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 – May 26, 2021

Contents

List of Abbreviations and Definitions

C. dubia – Ceriodaphnia dubia

CASA – California Association of Sanitation Agencies

CETIS – Comprehensive Environmental Toxicology Information System, (Tidepool Scientific Software)

CV – Coefficient of variation

DO – Dissolved oxygen

ELAP – State of California Environmental Laboratory Accreditation Program

FTP – File Transfer Protocol

IC – Inhibition Concentration is the toxicant concentration that would cause a given percent reduction in a non-quantal biological measurement for the test population.

IC25 – 25% inhibitory concentration

Inter-laboratory variability – the variability between laboratories, measured by comparing results from different laboratories using the same test method and the same test material.

Intra-laboratory variability – the variability within a laboratory, measured when tests are conducted using specific methods under constant conditions in the same laboratory. This variability includes within-test variability.

LC – Lethal Concentration is the toxicant concentration that would cause death in a given percent of the test population.

LC50 – 50% lethal concentration

QAP – Quality Assurance Plan

SCCWRP – Southern California Coastal Water Research Project

SD – Standard deviation

SOP – Standard Operating Procedure

SWRCB – State Water Resources Control Board

TST – Test of Significant Toxicity

U.S. EPA – United States Environmental Protection Agency

YCT – Yeast-Cerophyll-Trout chow

WET – Whole Effluent Toxicity

1. Introduction

The California State Water Board recently adopted Toxicity Provisions, which include numeric effluent limitations to protect California's enclosed bays, estuaries, and inland water bodies from contaminated discharges. The Toxicity Provisions also include a requirement to use the Test of Significant Toxicity (TST) statistical approach, which provides greater confidence in evaluating aquatic toxicity test data. Compared to other statistical approaches, the TST controls for both false positive and false negative error rates (Denton et al. 2011). This approach also "restated" the null and alternative hypotheses compared to traditional hypothesis tests; the null hypothesis of the TST being that the sample is toxic. Because of the controls on error rates and the restating of the null hypothesis, the TST approach is more likely to find a sample to be toxic if within-test variability is high. In using this approach, the State Water Resources Control Board (SWRCB) aims to incentivize dischargers to generate high-quality data (i.e., data with low within-test variability). However, dischargers have expressed concerns about the inherent variability in some of the Whole Effluent Toxicity (WET) tests included in the Toxicity Provisions such as the WET test for *Ceriodaphnia dubia* (*C. dubia*) chronic reproduction.

The *C. dubia* reproduction test is a well-established and validated method, and was first promulgated in October 1995 and finalized in 2002, nearly 20 years ago (U.S. EPA 2002a, b, c; U.S. EPA 2016). While the State Water Board has full confidence in the use of *C. dubia* for regulatory programs, they recognized that some laboratories may need to improve their implementation of the *C. dubia* method. For this reason, implementation of the monthly median effluent toxicity limitation for the *C. dubia* reproduction test has been delayed until January 1, 2024 for some dischargers as specified in the Toxicity Provisions. During this time, the State Water Board has committed to a study, in collaboration with stakeholders and laboratories, to evaluate laboratory performance, investigate factors that can lead to test variability and decrease confidence in assessments of toxicity or non-toxicity, and provide revised laboratory technique guidance to improve laboratory performance. The study provides an opportunity to build stakeholder and public confidence in the *C. dubia* method.

It should be noted that 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 18 ELAP accredited laboratories for conducting the *C. dubia* test for 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 the ELAP 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.

The WET test methods (U.S. EPA 2002a, b, c; U.S. EPA 2016) allow laboratories some flexibility when implementing certain laboratory techniques. For example, for the *C. dubia* test method 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. It has been hypothesized that these small differences between laboratories may lead to intra- or inter-laboratory 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 22 out of 122 *C. dubia* chronic tests did not meet test acceptability criteria for survival or reproduction (U.S. EPA 2001a, b). The invalid tests were confined to

10 out of 34 participating laboratories. The study reported intra- and inter-laboratory coefficients of variation (CVs) for the IC25 values of effluent and receiving water split samples at 17% and 28%, respectively for the reproduction endpoint. More recently, a smaller interlaboratory comparison 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" based on three factors including test acceptability, intra-laboratory precision, and inter-laboratory precision. 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. From their review, the authors found that either reducing the between replicate variability or increasing the number of replicates improved lab performance.

Various studies have focused on the causes of *C. dubia* test variability or ways to optimize the test. The main thrust of these studies has been on test water and organism feeding. Elphick et al. (2011) found that water hardness influenced the sensitivity of the organisms to chloride, with a decrease in toxicity observed as hardness increased. Other studies found acute toxicity associated with major ions (Na^+, K^+, Ca^{2+}, Mg2+, Cl^-, SO_4^{2-}, and HCO_3^-/CO_3^{2-}); salts can be a confounding factor both in natural waters and anthropogenically influenced waters (Mount et al. 2016, Erickson et al. 2017). Additionally, Mount et al. (2016) found that natural waters with low major ion concentrations caused *C. dubia* to be more sensitive to solutions of some salts. They noted that the dilute solutions appeared to be stressful to the organisms. How these results would translate to the reproductive endpoint is unknown but might be important if any laboratories are using natural waters for culture and dilution.

