Freshwater Sediment Standards

Science Panel

May 20, 2011

Presenters:

Russ McMillan
Department of Ecology

Teresa Michelsen
Avocet Consulting

DEPARTMENT OF ECOLOGY
State of Washington

Freshwater Sediment Standards Discussion Points For Today

- Provide a recap of proposed biological and chemical criteria and framework.
- Provide responses to Panel's questions.
- Provide an update on items that we were working on during our previous meeting.

Requirements for Proposed Bioassay Suite

- Bioassay suite to include at least:
  - 2 Species
  - 3 Endpoints
  - 1 Chronic Test
  - 1 Sublethal Endpoint
- Interpretation
  - SQS: Single SQS level hit
  - CSL: 2+ SQS level hits; 1+ CSL level hit

Bioassays: Acute and Chronic and Endpoint Effects Levels

Test	Acute Bioassays	Chronic Bioassays
Hyalella azteca		
10-day mortality	X	
28-day mortality		X
28-day growth		X
Chironomus dilutus		
10-day mortality	X	
10-day growth	X	
20-day mortality		X
20-day growth		X

Biological Standards Questions

Is the proposed bioassay suite scientifically defensible as being appropriately protective of the benthic community?

Is it scientifically defensible to base the designation of sediment quality on a suite of bioassays comparing test to control without the benefit of a reference sediment?

Science Panel's Response: Generally yes, but left Ecology with several questions and recommendations

Science Panel's Questions

What is the definition of the health of the benthic community?

Are the proposed bioassays/species a proven indicator of benthic community health?

- A healthy benthic community provides a number of important functions to the environment.
  - Prey base for other aquatic animals.
  - Process, integrate and recycle organic material in the sediments.

Science Panel's Questions

- *Objectives for development of biological standards.*
  - Protect the functions and integrity of the benthic community - Multiple species/sensitive life-history stages.
  - Discern a lower range of effects bounding minor adverse effects levels.
    - SQS is the minimum detectable difference
    - CSL is set as 10% or 15% above MDD.

Science Panel's Questions

How sensitive are these bioassays to the chemicals typically found at freshwater sites?

Do we have either sensitivity curves or regional data to confirm this?

- Literature reports sensitivities vary greatly for different chemicals and different species.
  - In comparisons to other species, our two were sometimes more, and sometimes less sensitive.
  - Neither was consistently less sensitive.

Science Panel's Questions

Are the bioassays appropriate for diverse types of waterbodies (i.e., different sizes of lakes, streams, intermittent waterbodies)?

Are there limitations based on factors such as hardness, pH/alkalinity, grain size, etc.?

- *Hyalella* and *Chironomus* have been selected for standardized tests because they are;
  - Widely distributed and found in a broad range of aquatic environments.
  - They have broad tolerance for varying physico-chemical characteristics of sediments.

Science Panel's Recommendations

- Maintain flexibility to use other species/endpoints where appropriate.
- Provide implementation ***guidance*** for site managers and consultants.
  - When/where the standards (chemical and biological) are applicable and when/what to consider alternate approaches.

Bioassay Suite Questions Posed to Expert Peer Reviewers

Is the proposed bioassay suite appropriately sensitive to protect the freshwater macro benthic community (i.e., typical taxanomic structure and functions such as prey base to endangered species like salmon)?

From your experience are there other bioassays/species that provide consistent, reproducible and sensitive results that should be considered for developing biological criteria?

Expert Peer Reviewers Responses

- “It is the best you can do at present.”
- “It would be preferable to have additional taxa and sublethal endpoints represented.”
- Recommendations were made for other species to consider.
  - A freshwater mussel bioassay is very close to being finalized, snails, and a mayfly test were also mentioned.
  - A biomass endpoint could be considered.

Questions?

