

# Water Quality Standards and Assessment

## Listing and Delisting Procedures: Binomial and Large Data Set Assessment Options

Dec. 06, 2017  
ODEQ HQ

Documents can be provided upon request in an alternate format for individuals with disabilities or in a language other than English for people with limited English skills. To request a document in another format or language, call DEQ in Portland at 503-229-5696, or toll-free in Oregon at 1-800-452-4011, ext. 5696; or email [deqinfo@deq.state.or.us](mailto:deqinfo@deq.state.or.us).

# Overview

## Basic Conundrum:

- In past, mostly small assessment data sets
- But now, assess datasets from 2 samples → 100's

## Improvement Goals:

- Update and clarify listing and delisting methodology
- Introduce statistical methods for assessing larger datasets

# Overview

## Process:

- Draft updates to the methodology document
  - Focus on statistical methods
  - Clarify data requirements
- Solicit peer review to refine draft and define acceptable options
- Discuss policy implications of acceptable options with workgroup
- Final proposal for methodology
- Address public comments

# Overview

## Today's Goals for Workgroup:

- Inform
  - Our recommended approach
  - About peer review of stats methodology specifics
  - Prepare workgroup to be able to weigh in following peer review
- Solicit Feedback
  - Input on peer-review charge questions
  - Review options from peer review (Winter 2018)
  - Public comment period for draft of methodology (Spring 2018)

# Topics

## Listing Considerations

- Addition to 303(d) list
- Address “large” data sets ( $n > 18$ )
- Statistical Methods: binomial test

## Delisting Considerations

- Removal from 303(d) list
- Clarify Data Requirements
- Statistical Methods: binomial approach

# Part 1: Listing Considerations

Category	Description
<b>Category 1</b>	<u>All</u> designated uses are supported. (Oregon does not use this category.)
<b>Category 2</b>	Available data and information indicate that <u>some</u> designated uses are supported and the water quality standard is attained.
<b>Category 3</b>	Insufficient data to determine whether a designated use is supported.
	Oregon further sub-classifies waters if warranted as: <b>3B:</b> Potential concern when data are insufficient to determine use support but some data indicate non-attainment of a criterion. <sup>5</sup>
<b>Category 4</b>	Data indicate that at least one designated use is not support but a TMDL is not needed. This includes:
	<b>4A:</b> TMDLs that will result in attainment of water quality standards have been approved.
	<b>4B:</b> Other pollution control requirements are expected to address pollutants and will result in attainment of water quality standards.
	<b>4C:</b> Impairment is not caused by a pollutant (e.g., flow or lack of flow are not considered pollutants).
<b>Category 5</b>	Data indicate a designated use is not supported or a water quality standard is not attained and a TMDL is needed. This category constitutes the Section 303(d) list that EPA will approve or disapprove under the Clean Water Act.



Category 5 waters are the 303(d) list.

# Oregon's Numeric Water Quality Criteria

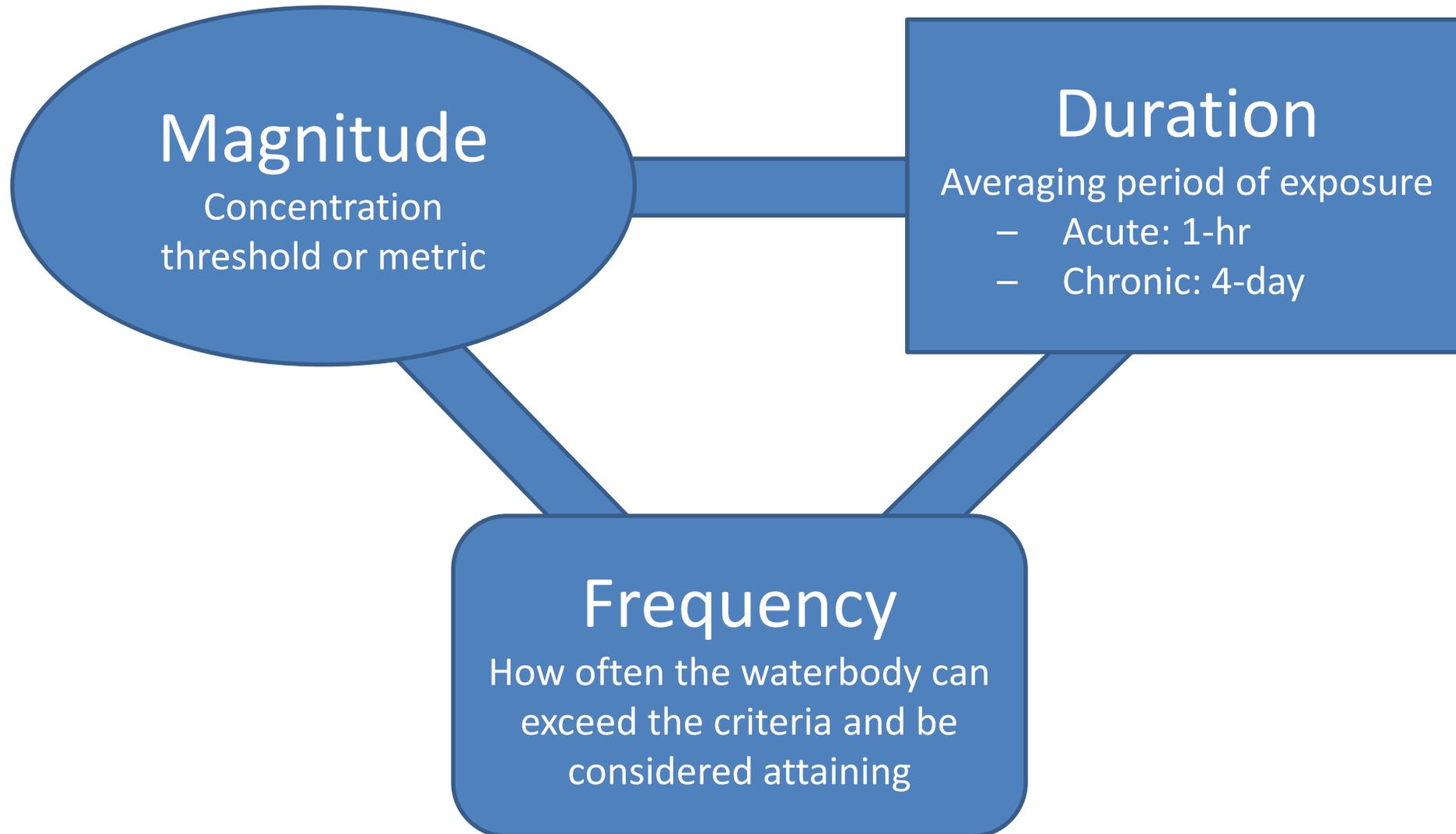
## Part of Oregon's WQS regulation (OAR 340-041-XXX)

- Toxic Substances for Protection of Aquatic Life (Table 30)
- Toxic Substances for Protection of Human Health (Table 40)
- Conventional Pollutants (i.e. pH, D.O.)