One California laboratory conducted multiple studies to reduce sources of variability in their own *C. dubia* tests. The lab tested multiple dilution water types and sources and found that synthetic versus natural water had the most impact on reproductive variability, but it was small compared to feeding related aspects (Briden et al. 2017). In a study on the effects of water hardness, the California lab also found it had no impact on long-term culture performance (Clark and Briden 2018). However, the lab found that organism source and control/dilution water hardness might have an impact on test results. In two of six samples where both a soft and moderately hard water control was used, the interpretation of toxicity differed depending on which control the sample was compared to. The laboratory also conducted two studies looking at the effects of food quality. In the first study, the lab found that quality of the food had an impact on the test performance even if inferior quality food was only fed to culture animals, but higher quality was used during the test period (Jorgenson et al. 2017). The study also found that the quality of the algae was the most important factor and was greater than control/culture water parameters, feeding density, food component, culture line, or analyst training. Source of the YCT did not appear to affect test control precision. In their second food study, the lab found that vendor sourced food was not necessarily of consistent quality (Prosser et al. 2018). The lab concluded that it was important to run QC tests before using the food in cultures or tests. Little difference was found in reproduction based on variable food components, but that when larger volumes of trout chow were digested a negative impact on test performance was observed. The lab also noted that the EPA recommendation of a 2-week shelf life for a *Selenastrum* batch may be too restrictive. Visual and olfactory observation of each batch were important to determine shelf life.

1.1. Objective of this study

The objective of this study is to build on previous efforts and investigate all possible sources of variability in the *C. dubia* reproduction test conducted by California-accredited laboratories. The goal is to provide laboratory technique guidance 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, while retaining the necessary flexibility for environmental relevance.

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 intra- and inter-laboratory technique differences used by ELAP accredited laboratories?
3) Does standardizing differences in the *C. dubia* chronic reproduction toxicity test laboratory techniques reduce intra- and inter-laboratory 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. It is also not expected that all variability will be eliminated from the test method. Finally, it should be noted that this study is not designed to address aspects of testing that may be more effectively dealt with by appropriate study design: e.g., ion, hardness, or conductivity controls in cases where those variables have the potential to affect test outcomes, but do not represent environmental risks.

1.2. Approach

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

1) Create a Governance structure
2) Analyze historical data and lab techniques to identify sources of variability
3) Optimize lab technique(s) and recommend lab technique guidance
4) Evaluate the revised lab technique guidance via split-sample testing
5) Provide final recommended guidance in a Final Report

These tasks are sequential with each one informing the details of the next. This project will be implemented using an iterative process and follow the Adaptive Management Approach, using the plan, do, evaluate & learn, and adjust cycle. Before moving on to the next task, the Stakeholder Committee and Expert Science Panel members will review the results and help define the approach to accomplish the next steps. The workplan will be updated as needed to include any revisions and new details or methods.

The first Task 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 Task will be comprised of two subtasks. First, an inventory of lab techniques used by ELAP accredited laboratories will be created. The inventory will elucidate the level of comparability and differences in test implementation. Historical testing data will also be collected from the ELAP accredited laboratories to quantify the level of variability within and among laboratories. The inter- and intra-laboratory variability will be assessed based on the reproductive endpoint of the test method (average number of neonates per female). Secondly, the differences in lab techniques will be compared to the lab test result variability. The goal of these data analyses is to indicate which lab techniques might account for the observed variability in the test outcomes.

An optional third subtask is under consideration to collect new data from split samples to assess intra- or inter-laboratory variability. The split sample analysis for this subtask would supplement historical data analyses and confirm possible sources of test variation, including possibly quantifying interlaboratory variability. The scope and study design for split sample testing will be dependent on the availability of additional funds and the outcome of subtasks 2.1 and 2.2. Therefore, it is not included in the current workplan.

The third Task will focus on targeted laboratory experiments by one or two laboratories to quantify the variability of lab techniques identified in Task 2. In the event that many variable-inducing parameters are identified in Task 2, the Stakeholder Committee and Expert Science Panel will prioritize the parameters requiring further optimization. The results of this Task will culminate in a draft recommended guidance on the lab technique that produces the least variability in test results.

The fourth Task will verify that draft recommended guidance from Task 3 does reduce variability, both within and among ELAP accredited laboratories. Split samples will be distributed to the ELAP accredited laboratories for testing using the draft recommended guidance for laboratories to follow.

The fifth Task will complete the study and include the final recommended guidance based on the results from Task 4, and a final report.

2. Detailed Methods

2.1. Task 1 — 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 and other key participants.

The ultimate decision-making body is the 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). A list of key SCCWRP personnel involved in the project is provided in Table 1. SCCWRP will be responsible for project

design (including this workplan), facilitation of the Stakeholder Committee and Expert Science Panel, project implementation (including interacting with ELAP accredited laboratories), and project reporting (including the final report).

SCCWRP has 50 years' experience culturing and toxicity testing in its own research laboratory. SCCWRP has tested more than 18 different freshwater, marine, and estuarine species including invertebrates (including *C. dubia*), fish, and algae. SCCWRP has experience toxicity testing with a variety of matrices including discharges, ambient water, sediment, and bioaccumulation, and evaluating a wide range of endpoints including mortality, reproduction, growth, normal development, behavior, physiological condition, and molecular activity. SCCWRP research has included a diversity of toxicity test method development, including methods in the USEPA's West Coast Methods Manual (USEPA 1995). Project staff serve in leadership roles on both regional and national chapters of the Society of Environmental Toxicology and Chemistry. As an independent research agency supporting regional and statewide monitoring programs, SCCWRP routinely conducts laboratory intercalibrations for toxicity testing in water and sediment, as well as field sampling, chemistry, microbiology, and biological taxonomic laboratory analysis. SCCWRP has also facilitated external reviews of the State's Environmental Laboratory Accreditation Program (ELAP). Cumulatively, lead project staff have over 138 person-years of experience. While SCCWRP has extensive experience in project attributes, the agency is not for hire to conduct routine compliance whole effluent testing.

The independent Expert Science Panel is 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 accordingly, and providing a consensus opinion on the final method guidance.

