C. dubia QA evaluation study

Expert Science Panel Report-Out

July 5, 2023

Agenda

- Background and goals of the study
- Review of study findings
  - Historical data analysis
  - Baseline intercalibration study
  - Lab visits and roundtable workshop
  - Second intercalibration study
- Preliminary recommendations and guidance

Background and Goals of the Study

- The *C. dubia* reproduction test, an EPA promulgated method, is one of the toxicity tests included in the California Toxicity Provisions
- While most laboratories can perform the test and meet acceptability criteria, stakeholders raised some concerns regarding consistency within and comparability among labs
- To address these concerns, the State Water Resources Control Board and the California Association of Sanitation Agencies funded a study to:
  - Investigate test conditions and factors that can be controlled to reduce intra- and inter-laboratory variability
  - Develop best practices guidance to improve laboratory's performance

Study Questions

1) What are the *C. dubia* test laboratory techniques used by California Environmental Laboratory Accreditation Program (ELAP) accredited laboratories?

2) How does variability in control reproduction and reference toxicant response compare amongst ELAP accredited laboratories?

3) Does standardizing select laboratory techniques reduce intra- and inter-laboratory variability in control reproduction and reference toxicant response?

Study Approach

- Create a governance structure
- Develop a list of potential sources of intra- and inter-laboratory variability via method inventory and analysis of historical data
- Describe potential sources of intra- and inter-laboratory variability via baseline intercalibration
- Conduct lab visits and roundtable workshop to standardize lab techniques that may increase variability in test outcome
- Evaluate efficacy of lab technique standardization via a second laboratory intercalibration
- Develop recommended guidance to minimize intra- and inter-laboratory variability

Stakeholder Advisory Committee

- Katie Fong (SWRCB)
- Amelia Whitson (EPA Region IX)
- Veronica Cuevas (RWQCB)
- Mitch Mysliwiec (Wastewater)
- Jian Peng (Stormwater)
- Sarah Lopez (Agriculture)
- Peter Arth (Private Laboratories)
- Josh Westfall (Public Laboratories)
- Annelisa Moe (NGO)
- Steven Boggs (ELAP)

Expert Science Panel

- Teresa Norberg-King (Formerly US EPA)
- Robert Brent (James Madison University)
- Howard Bailey (Nautilus Environmental)
- Leana Van der Vliet (Environment and Climate Change Canada)
- A. John Bailer (Miami University, Ohio)

Agenda

- Background and goals of the study

- Review of study findings
  - Historical data analysis
  - Baseline intercalibration study
  - Lab visits and roundtable workshop
  - Second intercalibration study

- Preliminary recommendations and guidance

Review of Project Findings: Historical Data

Method Inventory Approach

- Data collected from 17/18 accredited laboratories in California
  - Standard Operating Procedures (SOPs)
  - Quality Assurance Plans (QAPs)
  - Questionnaire
  - One-on-one phone interviews
- Historical data compiled (targeted 30 tests/3 years/laboratory)
  - 551 control samples
  - 452 reference toxicant dose-response samples