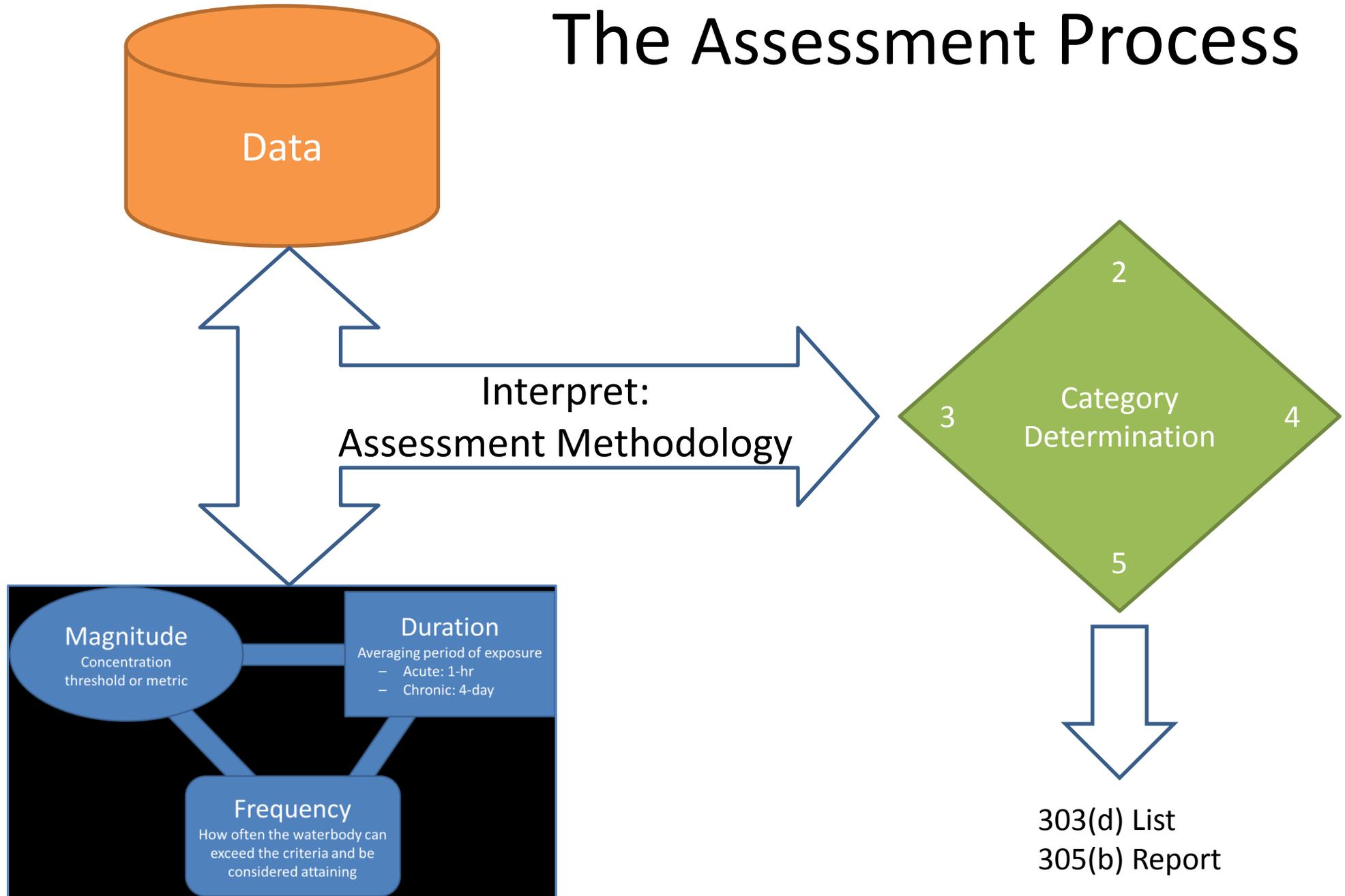
## Does not address:

- Standards with specified statistical methods in the criteria rule:
  - Bacteria (30-day mean)
  - Temperature (7-day average of daily max.)
  - Continuous D.O. (30-day mean min.)
  - Biocriteria (category % raw scores)

# Components of Numeric Criteria



# The Assessment Process



# Interpreting Numeric Criteria

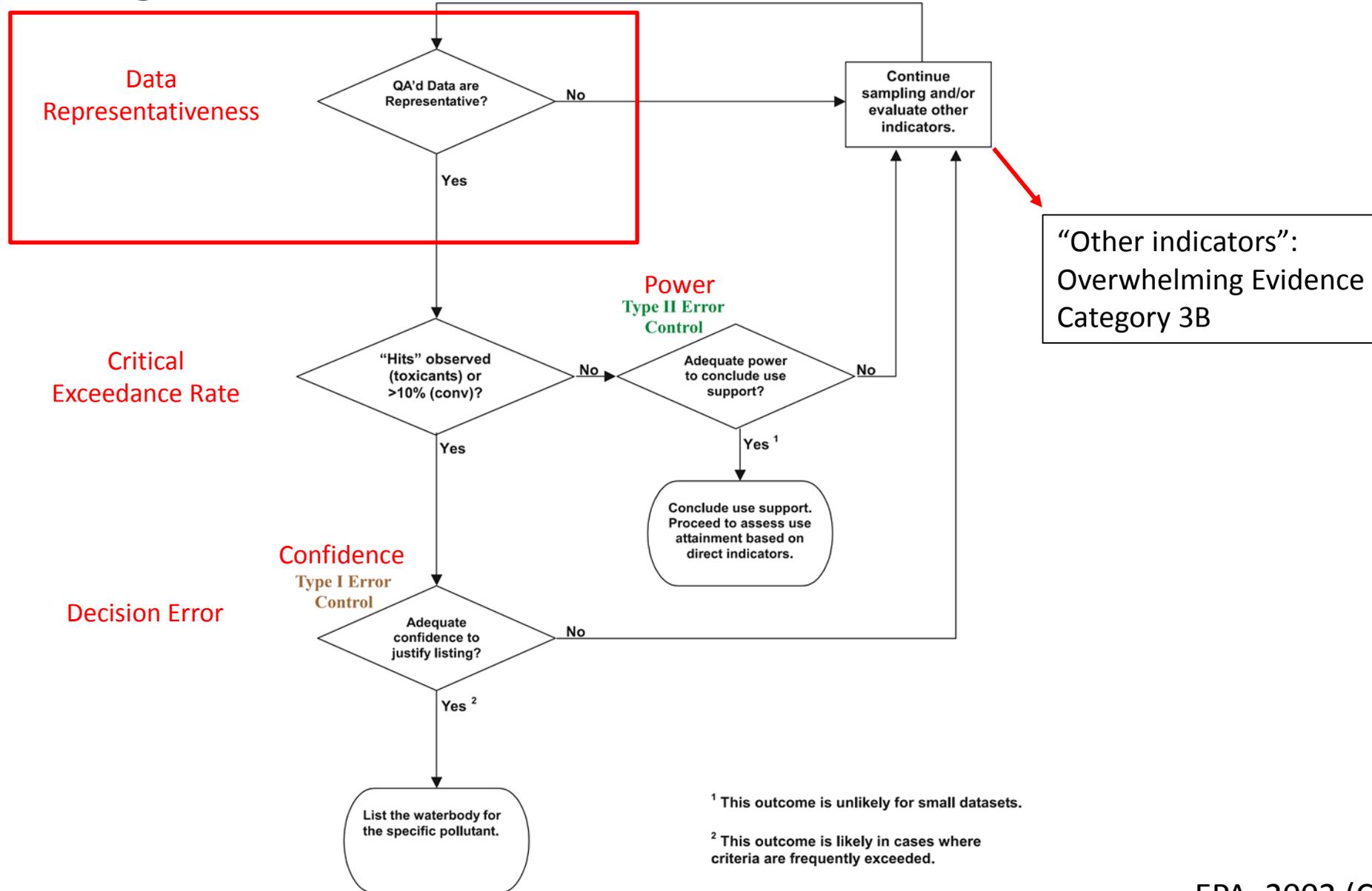
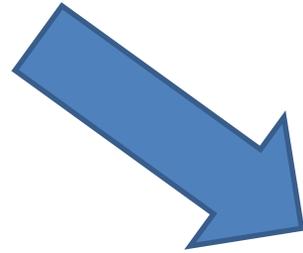


Figure 4-3. Sequential Decisionmaking. Making use support decisions, based on small sets of water column chemistry data, while balancing the risk of Type I (false positive) and Type II (false negative) decision errors.

EPA, 2002 (CALM)

# Data Representativeness



Does a sample represent conditions in the waterbody as a whole?

- Sample size
- Bimonthly monitoring
- Point samples
- Instantaneous grab samples

Does the concentration of the sample persist?

- For 1-hour (acute)?
- For 4-days (chronic)?



