If you have questions or comments regarding the draft work plan, please contact your representative on the Stakeholder Advisory Committee:

https://www.waterboards.ca.gov/water_issues/programs/state_implementation_policy/docs/c_dubia_study_stakeholder_meeting_1_stakeholder_committee_members.pdf"

Conceptual Workplan: Quality Assurance Evaluation of the *Ceriodaphnia dubia* Reproduction Test

Prepared by the Southern California Coastal Water Research Project

For the State Water Resources Control Board

Agreement #19-278-0780

Draft - March 18, 2021

Contents

Introduction

The California State Water Board recently adopted Toxicity Provisions, which include numeric effluent limitations for aquatic toxicity. Implementation of the monthly median effluent toxicity limitation for the *Ceriodaphnia dubia* (*C. dubia*) reproduction test has been delayed until January 1, 2024 for some dischargers as specified in the Toxicity Provisions. During this delay, the State Water Board has committed to a study, in collaboration with stakeholders and laboratories, to investigate factors that can lead to test variability. High test variability could lead to low statistical confidence in assessments of toxicity or non-toxicity.

The *C. dubia* test has been used for many decades and the currently promulgated protocol for the *C. dubia* reproduction test was established nearly 20 years ago (USEPA 2002). The promulgated method allows laboratories some flexibility when implementing certain laboratory techniques. For example, there are multiple options for preparing culture water or food. In some instances, the promulgated method is silent on test techniques, leaving laboratories to use their best professional judgement. To be clear, the *C. dubia* method is well-established and validated. However, these small differences between laboratories may lead to intra- or interlaboratory variability, which could influence test results.

Previous studies have assessed the variability of the *C. dubia* test results within and among laboratories. In the early 2000s, an interlaboratory comparison exercise performed by the EPA found that only 22 out 122 *C. dubia* chronic tests did not meet test acceptability criteria for survival or reproduction (USEPA 2001). The invalid tests were confined to 10 out of 34 participating laboratories. The study reported intra- and inter-laboratory coefficients of variation for the IC25 values of effluent and receiving water split samples at 17% and 28%, respectively for the reproduction endpoint. More recently, a smaller intercomparison exercise was conducted in California to evaluate the reliability of *C. dubia* chronic test for stormwater toxicity evaluation (Schiff and Greenstein 2016). Of the nine labs that tested split samples of dilution water, three were considered “low” comparability. Lack of comparability among a minority of laboratories testing split samples of dilution water was also identified by others (Moore et al. 2000; Diamond et al. 2008). Most recently, routine testing data generated by eight California-accredited laboratories was examined by Fox et al. (2019) and results indicated that intra-laboratory variability, particularly in controls, influenced whether test samples would be identified as toxic. Either reducing the between replicate variability or increasing the number of replicates improved lab performance.

The State of California Environmental Laboratory Accreditation Program (ELAP) accredits all laboratories conducting analysis for regulatory compliance purposes, including the *C. dubia* test. Currently, there are 17 ELAP accredited laboratories for conducting the *C. dubia* test in California (Appendix A). Accreditation is based on the demonstration that laboratories are following the testing protocols, properly training their staff, keeping accurate records, and demonstrating they can meet data quality objectives for reference toxicant and performance evaluation test samples. While this process demonstrates that a laboratory capably performs a test, it does not address test variability between laboratories or differences in lab techniques that are allowed by the protocols.

Objective of this study

The objective of this study is to evaluate sources of variability in the *C. dubia* reproduction test conducted by California-accredited laboratories and identify potential laboratory technique guidance and/or recommendations to: (a) improve the consistency of the execution of the C. dubia test method

to achieve improved precision (i.e., as measured by the control coefficient of variation) within each testing laboratory; and (b) improve the consistency and comparability of *C. dubia* test results among testing laboratories.

The study will seek to answer the following questions:

1) What are the *C. dubia* chronic reproduction toxicity test laboratory techniques used by Environmental Laboratory Accreditation Program (ELAP) accredited laboratories in the state of California?
2) How does variability in control reproduction and/or reference toxicant response in the *C. dubia* chronic reproduction toxicity test compare amongst laboratory technique differences used by ELAP accredited laboratories?
3) Does standardizing differences in the *C. dubia* chronic reproduction toxicity test laboratory techniques reduce variability in control reproduction and/or reference toxicant response?

Based on the results of this study, a list of suggested best-practices for the *C. dubia* reproduction test laboratory techniques will be developed.