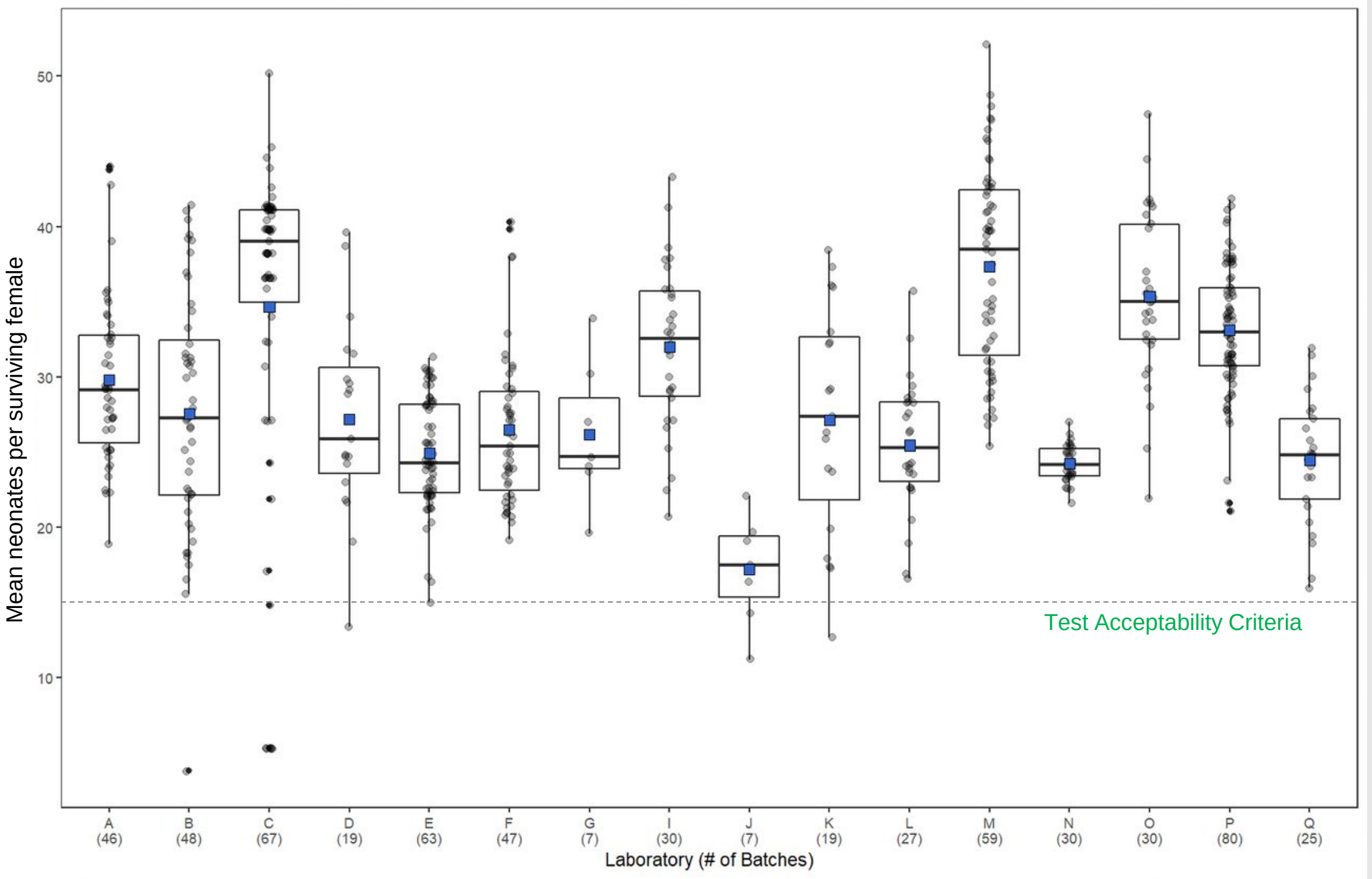
Method Inventory Findings

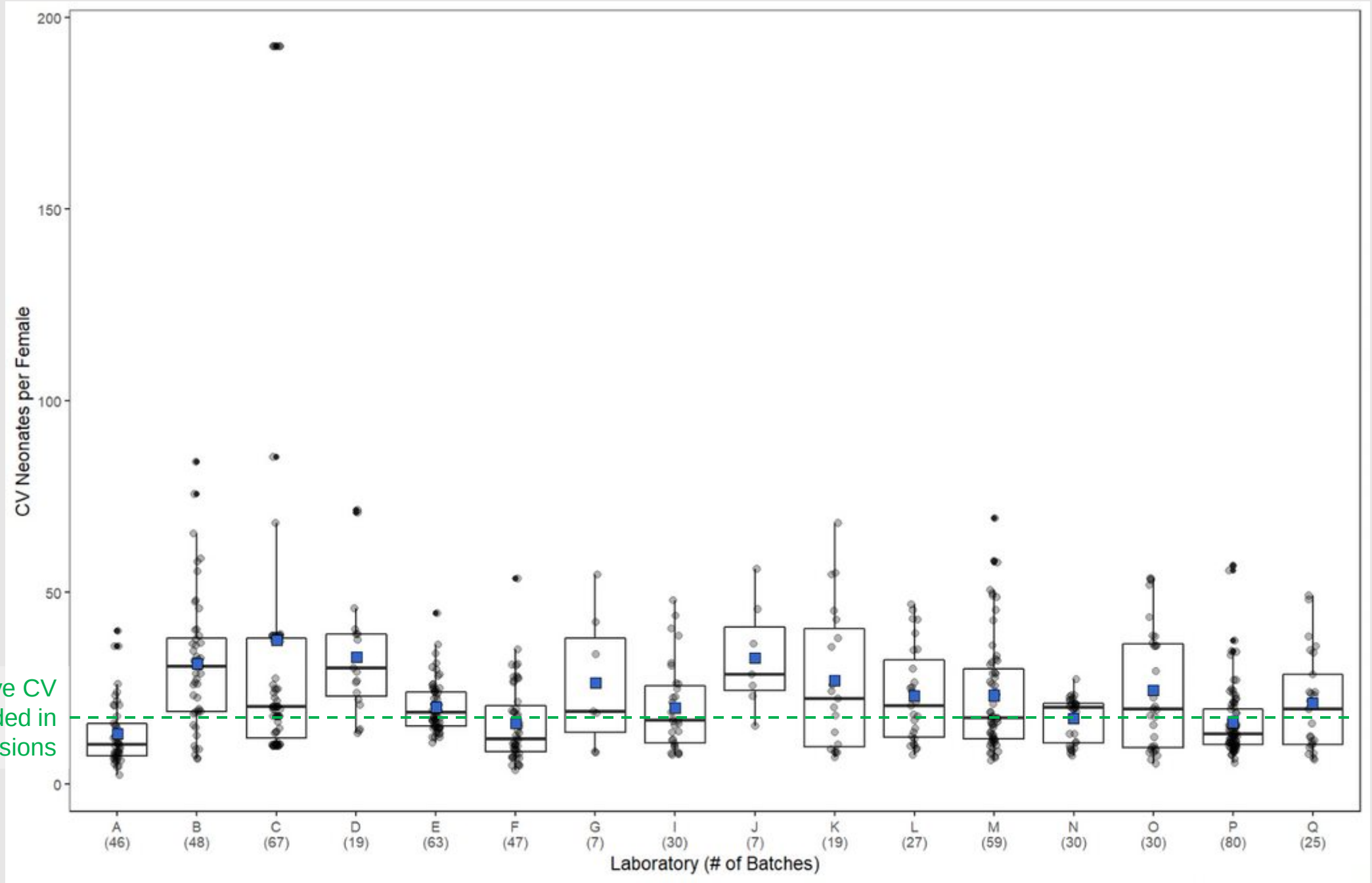
- Wide range of variability in test conditions and procedures

Lab	Dilution Water Formula	Brood Board Renewal	Test Termination Trigger	Chamber Material	Chamber Volume (ml)	Sample Volume (ml)	Light Source
A	EPAMH + vitamins + Se	daily	≥60% with 3 broods	glass	20	20	fluorescent
B	Modified EPAMH	daily	≥60% with 3 broods	glass	30	15	fluorescent
C	EPA Hard + vitamins + Se	daily	≥60% with 3 broods	polystyrene	29.57	15	fluorescent
D	EPAMH + Se	daily	≥60% with 3 broods	polystyrene	29.5	15	fluorescent
E	EPAMH	daily	7 days	polypropylene	29.5	30	fluorescent
F	80% DIW: 20% Perrier	daily	≥60% with 3 broods	polypropylene	29.57	15	fluorescent
G	80% DIW: 20% Perrier	daily	≥80% with 3 broods	glass	26	15	fluorescent
H	80% DIW: 20% Evian	daily	≥70% with 3 broods	polystyrene	26	Not sent	fluorescent
I	Hoheisl +vitamins + Se	every other day	≥60% with 3 broods	plastic	36.9	15	fluorescent
J	EPAMH	daily	≥60% with 3 broods	Not sent	Not sent	15	fluorescent
K	L1650% + vitamins + Se	daily	≥60% with 3 broods	glass	20	15	fluorescent
L	EPAMH + vitamins	daily	≥60% with 3 broods	plastic	29.57	15	fluorescent
M	Modified EPAMH + vitamins	every other day	≥60% with 3 broods	polystyrene	Not sent	15	fluorescent
N	EPAMH + Se	every other day	≥60% with 3 broods	polystyrene	29.57	15	LED
O	EPAMH + vitamins + Se	every other day	≥60% with 3 broods	polystyrene	36.97	15	LED
P	80% DIW: 20% Perrier	daily	≥60% with 3 broods	polystyrene	Not sent	30	fluorescent
Q	80% DIW: 20% Perrier	daily	≥60% with 3 broods	polystyrene	30	15	fluorescent

Historical Data Compilation Findings

- Some labs could not consistently achieve test acceptability criteria for number of neonates per female
- There are differences among labs in mean neonate per female per test
  - Coefficient of variation in neonate production per test
- There are differences among labs in reference toxicant response
- There was no statistical relationship between lab techniques and neonate production, variation in neonate production, or reference toxicant response


Mean neonates per surviving female
50
40
30
20
10
Test Acceptability Criteria
A (46)
B (48)
C (67)
D (19)
E (63)
F (47)
G (7)
I (30)
J (7)
K (19)
L (27)
M (59)
N (30)
O (30)
P (80)
Q (25)
Laboratory (# of Batches)



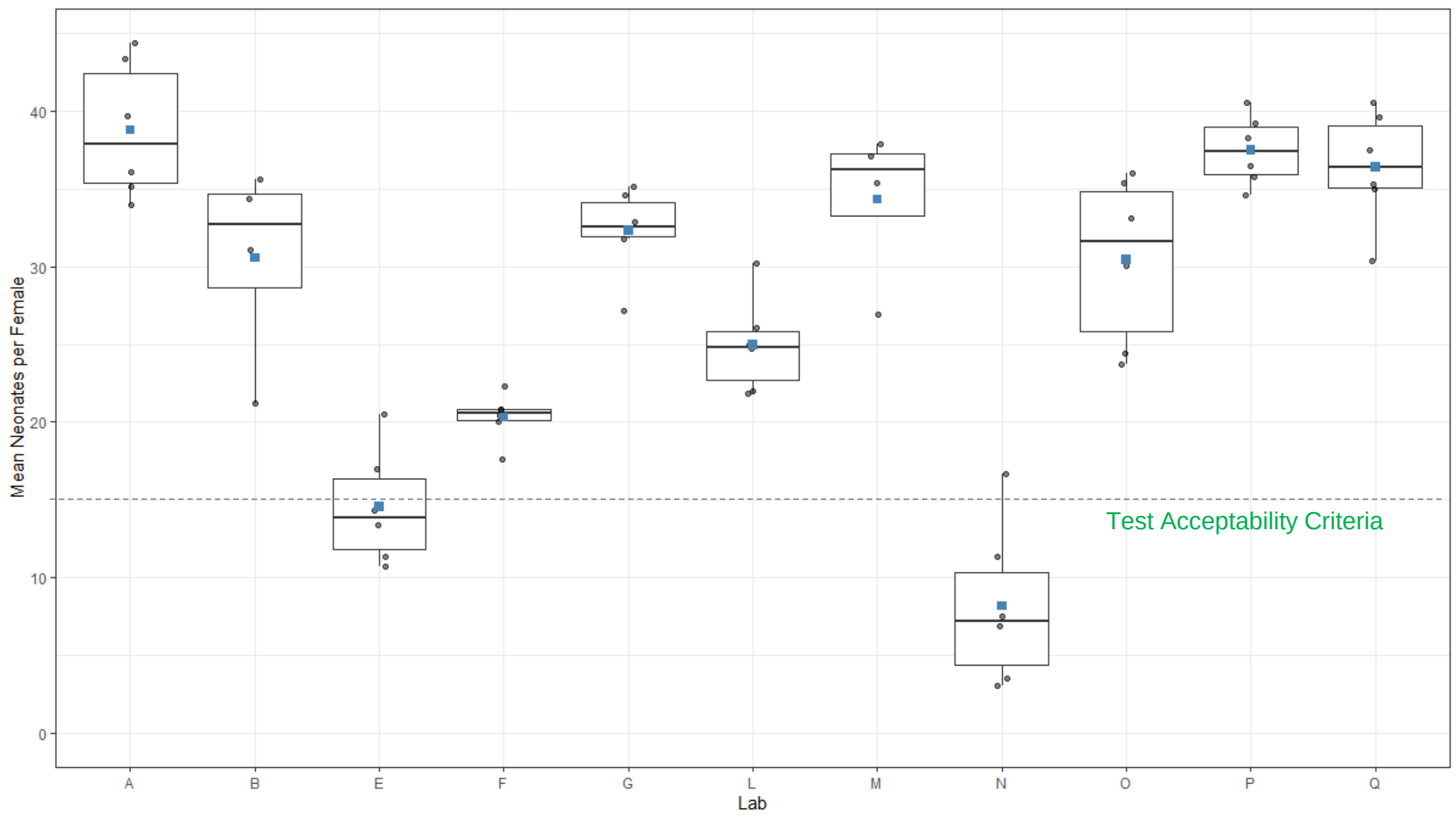
Long term ave CV recommended in toxicity provisions