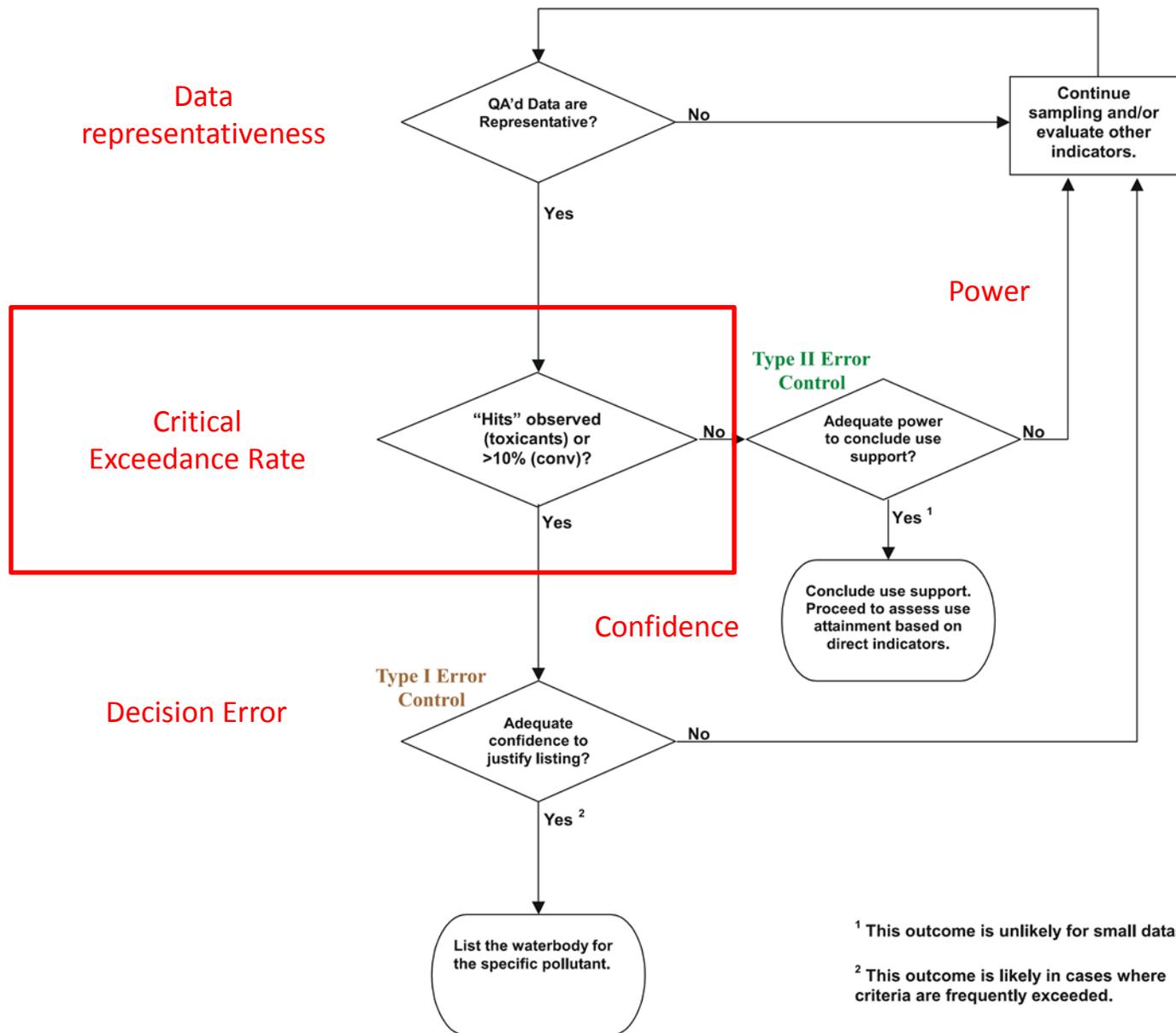


Figure 4-3. Sequential Decisionmaking. Making use support decisions, based on small sets of water column chemistry data, while balancing the risk of Type I (false positive) and Type II (false negative) decision errors.

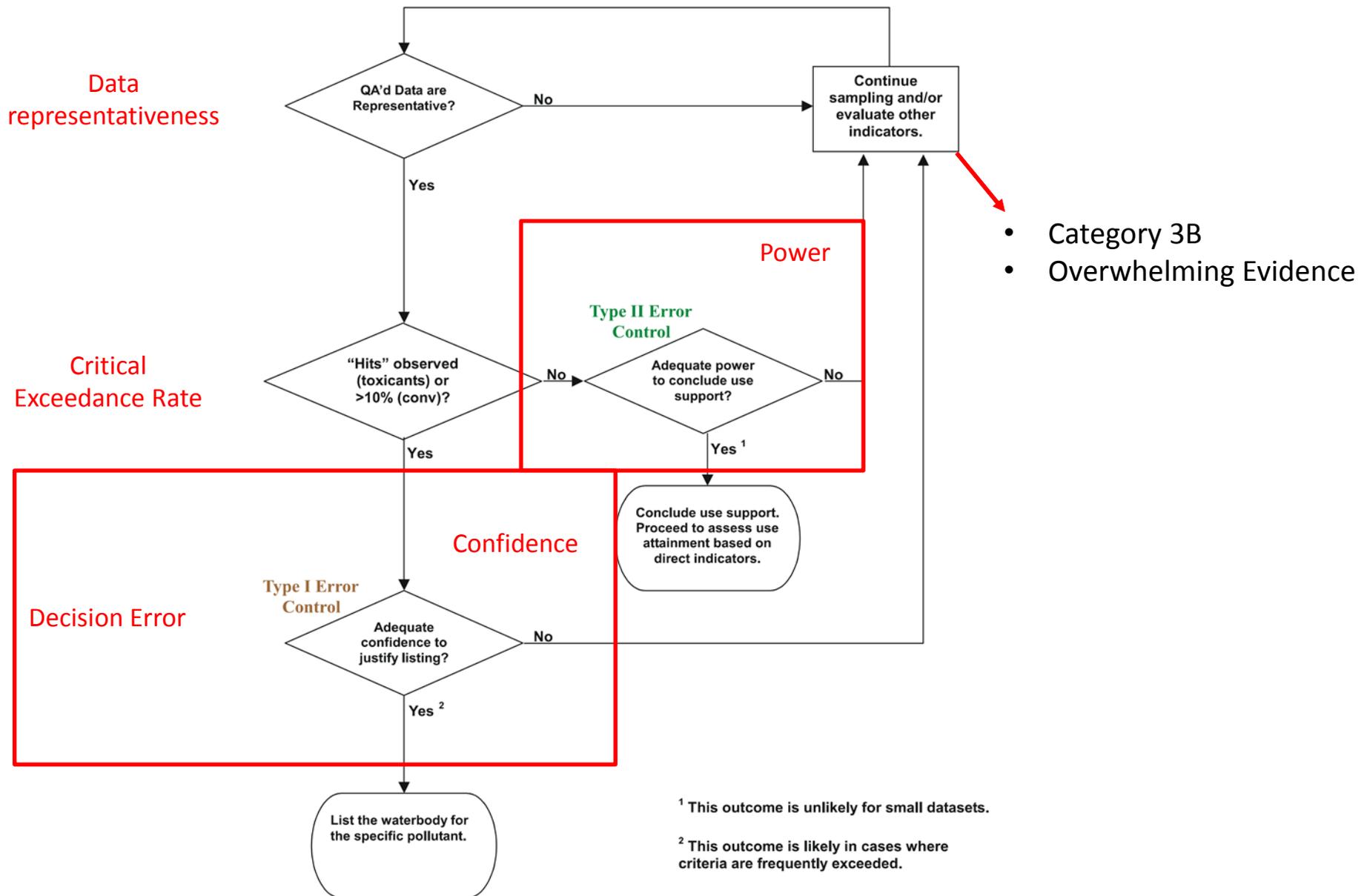
# Critical Exceedance Rate

- Used to evaluate whether the samples at hand indicate *the waterbody* is exceeding the criteria
- Addresses representativeness, uncertainty, and data quality
- A sample exceeding the magnitude is NOT the same as the waterbody exceeding the criteria
- The waterbody still cannot exceed the criteria >1-in-a-3 year period on average according to the method of evaluation
- Some criteria specify methods in rule (temperature, bacteria)

# Multiple Acceptable Rates for Interpreting Criteria

Critical Exceedance Rate	Source	Application
≤ 1 sample in 3 years	EPA, 1997**	Acute criteria.
>10%	EPA, 1997**	Acute criteria not supporting beneficial uses. Sampling and measurement error accounted for.
0.09% (1 sample out of 1,095)	EPA, 2002	Acute criteria. A 1-in-3-year frequency of daily averages.
0.36% (1 sample out of 274)	EPA, 2002	Chronic criteria. A 1-in-3-year frequency of 4-day averages.
5% plus a 15% effect size	EPA, 2002	Toxics criteria, equivalent to a 1-in-3-year frequency.
10%	EPA, 2003	For chronic criteria, and acute if justified. Binomial or raw score test.

