

Toxicity Test Error Rate Analyses

Steven Bay

Commission Meeting – March 2, 2012



BACKGROUND

- **State is developing a new toxicity policy for inland surface waters, bays, and estuaries**
 - Requires toxicity monitoring of municipal stormwater
 - Establishes numeric toxicity objective for POTWs
 - Specifies new statistical method (called “TST”) for determining compliance
- **Commission requested two previous presentations**
 - TST overview
 - Toxicity test error rates

ISSUES ASSOCIATED WITH ERROR RATES

- **False positive error could be large**
 - Estimates range 2-15%
 - Data lacking for some species
- **Water Board likely to revise toxicity limit in new policy**
 - Multiple samples
 - Effect on violations not documented

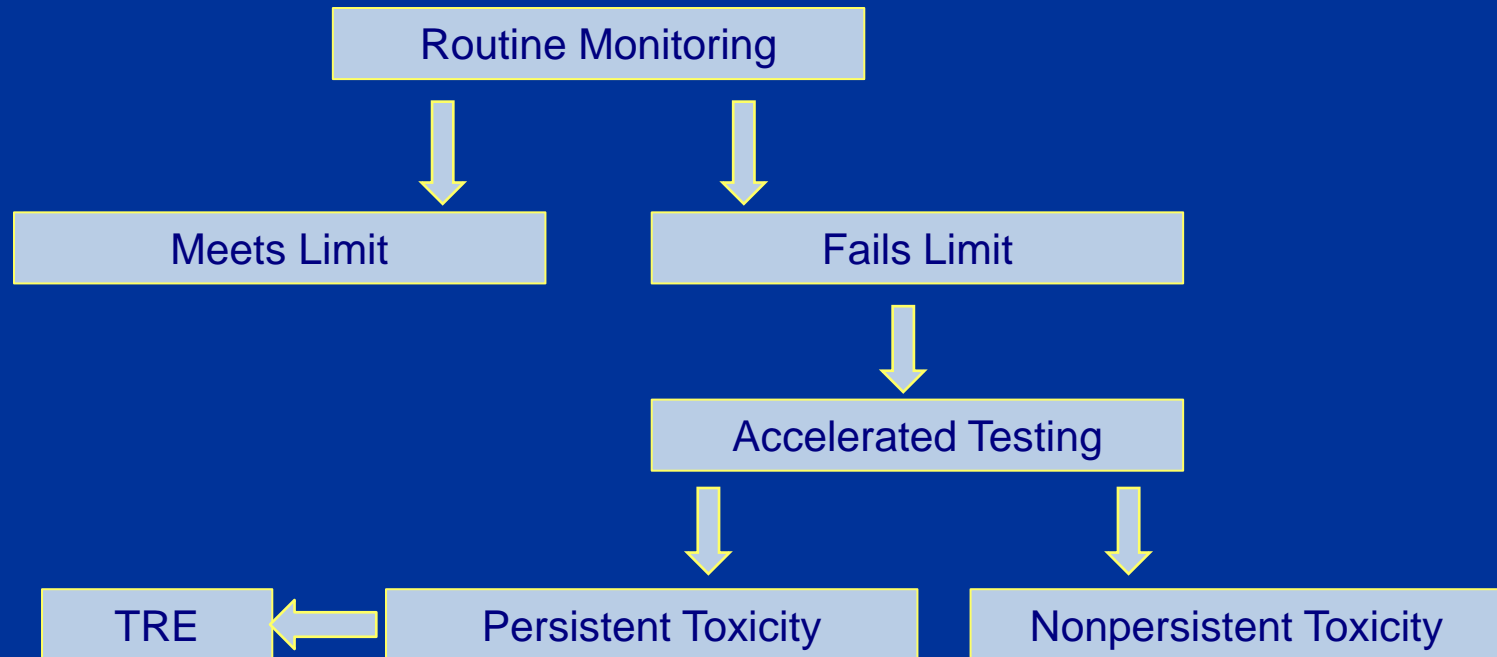
ADDITIONAL INFORMATION REQUESTED AT THE LAST COMMISSION MEETING

- **Effect of single sample vs. multiple sample effluent testing**
 - How does testing of multiple samples affect occurrence of false positives and false negatives?
- **Plan for a new study to improve error rate estimates**
 - What is the effort and cost involved with improving false positive error rate estimates?