The Stakeholder Committee is 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 3. The Stakeholder Committee provides 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 the governance structure. 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. Since this project is designed to be an open and transparent process, members of the public are also included. The public can interact in the governance structure through the Stakeholder Committee representative for their sector. Public comments or questions are included as an agenda item during every Stakeholder Committee meeting.

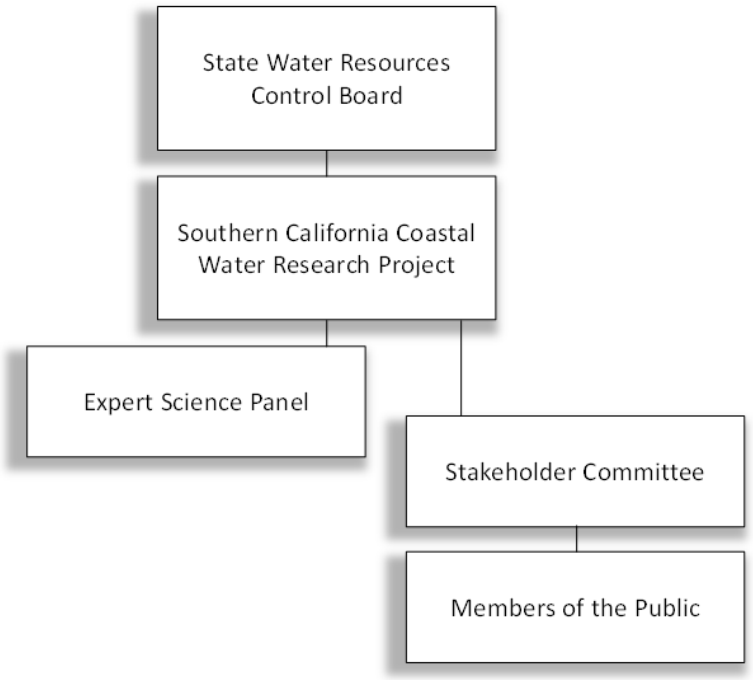


Figure 1. Project organization chart

Table 1. List of key SCCWRP personnel

Personnel	**Role**	**Total Years Experience**	**Years Experience in Study Specialty Areas**		
			Toxicology	**Laboratory Intercalibrations**	**Quality Assurance**
Ken Schiff	Co-Principal Investigator	35	35	25	30
Alvina Mehinto	Co-Principal Investigator	20	20	10	10
Ashley Parks	Quality Assurance Officer	13	13	8	10
Darrin Greenstein	Task Manager	35	35	25	32
David Gillett	Biostatistician	15	10	15	15
Paul Smith	Data Management Officer	20	10	15	20

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 (U.S. EPA)
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)

Table 3. 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)
Laboratory Accreditation	Steven Boggs (CA ELAP)
Private Laboratories	Jeff Miller (Aqua-Science Laboratories)
Public Laboratories	Josh Westfall (Los Angeles County Sanitation Districts)

2.2. Task 2 — Analyze historical data and lab techniques and identify sources of variability

2.2.1. Inventory of laboratory techniques and historical data

To date, no one has created an inventory of historical data and laboratory techniques used by ELAP accredited laboratories for the *C. dubia* reproduction test. Task 2 will create that inventory, focusing on three categories of information (see Table 4):

- Test conditions
- Performance data for control samples and reference toxicant
- Culturing information

Multiple approaches will be used to compile the data starting with a review of the laboratory documentation such as Standard Operating Procedures (SOPs), Quality Assurance Plans (QAP), bench sheets, CETIS reports and control charts. Performance data will focus on control samples and reference toxicant. Because control samples do not contain any toxicants, this should represent the lab's ability to best perform the test. Similarly, test requirements for the reference toxicant dictate specified lab performance such as precision of test organism response. Test sample data response will not be utilized because sample performance expectations are not known. The goal is to compile key raw and summary data for each test.

Compiled tests from the participating laboratories will include all control and reference toxicant tests regardless of whether they pass the test acceptability criteria or not. Any tests which do not pass QA will

be included but flagged so they can be differentiated during data analysis. Individual replicate and raw data will be collected (i.e., daily counts, daily water quality data from each treatment, etc.) in addition to the summary data from CETIS, and other relevant metadata.

Table 4 distills (1) the parameters to be inventoried in each of the three categories, (2) the data type to be collected, and (3) the method of collection. The test condition parameters correspond to parts of the promulgated method that allows for flexibility in their lab techniques. Data will be recorded as number, text, drop-down menu selection, or calculated by SCCWRP based on raw data obtained by the laboratories. All the individual, raw data will be available for review by the Expert Science Panel as part of the iterative process of this study.

Table 4. C dubia test parameters and data to collect from each laboratory.

***C. DUBIA* LAB TECHNIQUES AND DATA**	**DATA TYPE**	**SOP, QAP, bench sheets, supporting documents**	**CETIS, raw data, control charts**
Test conditions			
Dilution water recipe	drop-down*	x	
Dilution water recipe modifications	text	x	
Source water	drop-down	x	
Dilution water shelf-time (weeks)	number	x	
Measured ions concentration (mg/L)	number	x	
Daily water hardness (mg/L)	number	x	x
Daily conductivity (µS/cm)	number	x	x
Daily pH	number	x	x
Daily temperature (°C)	number	x	x
Daily DO (mg/L)	number	x	x
Sample volume in test chamber (mL)	number	x	
Test chamber material	drop-down	x	
Test chamber volume (mL)	number	x	
Test chamber diameter (cm)	number	x	
Photoperiod	drop-down	x	
Light source	text	x	
Light intensity (min and max; foot-candles)	number	x	
Lab air temperature (range, °C)	number	x	
YCT vendor	text	x	
YCT concentration in chamber (mg/L)	number	x	
YCT shelf-time (weeks)	number	x	
Algal species	text	x	
Algal vendor	text	x	
Algal culture media	text	x	
Algae concentration in chamber (mg/L)	number	x	
Algae shelf-time (weeks)	number	x	
Feeding frequency (count/day)	drop-down	x	
Age window at test initiation (hrs)	drop-down	x	
Reference toxicant used	drop-down	x	
Number of replicates	calculated †		x
Time to reproduction (days)	calculated		x

*Drop-down indicates a constrained list of responses; † Calculated indicates data that will be generated based on raw data from the individual laboratories

Table 4 cont. C dubia test parameters and data to collect from each laboratory.
Performance data will be collected for both control samples and reference toxicant, unless stated otherwise.