\*\*Information contained in the most recent EPA guidance for assessment, listing, and reporting requirements was intended to supersede previous guidance.



- Category 3B
- Overwhelming Evidence

<sup>1</sup> This outcome is unlikely for small datasets.

<sup>2</sup> This outcome is likely in cases where criteria are frequently exceeded.

Figure 4-3. Sequential Decisionmaking. Making use support decisions, based on small sets of water column chemistry data, while balancing the risk of Type I (false positive) and Type II (false negative) decision errors.



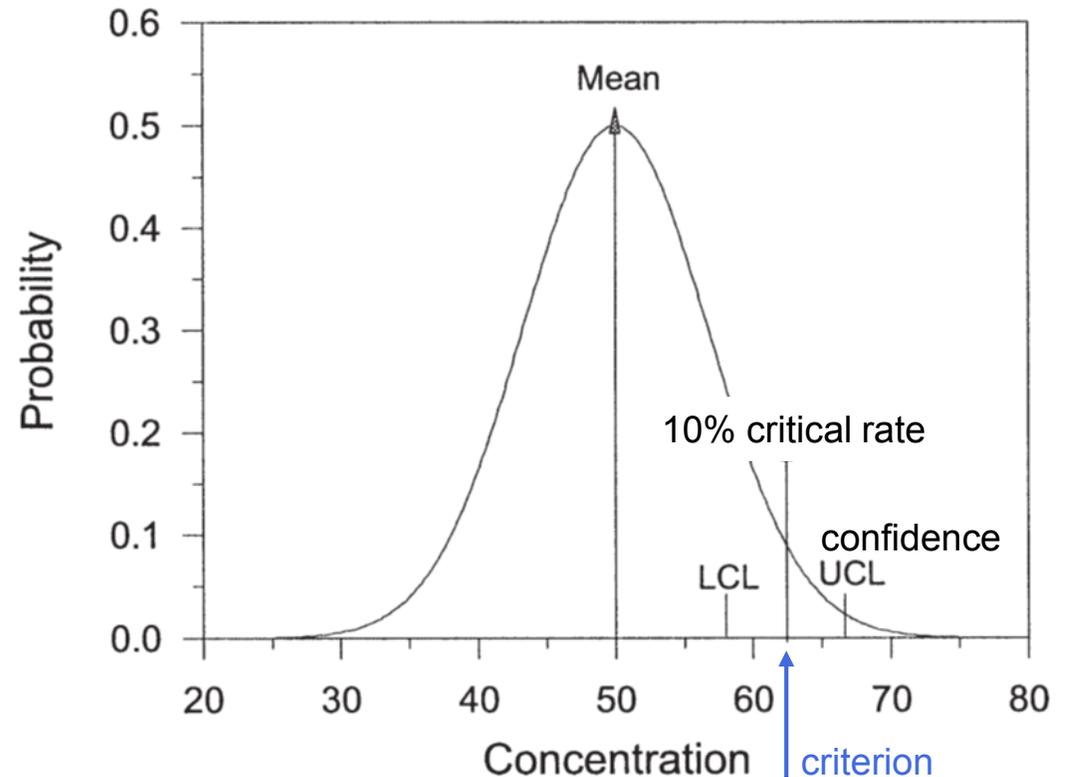
# Statistical Testing

## Hypothesis Testing

- Null Hypothesis: Waterbody attains
- Alternate Hypothesis: Waterbody is impaired → 303(d) list
- Need data to “accept” or “reject” the null

## Test Parameters:

- Critical Exceedance Rate
- Effect Size
- Confidence Level
- Explicit Error Rates



# Error Rates

## Null Hypothesis:

- Waterbody is attaining
- List if there is sufficient evidence of impairment
- Probabilities of making a listing error

Types of Decision Error		
Decision	Reality	
	Attaining	Impaired
Attaining (Accept Null)	Correct Decision (Confidence level)	Type II Error (False Negative)
Impaired (Reject Null)	Type-I Error (False Positive)	Correct Decision (Power/sample size)

# Statistical Testing

- The assessment methodology prescribes a specific technique for evaluating whether waterbodies comply with the appropriate standards
- Water quality standards are not changed by the selection of parameters used in statistical analysis.
- The EPA provides for a number of statistical approaches to evaluate attainment of standards for different types of pollutants and parameters (see white paper, Figure 1 and EPA, 2002 CALM)

# Oregon DEQ Current Practice

Critical exceedance rates for numeric criteria from DEQ's 2012 Assessment Methodology					
	Chronic			Acute	
	Attaining	Exceeding	Min. confidence interval	Attaining	Exceeding
Aquatic Life Toxics	0 samples	≥2-samples	NA	0 samples	≥2-samples
Conventional Pollutants	Raw score ≤10%	Raw score >10%	NA	NA	NA
Human Health Criteria*	0 samples	≥2-samples	NA	NA	NA

\*Not generally assessed in 2012. Listings based on most stringent criteria (usually aquatic life).

# Approach #1: >1-sample-in-3-years

- Based on 7Q10 return interval (10% probability per year)
- Based on community recovery from acute disturbances (fire, spill)
- Assumes as few as 2 samples adequately represent the waterbody
- Assumes instantaneous grab samples represent both 1-hour (likely) and 4-day (uncertain) average concentrations
- No way to specify confidence levels
- Very high false positive error probability, to reduce false-negative prob
- Error probability increases with sample size
- Regulatory and economic burden of false-positive Category 5 listings
- Disincentive for independent monitoring and data sharing by 3<sup>rd</sup> parties

# Approach #2: Raw Score Method

- Used for conventional pollutants (10% critical rate)
- “No more than x% of samples over the magnitude”
- Scales with number of samples
- Recognizes short-term excursions not likely to harm aquatic life
- Allows for natural variability and sampling error
- Assumes samples = population
- Relatively high false positive error rate, high uncertainty

# Approach #3: Statistical Tests

- Samples are a subset of a population (waterbody)
- What do the samples on hand tell us about the waterbody as a whole?
- Probability that *sample* shows the *waterbody* exceeds the critical rate
- Scales with sample size
- Proposed for conventional pollutants and chronic toxics criteria
- Directly controls for probability of making a false-positive error
- Quantify probability of false-negative error (indirect, sample size)
- Tradeoff: Slightly higher false-negative error rate