Review of Project Findings: Baseline Intercalibration

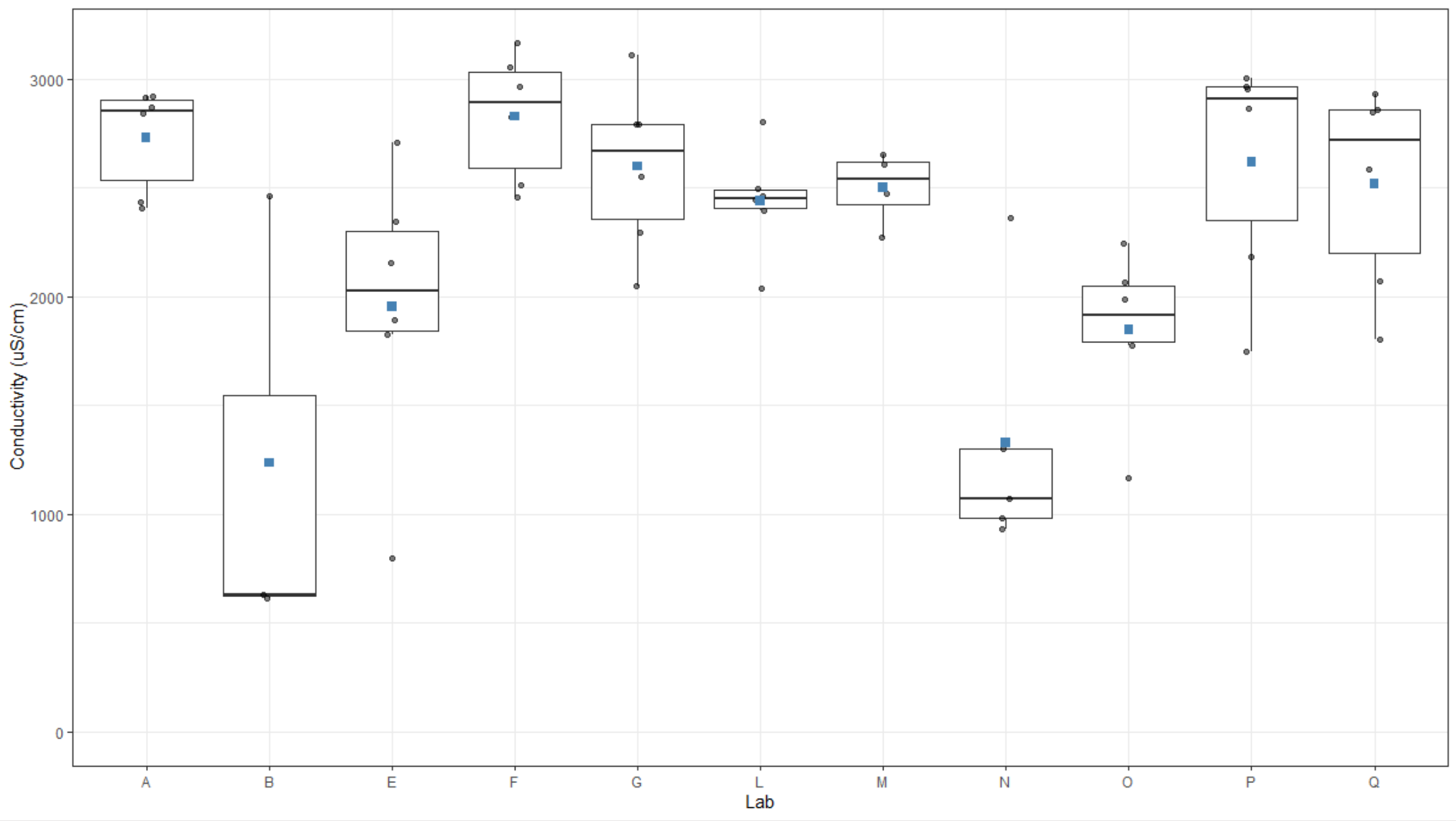
Baseline Intercalibration Approach

- 11 accredited laboratories participated
- Provided homogenized split samples to every lab
  - Unspiked samples: EPA MHW, Dilute Perrier
  - Spiked samples: Sodium chloride (NaCl) in dilute Perrier (5 concentrations), and as powder for lab to dilute with their own dilution water
- Split samples tested three times
- Labs tested samples using existing in-house protocols and techniques
  - But additional data requested (e.g., brood board health)

Mean Neonate Production for Unspiked Perrier and MHW Dilution Waters During the Baseline Intercalibration (N=6 per lab)



IC50 for NaCl spiked Perrier and MHW Dilution Waters During the Baseline Intercalibration (N=6 per lab)



Review of Project Findings:
Lab Visits and Roundtable Workshop

Lab Visits and Roundtable Workshop Approach

- Subset of Expert Science Panel members with testing expertise visited a subset of labs comprising a range of variability
  - Identify potential sources of variability not recorded in bench sheets or SOPs
- Roundtable Workshop brought together all of the accredited labs over two days to identify what lab techniques to standardize
- The group came to consensus agreement on eight items to standardize for the second intercalibration study

