Test of Significant Toxicity (TST)

Proposed Statistical Approach for Analyzing Toxicity Test Data

1 Take the Sample

2 Run the Tests

3 Record the Biological Response (Data)

4 Analyze Data and Make a Decision

Toxic vs. Non-Toxic

TST

5 Reasonable Potential Determination

6 Permit Requirement(s) Decision

Permit Limit or Monitoring

TST is NOT a Change to the WET Test Methods

Labs still conduct the same approved test methods:

- Same organisms
- Same food
- Same testing procedures
- Same test acceptability criteria

What is the question we want to answer using WET testing?

Is the effluent or sample toxic?

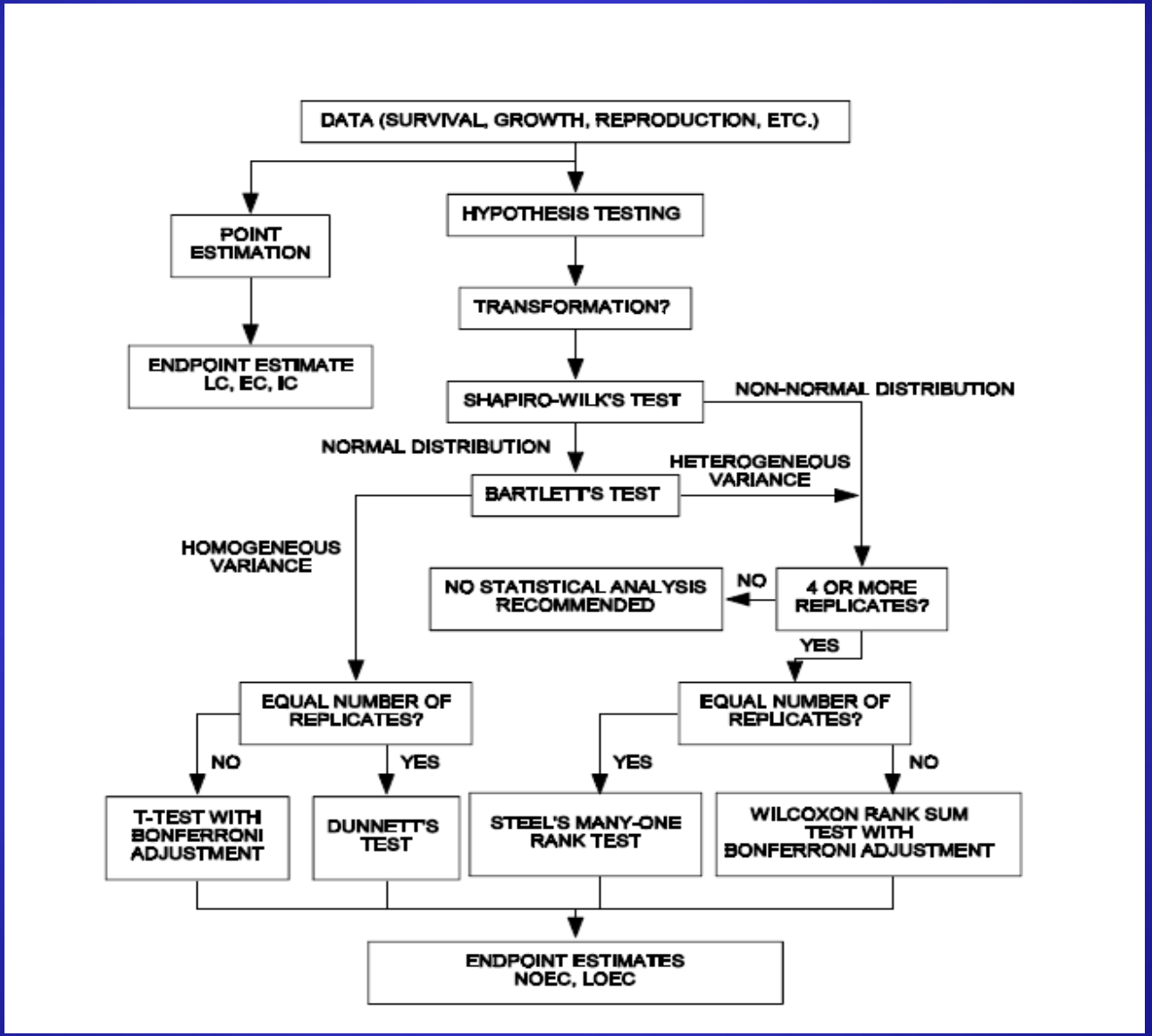
WET is not an experimental research program – WET is a regulatory program.

Statistics should give you a yes or no answer.

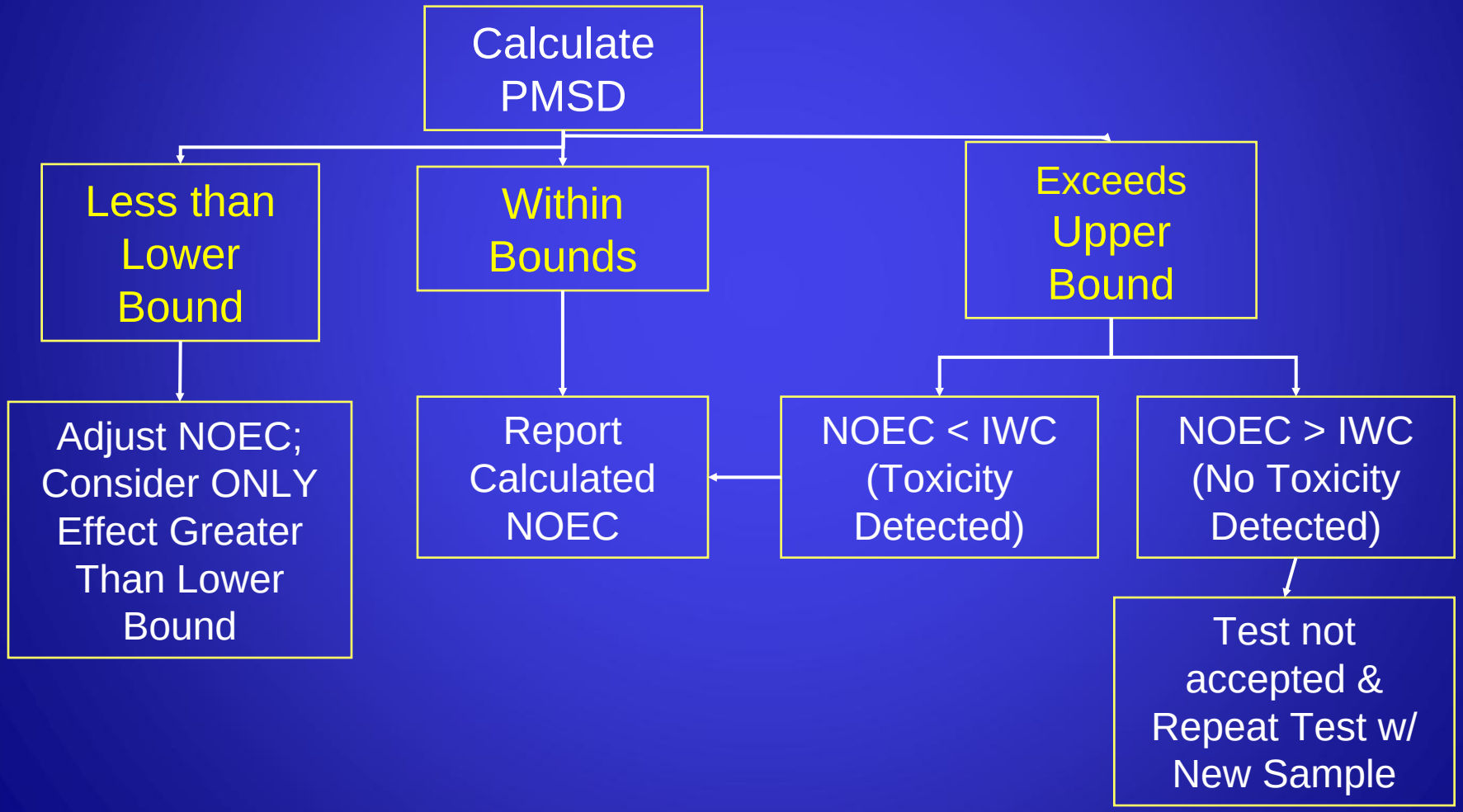
TST is designed to give a yes or no answer using rigorous, peer-reviewed statistics.