Just as importantly, this study is not designed to address or quantify false negative or false positive rates for detecting toxicity from known or unknown samples.

Approach

A five-step design will be used to address the study objectives:

1) Create a Governance structure
2) Analyses of historical data and lab techniques provided by ELAP-accredited laboratories to identify sources of variability
3) Dose-response testing to optimize lab technique(s) and recommended lab technique guidance
4) Evaluation of the revised lab technique guidance via split-sample testing by accredited laboratories
5) Final report with final recommended guidance

These steps are sequential with each one informing the details of the next.

The first step will create a two-tiered governance structure to ensure transparency and technical rigor. One tier will be a Stakeholder Committee comprised of representatives from sectors potentially impacted by the study results. The second tier will be an independent Expert Science Panel comprised of scientists with different areas of expertise and no potential conflict with study results.

The second step will be comprised of two subtasks. The first subtask will create an inventory of lab techniques used by ELAP accredited laboratories. The inventory will elucidate the level of comparability and differences in test implementation. The second subtask will collect historical testing data from the ELAP accredited laboratories. The historical data will be analyzed to quantify the level of variability within and among laboratories. Finally, the differences in lab techniques will be compared to the lab test result variability. The goal of this data analysis is to indicate which lab techniques might be accounting for the observed variability. An optional subtask is to collect new data to assess intra- or interlaboratory variability using split samples, to confirm possible sources of test variation that historical data does not provide. This split sample testing will be dependent on the availability of additional funds.

The third step will focus on dose-response testing procedures to quantify the variability of lab techniques identified by the historical analysis in Step 2. There may potentially be many variable-inducing differences in lab techniques from Step 2. So, a prioritization of which techniques require dose-response testing will need to be evaluated by the Stakeholder Committee and Expert Science Panel formed in Step 1. The results of the dose-response testing will culminate in a draft recommendation on the lab technique guidance that produces the least variability in test results.

The fourth step will verify that draft recommendation from Step 3 does reduce variability, both within and among laboratories. To accomplish this verification, split samples will be distributed to ELAP accredited laboratories for testing, using the draft recommended guidance for laboratories to follow.

The fifth step will complete the study including final recommended guidance, results from the split sample testing in Step 4, and a final report.

Detailed Methods

Create a governance structure

This task will create a multi-tiered governance structure to ensure transparency and technical rigor. Figure 1 illustrates the governance structure.

The ultimate decision-making body is the State Water Resources Control Board (SWRCB). The SWRCB staff is charged with making the final recommendations to the SWRCB about the need for implementing any recommended guidance on lab technique for the *C. dubia* reproduction test.

The project facilitator is the Southern California Coastal Water Research Project (SCCWRP). SCCWRP will be responsible for project design (including this workplan), project implementation (including interacting with ELAP accredited laboratories), and project reporting (including the final report).

One tier will be a Stakeholder Committee comprised of representatives from sectors potentially impacted by the study results. A list of the Stakeholder Committee sectors, and their designated representatives is included in Table 1. The goal of the Stakeholder Committee is to ensure there is a formal mechanism for input and feedback to the project design, planning, conclusions, and recommendations. While not a decision-making body, the Stakeholder Committee is a crucial piece of governance. The Stakeholder Committee will review any study design, results, and recommended guidance first, prior to the Expert Science Panel, to make sure the study is rooted in applicable and achievable guidance.

The second tier will be an independent Expert Science Panel comprised of scientists with different areas of expertise and no potential conflict with study results. A list of the Expert Science Panel disciplines and the designated scientists are listed in Table 2. The Expert Science Panel is a decision-making body and is tasked with reviewing the study design and approving the Workplan, reviewing intermediate work products and refining the study design, and reviewing the recommended changes to lab techniques and providing a consensus opinion on the final method guidance.