Recommended Standardized Test Parameters for the Second Intercalibration Study

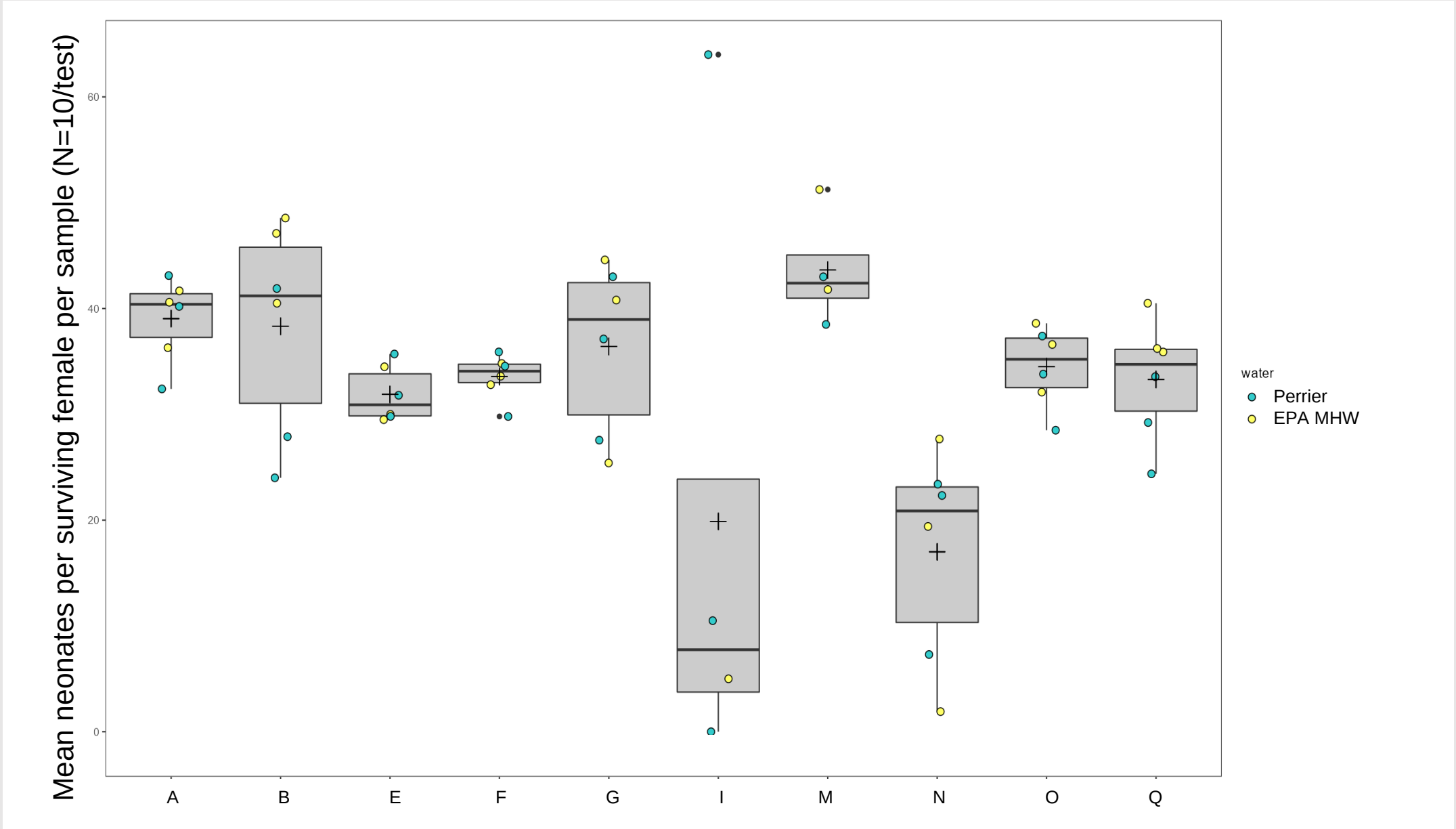
- Evaluate brood board health for 2 weeks prior to testing using a common set of health criteria.
- Limit age of adults used to start the test to 6-10 days old.
- Quantify density of food in test chambers.
- Follow holding times for YCT ≤ 7 days after thawing and algae ≤ 21 days.
- Renew test boards daily within + or – 1 hour window from test initiation.
- Run test for 8 days AND record when labs would have ended the test.
- Document split broods on bench sheets at the time of observation.
- Require randomization by blocking of known parentage AND randomization of cups

Review of Project Findings: Second Intercalibration

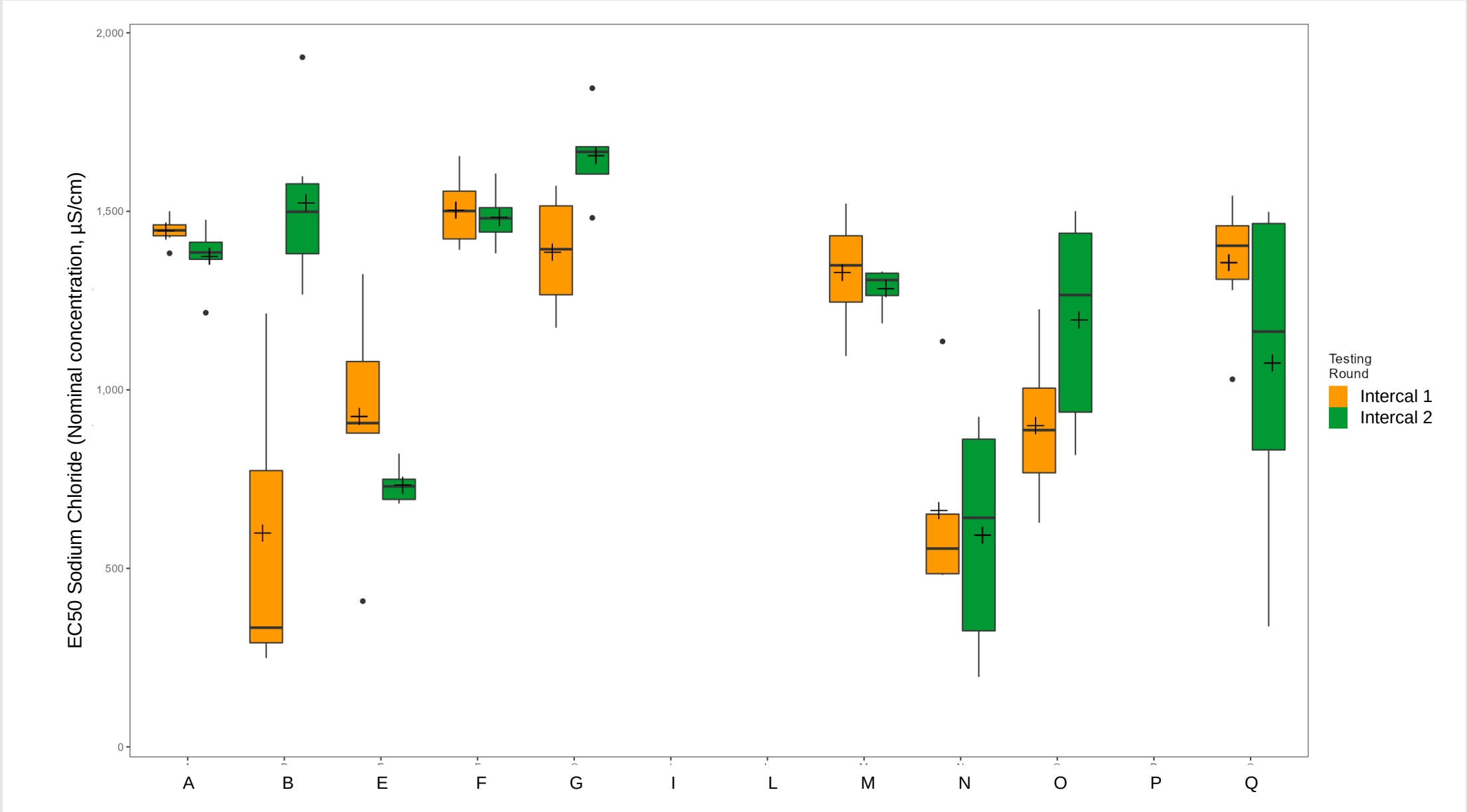
Second Intercalibration Study - Approach

- 10 accredited labs participated
  - 9 repeated from the baseline intercalibration
- Used the same study design as the baseline intercalibration
  - 2 water types, unspiked and spiked samples, conducted 3 times
- Standardized the items identified in the Roundtable Workshop
  - Pre-testing training on ensure consistency on standardized techniques
  - Enhanced documentation to ensure lab techniques are recorded
- Assessed whether implementing standardized parameters reduced intra- and inter-laboratory variability in 2nd intercalibration compared to 1st baseline intercalibration.