Data Analysis Using TST is More Straightforward, Streamlined, and Simpler to Use than Current Approaches

EPA Chronic NOEC Analysis



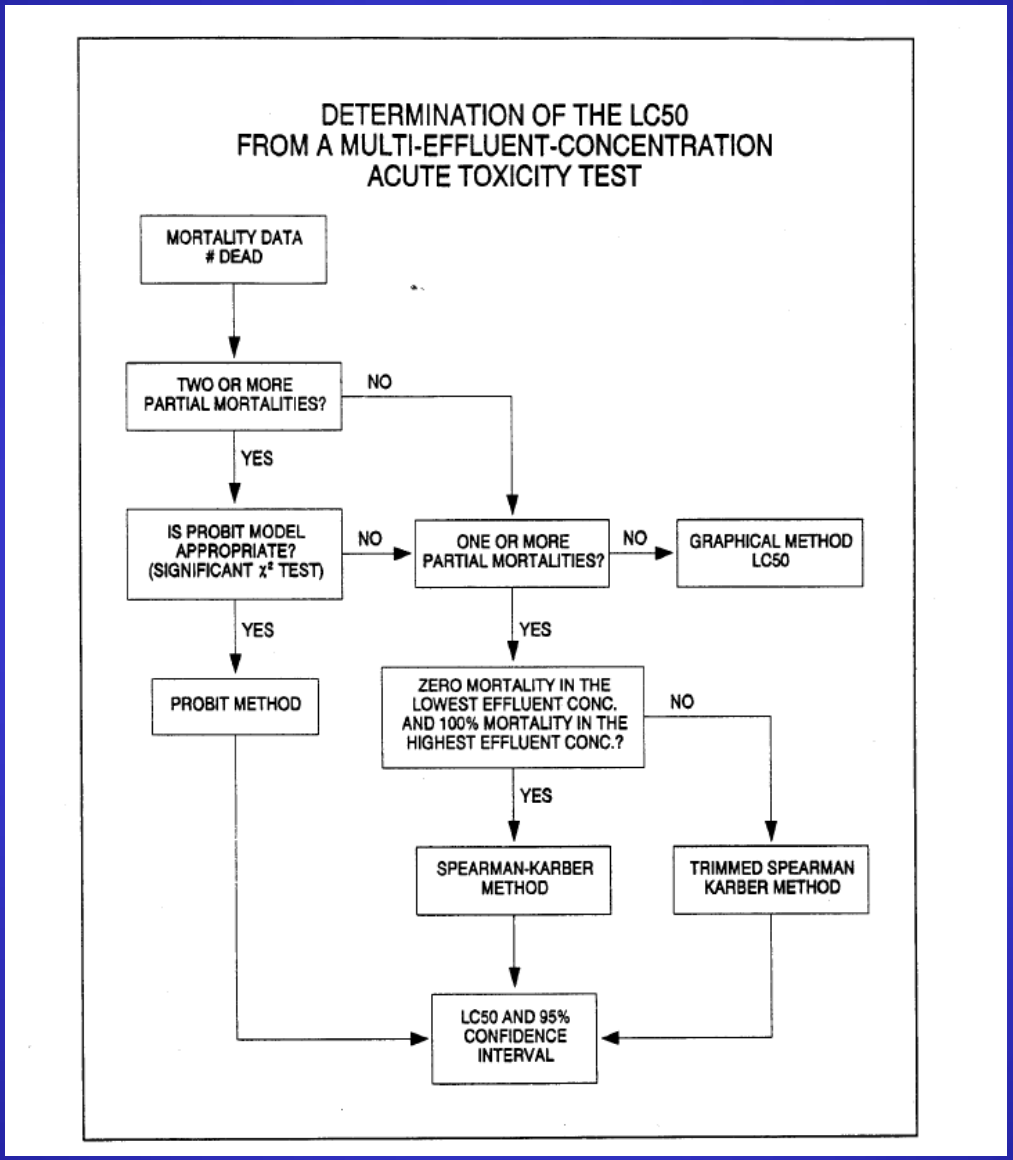
Percent Minimum Significant Difference (PMSD)



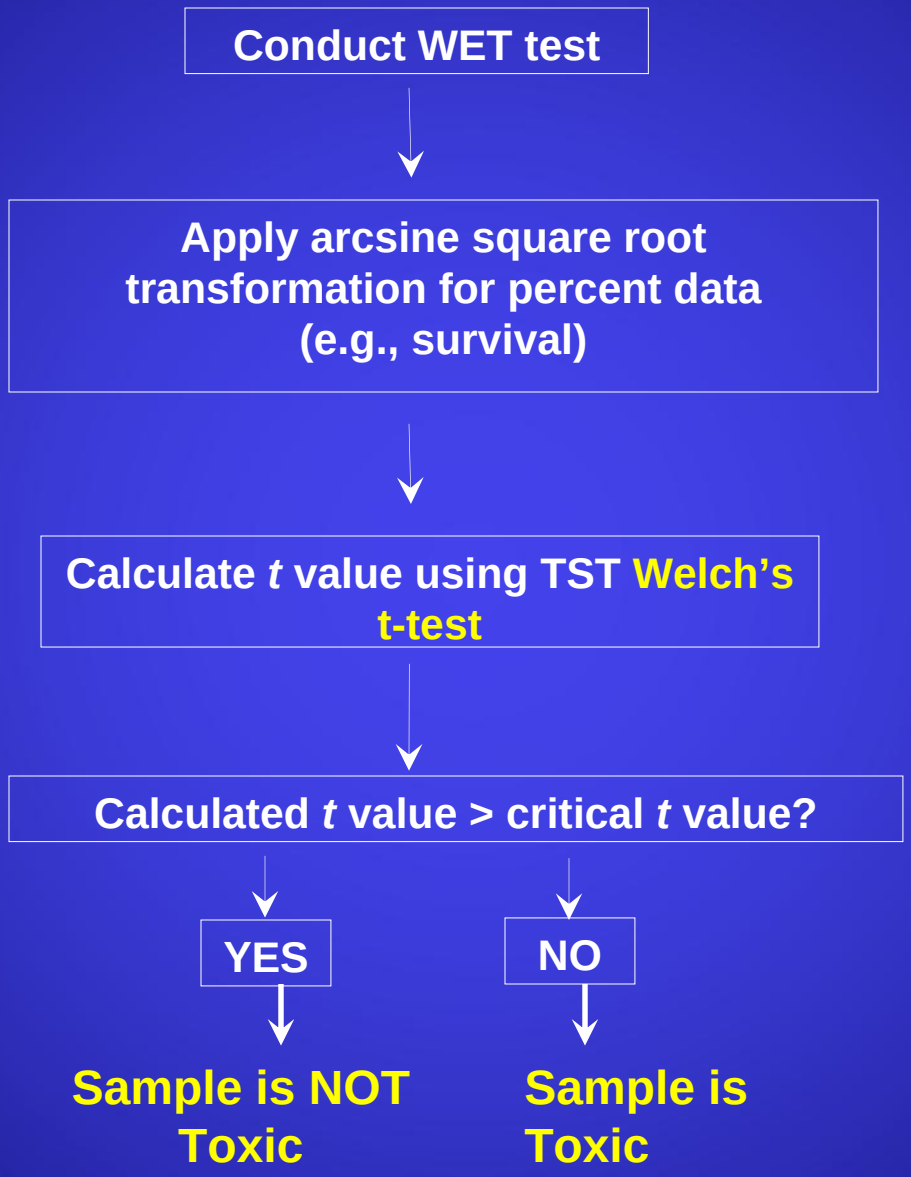
EPA Chronic IC25 Analysis

- Linear interpolation method recommended
- Not appropriate for non-linear responses
- Point estimate may not be correct depending on within-test variability
- Confidence intervals may not be calculated due to inappropriate data
- Wide disagreement on the correct model(s) to use for point estimates