# Approach #3: Binomial Statistical Test

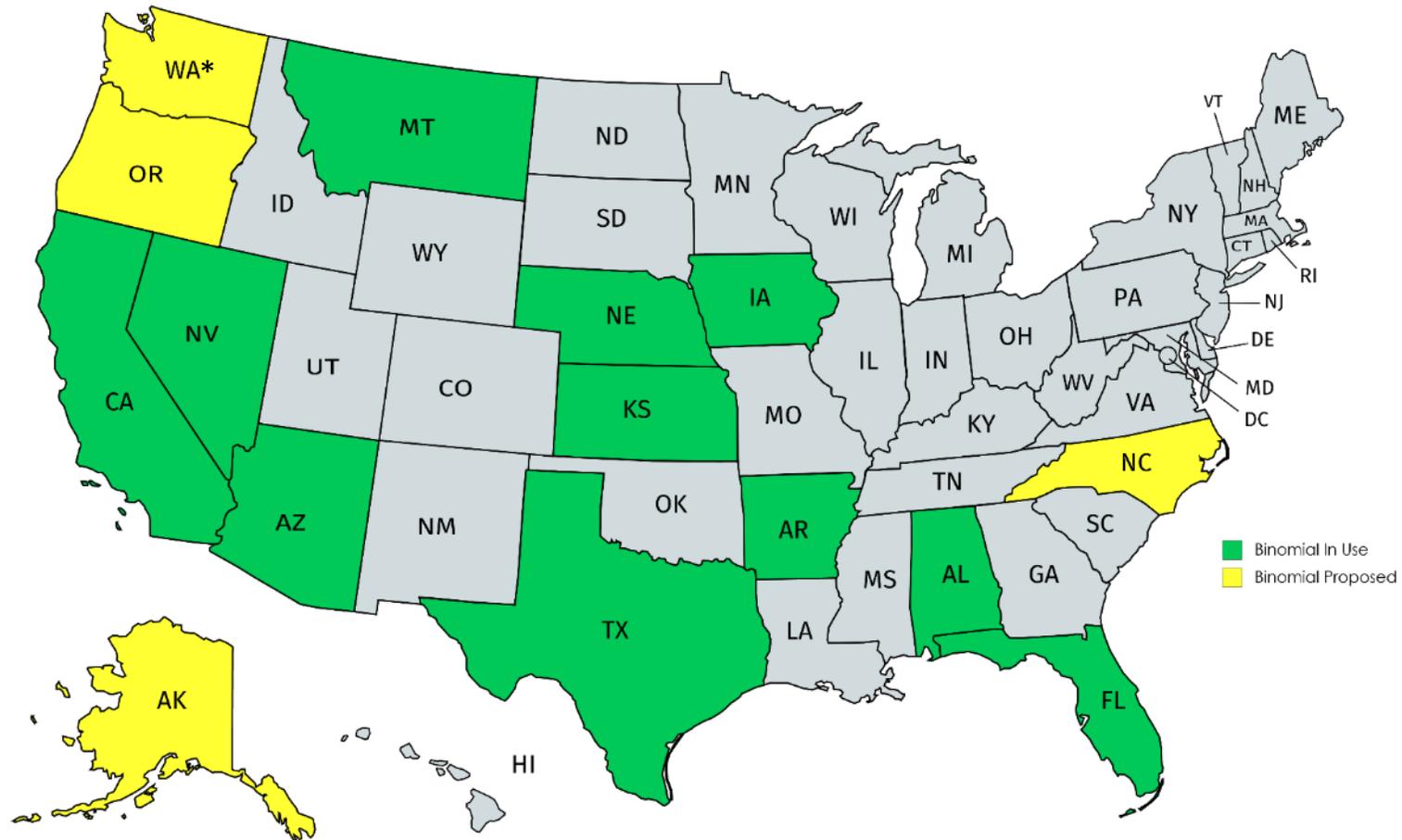
- Coin-flips, heads or tails.
- Samples “pass” or “fail” the criterion magnitude
- Do samples indicate the proportion of “fails” in the waterbody would exceed the critical rate?

This test requires the choice of :

- (1) the critical exceedance rate
- (2) the desired confidence level (direct Type-I error rate)
- (3) the minimum sample size allowed (indirect Type-II error rate)

***\*Note - We are recommending test parameters for external peer review.***

# States Using or Proposed Binomial Test



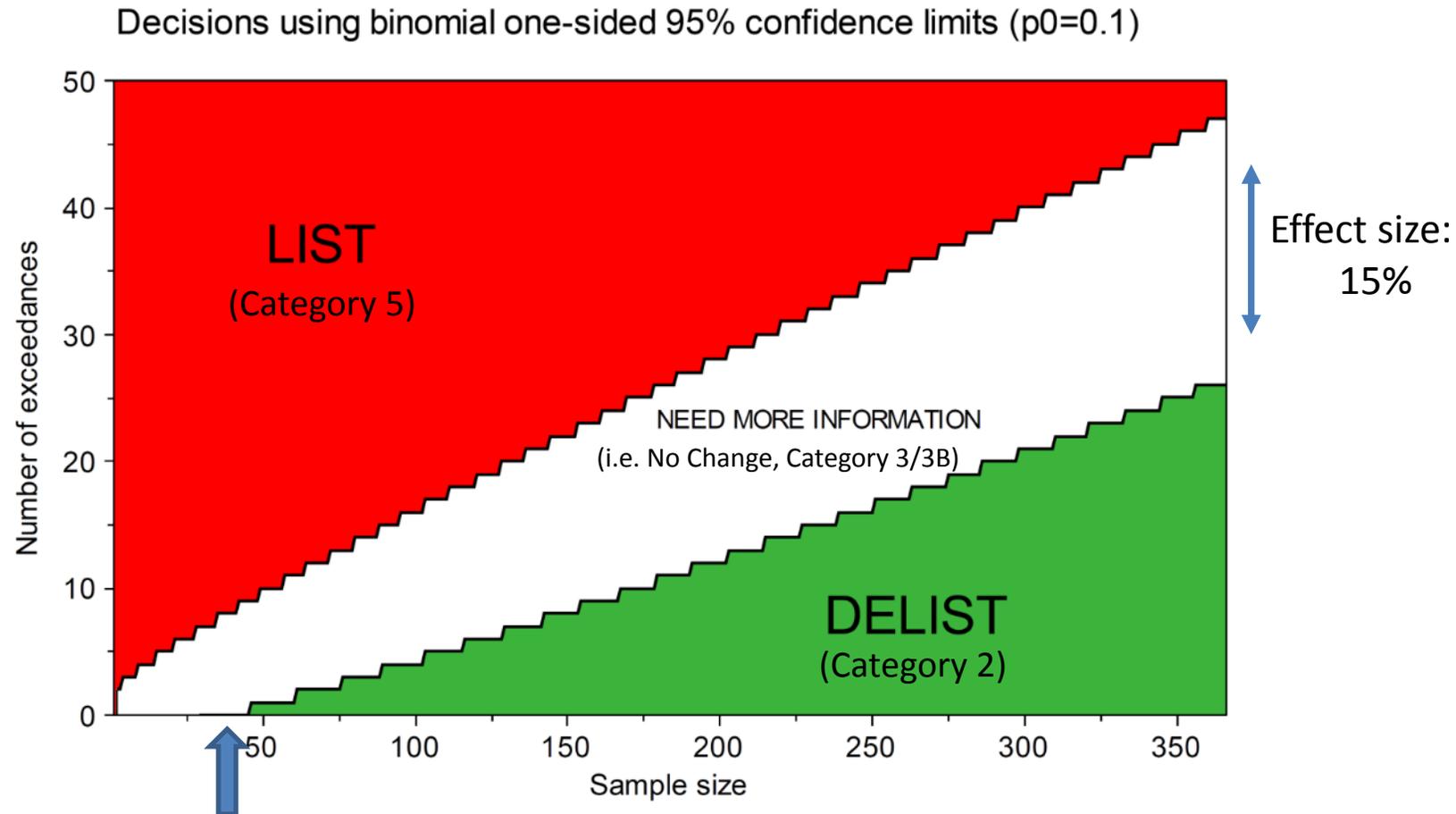
\* Withdrawn for hypergeometric test

Not an exhaustive survey

# Example Binomial Test

## Test Parameters Shown

- Critical Exceedance rate: 10%
- Effect Size: 15%
- Confidence Level: 95%
- Error Rates (Type-I) :  $\leq 5\%$



*\*Note – Our recommended test parameters will differ.*

Figure 1. Sample sizes and numbers of exceedances required for listing/delisting waterbodies based on binomial one-sided 95% confidence intervals, assuming  $p_0 = 0.1$ .