Distributions of Mean # neonates/surviving female/test by laboratory in unspiked dilute Perrier or EPA Moderately Hard Water (N=6 per lab)

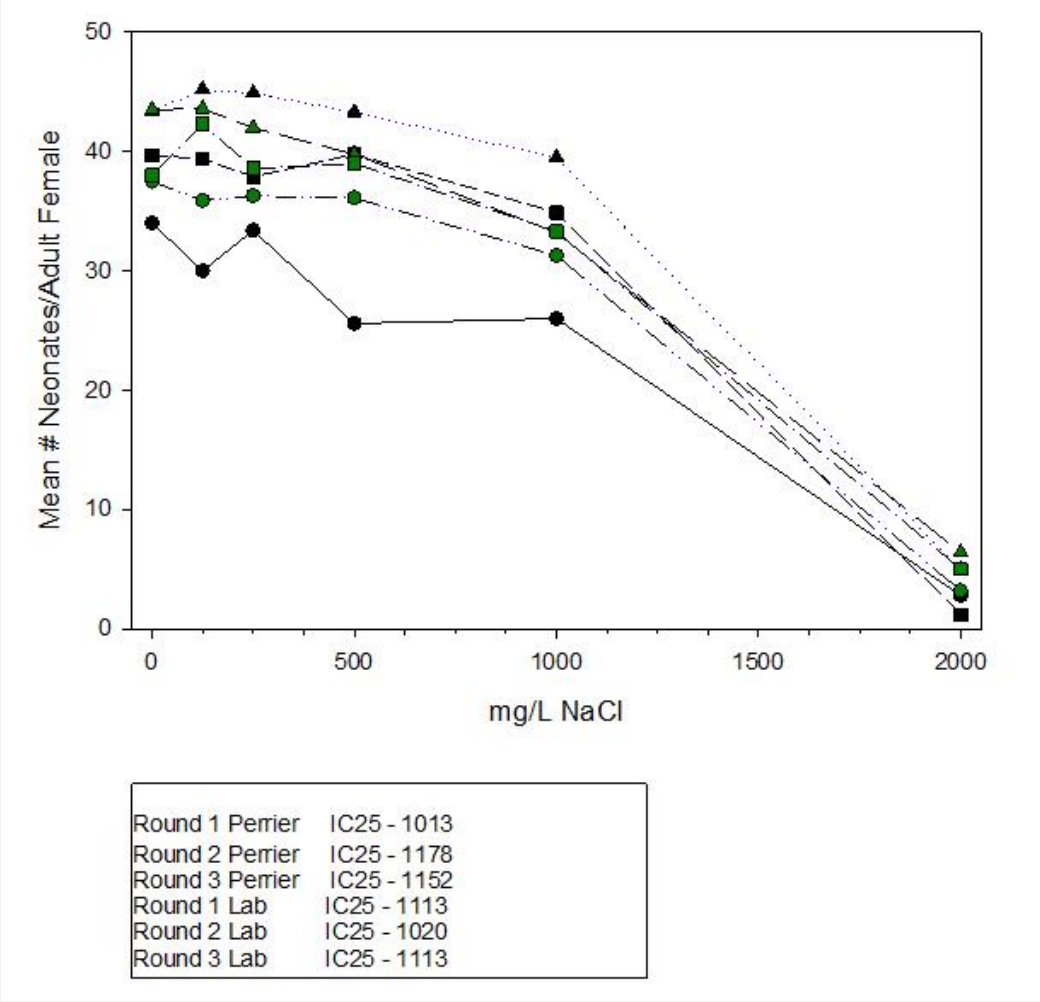


IC50 for NaCl spiked Perrier and MHW Lab Dilution Waters During the Second Intercalibration (N=6 per lab)