Figure 1. Project governance structure

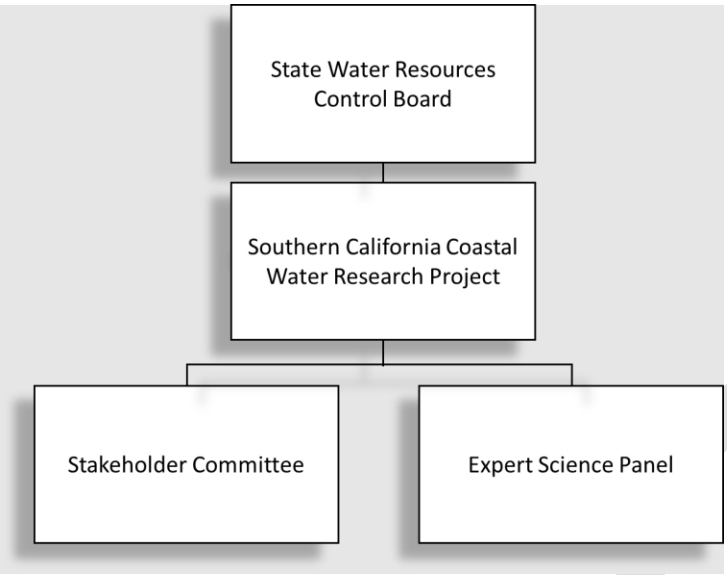


Table 1. List of Stakeholder Committee sectors and designated representatives

SECTOR	**Representative**
Local Government	John Wheeler (SWRCB)
Federal Government	Debra Denton (EPA Region IX)
Regional Water Board; NPDES Permitting	Veronica Cuevas (RWQCB4)
Wastewater Agencies	Mitch Mysliwiec (LWA representing CASA)
Stormwater Agencies	Jian Peng (CASQA)
Agricultural Coalition	Sarah Lopez (Central Coast, CCWQP)
Non-Governmental Organization	Kaitlyn Kalua (CA Coastkeeper)
Private Laboratories	Jeff Miller (Aqua-Science Laboratories)
Public Laboratories	Josh Westfall (Los Angeles County Sanitation Districts)

Table 2. List of Expert Science Panel disciplines and designated scientists.

Expertise	**Representative**
Freshwater Toxicology- Academic	Robert Brent (James Madison University)
Freshwater Toxicology- Government	Teresa Norberg-King (USEPA)
Freshwater Toxicology- Industry	Howard Bailey (Nautilus Environmental)
Biostatistics	A. John Bailer (Miami University)
Data Quality Objectives for WET testing	Leana Van der Vliet (Environment Canada)

Inventory of lab techniques and analyses of historical data

Inventory of Laboratory techniques

To date, no one has created an inventory of laboratory techniques used by ELAP accredited laboratories for the *C. dubia* reproduction test. This task will create that inventory, some of which will be easily accessible and expected based on the promulgated method, and some of which is expected to be more difficult. The inventory will focus on four types of lab techniques including:

- Dilution water
- Food
- Culturing
- Technician training and laboratory's level of experience

Table 3 distills the factors within each of the four types of data to be inventoried, and how they will be collected. For the most part, these categories correspond to parts of the promulgated method that allows for flexibility in their lab techniques.

Multiple approaches will be used to compile the lab technique information including Standard Operating Procedures (SOPs), supporting documents such as bench sheets and quality assurance plans, and a survey questionnaire. The questionnaire will be created after reviewing SOPs, ensuring that all of the relevant information can be collected. If necessary, follow-up one-on-one interviews with lab managers or lab directors may be conducted to verify lab techniques and fill in any missing information.

Compile historical test data

Historical testing data will be compiled from ELAP accredited laboratories. Testing data will focus on controls and each laboratory's reference toxicant results. Test acceptability requirements for controls dictate minimum lab performance and, because there is a lack of toxicant exposure, it is assumed that this represents each lab performing the test to the best of their ability. Similarly, reference toxicant test requirements dictate specified lab performance, especially in precision of test organism response. Test sample data response will not be utilized because test performance expectations are not known.

The goal is to compile the daily number of neonates per replicate for each test and supporting data. Table 3 lists the factors to be calculated from compiling this historical test data including:

- Average number of neonates/female
- Average number of broods/female
- Age of females at test start
- Control variability as standard deviation (SD), coefficient of variation (CV)
- Reference toxicant 50% lethal concentration (LC50) variability as SD and CV
- Reference toxicant 25% inhibitory concentration (IC25 for reproduction) variability as SD and CV
- Test water quality data
- Number of replicates tested

In addition, laboratories may be queried for testing details including:

- Procedure for determining mortality
- Procedure to exclude 4th broods
- Frequency of test failures (if data not provided)

Table 3. Categories of information types to be collected and their likely method of collection.