EPA Acute LC50 Analysis



TST Analysis Flowchart



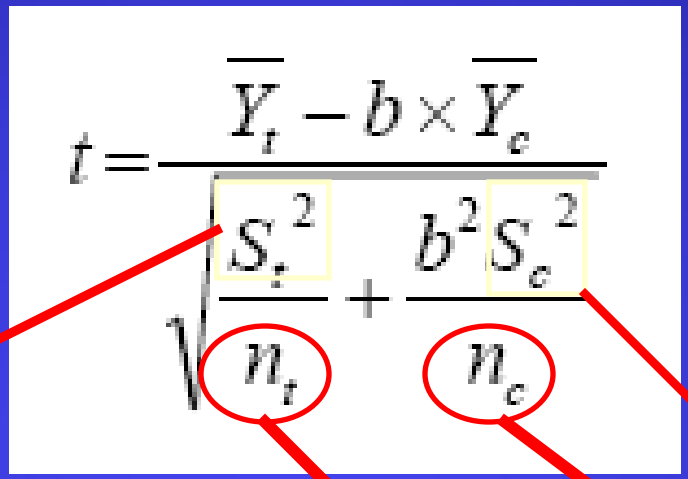
The t-Test approach is nothing new in WET analysis

EPA recognized the t-test approach in its promulgated methods:

APPENDIX H of Chronic Manuals: "SINGLE CONCENTRATION TOXICITY TEST - COMPARISON OF CONTROL WITH 100% EFFLUENT OR RECEIVING WATER
To statistically compare a control with one concentration, **such as 100% effluent or the instream waste concentration, a t-test is the recommended analysis. "**

Welch's t-test is a generalized form of the t-test that is robust when there are unequal variances or unequal sample sizes. Welch's t-test has been around since 1947.

TST Formula



$$t = \frac{\overline{Y_t} - b \times \overline{Y_c}}{\sqrt{\frac{S_t^2}{n_t} + \frac{b^2 S_c^2}{n_c}}}$$

- Labs can control n_t and n_c
  - Number of Control and Treatment Replicates
  - Doing more than the minimum required can help
- And S_t^2 and S_c^2
  - Control and Treatment Variance
  - Good lab QA/QC helps

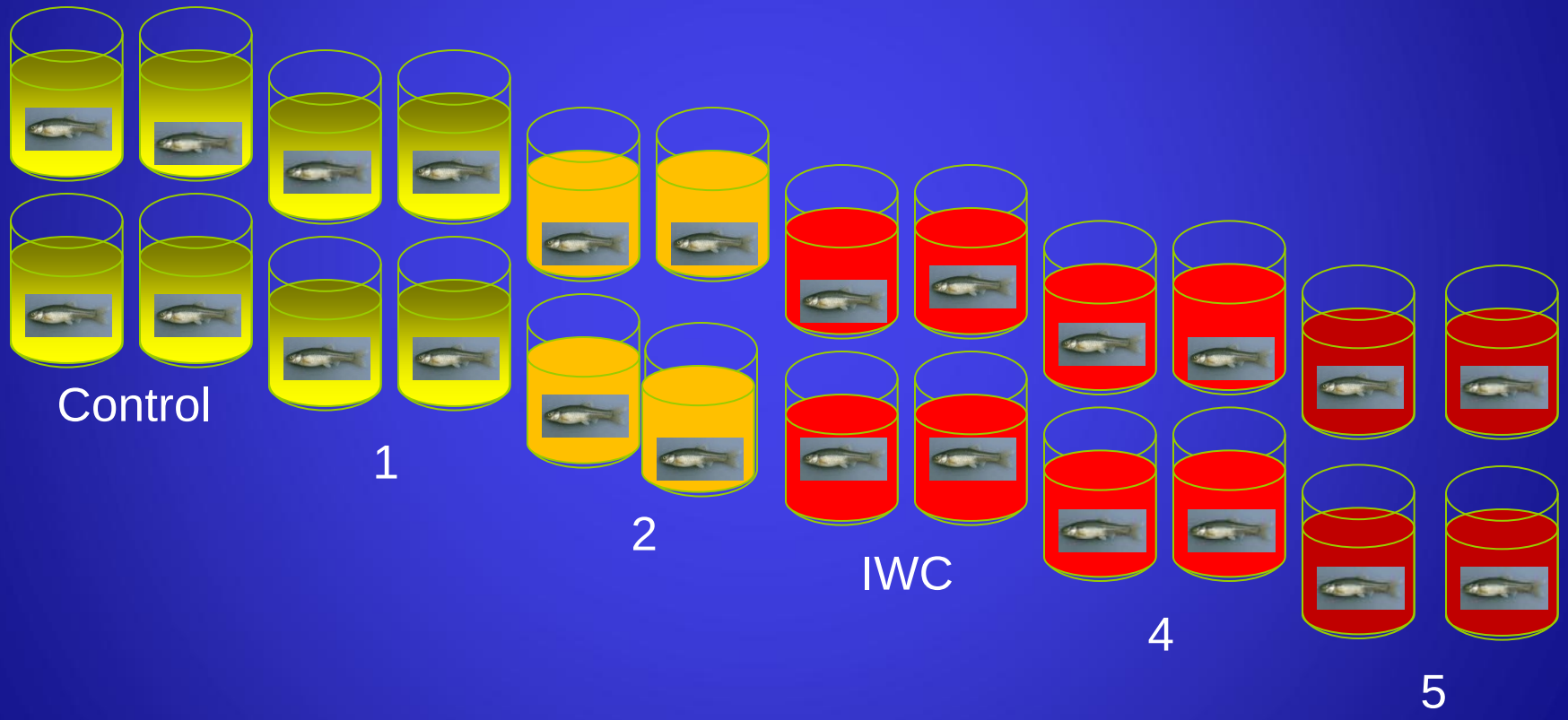
Only need a two concentration test design using TST: Control and the IWC

- About 50% less costly than multi-concentration tests used with NOEC or point estimate approaches