***C. DUBIA* LAB TECHNIQUES AND DATA**	**DATA TYPE**	**SOP, QAP, bench sheets, supporting documents**	**CETIS, raw data, control charts**
Performance data			
Daily neonate counts per replicate	number		x
Number of neonates per female (mean, SD, CV)	calculated		x
Number of broods per female (mean, SD, CV)	calculated		x
Reference toxicant 50% lethal concentration (mean, SD, CV)	calculated		x
Reference toxicant IC25 for reproduction (mean, SD, CV)	calculated		x
Culturing Information			
Origin of brood stock	text	x	
Water hardness range (min and max; mg/L)	text		x
Conductivity range (min and max; µS/cm)	text	x	
pH range	text		x
Water temperature range (min and max; °C)	number		x
Photoperiod	drop-down		
Light source	text		
Light intensity (min and max; foot-candles)	number		
Percentage of males	number	x	
Percentage adult mortality	number	x	
Percentage of unhealthy/small adults	number	x	
Percentage of neonate mortality	number	x	
Percentage of unhealthy neonates	number	x	
YCT concentration in culture chamber (mg/L)	number	x	
Algae concentration in culture chamber (mg/L)	number	x	

A follow-up survey questionnaire will be developed and submitted to the laboratories to ensure that all relevant information described in Table 4 has been collected. Thus, the specific survey questions are not part of the current workplan and will be identified after reviewing the data collected from lab documentation. The survey will include both targeted and open-ended questions aimed at filling in data gaps and learning more about each laboratory's practices and challenges. In the absence of in-person visits due to COVID-19, one-on-one phone calls will be scheduled with lab managers or lab directors to collect the responses to the survey questions.

Below is a sample of possible survey questions:

- Which brood stock animals were used in tests?
- What is the specific age window of animals at test start?
- What is the feeding frequency and concentration during the test?
- What are your procedures for determining mortality?
- What is your procedure to exclude 4th broods?
- What is your annual percentage of test failures and for what reason(s)?
- What is your experience with regards to reducing test variability and improving performance?
- What, if any, deviations have been made from your SOP, and how were these noted?
- How many times has your lab had to restart your culture in the last 3 years?
- How do you treat outlier data?
- What is your percentage of data audited by the QA officer?
- What is your testing capacity (tests/year)?
- How many years has your lab conducted the WET test?
- How many years of experience does your lead technician have?
- How many tests has your lead technician conducted?
- For new technician training, how many practice tests are required as part of the training process?
- How frequently are cultures monitored to assess general health?

This is not the full list of survey questions and a more detailed survey will be provided to the Stakeholder Committee and Expert Science Panel after analyses of laboratory data have been completed.

In addition to the phone interviews, SCCWRP and the Expert Science Panel will schedule a group meeting with the participating laboratories to further discuss their experience conducting the *C. dubia* reproduction test.

2.2.2. Number of tests to collect from participating laboratories

The number tests to be collected is a critical element of this study to assess within and inter-lab variability. Currently, there are 18 ELAP accredited laboratories conducting the *C. dubia* test for California. This study will seek participation of every ELAP accredited laboratory to ensure maximum sample size for the assessment of inter-laboratory differences. To evaluate within-laboratory variability, the last 30 tests will be compiled or the last 3 years of test data from every ELAP accredited laboratory, whichever comes first. The datasets will include tests that pass or fail test acceptability criteria. 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 potential variability-inducing parameters such as evolving lab techniques, turnover in personnel, and other challenges.

A simplistic pre-survey simulation was conducted to help assure the estimated 30 tests per laboratory is an appropriate sample size for assessing confidence in a laboratory's average control precision. This simulation focused on precision defined as the coefficient of variation (CV) in control brood size per female. Lower average CVs have greater precision than larger average CVs. The assumptions associated with this simple simulation are:

- A laboratory was randomly assigned an average number of neonates per female for its control for a single test.
- The average number of neonates per test was constrained between 15 – 30. This range meets the promulgated method minimum and mirrors the example used in the promulgated method manual (U.S. EPA 2012).
- This was repeated for a single laboratory assuming they ran different numbers (sample sizes) of tests ranging from 5 to 50 tests.
- The simulation at each sample size was run for 10 iterations.
- The CV and 95% confidence interval were averaged across the ten iterations at each number of tests.

The goal of this simple simulation was to assess at what point do we optimize confidence in CV estimates (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 tend to support the study goal of 30 tests per laboratory. If a lab has conducted 20 tests in 3 years, the loss in confidence is not great (confidence intervals similar in size). However, investing additional effort to inventory more than 30 tests does not yield much greater confidence either. Since the CVs in this simulation are on the bounds of acceptable CV (between 0.020 - 0.25), we expect even smaller sample sizes may be necessary for better performing laboratories who maintain CV < 0.20. Importantly, this simulation does not assess power to detect difference between laboratories. As discussed in section 2.2.5 below, detecting statistically significant differences in control CV between laboratories is not a goal of this study.