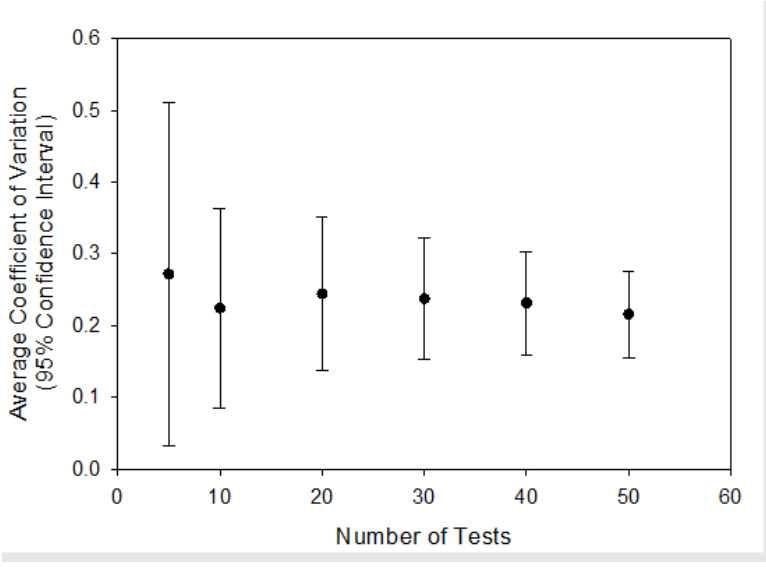
	SOP and supporting documents (e.g. QAP, bench sheets)	Survey and/or phone call	CETIS report and raw data
Dilution water			
Recipe (incl. supplements), vendor	x	x	
Source water	x		
Shelf-time	x	x	
Food			
YCT recipe , vendor	x	x	
Shelf-time	x		
Algal species, source, culture media	x	x	
C. dubia culture			
Frequency of restart/turnover		x	
Frequency of culture failure		x	
Photoperiod	x		
Culture water quality data (e.g. hardness)	x		
Testing procedure/ historical data			
Control variability			x
Age of females at test initiation	x		x
# of neonates/female in controls			x
# of broods/female in controls			x
Daily number of neonates			x
Reference toxicant variability			x
Frequency of test failures	x		x
Test water quality data		x	x
Number of replicates			
Treatment of data outlier		x	
Procedure to exclude 4th broods	x	x	
Time to reproduction	x		
Technician Experience			
Training protocols		x	
Technical experience	x		

Sample size for the historical data compilation is a critical study element. Currently, there are 17 ELAP accredited laboratories conducting the *C. dubia* test in California. Thus, sample size to compare differences among labs is already truncated. The study goal is to collect data from every ELAP accredited laboratory.

Sample size for within laboratory variability assessment is also critical. The study goal is to compile at least 30 tests or the last 3 years of test data from every ELAP accredited laboratory, whichever comes first. Knowing that laboratories do not have the same testing frequency, this target sample size is based on weighing two competing factors that could influence variability assessments. The first factor is the desire to have as much data as possible to have confidence in the variability quantification. The second factor is the desire to keep data as current as possible to reduce the effects of potentially variable-inducing parameters such as evolving lab techniques, turnover in personnel, and other challenges.

Figure 2 presents the results of a simulation to assess confidence in control variability at various sample sizes. This simulation focused on the coefficient of variation (CV) in control brood size per female. Lower CVs have less variability than larger CVs, and Fox et al (2019) targets CVs<0.25 as preferred for reducing errors using the TST. For this simulation, the number of neonates per female ranged from 15 – 30, which meets the promulgated method minimum and approximates the example used in USEPA (2012). The number of neonates per female was randomly assigned for 10 replicates per test and the control CV calculated. This was repeated for sample sizes ranging from 5 to 50 tests. The average and 95% confidence interval of brood size CV per test was calculated for varying sample sizes in Figure 2. This simulation illustrates two important points; a) the average CV tends to stabilize after about 20 tests, and b) the 95% confidence interval about the average CV continues to get smaller with more tests (as expected), but large gains in confidence subside after sample sizes >30. These two results support the study goal of 30 tests or 3 years, whichever comes first.

Figure 2. Simulation to assess average and 95% confidence intervals in control coefficient of variation (CV) at various sample sizes.



Data management, quality assurance and analysis

Data workflow will follow the schema in Figure 3. Input data will come in two categories mirroring the subtasks for this step: laboratory techniques and historical data.