Fathead minnow chronic test

Current approach

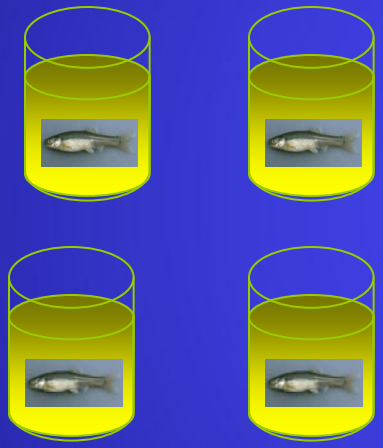
240 fish total



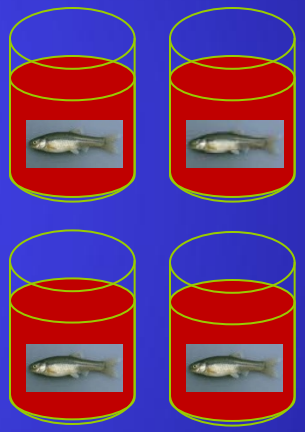
Fathead minnow chronic test

TST approach

80 fish total



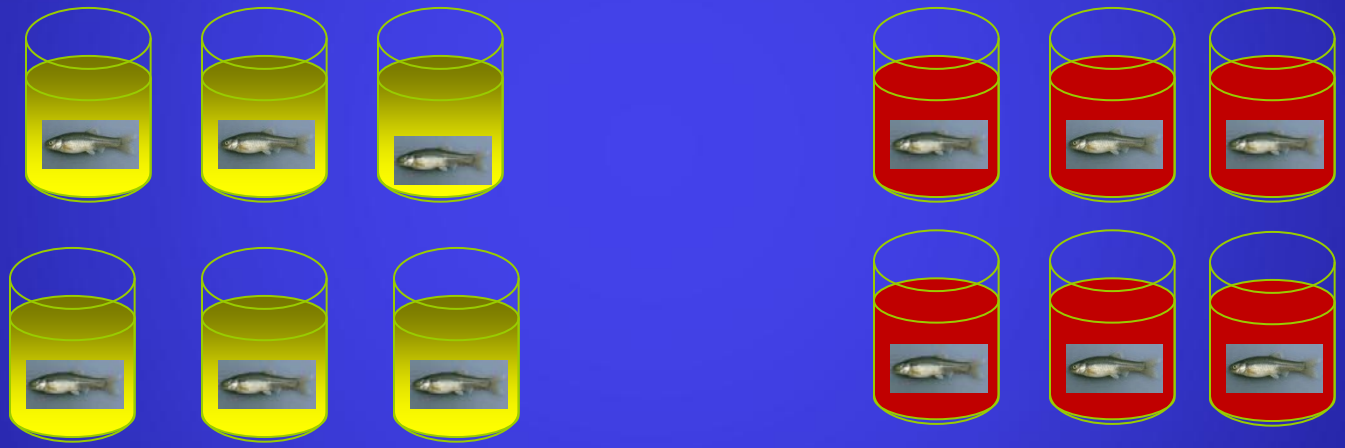
Control



IWC

Fish Chronic Test Design with greater test power

2 more reps for the IWC and control
120 fish total



Control

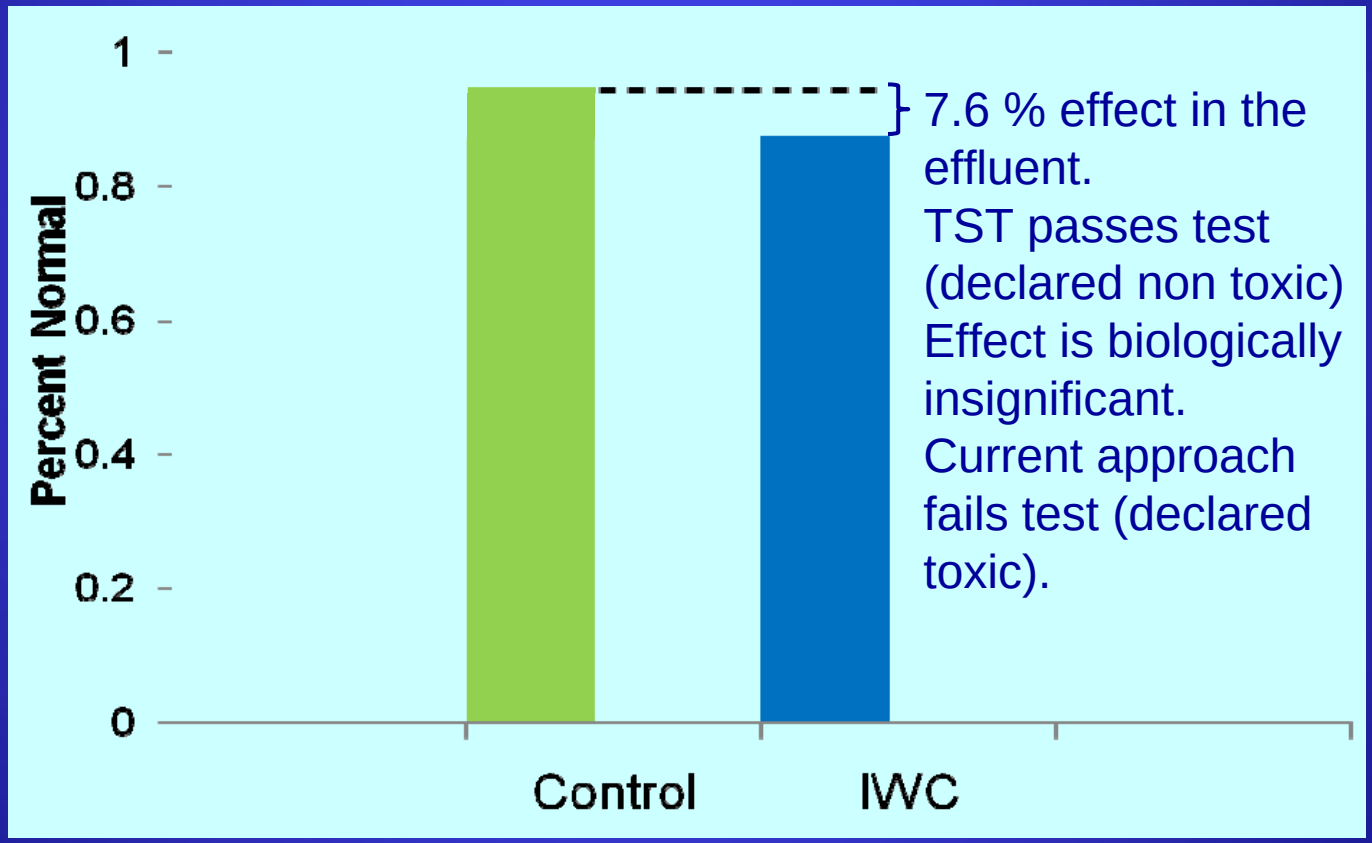
IWC

TST Addresses Error to the Permittee as well as Errors to the Environment