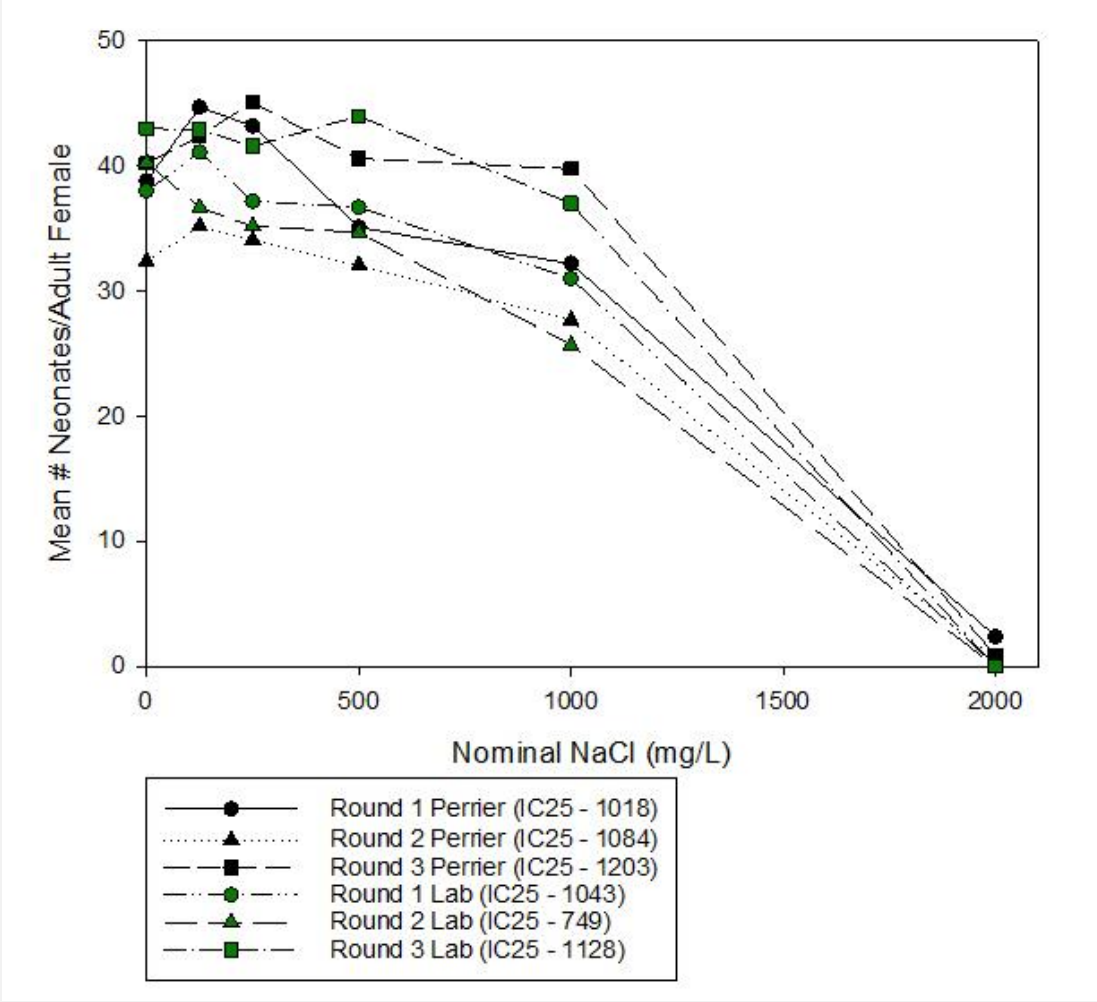


Comparisons of Dose-Response Curves within a Lab

Baseline ILS



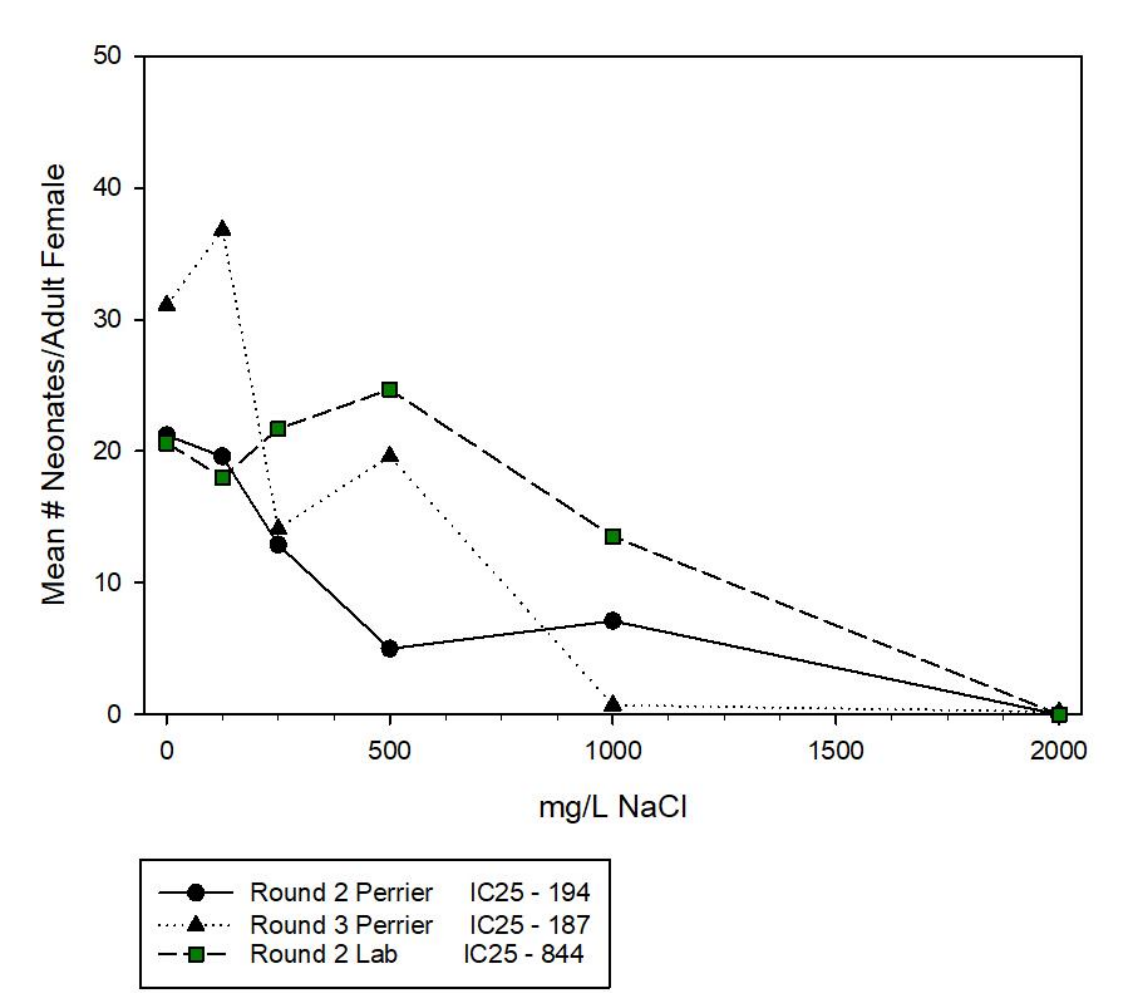
Second ILS



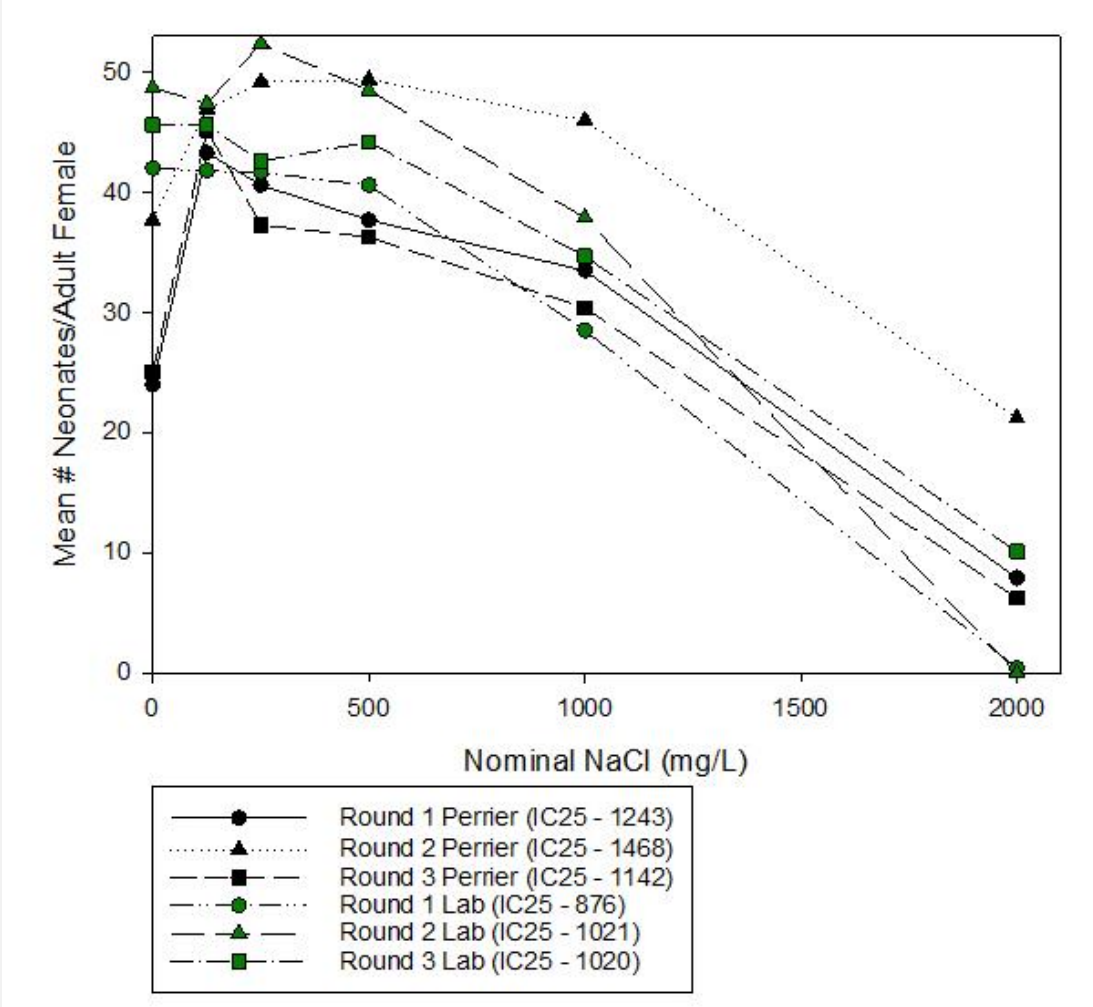
Lab A, considered a lab with consistent quality

Comparisons of Dose-Response Curves within a Lab

Baseline ILS



Second ILS



Lab B, previously considered inconsistent quality, showed improvement between baseline and second ILS

Agenda

- Background and goals of the study
- Review of study findings
  - Historical data analysis
  - Baseline intercalibration study
  - Lab visits and roundtable workshop
  - Second intercalibration study
- Preliminary recommendations and guidance