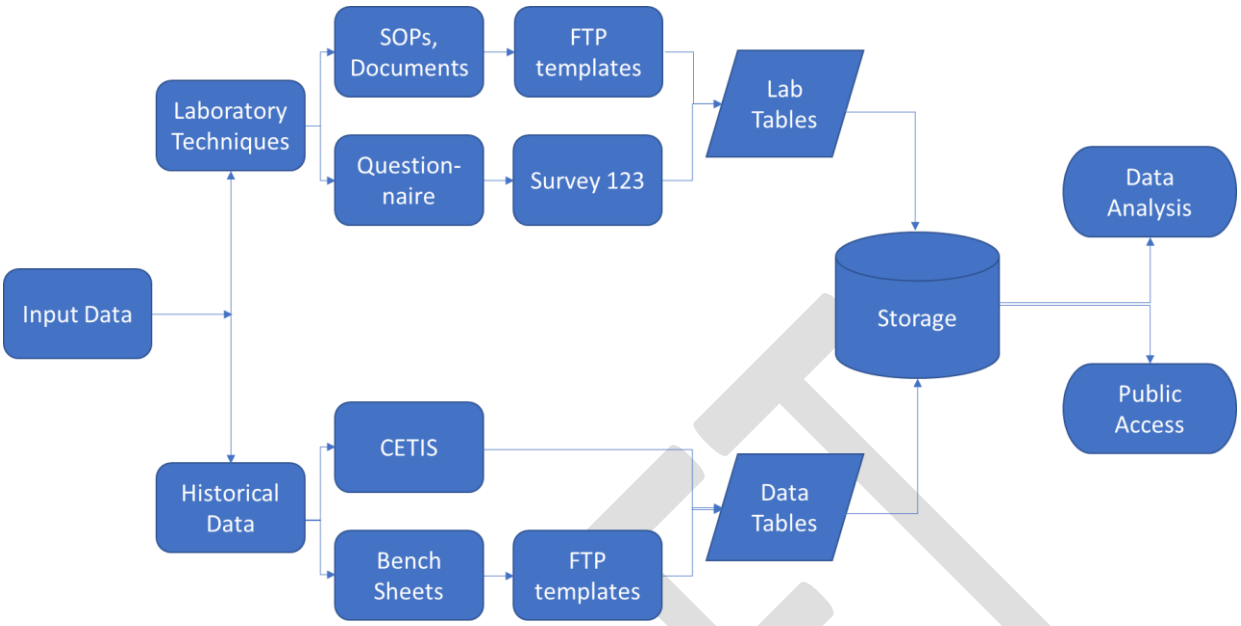
Laboratory technique data will be gathered in two different formats. The first format will be document extraction. This includes documents such as SOPs, Quality Assurance Plans, and other supporting documentation. These data will be categorical, and input using file transfer protocol (ftp) data entry templates by SCCWRP staff. At least 20% of the hand data entered will be checked manually by supervisors and, if errors are found, the remaining 80% of data from that data entry personnel will be checked and accompanied by additional personnel training. The ftp process will also include automated data checkers prior to storage for a 100% data audit of completeness, redundancy, syntax, and look up list errors. The second format will be the laboratory survey, where directed questions will be asked of lab managers or lab directors. The electronic survey will be created in Survey 123, with pre-defined look up lists to eliminate any data entry errors. After automated data checks for completeness and redundancy, these data will be loaded into storage.

Historical data will be more data management intensive than laboratory technique data. The preferred route for compiling historical data is to utilize CETIS (Comprehensive Environmental Toxicology Information System), a software package utilized by regulated agencies for submitting compliance toxicology data. CETIS provides all of the data necessary for this project including daily counts of surviving females and neonate production per replicate, as well as water quality monitoring. SCCWRP will create an ftp for receiving exported CETIS files and transforming into the necessary formats for this project. Automated data checkers prior to storage will create a 100% data audit of completeness, redundancy, syntax, and look up list errors. Where errors occur in exported CETIS data, SCCWRP will query the laboratory of origin for the missing data.

For historical data that is not in CETIS, SCCWRP will need to hand enter data directly from bench sheets. Similar to the workflow for laboratory techniques, these continuous data will utilize file transfer protocol (ftp) data entry templates input by SCCWRP staff. At least 20% of the hand data entered will be checked manually by supervisors and, if errors are found, the remaining 80% of data from that data entry personnel will be checked and accompanied by additional personnel training. The ftp process will also include automated data checkers prior to storage for a 100% data audit of completeness, redundancy, range checks, syntax, and look up list errors.