Result is better decision-making for WET

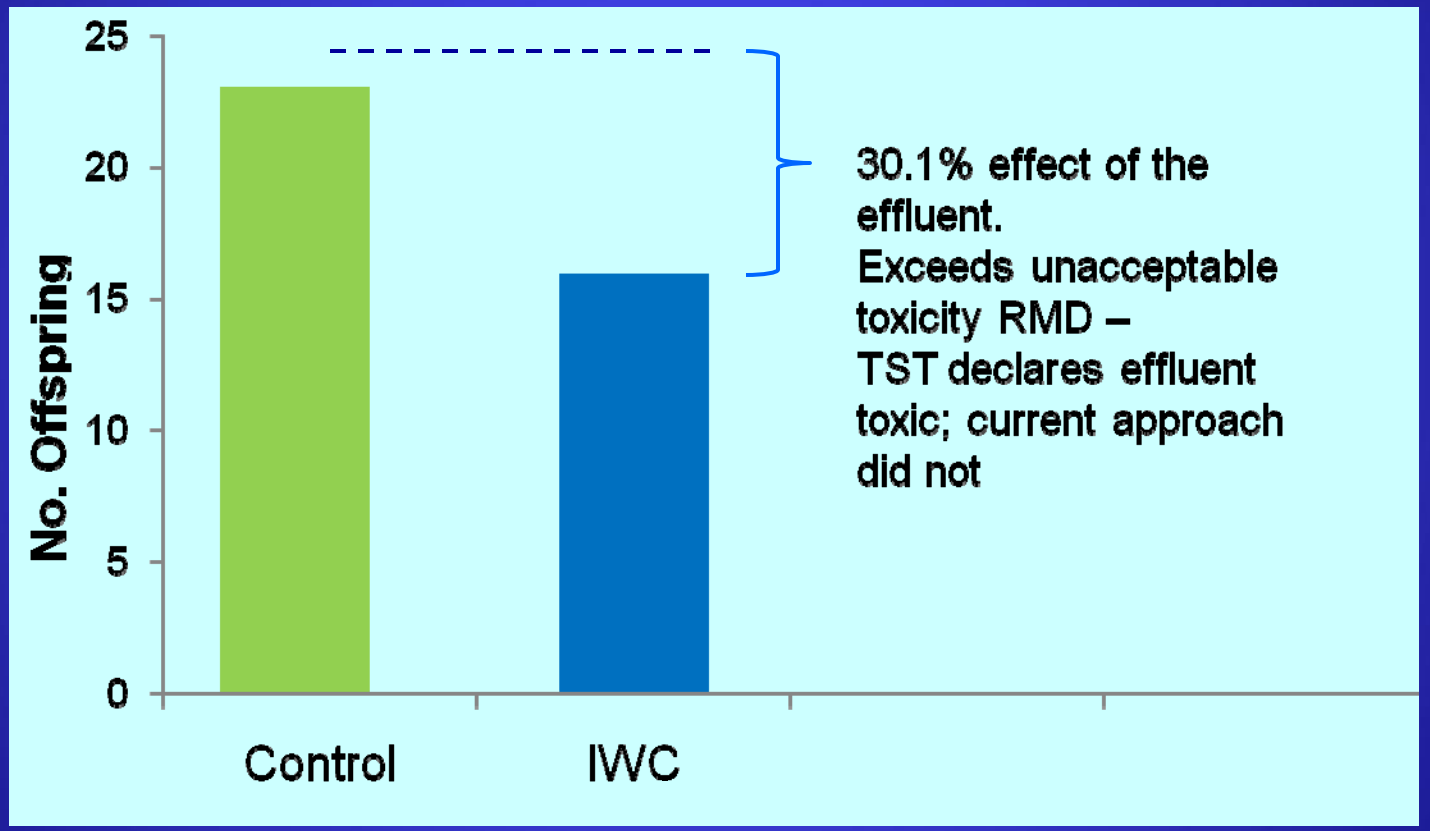
TST rewards high quality WET data

Red Abalone Larval Development Test



TST does not reward poor quality data if toxicity is unacceptable

Ceriodaphnia reproduction test



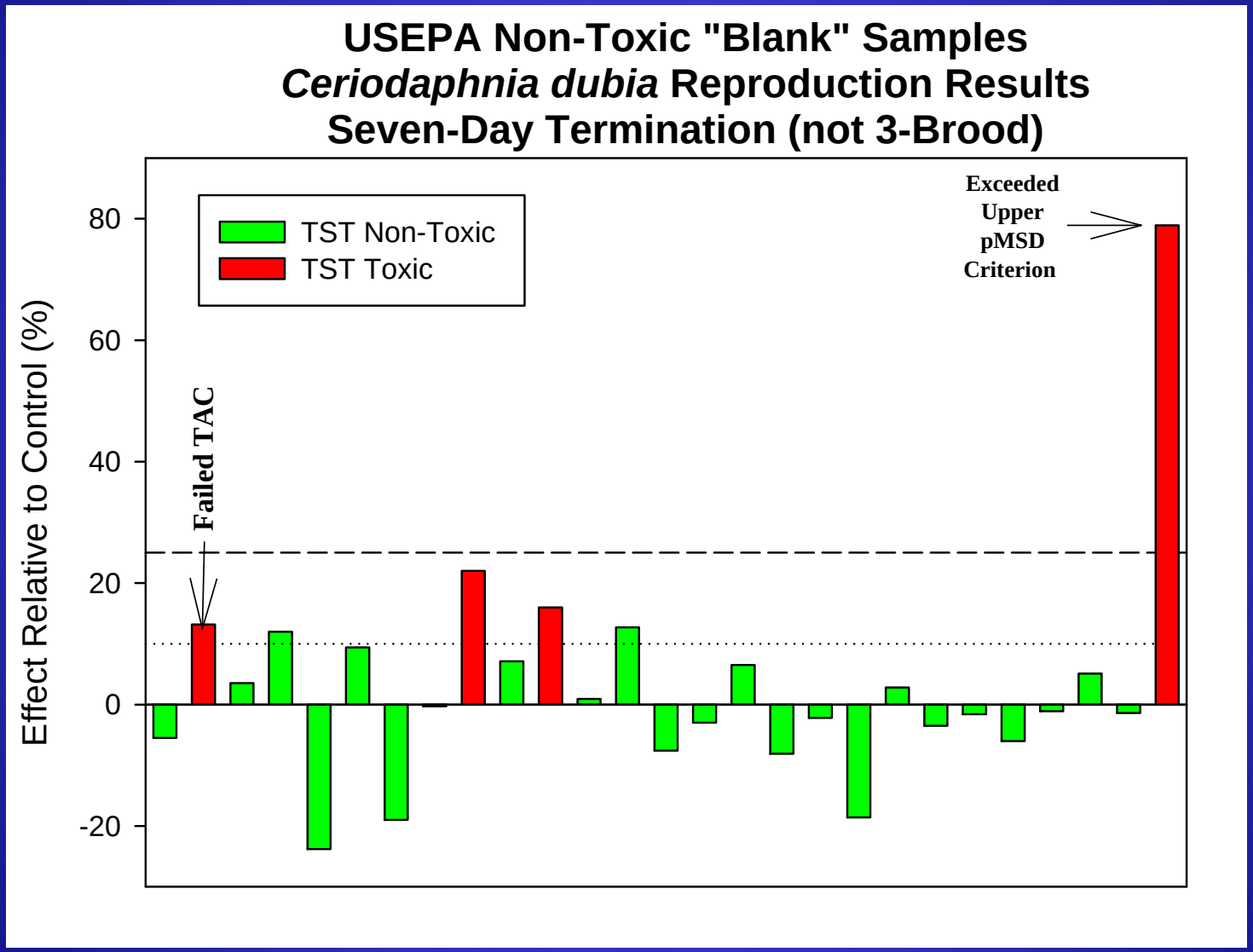
Answers to Technical Comments Regarding TST

Claim: False Positive Rate of TST is > 5% (14.8%) based on EPA 1999 "Blank" Study