ERROR RATE ANALYSIS APPROACH

- **Analyzed 2 alternatives for effluent limits**
 - Current SWRCB proposal (2010)
 - Likely revision (based on Water Board staff input)
- **Evaluate two false positive outcomes**
 - Incorrect violation
 - Unnecessary Toxicity Reduction Evaluation (TRE)
- **Evaluate false negative outcomes**
 - Failure to detect a violation

EFFLUENT TOXICITY MONITORING



ALTERNATIVES EVALUATED FOR EFFLUENT TOXICITY MONITORING

- **Violation occurs when effluent limit is not met**
 - One fail of TST (current draft policy)
 - 2/3 of samples fail TST (staff alternative)
- **TRE required for persistent effluent toxicity**
 - Six tests over 12 weeks (current draft policy)
 - Any TST failure triggers TRE

FALSE POSITIVE ANALYSIS ASSUMPTIONS

- **Used range of false positive rates from prior studies**
 - High rate = 15% (*Ceriodaphnia*; EPA's blank study)
 - Low rate = 2% (marine tests; Water Board TST test drive)
- **Monthly toxicity testing**
 - Current draft policy for POTWs >1MGD
- **TRE triggered by one TST fail out of six subsequent tests**
- **Results expressed relative to 5-year permit cycle**
 - Number of violations due to false positives
 - Number of TREs

FALSE POSITIVE RESULTS

Number of Occurrences / 5 yrs

Error Rate	Violation	
	1 TST Limit	2/3 TST Limit
2%	1.2	0.05
15%	9	2.5

- A multiple sample limit greatly reduces chance of false violations

FALSE POSITIVE RESULTS

Number of Occurrences / 5 yrs

Error Rate	Violation		TRE	
	1 TST Limit	2/3 TST Limit	1 TST Limit	2/3 TST Limit
2%	1.2	0.05	0.1	0.01
15%	9	2.5	5.6	1.6

- A multiple sample limit greatly reduces chance of false violations
- Chance of unnecessary TREs also reduced
- Magnitude of error rate has large effect

OTHER CONSIDERATIONS

- **Revised policy may look different than our assumption**
- **We evaluated false positive rates from a single species**
 - False positive rates for other species may differ
- **We evaluated false positive rates for a single discharge**
 - Cumulative number of false positives will increase with multiple discharges
- **There are additional options to refine the analysis**

ERROR RATE ANALYSIS APPROACH

- **Analyzed two alternatives for effluent limits**
 - Current SWRCB proposal (2010)
 - Likely revision (based on Water Board staff input)
- **Evaluate two false positive outcomes**
 - Incorrect violation
 - Unnecessary Toxicity Reduction Evaluation (TRE)
- **Evaluate false negative outcomes**
 - Failure to detect a violation

FALSE NEGATIVE ANALYSIS ASSUMPTIONS

- **Used a range of false negative rates from Water Board's TST documentation**
 - High rate = 25% (fathead minnow)
 - Low rate = 5% (marine tests)
- **Monthly toxicity testing**
 - Current draft policy for POTWs >1MGD
- **Results expressed over various time scales**
 - Chance of failing to detect a toxic discharge
 - Over 1, 3, 12 months of testing

FALSE NEGATIVE RESULTS

Chance of Missed Violation

Error Rate	1 TST Limit		
	1 mo.	3 mo.	12 mo.
5%	5%	0.01%	<0.001%
25%	25%	1.6%	<0.001%

- **False negative rate has a small effect on the ability to detect a toxic discharge**
 - Slight delay in time to detect violation

FALSE NEGATIVE RESULTS

Chance of Missed Violation

Error Rate	1 TST Limit			2/3 TST Limit		
	1 mo.	3 mo.	12 mo.	1 mo.	3mo.	12 mo.
5%	5%	0.01%	<0.001%	5.2%	0.01%	<0.001%
25%	25%	1.6%	<0.001%	29.7%	2.6%	<0.001%

- **False negative rate has small effect on the ability to detect a toxic discharge**
 - Slight delay in time to detect violation
- **Little difference in missed violations with multiple vs. single sample limit**

ADDITIONAL INFORMATION REQUESTED AT THE LAST COMMISSION MEETING

- **Effect of single sample vs. multiple sample effluent testing**
 - How does testing of multiple samples affect occurrence of false positives and false negatives?
- **Plan a new study to improve error rate estimates**
 - What is the effort and cost involved with improving false positive error rate estimates?

Number of Occurrences / 5 yrs

Error Rate	Violation		TRE	
	1 TST Limit	2/3 TST Limit	1 TST Limit	2/3 TST Limit
2%	1.2	0.05	0.1	0.01
15%	9	2.5	5.6	1.6

POTENTIAL BENEFITS OF A NEW FALSE POSITIVES STUDY

- **More precise estimate of error rates**
- **Stronger connection to CA labs and current methods**
- **Information for additional species**

COSTS AND SCHEDULE

- **Estimated cost: about \$400,000**
 - Pay contract labs to test samples
 - Substantial planning, coordination, and communication activities
- **Possible cost-leveraging opportunities**
 - Stormwater Monitoring Coalition toxicity intercalibration study
 - Commission member labs
- **Will take at least 12 months to complete**

PROPOSED STUDY DESIGN

- **50 blank samples per test method**
 - Multiple rounds of testing by approximately 18 labs
 - Test blanks and reference toxicants
- **Two toxicity test methods**
- **Screening process to select labs**
 - Representative of effluent testing labs

WHERE TO GO FROM HERE?

- **Would you like additional data analyses?**
 - Alternative scenarios or assumptions needed?
- **Is a new study worth it?**
 - Is the extra precision worth the cost?
 - Are member agencies willing to contribute time and money?
 - Is coordination with SMC study the best option?