Once completed, the central data storage will link the laboratory technique and historical data through unique key fields to bind the tests specific to each laboratory. All laboratories will remain anonymous for this study to enhance laboratory participation. The central data storage will be located behind firewalls on SCCWRP servers and backed up twice daily in at least two locations to ensure data security. The laboratory technique and historical data will become publicly accessible data at the conclusion of the study. However, all laboratory identifiers will be kept anonymous to ensure that laboratories are free to participate without fear of being singled out.

Figure 3. Data workflow for this study.



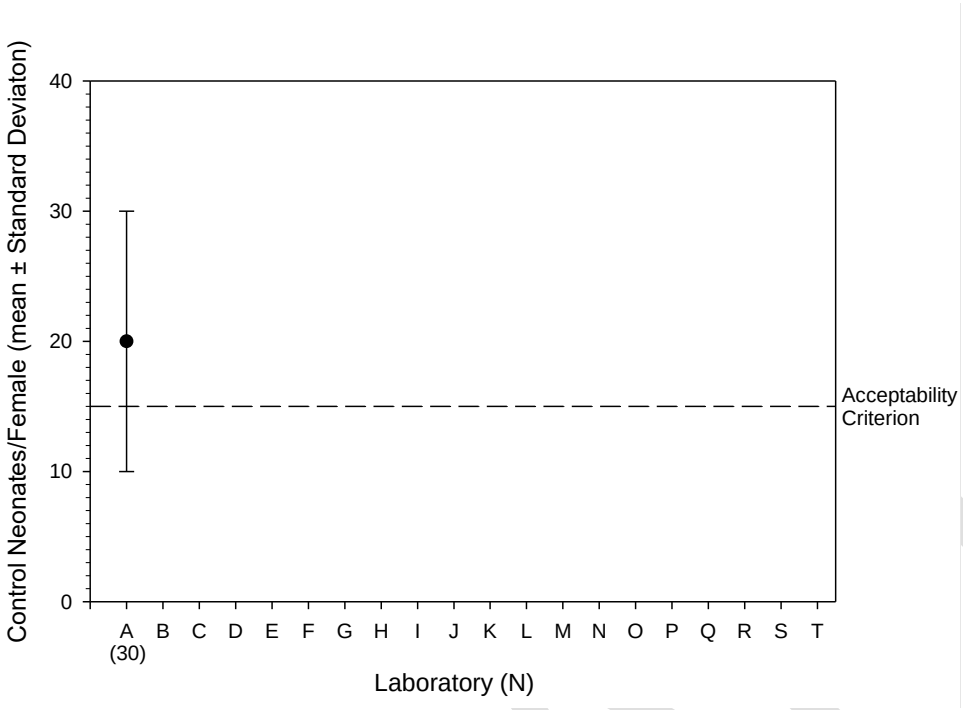
Data analysis will focus on three aspects for this task, consistent with the study questions:

1. Lab technique inventory
2. Comparison of lab performance among laboratories based on historical data
3. Relationships between lab performance and historical data

The lab technique inventory will be a table that quantifies the number of labs utilizing each technique.

Comparison of lab performance among laboratories will utilize pairwise or multiple pairwise (e.g., ANOVA) analysis between laboratories for key test performance metrics. Key test performance metrics will focus on the number of neonates per female (average, standard deviation, CV) in controls and reference toxicant effect concentrations (LC50 and IC25, average, standard deviation, CV). An example of one keystone graphic to compare lab performance among laboratories is Figure 4. To be clear, while statistical analysis will be applied to these data, statistically significant differences are not the goal of this analysis. Especially with so few labs and expected unequal sample sizes among labs, it is anticipated that patterns will be just as important as statistical testing. Where patterns emerge, the lab techniques for the best and worst performing laboratories can be more carefully examined.