Need to distinguish statistics from test performance

Statistical error is NOT the same thing as measurement error

EPA Inter-lab Blank Results



One test exhibited an 80% effect of the “Blank” sample

Mean control reproduction = 19.4
Mean sample reproduction = 4.1
NOEC = 12.5%

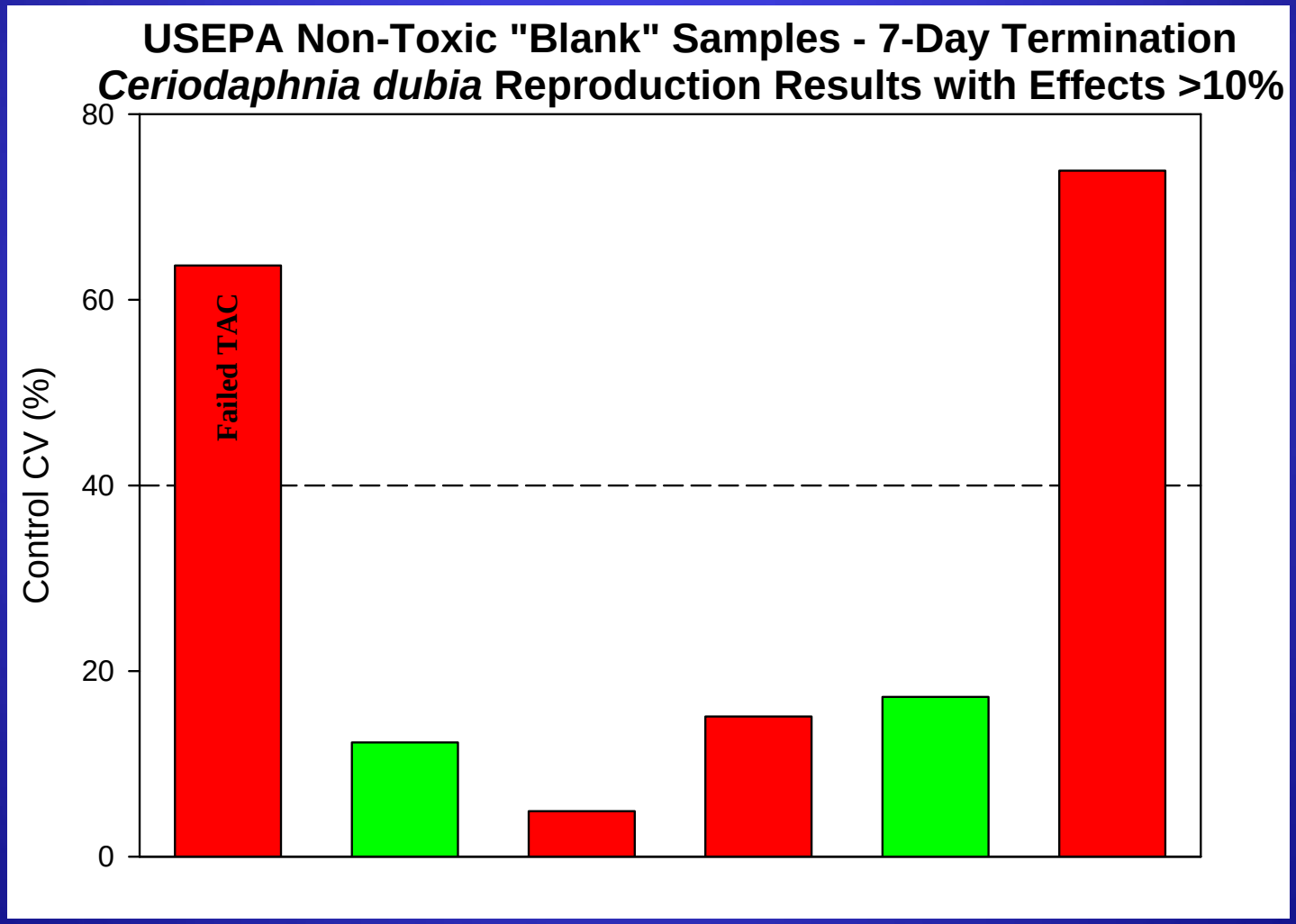
Either the lab received a really toxic sample, or there was some type of measurement error

No statistics will help this situation

A second test failed EPA's test acceptability criteria

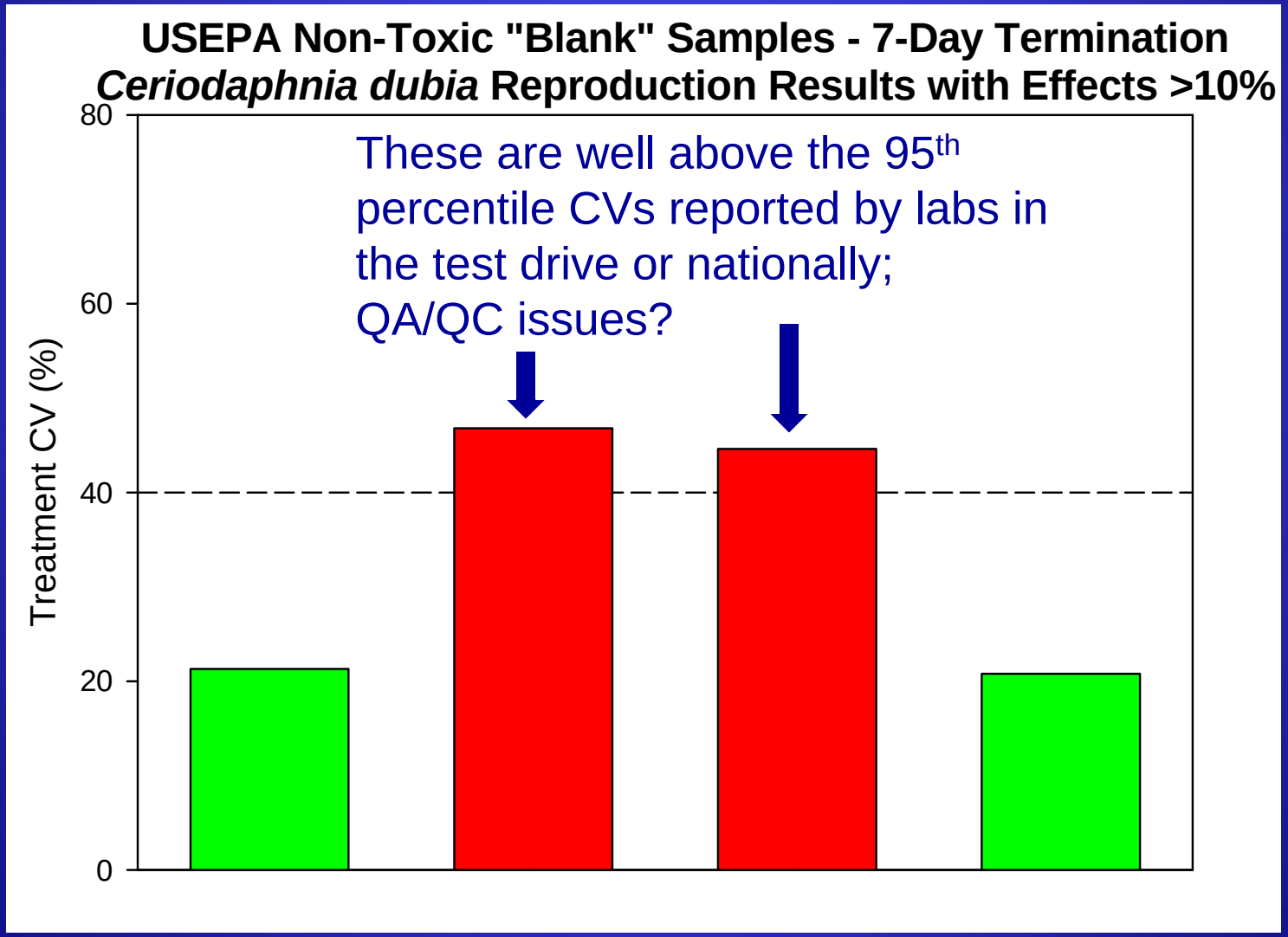
- Test is invalid – Lab QA/QC issue
- Also indicated by extremely high control CV (variability)

EPA Inter-lab Blank Results: Control Variability



What about the remaining 2 tests declared toxic using TST?

EPA Inter-lab Blank Results: Sample CV



Conclusions regarding alleged false positive rate of TST

- Claim is incorrect and overblown; 2 of the 4 tests that would be declared toxic using TST were either invalid or demonstrated high toxicity **This is <u>Not</u> a statistical issue**
- The remaining 2 tests also had QA/QC issues and are suspect.
- Even so, 2 out of 27 blank tests = 7.4%, well within the population error rate of 5% given such a small sample size (27 tests)

Error rates refer to population statistics, not single tests

"It may also be worthwhile to remind the readers that we erroneously tend to use these error rates (P-value) as some sort of statistical evidence obtained FROM A GIVEN TEST RESULT as applicable to THAT PARTICULAR TEST ALSO. The value of these error rates is only in the sense of "long run frequency" of repeated sampling (or WET testing in the present context) as envisaged by Neyman and Pearson in their classic paper on Testing of Hypothesis (also termed as Acceptance sampling)."

- From an anonymous reviewer for paper submitted to ***Integrated Environmental Assessment and Management***

TST Test Drive

Purpose of Test Drive

- Address concerns raised at the November 2010 Board workshop
- Determine whether TST would result in a significant change in WET data interpretation as compared to current approach (NOEC)

Who was involved in the Test Drive?

