H. azteca* potential variability. The importance of determining if this site is subject to variability is due to the fact that it is only 303(d) listed for toxicity, no other constituents.

Selenastrum capricornutum

The algae species *Selenastrum capricornutum* was not included in the intercalibration study; however, it was found toxic in all but one site in this analysis. Thus, a closer look at the impacts of potential variability similar to the intercalibration species should be considered.

An assessment of the ecological health of the 11 San Diego Region hydrologic units conducted in 2002-2003 included a review of water chemistry, toxicity, and fish. Toxicity was found at nearly all the sites, predominantly due to this algae species, with occasional toxicity to *H. azteca* or *C. dubia*. Based on the variability in the intercalibration study for other species and the frequency of toxicity observed for *S. capricornutum,* intercalibration for algal species should be considered. This is particularly important because multiple species can be required to be tested for toxicity, and toxic results from only one species are required for listing. Therefore, many of these water bodies would have been on the 303(d) list regardless of whether they showed toxicity to *C. dubia* or *H. azteca*.

Known interferences that can result in *C. dubia* variability includes contamination by glassware, testing equipment, sample handling, pathogenic or predatory organisms in the dilution water, food, and especially pH drift. These same interferences can be found in *S. capricornutum* testing, with additional interferences such as the frequency of shaking the algae samples (2x/day or continuous), which can have a significant outcome on test results due to suspension and growth; type of cell count applied (flourometer, microscope); internal competition by native algae; and conductivity. Effluent-dominated systems can be highly affected by interferences due to the polymers and clarifiers used in the coagulation process that can produce false toxicity. The presence of metals can also have differing effects on the results. These additional confounding issues may require additional guidance for laboratories as well.

Discussion

The review of this study's analysis shows that a significant number of the individual toxicity test could potentially change their results based on the variability observed in the intercalibration study. The worst of these tests was the *C. dubia* reproduction test in which as high as 91% of the passing results exhibited possible changes to failures and 76% of the failing results changing to passing results based on the intercalibration Cu spiking variability. Further this study has revealed that based on these changes that seven of the eight water bodies considered in this study contained individual tests that could have changed their toxic status based on intercalibration variability. These seven water bodies could have potentially not been included on the 303(d) list based on *C. dubia* alone, although algal toxicity was observed in all of these sites.

These results present an important potential impact to regulatory impairment status. This is particularly impactful when considering that listing a water body only requires two samples with toxicity exceedances and delisting requires 28 samples with no more than two exceedances. This means that based on the variability of 22 – 46% used in this study there is a high probability that almost any sample is susceptible to switching results and thus of at least one sample used for delisting could be effected. This highlights the potential expenditure of significant costs to assess whether these sites are toxic, or if the toxicity was a result of interferences in testing or variability in laboratory methods or laboratories. Therefore, sample collection, laboratory analysis, and variability must be assessed to ensure that sites are correctly listed, or can reasonably be delisted. This variability makes it difficult to determine if toxic results are due to the sensitivity of the different species, an artifact due to variability in testing procedures, or perhaps the sites are truly exhibiting toxicity. If the toxicity was due to methodology or laboratory variability, re-evaluation of these sites using the updated standardized protocols could be considered as part of the anticipated Total Maximum Daily Load (TMDL) creation within the next 5 years. This reevaluation, however, would have the caveat of being reflective of current conditions and not the time of original sampling.

This uncertainty affects the management and regulatory actions on water bodies throughout California, not just to SMC organizations. This should be a primary concern to dischargers and regulatory agencies that rely on toxicity testing for management responses such as permit compliance, toxicity identification evaluations, or TMDLs. According to the Regional Board (SWAMP 2011), in monitoring conducted between 2001 and 2010, greater than 50 percent of the collection sites have shown some degree of toxicity in fresh water. Correlation analyses and toxicity identification evaluations (TIEs) used to determine the likely causes of surface water toxicity suggested that toxicity to invertebrate test species was most often caused by pesticides. Results of statewide water toxicity tests with the three standard EPA test species show that more samples were toxic to fish larvae and algae than to water fleas (*C. dubia*) (Water Board, 2011). The vast majority of studies in California have included TIE evidence that demonstrated water toxicity is caused by pesticides. It is of interest that the findings have significant algae toxicity results, indicating possible herbicides or metals present, as well as pyrethroid pesticides. Because most 303(d) listings due to toxicity are based on an "unknown source," and no TIE or TMDL has been conducted, it is hard to attribute cause.

Outcomes

The continued variability in observed in the intercalibration study has been shown here to have the potential to directly affect the listing or non-listing of waterbodies as impaired for toxicity on the 303(d) list. This presents difficult implications and costs to SMC members and the broader storm water community. These revelations present a clear need to address this variability as the status quo represents uncertain results. The community has two choices to discontinue the use of this test or improve them.

The choice to discontinue the use of *C. dubia* reproduction toxicity means that results of other toxicity tests would need to carry more weight in future tests as well as the historical record. *C. dubia* does represent the most common species utilized for toxicity by SMC members over the past 20 years and thus its continued use would be valuable. This could be accomplished through the survival test which using lessons learned in the intercalibration study has shown significant increase in comparability.

The second choice to improve this test offers opportunities to understand the factors causing the observed variability and thus interpret historical results. This could potentially allow for confidence in historical records or the culling of samples suspected to be affected. Similar efforts to those undertaken following the first round of the intercalibration study to reduce variability in *C. dubia* survival and *H. Azteca* can be applied to *C. dubia* reproduction toxicity.

References

Stormwater Monitoring Coalition Laboratory Guidance Document. 2010. Gossett, R. and Schiff, K. 3rd Southern California Coastal Waters Research Project, Technical Report 615.

Stormwater Monitoring Coalition Laboratory Guidance Document. 2004. R Gossett, D Renfrew, K Schiff. Southern California Coastal Water Research Project, Technical Report 420. Westminster, CA.

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

Policy for Implementation of Toxics Standards for Inland Surface Waters, Enclosed Bays, and Estuaries of California (2005). http://www.waterboards.ca.gov/board_decisions/adopted_orders/resolutions/2015/020315_8_amendment_clean_version.pdf

Toxicity in California Waters. 2011. Surface Water Ambient Monitoring Program, State of California State Water Resources Control Board, Sacramento, California.

Toxicity Test Reference: U.S. EPA. 1994. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to freshwater organisms. Third edition. U.S. EPA Environmental Monitoring Systems Laboratory, Cincinnati, Ohio. EPA/600/4-91-002.

Water Quality Control Policy for Developing California's Clean Water Act Section 303(d) List. (2015). State of California State Water Resources Control Boards/workplans/rb3_methods_paper.pdf