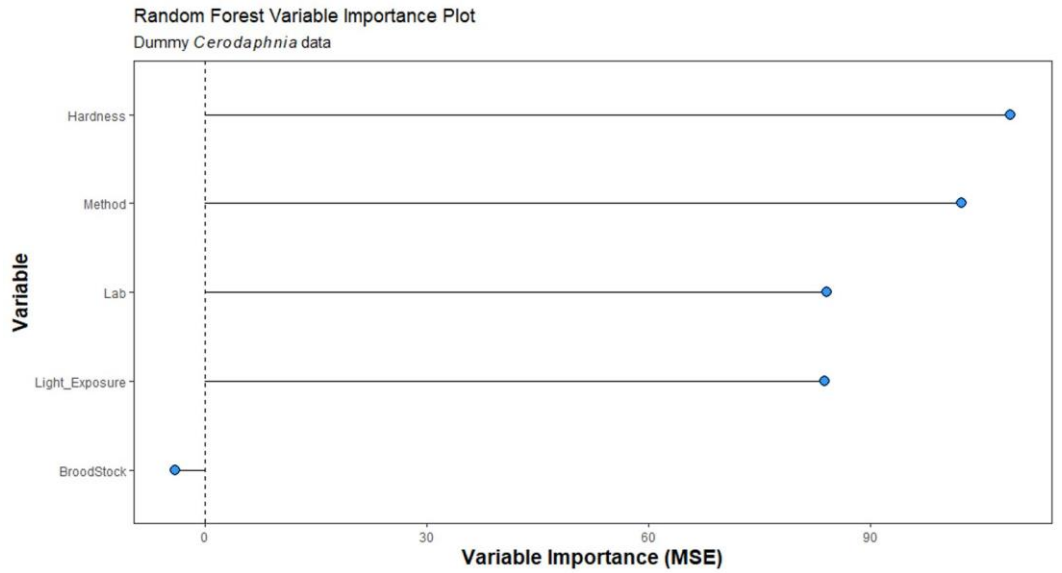
Figure 4. Example keystone graphic for comparing control performance among laboratories. Data for Lab A is not real and plotted only as an example. Data is not plotted for Labs B-T.



Relationships between lab techniques and historical data is the ultimate goal of this task. Multivariate statistics will be used to compare categories and subcategories of lab techniques and the historical data test metrics to determine which lab technique(s) is the source of the most variability (e.g., dilution water recipe versus food source, versus laboratory experience, etc.). The multivariate analysis will include Random Forest and/or Linear Mixed Effects Models, which will attempt to identify variables of greatest importance. An example keystone graphic may look like Figure 5, where mean square error is used to rank variable importance. The data in figure 5 are not real and should be used only for visualization of example end product.

From this series of analyses, a set of proposed revisions or standardizations to the laboratory techniques will be developed. If it is determined that it would be useful to gather more data from the laboratories to verify what was established during the data analysis, then round-robin testing may be proposed if there is sufficient funding. This exercise would seek the participation of all ELAP accredited laboratories to analyze select split samples (e.g., dilution water with different hardness).

Figure 5. Example of keystone graphic illustrating results of random forest. Data is not real, and is for illustration purposes only.



Dose-response testing to optimize lab techniques

The third of five steps in this study will focus on dose-response testing procedures to quantify the variability of lab techniques identified by the historical analysis in Step 2.

There may potentially be many variable inducing differences in lab techniques from Step 2, and not all of them may be followed up in this Step 3. So, a prioritization of which lab techniques require dose-response testing will need to be evaluated by the Stakeholder Committee and Expert Science Panel formed in Step 1. The criteria for prioritization should include:

- Lab techniques that appear to contribute the greatest within or among test variability
- Lab techniques that have multiple options across the most laboratories
- Techniques that have optional approaches in the promulgated method guidance
- Techniques that are not defined in the promulgated method guidance
- Others as agreed upon by the Stakeholder Committee

The dose-response testing will be conducted by one (or a limited number) of laboratories to remove confounding interlaboratory variability. This should help to isolate the variability associated with only the lab technique. Therefore, the laboratory(ies) selected should be amongst the most competent and experienced to ensure capability, minimize intra-laboratory variability, and need not be ELAP accredited.

The results of the dose-response testing will culminate in a technical memorandum that describes the variations of the lab technique selected to be tested, the results of the dose-response testing, and a draft recommended guidance on the optimized lab technique to produce the least variability in test results.

Evaluation of the revised lab technique

The fourth of five steps in this study will be for all of the ELAP accredited laboratories to participate in a round-robin split-sample exercise using the draft laboratory technique guidance developed from Step 3.

Since the project builds from step-to-step, the exact sample types, number of laboratories, and the laboratory technique guidance are currently unknown. The sample types will be chosen so that the effect of the draft laboratory technique guidance can best be tested and quantified. It is presumed that the challenge samples may include a variety of blank samples, as well as some spiked samples. However, the exact number of samples will be agreed upon after Step 3. The number of laboratories will be dependent upon the amount of additional resources available.