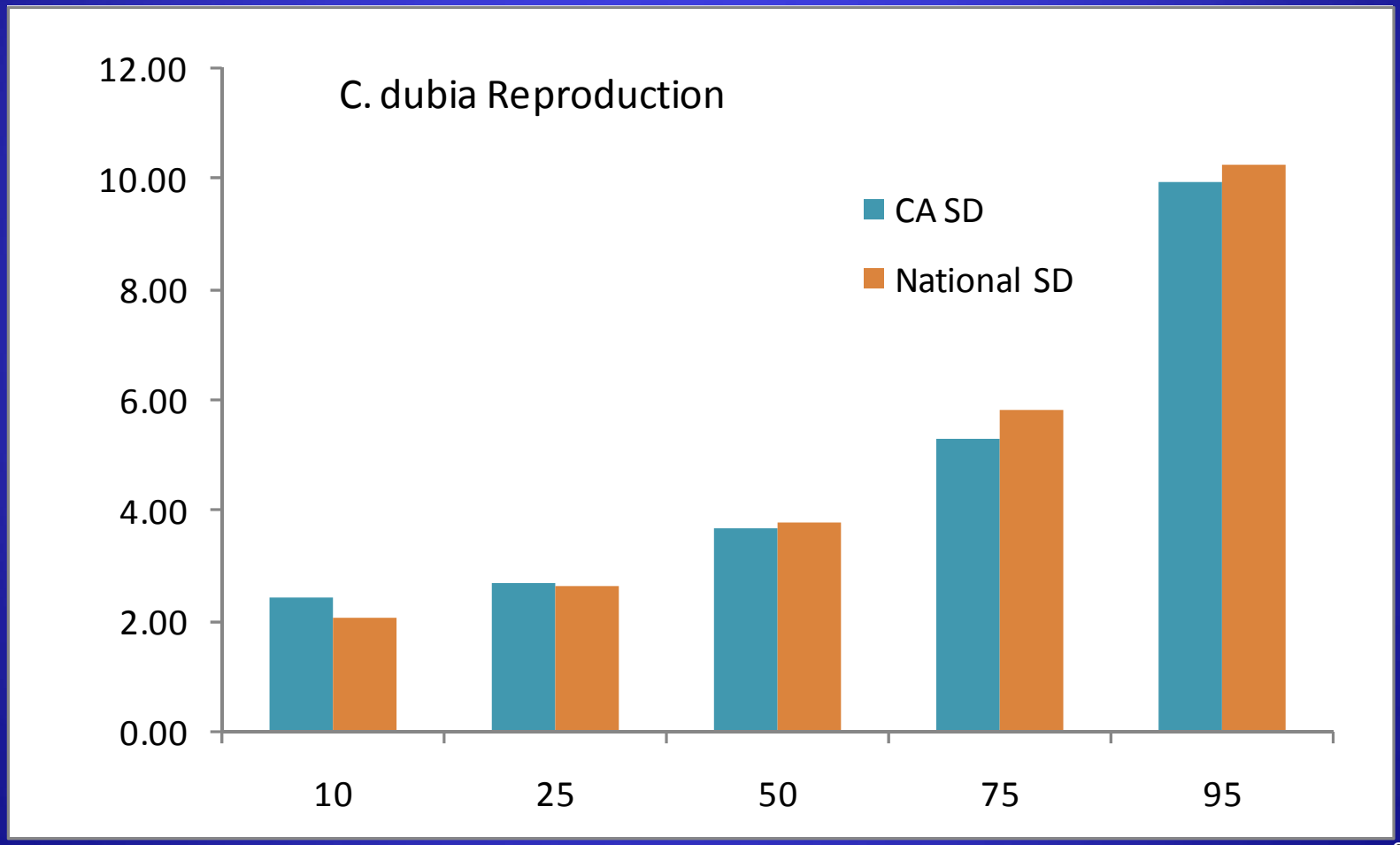
- 18 dischargers in California
- More than 8 laboratories
- Several small, underprivileged communities

What was evaluated in the Test Drive?

- All of the WET methods commonly used in California
- Total of 775 valid WET tests
- Results compared using TST versus the NOEC approach for each test
- Effects of test performance on results using TST and NOEC

Test Drive Results

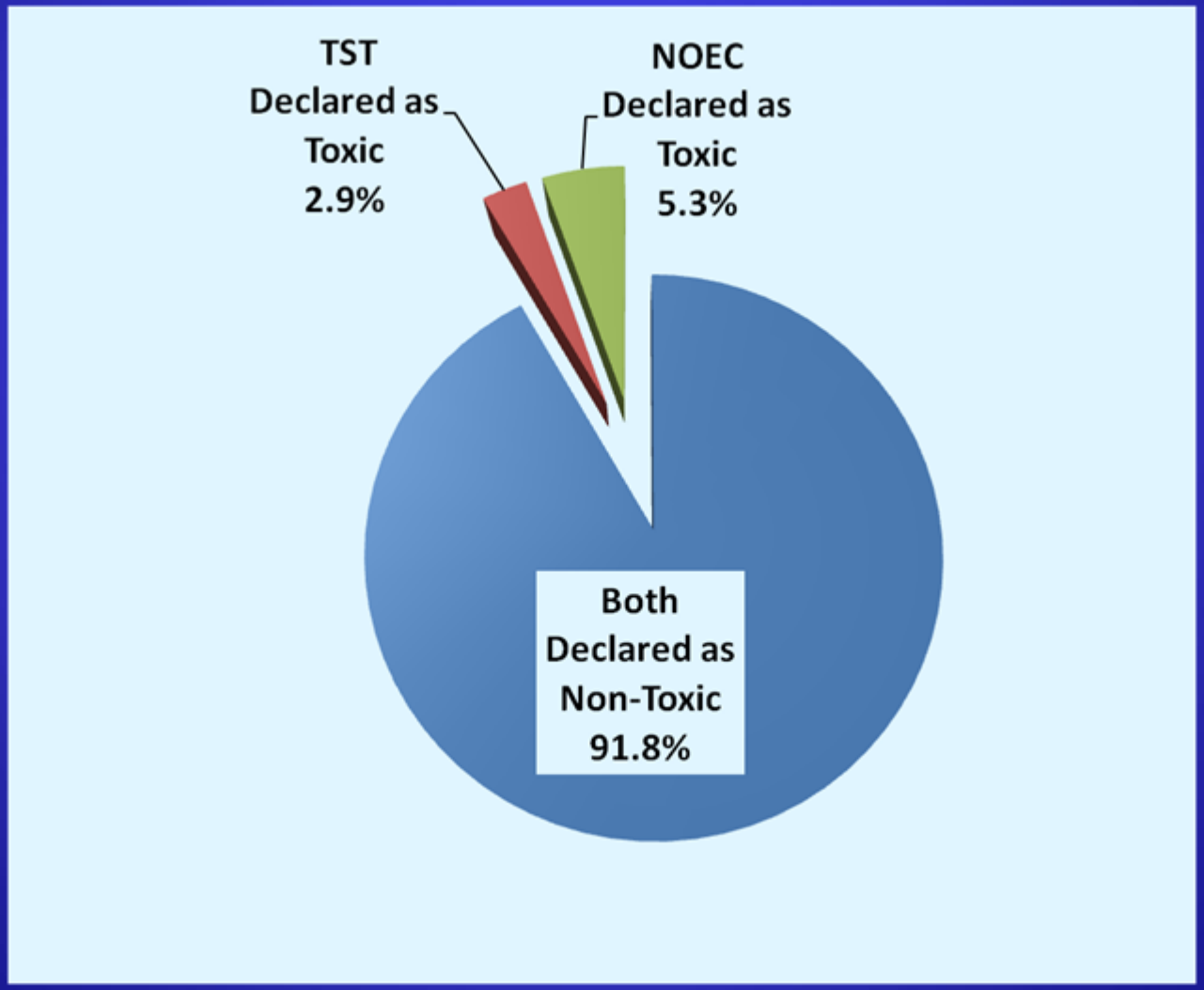
California labs tracked well with the national findings