# DEQ Recommendations:

## Listing Determinations (Category 5)

	Chronic			Acute	
	Attaining	Exceeding	Min. confidence interval	Attaining	Exceeding
Aquatic Life Toxics	Binomial $H_0$ : $\leq 5\%$ of samples exceed the 4-day chronic criterion value	Binomial $H_A$ : $>15\%$ of samples exceed the 4-day chronic criterion value	90%	$\leq 1$ -sample-in-3-years	$\geq 2$ -sample-in-3-years
Conventional Pollutants	Binomial $H_0$ : $\leq 10\%$ of samples exceed the criterion value	Binomial $H_A$ : $>20\%$ of samples exceed the criterion value	90%	NA	NA
Human Health Criteria	Geometric mean sample concentration $\leq$ criterion	Geometric mean sample concentration $>$ criterion	NA	NA	NA

# DEQ Recommendations:

## Listing Determinations (Category 5)

### Chronic Aquatic Life Toxics Criteria and Conventionals

- Consistent with CALM
- Apply a binomial test at 90% confidence to determine category 5
  - $p = 0.05$  Toxics
  - $p = 0.10$  Conventionals
- Retain  $\geq 2$ -samples for small sample sizes (avoid higher type-II error)
  - No de-facto change in listing threshold for sample sizes  $< 18$
- Consult technical review panel to select appropriate confidence intervals and null hypotheses

# Lookup table toxics example

<b>minimal number of sample excursions to list as impaired for toxic substances</b>	
Null Hypothesis: Actual excursion proportion is $\leq 5\%$ Alternate hypothesis: Actual excursion proportion is $> 15\%$ Minimum confidence level is 90%	
<b>Sample Size</b>	<b>Excursions to list</b>
2-18	2*
19-22	3
23-35	4
36-49	5
50-63	6
64-78	7
79-92	8
93-109	9
110-125	10
126-141	11
142-158	12
159-171	13
179-191	14
192-200	15
* sample sizes $< 18$ use $> 1$ -samples-in-3-year critical exceedance rate	

# DEQ Recommendations:

## Listing Determinations (Category 5)

### Acute Aquatic Life Toxics Criteria

- Continue to assign category 5 with  $\geq 2$ -samples-in-3-years as critical exceedance rate
  - grab sample likely to reflect 1-hour duration
- Reduces ecological risk of type-II errors for small samples
- Frequency matches community recovery interval of studies
- Avoid significant disturbances to aquatic communities

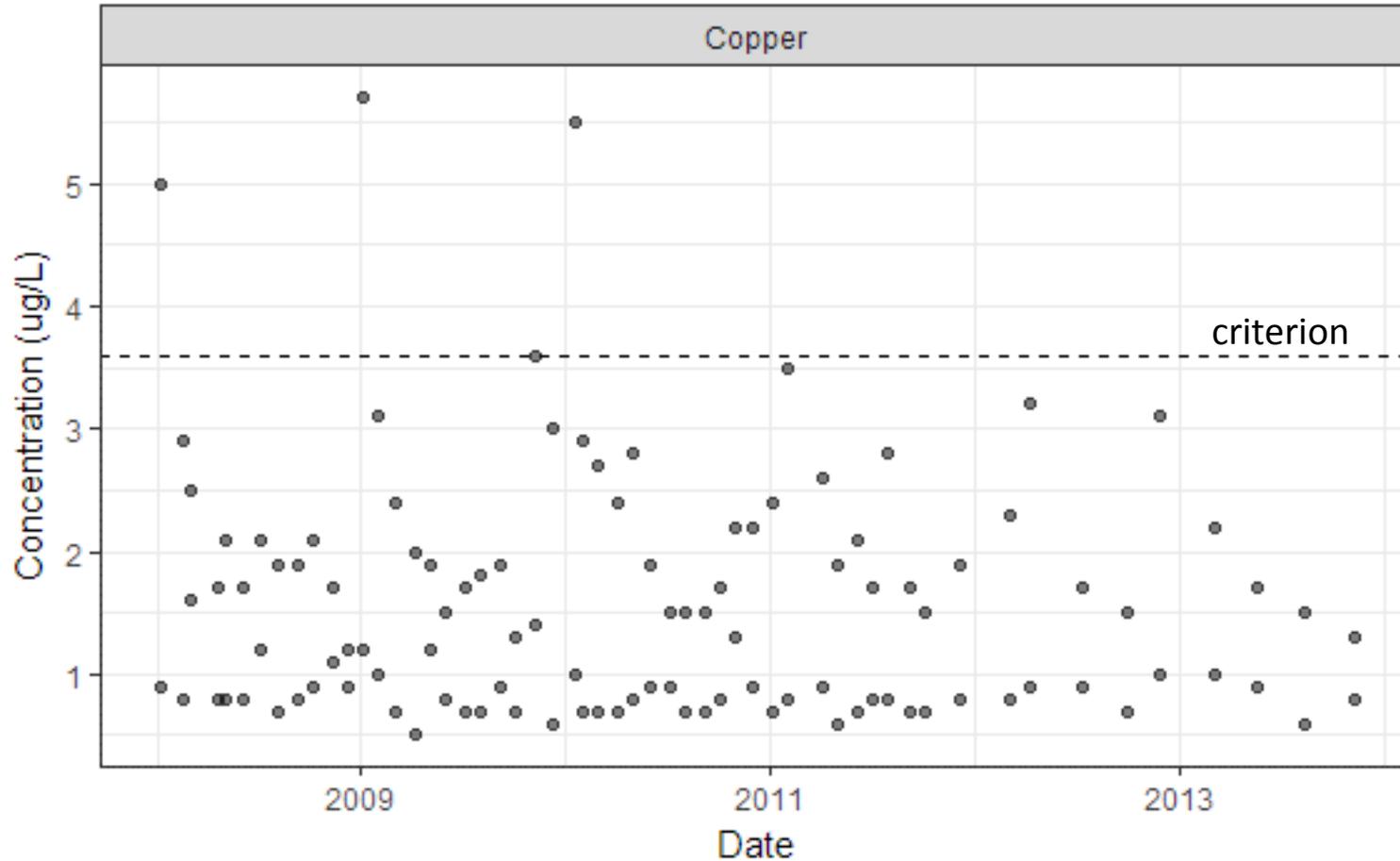
# DEQ Recommendations:

## Listing determinations (Category 5)

### Human Health Criteria

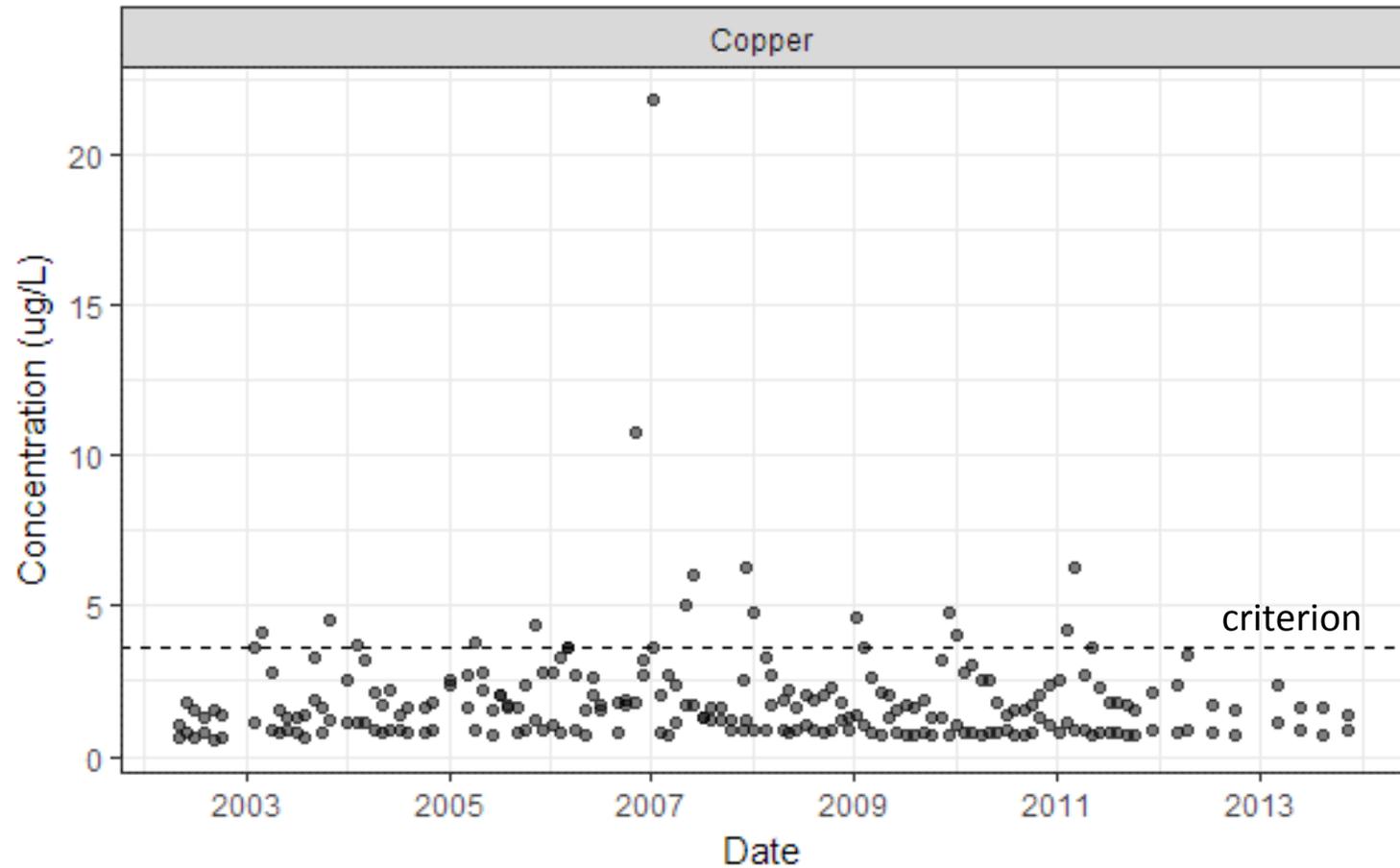
- Not frequently the cause of listings in the past
- Differs from aquatic life criteria
- Compare geometric mean of concentration within assessment window
  - Better align assessment methods to the duration of exposure
  - Generally for long-term (lifetime) exposures
  - E.g. Carcinogens: 70 years
  - Some exceptions based on criteria for individual pollutants