Whatever criteria are used to select samples for testing, they first will be vetted by the Stakeholder Committee and then approved by the Expert Science Panel.

The results of the split-sample testing will culminate in a technical memorandum that describes the split-samples created and distributed to laboratories, the results of the split-sample testing, and an assessment of the final draft recommended guidance on the optimized lab technique to produce the least variability in test results.

Final report with final recommended guidance

The final report will summarize the study objectives, methods, results, and a discussion of the findings and limitations of the study. The final report will include the interim deliverables contained within the Technical Memos from Steps 2-4. The final report will also be published documentation to accompany the project database.

Most importantly, the final report will contain the vetted recommended guidance for laboratory activities to optimize variability implementing the *C. dubia* reproduction test.

The Stakeholder Committee will have multiple opportunities to review and provide input on the final report. The Expert Science Panel will also review the final report and provide a consensus opinion on the recommended laboratory technique guidance for implementing the *C. dubia* reproduction test.

SWRCB staff will be responsible for deciding the final disposition of the recommended laboratory technique guidance, and the final recommendation to the State Water Board.

Schedule

Task	Product	Deadline
Study Workplan		
Draft	Draft Workplan to identify potentially variable-inducing lab techniques	3/1/21
Final	Final Workplan approved by Expert Science Panel	5/1/21
Historical Data Analysis		
Lab Data Analysis	Technical Memo identifying potentially variable-inducing lab techniques	7/1/21
Split Sample Analysis (if conducted)	Technical Memo quantifying within and among lab variability	1/1/22
Optimization Testing	Technical memo with draft recommended guidance to reduce within and among lab variability	3/30/22
Interlaboratory Testing	Technical Memo quantifying within and among lab variability	7/31/22
Final Report		
Draft	Draft Report with final recommended guidance	11/1/22
Final	Final Report approved by Expert Science Panel	12/31/22

References

Diamond, J., P. Stribling, M. Bowersox, and H. Latimer. 2008. Evaluation of effluent toxicity as an indicator of aquatic life condition in effluent dominated streams: A pilot study. *Integrated Environmental Assessment and Management* 4:456-470.

Fox, J.F., D.L. Denton, J. Diamond, and R. Stuber. 2019. Comparison of false-positive rates of 2 hypothesis-test approaches in relation to laboratory toxicity test performance. *Environmental Toxiciology and Chemistry* 38:511-523.

Moore, T.F., S.P. Canton, and M. Grimes. 2000. Investigating the incidence of type I errors for chronic whole effluent toxcitiy testing using *Ceriodaphnia dubia*. *Environmental Toxiciology and Chemistry* 19:118-122.

Schiff, K.C. and D. Greenstein. 2016. Stormwater monitoring coalition toxcity testing laboratory guidance document. Southern California Coastal Water Research Project. Costa Mesa, CA.

USEPA. 2001. Final report: Interlaboratory variability study of EPA short-term chronic and acute whole effluent toxicity test methods, Vol. 1. U.S. Environmental Protection Agency. Washington, DC.

USEPA. 2002. Short-term methods for estimating the chronic toxicity of effluents and receiving water to freshwater organisms. EPA-821-R02-013. U.S. Envirnmental Protection Agency. Washington DC.

APPENDIX A – List of ELAP Accredited Laboratories

Lab Name	**Lab Type**
ELAP accredited laboratories in California	
49er Water Laboratory	Private
Aqua-Science	Private
Aquatic Bioassay & Consulting Laboratories, Inc.	Private
Aquatic Testing Laboratories	Private
Aquatic Toxicology Laboratory, Aquatic Health Program	Academic
Enthalpy Analytical, LLC (Nautilus)	Private
Environmental Monitoring Div. (EMD) Lab. at Hyperion Treatment Plant	Public
Granite Canyon -- UC Davis Marine Pollution Studies Laboratory	Academic
Inland Empire Utilities Agency Laboratory	Public
MBC Aquatic Sciences	Private
McCampbell Analytical, Inc.	Private
Pacific EcoRisk	Private
San Jose Creek Water Quality Laboratory	Public
Wood Environment & Infrastructure Solutions, Inc.	Private
ELAP accredited laboratories outside of California	
Eurofins TestAmerica - Corvallis (ASL)	Private
GEI Consultants, Inc.	Private
Ramboll	Private
Tetra Tech's Ecological Testing Facility	Private