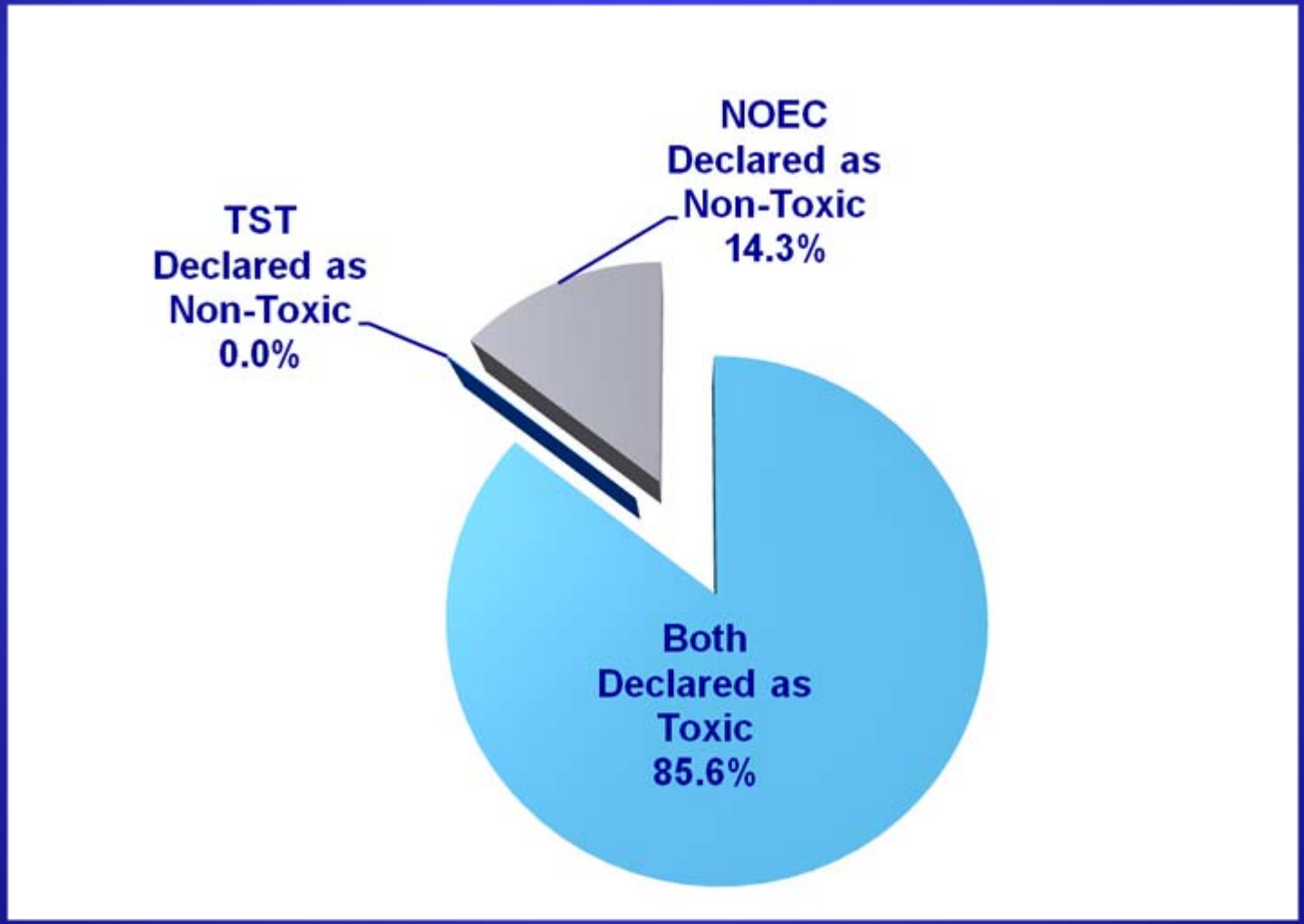
TST and NOEC Results Agreed for most tests

WET Method Type	Percent of Tests Declared Non-Toxic		Percent of Tests Declared Toxic	
	TST	NOEC	TST	NOEC
Chronic Marine	89.4	83.6	10.6	16.4
Chronic Freshwater	80.1	82.3	19.9	17.7
Acute Marine	100	100	0	0
Acute Freshwater	96.8	98.9	3.2	1.1
All Methods	**88.6** (687)	**87.2** (676)	**11.4** (88)	**12.8** (99)

TST Declared Fewer Samples as Toxic That Were Below the RMDs
≤ 25% Effect (chronic); ≤ 20% Effect (acute)



TST Declared More Samples As Toxic When the Mean Effect at the IWC was ≥ 25% (Chronic) or ≥ 20% (Acute)



For the few tests where results were uncertain due to high within-test variability, a few additional replicates would often make the difference using TST

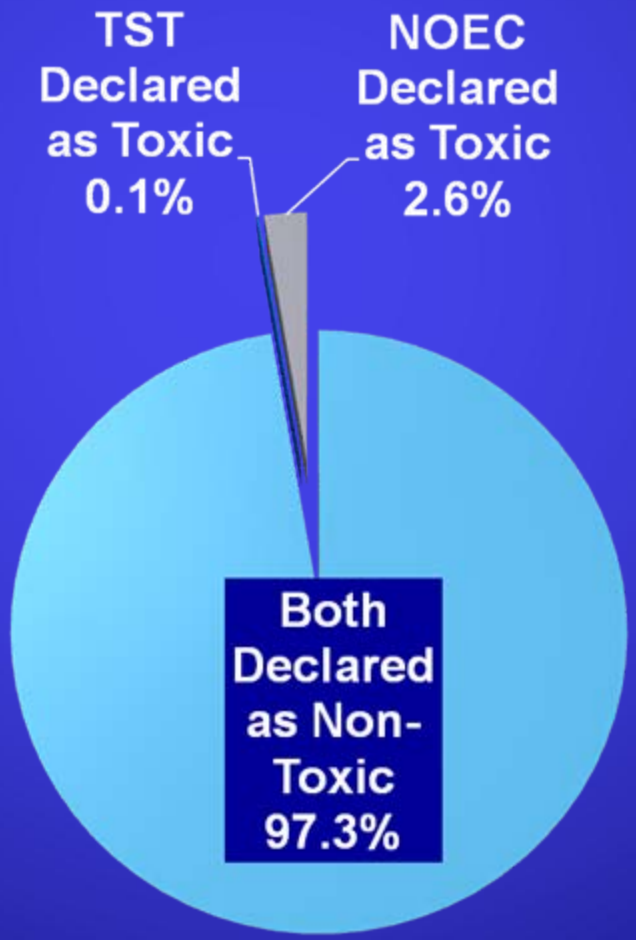
Examples from chronic tests observed in the Test Drive

Test	Percent Effect	# Additional Reps Needed to Declare Test Not Toxic
Red Abalone	15.4%	1
Urchin fertilization	15.9%	2
Topsmelt	19.1%	2
Ceriodaphnia	20.6%	7
Fathead minnow	17.4 %	1

Additional Observations from the Test Drive

Effluents that demonstrate biologically trivial effects are rarely declared toxic using TST, consistent with EPA's evaluation

Tests having ≤ 10% effect at the IWC



Summary of Test Drive

- Test results using both TST and the current NOEC approach were very similar
- Samples having biologically trivial effects were declared non-toxic more often using TST than the current approach
- Samples exhibiting significant toxicity effects at the IWC were declared toxic more often using TST than the current approach
- For samples exhibiting effects in the “gray area”, addition of a few extra replicates to these tests would likely result in the sample being declared non-toxic