# Potential Impact: Attaining Stream



n=110  
excursions = 3

# Potential Impact: Impaired Stream



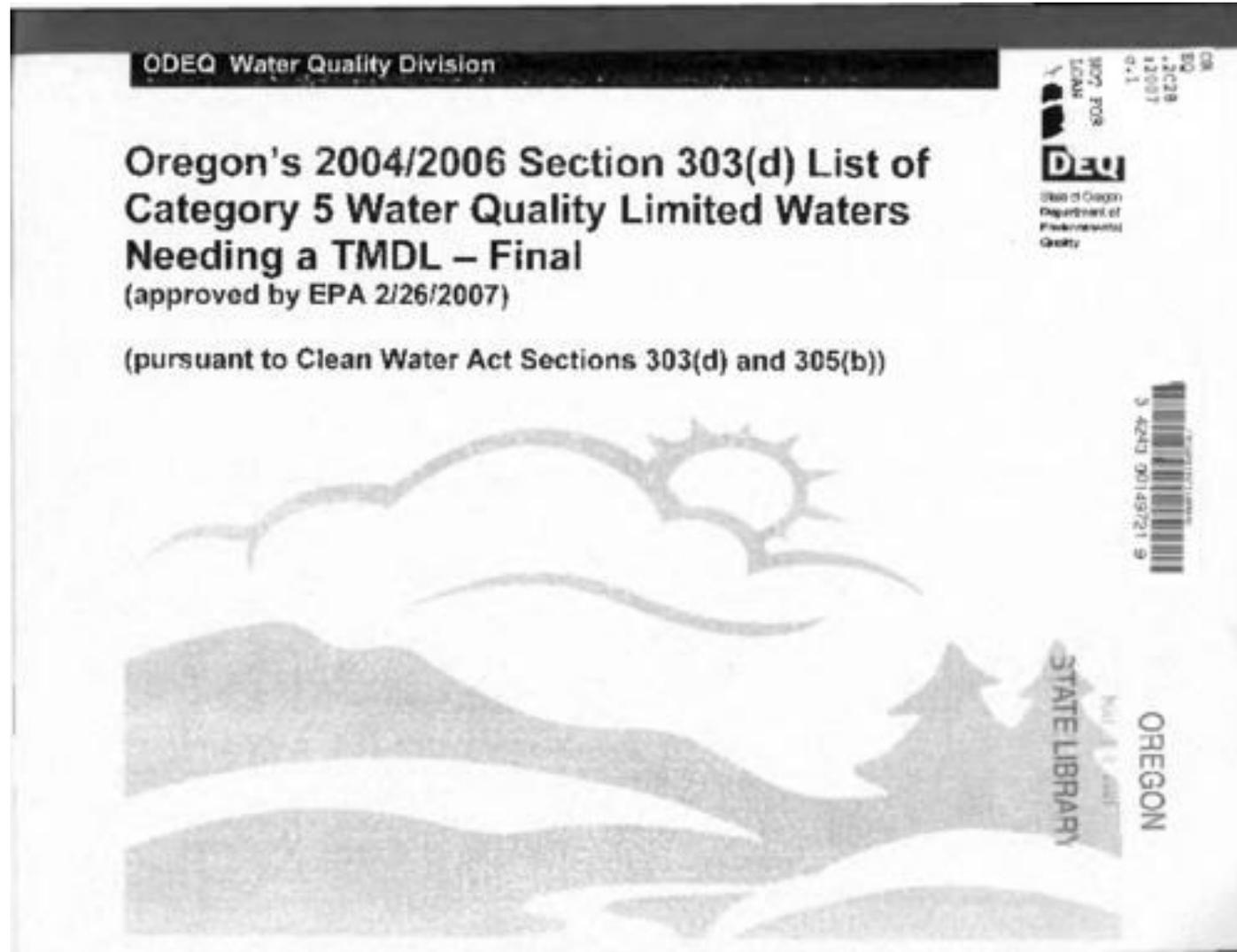
n=229

excursions = 21

# Listing Considerations - Summary

- Conventional Pollutants – Div. 41 -016 to -031
  - Apply an exact binomial test instead of raw score for sample sizes ( $n \geq 10$ )
- Fish and Aquatic Life Toxics – DEQ Table 30
  - No de-facto change to assessment of small data sets and acute toxics ( $n < 18$ )
  - Apply an exact binomial test to assess large data sets ( $n > 18$ )
- Human Health Toxics – DEQ Table 40
  - Use geometric mean concentration
    - Better match criteria duration

# Listing Questions?



# Part 2: Delisting Considerations



Category	Description
<b>Category 1</b>	All designated uses are supported. (Oregon does not use this category.)
<b>Category 2</b>	Available data and information indicate that <u>some</u> designated uses are supported and the water quality standard is attained.
<b>Category 3</b>	Insufficient data to determine whether a designated use is supported.
	Oregon further sub-classifies waters if warranted as: <b>3B:</b> Potential concern when data are insufficient to determine use support but some data indicate non-attainment of a criterion. <sup>5</sup>
<b>Category 4</b>	Data indicate that at least one designated use is not support but a TMDL is not needed. This includes:
	<b>4A:</b> TMDLs that will result in attainment of water quality standards have been approved.
	<b>4B:</b> Other pollution control requirements are expected to address pollutants and will result in attainment of water quality standards.
	<b>4C:</b> Impairment is not caused by a pollutant (e.g., flow or lack of flow are not considered pollutants).
<b>Category 5</b>	Data indicate a designated use is not supported or a water quality standard is not attained and a TMDL is needed. This category constitutes the Section 303(d) list that EPA will approve or disapprove under the Clean Water Act.

# Delisting

- Delisting is removing from the 303(d) list
- Still part of 305(b) report
- Anything changed from Category 5 to another status
- Possible Scenarios:
  - Still Impaired:
    - Category 4, 4A,4C: TMDL approved/not needed
    - Category 4B: other pollution control
  - Attaining:
    - Category 2

# Reasons for Status Changes for Delisting from 303(d)

1. Current information shows water quality standards are attained
2. Category 5 status was assigned in error
3. Water quality standards changed and/or no longer apply
4. TMDLs approved for waterbody
5. Other pollution control requirements in place
6. Impairment is not caused by a pollutant (i.e. habitat or flow modification)

# New Data: Category 5 → 2

## Not Impaired

1. Current/new information (data) shows water quality standards are attained
  - Indicates an improvement / change in water quality

# Current Practice: New Data

A water body may be delisted and assigned to Category 2 under Section (a) if there is:

“sufficient information from the current assessment to evaluate the pollutant or parameter and the information demonstrated that currently applicable water quality standards were being met... Generally, ***similar data were required to delist a water body as initially used*** to place the water body on the 303(d) list. For example, ***if the listing was based on two successive years of a standard not being met, DEQ looked for at least two successive years of data indicating that the standard is being met.***”

# Current Practice: New Data

- Vague data requirements for delisting
- Arbitrary time limits
  - Excursions no longer within assessment data window
  - An equal number of samples are collected
  - Attaining for as long as was exceeding
- Inconsistent decision making
- Difficult to credit successful restoration / pollution control