C. dubia Testing Recommendations

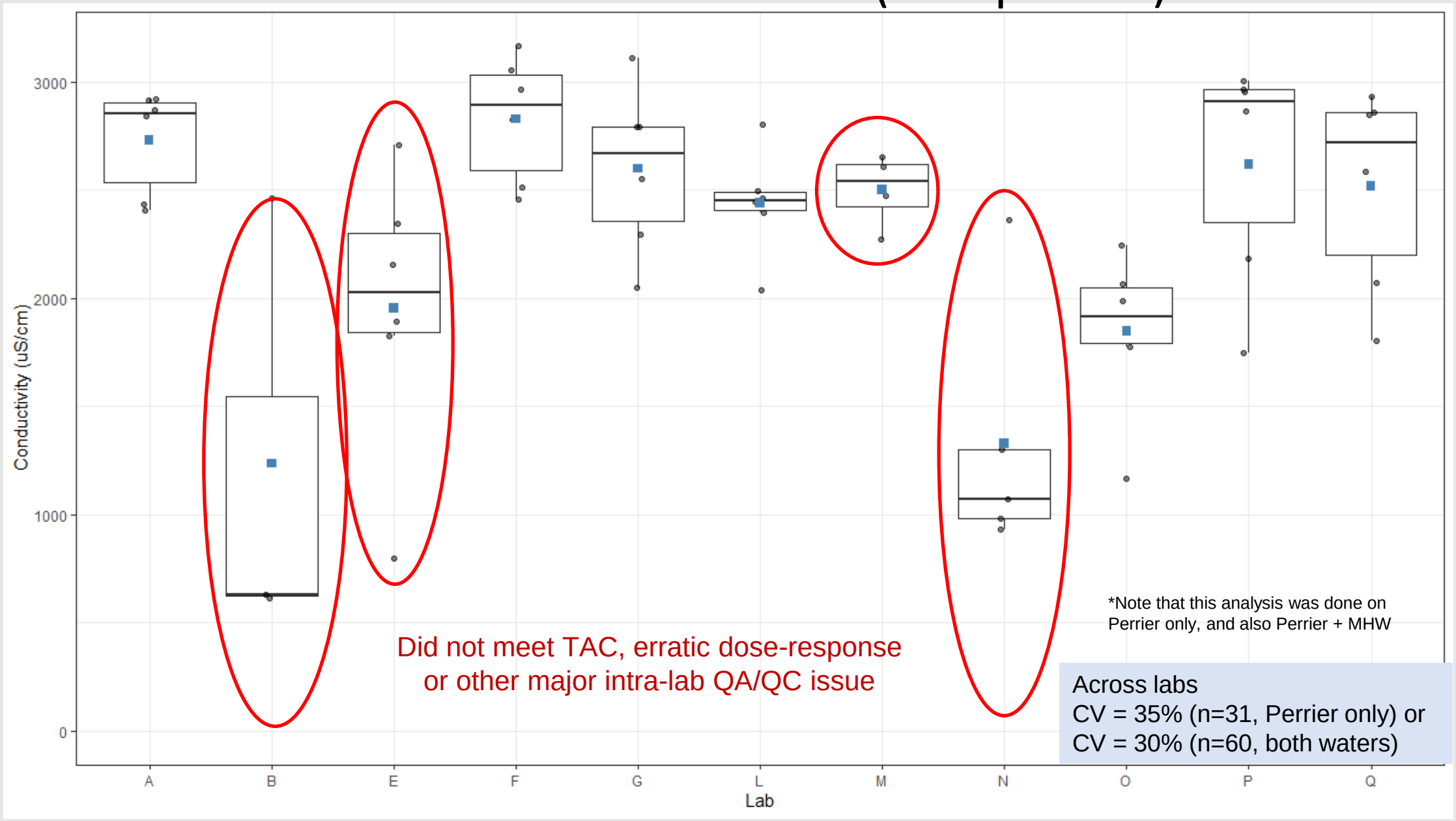
Recommendations Come in Four Categories

- Accreditation
- Method best practices
- Training
- Future studies

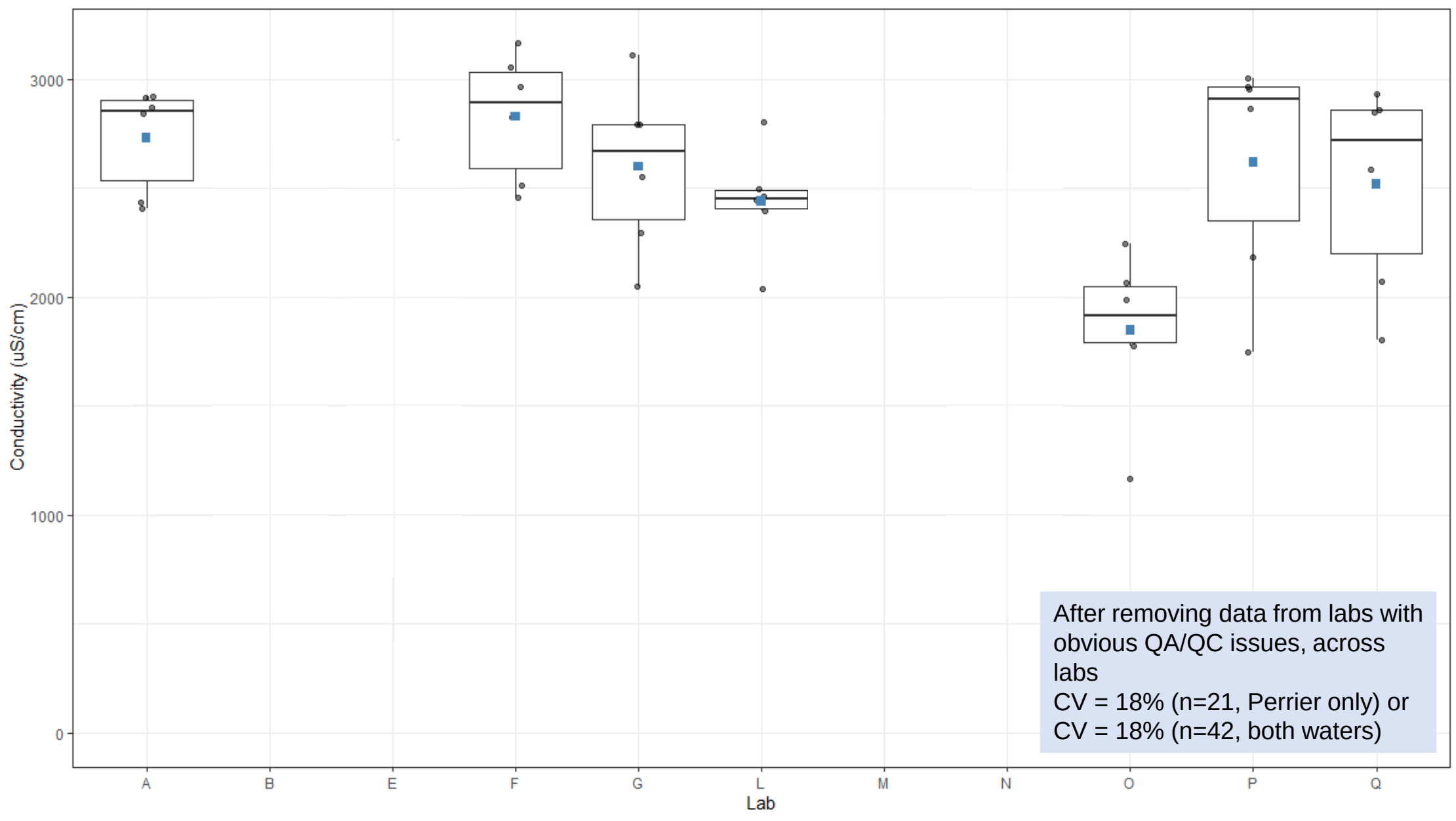
Rationale for Accreditation as a Category

- Some inconsistencies between lab records from the historical records review was observed suggesting internal QA review was uneven.
- ELAP staff seem to be stretched for resources for site visits.
- Quality issues from the baseline intercalibration which normally should have been "caught" at the lab level (e.g., failed TAC) were not caught.
- Goals of existing proficiency testing (PT) program don't seem to fully address the concerns with inter-laboratory variability.

IC50 for NaCl spiked Perrier and MHW Dilution Waters* During the Baseline Intercalibration (N=6 per lab)



IC50 for NaCl spiked Perrier and MHW Dilution Waters During the Baseline Intercalibration (N=6 per lab)



Accreditation Recommendations

- Increase number and /or frequency of PT samples per year
- Submit additional data associated with the PT sample
- Optimize in-lab audits to ensure effective best practices and consistent laboratory procedures

Increase number and /or frequency of PT samples per year

Findings:

- Labs are required to test one PT sample per year

Conclusions:

- Increased frequency and/or sample size will provide necessary data to monitor both inter- and intra-lab performance
- Building a comparative database will help evaluate interlab comparability
- A possible alternative to more PT samples is for all labs to use the same reference toxicant and use that data for interlab performance evaluation

Submit data associated with the PT sample

Findings:

- PT data submitted for accreditation determination is limited to No observed effect concentration (NOEC), Lowest observed effect concentration (LOEC) and IC
- Little is done with the data other than pass/fail response for individual labs

Conclusions:

- Percent minimum significant difference (PMSD) and CV must be included
- Additional data on control performance, culturing, and testing must be provided and evaluated to help identify remedial actions
- PT results should be made public (lab anonymous) in a timely fashion; one benefit would be allowing labs to self-correct

Optimize in-lab audits to ensure effective best practices and consistent laboratory procedures

Findings:

- ELAP audit checklist may be incomplete
- Auditors may not apply the current checklist consistently across laboratories

Conclusions:

- ELAP checklist needs to be updated with performance goals in mind, and consider both method-specific as well as quality system issues
- ELAP should follow-up to verify that changes are implemented appropriately and consistently
- Consider expanding the number of auditors
- Consider regular meetings or workshops to discuss assessment findings and promote communication

Recommendations Come in Four Categories

- Accreditation
- **Method best practices**
- Training
- Future studies

Rationale for Best Practices Recommendations

Phone interviews, site visits and roundtable discussion have shown that laboratories have different interpretations of the method requirements.