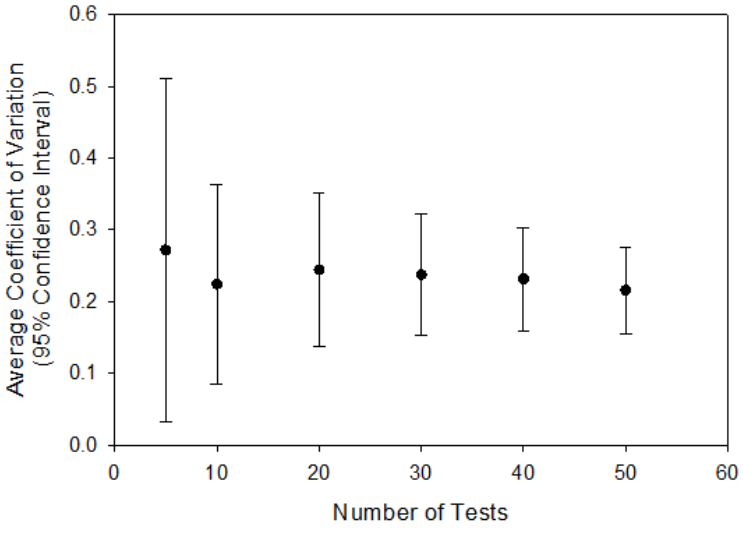


Figure 2. Simple simulation to assess average and 95% confidence intervals in control coefficient of variation (CV) after running varying numbers of tests per laboratory.

2.2.3. Data management

Data workflow will follow the schema in Figure 3. As described above, input data will come in two categories: laboratory techniques and historical data.

Laboratory technique data will come in three forms, consistent with Table 4: a) SOPs, QAPs, bench sheets, b) electronic survey questionnaires, c) lab interviews. Data extracted from existing documents such as SOPs, QAPs, bench sheets and other supporting documentation will largely be categorical or from constrained lists. In some cases, they may require additional text. These data will likely be hand entered by SCCWRP staff from hard-copy documents or pdfs. The hand-entered data will be input to the project database using file transfer protocol (ftp) data entry templates.

Electronic surveys and lab interviews will utilize ESRI Survey 1-2-3 to enter laboratory technique data. Survey 1-2-3 will be constructed to collect data using constrained lists, where appropriate, and the export format linked directly to database structure for seamless uploading.

Historical data will require more intensive data management 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 most 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. For historical data that is not in CETIS electronic format, 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.

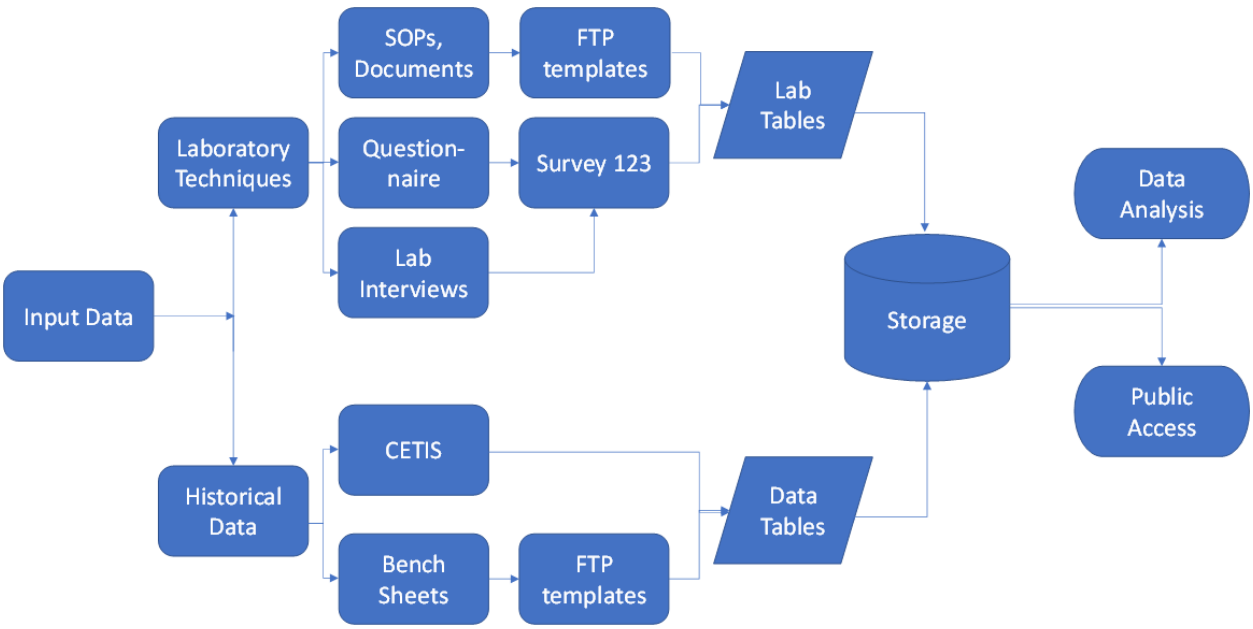


Figure 3. Data workflow for this study.

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. 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.

All laboratories will remain anonymous for this study to enhance laboratory participation. The compiled historical data will become publicly accessible at the conclusion of the study, after the Expert Science Panel has approved the final report and it has been submitted to the SWRCB. However, all laboratory identifiers will be kept anonymous to ensure that laboratories are free to participate without fear of being singled out.

2.2.4. Quality assurance plan

Data Quality Objectives (DQO) refer to the historical data and associated descriptive information to be collected from the laboratories as part of Tasks 2.1-2.2. DQOs for any newly generated data in later Tasks 2.3-4.0 will be set once the study design of that testing has been established.