# Defining “Similar Data”

## Delist *if*:

- A minimum number of additional samples has been collected
- There are sufficient valid samples to assess the waterbody
- Waterbody meets Category 2 requirements

# Administrative: Category 5 → 2

## Not Impaired

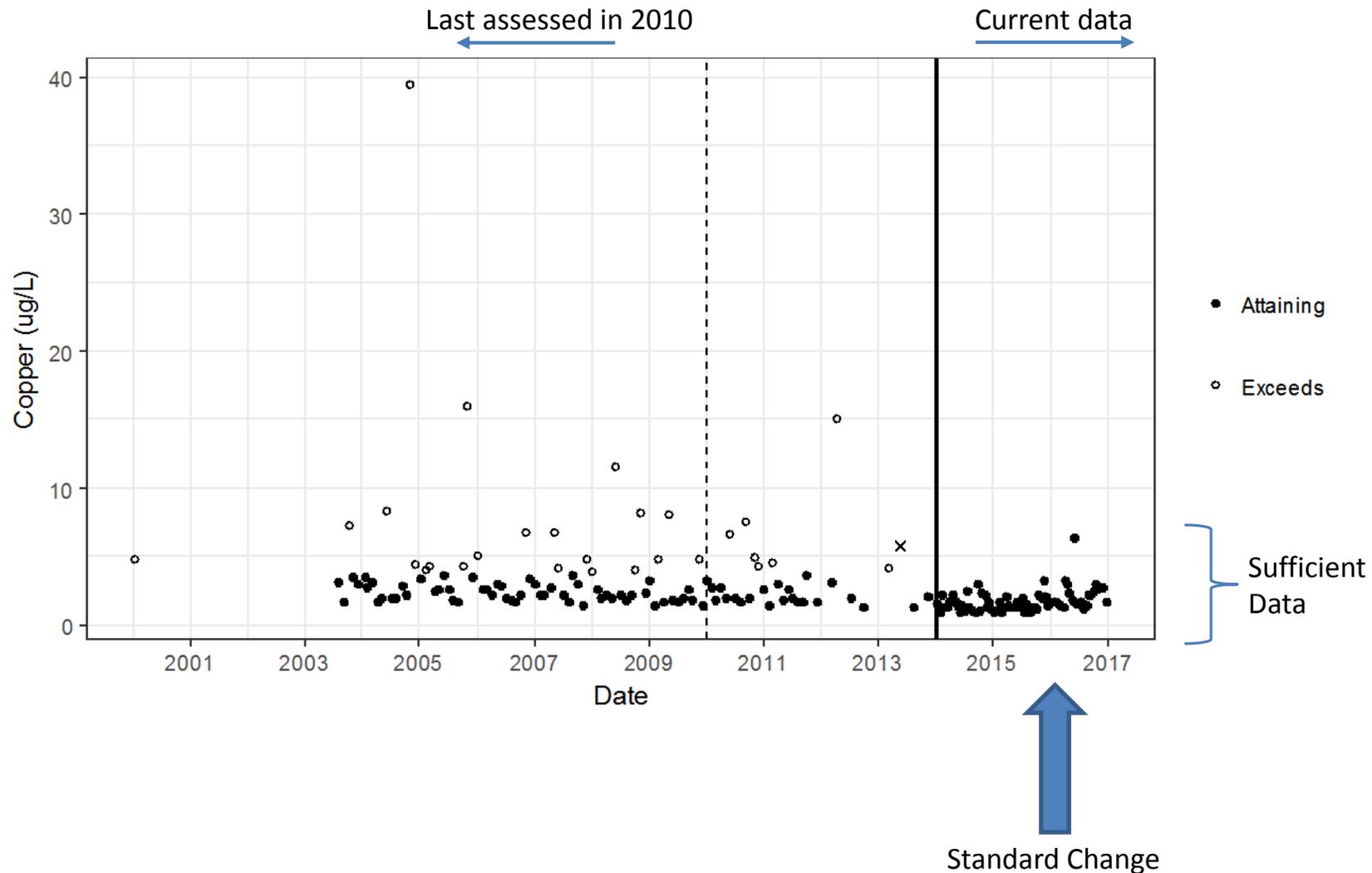
2. Current information shows an error in the Category 5: 303(d) listing – which indicates the 303(d) Category 5 status was assigned in error.
3. Water quality standards have changed or no longer apply in certain waterbodies.

# Administrative: Category 5 → 4

Still impaired, but no TMDL needed → 305(b) List:

4. TMDLs approved for waterbody and pollutant - if a TMDL is developed for a pollutant on a watershed scale, all waterbody waterbodies listed for that pollutant criteria within the watershed are delisted and placed in Category 4A.
5. Other pollution control requirements in place – when pollution control measures are implemented and are expected to result in attainment of water quality standards, the water bodies will be delisted from Category 5: 303(d) and placed in Category 4B.
6. Impairment is not caused by a pollutant - when data or information indicate that waterbody impairment is being caused by pollution (i.e. habitat or flow modification), not pollutants (e.g. toxic substances), the waterbody is moved to Category 4C.

# Example: Justification 1 - New Data Showing Attainment



# Review of Delisting Justifications

## Need for Clarification

- Under what circumstances do we reconsider an impaired water to be attaining?
- Administrative reasons are implementable as written
- What is sufficient “similar data” for delisting?

## New Methods Should

- Match methodology used for listing
- Use statistical methods to address larger sample sizes
  - Send for external peer review

# Delisting: Exact Binomial Test

- Greater burden to demonstrate attaining once already listed
- Reduce risk of removing a waterbody that *is* actually impaired
- Different null hypothesis
  - Assume *impaired*
  - Requires sufficient power / sample size to conclude attainment

# DEQ Recommendations:

## Delisting determinations (Category 2)

	Chronic			Acute	
	No change	Delist	Min. confidence interval	No change	Delist
Aquatic Life Toxics	Binomial $H_0$ : $\geq 15\%$ of samples exceed the 4-day chronic criterion	Binomial $H_A$ : $< 5\%$ of samples exceed the 4-day chronic criterion value	90%	$\leq 1$ -sample-in-3-years	$\geq 2$ -sample-in-3-years
Conventional Pollutants	Binomial $H_0$ : Actual excursion proportion $\geq 20\%$	Binomial $H_A$ : Actual excursion proportion $< 10\%$	90%	NA	NA
Human Health Criteria	Geometric mean sample concentration $\geq$ criterion	Geometric mean sample concentration $<$ criterion	NA	NA	NA

# Lookup table toxics example

Chronic Toxics: Minimal number of sample excursions to delist	
Null Hypothesis: Actual excursion proportion is $>15\%$ Alternate hypothesis: Actual excursion proportion is $\leq 5\%$ Minimum confidence level is 90%	
Sample Size	Excursions must be less than
10-15	1
16-25	2
26-33	3
34-42	4
43-51	5
52-59	6
60-67	7
68-76	8
77-84	9
85-92	10
93-99	11
100-107	12
108-115	13
116-123	14

# When do we consider revisiting old listings?

- Where sufficient data is available for our 2018/2019 assessment cycle
- When current data window overlaps with previous assessments.
- When standards have been revised since last assessment cycle.

# Delisting Summary

- Assumes initial listing is correct, but water quality has improved
- Apply an exact binomial test (mirror listing method)
- Higher bar to delist than to make an initial assessment
- Encourages better waterbody characterization (more data)
- Prevent “bouncing” on/off 303(d) list

# Conclusions

- Update to new statistical methods for making listing / delisting determinations
- Consistent with national guidelines
- Allows to plan for risk and choose tolerance for error in both attainment and impairment
- More transparent data requirements

# Next Steps

- Convening an independent peer review panel to refine methodology
  - Selection of Null and Alternate hypotheses
  - Setting parameters of the hypothesis tests
  - Error balancing
  - Identification of critical values (tables)
- Peer review in Jan. 2018

# Next Steps: External Peer Review

- Affirm foundational selection of binomial test as appropriate statistical test
- Affirm recommendation not to apply it to the acute criteria
- Formation of hypotheses for testing impairment, attainment
- Affirm selection of test parameters
- Tables of critical values

# Questions?



[www.portlandoregon.gov/bes/](http://www.portlandoregon.gov/bes/)