Best practices recommendations are intended to

- Reinforce the ‘must do’ which are requirements of the test method
- Provide guidance for the ‘should do’

Best Practices Recommendations

Must do

- End the test when 60% of surviving control females have produced 3rd brood, within a 2-hr window of test initiation
- Randomize test chambers on the test board
- Independently quantify food density and loading rate added to test chambers
- Store reagents appropriately (e.g., salts must be in a desiccator)
- Renew test solutions daily within + or - 1 hour of test initiation time
- Type 2 source water must be produced according to the manual's instructions.

Best Practices Recommendations

Should do

- Conduct ongoing detailed quantitative assessments of brood board health
- Document split brood on bench sheets and record supporting observations
- Use different rooms to maintain cultures and conduct the tests
- Use appropriate surrogate to measure water quality parameters
- Improve record-keeping of lab practices

Conduct ongoing detailed quantitative assessments of brood board health

Findings:

- Level of details documented for brood board health is highly variable among labs, preventing any comparisons among labs
- Culture health was one of the critical factors influencing performance and variability in the intercalibration study

Conclusions:

- Labs should record the daily number of young per female, unhealthy or dead females, unhealthy or dead young, and presence of males.
- This should be an ongoing assessment, but at a minimum should be monitored for a two-week period prior to testing.

Document split brood on bench sheets and record supporting observations

Findings:

- Review of bench sheets indicated that broods are misidentified
- Some labs conduct their determination of split brood after the test solely based on numbers of young
- Accurate tracking of broods is key to determining test termination criteria and test acceptability criteria

Conclusions:

- Labs should document the size of young and appearance of female brood pouch on bench sheets at the time of the assessment

Use different rooms to maintain cultures and conduct the tests

Findings:

- Site visits revealed that several labs are housing their culture (and dilution water) in the same room they are conducting the test
- Labs also frequently conduct other types of testing in-house, including hazardous waste samples that may contain volatile contaminants

Conclusions:

- Separate rooms are needed to ensure that volatile compounds found in test samples do not contaminate the cultures or dilution water stock

Use appropriate surrogate to measure water quality parameters

Findings:

- Many labs do not use a true surrogate of the same size and volume as the test chambers
- Surrogates for water quality are not always treated like the test chambers (e.g. no food added)

Conclusions:

- Water quality should be recorded under conditions as similar to the test samples as possible (e.g. size and volume of test chambers, addition of food, storage with test samples)
- Manipulating samples (e.g., pouring) prior to taking measurements may influence the results and should be avoided

Improve record keeping of lab techniques

Findings:

- Several lab techniques that could influence test outcome and data quality are not routinely documented. Such information was retrieved during phone interviews and lab visits

Conclusions:

- Labs should keep records of dilution water and food preparation and holding times

Recommendations Come in Four Categories

- Accreditation
- Method best practices
- **Training**
- Future studies

Rationale for Training as a Category

- Communication through training on the WET testing could ensure that the permittee or permitting authority have the information that they need to make informed decisions.

Training Recommendations

- Augment auditors' knowledge of the method requirements
- Training will aid the development and implementation of training performance goals for laboratory personnel
- Formalize training documentation for standard testing among labs
- Provide guidance to regulated parties to evaluate WET test data

Improve auditors' knowledge of the method requirements

Findings:

- This study identified lab practices that are not consistent or violate the EPA test methods. Protocols may not have followed the promulgated methods and documented in the labs' SOP.
- Auditors may not consistently address deviations from testing procedures methods across labs (from collection of samples to completing the test and analyzing the data).

Conclusions:

- An ongoing training program will help auditors understand method requirements vs suggestions, and provide a forum for discussion
- Training will help ensure that all auditors perform their evaluation more uniformly

Develop and implement training performance goals for laboratory personnel

Findings:

- Experience and turnover rate vary among laboratories
- Training programs are lab-specific with different evaluation criteria

Conclusions:

- Training should be conducted for main tasks of the job description using clear and quantifiable performance goals (e.g., balance procedures, food preparation and feeding, surrogates, target mean or CV neonate production, correct identification of split brood, data quality issues, etc.)
- Training data generated, and frequency should be documented

Formalize training documentation to be standard among labs

Findings:

- There is no easy way to evaluate training and experience across labs because they implement and document their training practices differently

Conclusions:

- Establish an on-going workgroup with ELAP and labs to develop a standard approach to developing and documenting lab training practices
- Develop method-specific training guidance to remind labs of key factors influencing test performance (e.g., water type, test set up and termination)

Provide guidance to regulated parties to evaluate WET test data

Findings:

- Regulated parties have different levels of familiarity with the test method
- They have also expressed concerns in the consistency of the test method implementation among labs
- There may be misconceptions in what is unacceptable variability and how to account for environmental factors influencing the test

Conclusions:

- Stakeholder education will facilitate their ability to review lab performance before selecting a lab and discuss data produced
- Produce interpretation guidance with data submitted to clients explaining test design, results and QA measures

Recommendations Come in Four Categories

- Accreditation
- Method best practices
- Training
- **Future studies**

Rationale for Future Studies as a Category

- Future studies would need to add significant value to the findings of the current study.
- In general, this study's findings were comprehensive relative to the study objectives; i.e., an evaluation of the extent and sources of variability in the test method.
- Consequently, pending detailed analysis of the second intercalibration study, no additional studies are being recommended at this time.

Caveat: Evaluate the impact of dilution waters on test outcomes

- Stakeholders have raised this issue in the past, particularly in the context of ion imbalance and hardness.
- This issue is related to specific matrix effects, and is not a broad source of variability across laboratories.
- Moreover, the current level of variability in test performance among laboratories likely confuses this issue.
- Once laboratories have achieved a uniformly acceptable level of performance, it may be appropriate to address this concern in greater detail.

Preliminary Summary of Findings

Study Approach - Status

- Create a governance structure **- COMPLETED**
- Develop a list of potential sources of intra- and inter-laboratory variability via method inventory and analysis of historical data **- COMPLETED**
- Describe potential sources of intra- and inter-laboratory variability via baseline intercalibration **- COMPLETED**
- Conduct lab visits and roundtable workshop to standardize lab techniques that may increase variability in test outcome **- COMPLETED**
- Evaluate efficacy of lab technique standardization via a second laboratory intercalibration **– IN PROCESS**
- Develop recommended guidance to minimize intra- and inter-laboratory variability **– IN PROCESS**

Study Questions – Preliminary Answers

1) What are the *C. dubia* test laboratory techniques used by California Environmental Laboratory Accreditation Program (ELAP) accredited laboratories?
   Observed a wide range of variability in test conditions and procedures

2) How does variability in control reproduction and reference toxicant response compare amongst ELAP accredited laboratories?
   Based on test quality metrics such as neonate production (mean, CV), consistent dose-responses within a lab and comparability among labs, there are several high-quality labs available for compliance testing in California

3) Does standardizing select laboratory techniques reduce intra- and inter-laboratory variability in control reproduction and reference toxicant response?
   While data analysis is incomplete, initial results suggest improvement for some labs between the baseline and the second intercalibration studies.

Next Steps

Project Deliverables

- Review second intercalibration results — *Scheduling for before **Jul 30***
- Main Guidance Document — *draft for review **Aug 30***
  - Provides the rationale/discussion/limitations for the recommendations
- Appendices — *draft for review **Aug 30***
  - Historical data analysis
  - Intercalibrations results
- Final compiled package (Panel consensus, Stakeholder review) — **Sep 30**