For the accuracy and precision DQOs for the historical data, two assumptions are being made: 1) We are using data not generated by SCCWRP, and that all data submitted to SCCWRP is complete and accurate, and 2) All data provided by the labs has passed its internal quality assurance and quality control. Consequently, where data sets from laboratories are incomplete, that data does not exist and is not available to SCCWRP for analysis. Therefore, DQOs refer to the data provided to SCCWRP, and the accuracy, precision, and completeness of SCCWRP's data management system.

The DQOs outlining accuracy, precision, and completeness are listed in Table 5.

At least 20% of all the hand-entered data (laboratory techniques and historical data) will be checked manually by supervisors. If data entry 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 include automated data checkers prior to storage for a 100% data audit of completeness, redundancy, syntax, and look up list errors.

Survey 1-2-3 data will have pre-defined look up lists to eliminate any data entry syntax errors. Automated data checkers will be utilized for a 100% data audit of completeness and redundancy.

Automated data checkers for exported CETIS data 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 data resolution.

Once quality assurance for laboratory technique and historical data has occurred, and all of the data has been uploaded to central data storage, any future errors detected and revised during data analysis will be logged and kept with the final data set.

Table 5. Data Quality Objectives for data supplied by participating laboratories.

Data Quality Indicator	**Explanation**	**Data Type**	**QC Action / Frequency**	**Acceptance Criteria**	**Corrective Action**
Accuracy	Overall agreement between a registered value in the data set and the true reported value	CETIS extracted data	100% of data checked for accuracy by automated data checkers prior to data upload	100% accuracy	Correct error and re-run automated data checkers
		Hand-entered data	Audit of 20% of hand-entered data	100% accuracy	100% audit of hand-entered data; correct errors
		Electronic survey data	100% of data checked for accuracy by automated data checkers prior to data upload	100% accuracy	Correct error and re-run automated data checkers
Precision	Agreement among repeated measures of the same property	Not applicable to collection of secondary data			
Representativeness	Degree to which data accurately represents a characteristic of a population	Labs	Study is representative of laboratories conducting the C. dubia test in California because all 18 ELAP accredited laboratories are invited to participate in the study	≥ 75% of all labs	Ask stakeholders to request participation from non-participating labs
		Laboratory test conditions and culturing information	Electronic surveys and personal communication with labs will ensure that reported data is truly representative of laboratory practices	100% agreement between written SOPs and verbal description of laboratory practices	Correct reported data based on verbal communication

Table 5 cont. Data Quality Objectives for data supplied by participating laboratories.

Data Quality Indicator	**Explanation**	**Data Type**	**QC Action / Frequency**	**Acceptance Criteria**	**Corrective Action**
Comparability	Degree to which one data set can be compared to another	Data comparability within the study	All data collected according to this workplan	100% data collected according to workplan	Correct deviation if unintended or revise workplan to accommodate intentional modifications
		Data comparability with other studies	All data archived in publicly available database; Full detailed report at the end of the project	100% data housed in publicly available database (with laboratory identities anonymized)	Post any missing data to the publicly available database
Completeness	Amount of valid data needed to meet the needs of the study	Labs	All 18 ELAP accredited laboratories conducting the *C. dubia* test for California invited to participate in the study	≥ 75% of all labs	Ask stakeholders to request participation from non-participating labs
		Tests	30 tests or 3 years of tests requested from labs	Minimum 30 tests	Exclude laboratory from data analysis if minimum test number not met
		Laboratory test conditions and culturing information	All information in Table 4 requested from each lab	100%	Follow-up on any missing data through personal communication with labs
Sensitivity	Capability of a method to discriminate between measurement responses	Not applicable to collection of secondary data			

2.2.5. Data analyses

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

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

The analyses for each of these aspects are briefly described below. A more detailed description of data analysis methods will be added in a revised version of this workplan after Task 2 (section 2.1) the data collection task is completed.

The lab technique inventory will consist of a table that quantifies the number of labs utilizing each technique. The goal of this analysis is to visually assess how similar or different the lab techniques are among the various accredited laboratories. This analysis will help to highlight which lab techniques are dissimilar between laboratories and may require further investigation.

Qualitative comparisons of lab performance among laboratories based on historical data are a simple first step towards assessing the similarities and differences in lab variability. The inter- and intra-laboratory variability will be assessed based on the reproductive endpoint of the test method (average number of neonates per female). It must be stressed here that the goal of this analysis is not to assess each laboratory's performance critically – all the labs in this study are already accredited – but rather to simply observe if the labs have comparable levels of average neonate production and, perhaps even more importantly, the variability in average neonate production. This analysis will be the first step towards determining if there are certain lab techniques specific to the labs with more (or less) variability in neonate production.

The qualitative comparisons of lab performance among laboratories will focus on historical data including average and standard deviation of neonate production in control samples. The analysis can also include other key test performance metrics including CV of neonate production in controls and reference toxicant effect concentrations (LC50 and IC25, average, standard deviation, CV). An additional data table may be useful examining the distribution of CVs (e.g., min, max, percentiles, etc.) among laboratories. To be clear, while statistical analysis will likely be applied to these data - pairwise or multiple pairwise (e.g., ANOVA) analysis - statistically significant differences between laboratories are not the goal of this analysis. The focus is on patterns and relationships with laboratory techniques (see next analysis). Where patterns emerge, the lab techniques that reduce variability or improve neonate production can be more carefully examined.

Preliminary data analysis will be implemented to help guide the approach to assess intra- and inter-laboratory variability. Performance data from laboratories that use the same reference toxicant will be compared to determine whether the variability in reference-toxicant performance is independent of control performance variability, and may guide the inclusion or exclusion of other test conditions (e.g. diet, photoperiod, test chambers, etc.).

Relationships between lab techniques and historical data are the ultimate goal of this task. These analyses are the critical piece of information for deciphering which laboratory techniques will be further quantified and optimized in the next task. To accomplish this analysis, we will use multivariate approaches to relate categories and subcategories of lab techniques to the historical data test metrics.

The multivariate analysis will include Random Forest and/or Linear Mixed Effects Models, which will attempt to identify variables of greatest importance to the overall variability.

2.2.6. Optional split sample testing

To better assess inter-laboratory variability, the governance may decide to gather new data from laboratories. If sufficient funds are available, a round-robin exercise may be conducted with the participation of all ELAP accredited laboratories to analyze selected split samples. Ideally, such a study would be designed after analyses of the historical data and lab techniques is conducted to identify where additional data is necessary.

2.3. Task 3 — Optimize lab technique(s) and provide lab technique guidance

The third Task of this study will focus on the optimization of lab techniques deemed responsible for the intra- and inter-laboratory variability observed as part of Task 2 data analyses (Section 2.2). The current workplan describes the conceptual approach and a detailed study design will be developed based on the findings of Task 2. Task 3 will result in proposed revisions or standardizations of the laboratory techniques, which will be evaluated in Task 4.

Based on the results of Task 2 (Section 2.2), a number of parameters will be identified as candidate sources of intra- and inter-laboratory variability. Stakeholders and the Expert Science Panel will prioritize those requiring additional optimization through *C. dubia* testing. The criteria for prioritization will include:

- Lab techniques that appear to contribute the greatest within or among test variability as identified by the statistical analyses (e.g., Figure 5)
- 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 and the Expert Science Panel

The toxicity testing will be conducted by one (or a limited number) of laboratories to remove or reduce the confounding factor of inter-laboratory variability. Therefore, the laboratories selected should be amongst the most experienced to ensure capability and minimize intra-laboratory variability.

The laboratories will conduct a series of *C. dubia* tests using controls or reference toxicants quantifying the variability associated with the lab techniques prioritized by the Expert Science Panel during their review of the Task 2 results and analysis (i.e., different dilution water recipes, different food concentrations, or different age windows at test initiation). The lab techniques that consistently reduce intra-laboratory variability will be documented in a technical memorandum. The optimized lab techniques will also be included in a draft recommended guidance.

Since the specific details for this portion of the Workplan are undefined, an amendment to this workplan, which will consist of a detailed Sampling and Analysis Plan, will be produced for the Stakeholder Committee, the Expert Science Panel, and the participating laboratories.

2.4. Task 4 — Evaluate the revised lab technique guidance via split-sample testing

Task 4 of this study will be for all the ELAP accredited laboratories to participate in a round-robin split-sample exercise using the draft laboratory technique guidance developed in Task 3.

Since the project builds from task-to-task, 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 round robin 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 Task 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 will first be vetted by the Stakeholder Committee and then approved by the Expert Science Panel. An amendment to this workplan, which will consist of a detailed Sampling and Analysis Plan, will be produced for the Stakeholder Committee, the Expert Science Panel, and the participating laboratories.

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.

2.5. Task 5 — Provide final recommended guidance in a Final Report

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 Tasks 2-4. The final report will also serve as 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.

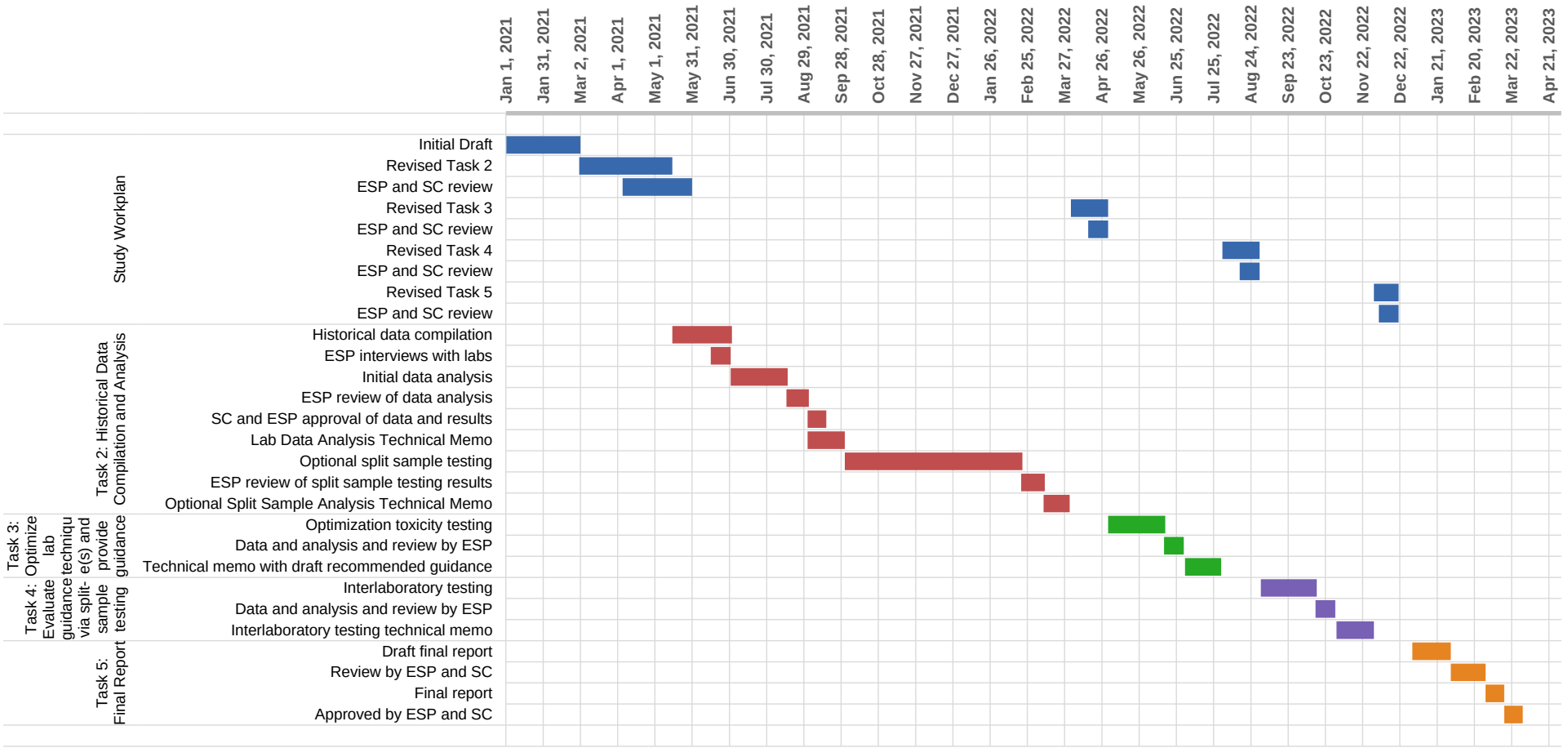
SWRCB staff will be responsible for deciding the final disposition of the recommended laboratory technique guidance, and the final recommendation to the State Water Board.

3. Deliverables Schedule and Project Timeline

3.1. Deliverables Schedule

Task	Product	Deadline
Study Workplan		
First Draft	Draft Workplan to identify potentially variable-inducing lab techniques	3/1/21
Revised Workplan	Workplan Version 2 discussed by the Expert Science Panel	5/1/21
Iterative Review of the Workplan	Revised document with additional study design, data analyses and/or QAP before each new Task	
Historical Data Analysis		
Lab Data Analysis	Technical Memo identifying potentially variable-inducing lab techniques	9/30/21
Split Sample Analysis (if conducted)	Technical Memo quantifying within and among lab variability	3/30/22
Optimization Testing	Technical memo with draft recommended guidance to reduce within and among lab variability	7/30/22
Interlaboratory Testing	Technical Memo quantifying within and among lab variability	11/30/22
Final Report		
Draft	Draft Report with final recommended guidance	1/31/23
Final	Final Report approved by Expert Science Panel	3/30/23

3.2. Project Timeline



4. References

Briden, A.M., S.L. Clark, B.C. Jorgenson, R.S. Ogle, and J. Cotsifas. 2017. Ceriodaphnia dubia chronic toxicity test variability: Evaluation of water as a source of test variability. Society of Environmental Toxicology and Chemistry North America 38th Annual Meeting. Minneapolis, MN.

Clark, S.L. and A.M. Briden. 2018. Effect of Ceriodaphnia dubia culture hardness on laboratory control test performance. Society of Environmental Toxicology and Chemistry North America 39th Annual Meeting. Sacramento, CA.

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: 1117-1126.

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.

Elphick, J.R.F., K.D. Bergh, and H.C. Bailey. 2011. Chronic toxicity of chloride to freshwater species: Effects of harness and implication for water quality guidelines. Environmental Toxiciology and Chemistry 30:239-246.

Erickson, R.J., D.R. Mount, T.L. Highland, J.R. Hockett, D.J. Hoff, C.T. Jenson, T.J. Norberg-King, and K.N. Peterson. 2017. The acute toxicity of major ion salts to Ceriodaphinia dubia: II. Empirical relationships in binary salt mixtures. Environmental Toxiciology and Chemistry 36:1525-1537.

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.

Jorgenson, B.C., S.L. Clark, A.M. Briden, R.S. Ogle, and J. Cotsifas. 2017. Ceriodaphina dubia chronic toxicity variability: Evaluation of food as a source of test variability. Society of Environmental Toxicology and Chemistry North America 38th Annual Meeting. Minneapolis, MN.

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.

Mount, D.R., R.J. Erickson, T.L. Highland, J.R. Hockett, D.J. Hoff, C.T. Jenson, T.J. Norberg-King, K.N. Peterson, Z.M. Polaske, and S. Wisniewski. 2016. The acute toxicity of major ion salts to Ceriodaphnia dubia: I. Influence of background water chemistry. Environmental Toxiciology and Chemistry 35:3039-3057.

Prosser, K.N., A.M. Briden, S.L. Clark, and B.C. Jorgenson. 2018. Assessing food preparation and storage as sources of Ceriodaphnia dubia culture quality and chronic test variability. Society of Environmental Toxicology and Chemistry North America 39th Annual Meeting. Sacramento, CA.

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

U.S. EPA. 1995. Short-term methods for estimating the chronic toxicity of effluents and receiving waters to west coast marine and estuarine organisms. Office of Research and Development. Cincinnati, OH. EPA/600/R-95/136.

U.S. EPA. 2001a. Final report: Interlaboratory variability study of EPA short-term chronic and acute whole effluent toxicity test methods. Volume 1. Office of Water, Office of Science and Technology. EPA 821-B-01-004.

U.S. EPA. 2001b. Final report: Interlaboratory variability study of EPA short-term chronic and acute whole effluent toxicity test methods. Volume 2. Office of Water, Office of Science and Technology. EPA 821-B-01-005.

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

U.S. EPA. 2002b. Guidelines establishing test procedures for the analysis of pollutants: Whole effluent toxicity test methods; Final rule. Environmental Protection Agency. 40 CFR Part l36 [FRL-7408-6].

U.S. EPA. 2002c. Guidelines establishing test procedures for the analysis of pollutants: Whole effluent toxicity test methods; Final rule. 67 Fed. Reg. 69952-69972 (November 19, 2002).

U.S. EPA. 2016. Whole effluent toxicity methods errata sheet. Office of Water, Environmental Protection Agency. 821-R-02-012-ES.

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
EcoAnalysts, Inc.	Private
Tetra Tech's Ecological Testing Facility	Private