Photo courtesy of Derek Yorks

Freshwater SQGs

Teresa Michelsen
Avocet Consulting

Outline

- Review questions 1 and 2 discussed previously
- Presentation and discussion of questions 3 and 4
- New information:
  - Additional statistical analyses
  - National peer review

Question 1

General Approach:

- Is it appropriate to use sediment bioassays to represent effects to the benthic community?

  ⇨ Discussed in bioassay talk; no benthic data

- Is the use of a multivariate model to empirically derive chemical SQVs scientifically defensible?

  ⇨ Empirical methods are used by all states and provinces; multivariate model provides better reliability

Freshwater vs. Marine

Methods

- Different models
- Univariate vs. multivariate
- Not organic-carbon normalized
- Fewer bioassay choices

Environment

- Various geochemical/biological systems
- Different source types/chemicals

⇨Differences in the chemicals and values

Question 2

Data Issues:

- Is the data set sufficiently robust and representative?

  ⇨ Could have more data on E side but predictiveness is good

- Has appropriate data screening and QA been conducted?

  ⇨ Yes

Question 3

Reliability Testing:

- Is the reliability testing that was conducted an appropriate method for evaluating SQVs?

- Is the comparative reliability analysis that was conducted an appropriate way of making decisions while fine-tuning the approach?

- Are the reliability measures that were used the right ones and were the relative weights given to them appropriate?

- Are there alternative methods of validation that could have been used without collecting additional data?

Reliability

- Sensitivity (100% – false negatives)
- Efficiency (100% – false positives)
- Predicted no-hit reliability
- Predicted hit reliability
- Overall reliability

All measures of reliability were used for ALL effects levels – endpoints given greater weight are shown in yellow

Freshwater Standards Reliability

No Effect Level

	% False Negatives	% False Positives	% Overall Reliability
acute endpoint (mortality)			
FPM	19	22	79
others	7	86	33
sub-lethal endpoints (growth)			
FPM	20	18	82
others	8	88	29

Values are averages across relevant assays

Minor Effect Level

	% False Negatives	% False Positives	% Overall Reliability
acute endpoint (mortality)			
FPM	22	15	84
others	36	36	64
sub-lethal endpoints (growth)			
FPM	18	13	87
others	40	37	63

Comparative Reliability Analysis

- East side vs. west side vs. combined
- TPH vs. PAH vs. combined
- Microtox – include?
- *Hyalella* growth – include Portland Harbor?
- Ammonia and sulfides issues
- N-qualified pesticides
- Blank-correction standardization
- Comparison to control vs. reference

SQV Validation

- RSET made a decision early on to not withhold part of the data set for validation
  - For such a large heterogeneous area, we needed all the data to develop the best possible model
  - Most other SQV sets have followed the same approach, with independent validation following after
- Independent validation will require a large, representative data set, not just a few projects
- Other validation methods available?

New Statistical Analyses

Designed to addressed issues of prevalence in the data set – more no-hits than hits (EPA/peer reviewers)

- Bias = predicted toxicity / actual toxicity
  - Bias > 1 indicates SQGs overpredict toxicity
  - Individual endpoint bias = 1.2 – 2.8
  - Overall SQS bias = 1.7; CSL bias = 1.1
  - Level of bias is appropriate

Statistical Analysis, cont.

- Odds Ratio = the odds that a prediction of toxicity is correct; higher is better
  - Individual endpoints OR = 9 – 30 for SQS, 9-100 for CSL
  - Overall SQS OR = 4.2, CSL OR = 3.7
  - Approximately an 80% chance that a prediction of toxicity will be correct
  - In accordance with Ecology's goals for the SQGs

Statistical Analyses, cont.

- Hanssen-Kuipers Discriminant is a goodness-of-fit kappa measure not affected by prevalence
  - HKD = 0.48 – 0.77 for individual endpoints, “moderate to substantial” agreement
  - HKD = 0.30 and 0.27 for SQS and CSL values (not developed by the model)
  - HKD is acceptable for modeled endpoints, not as good when other approaches are used to select values

Question 4

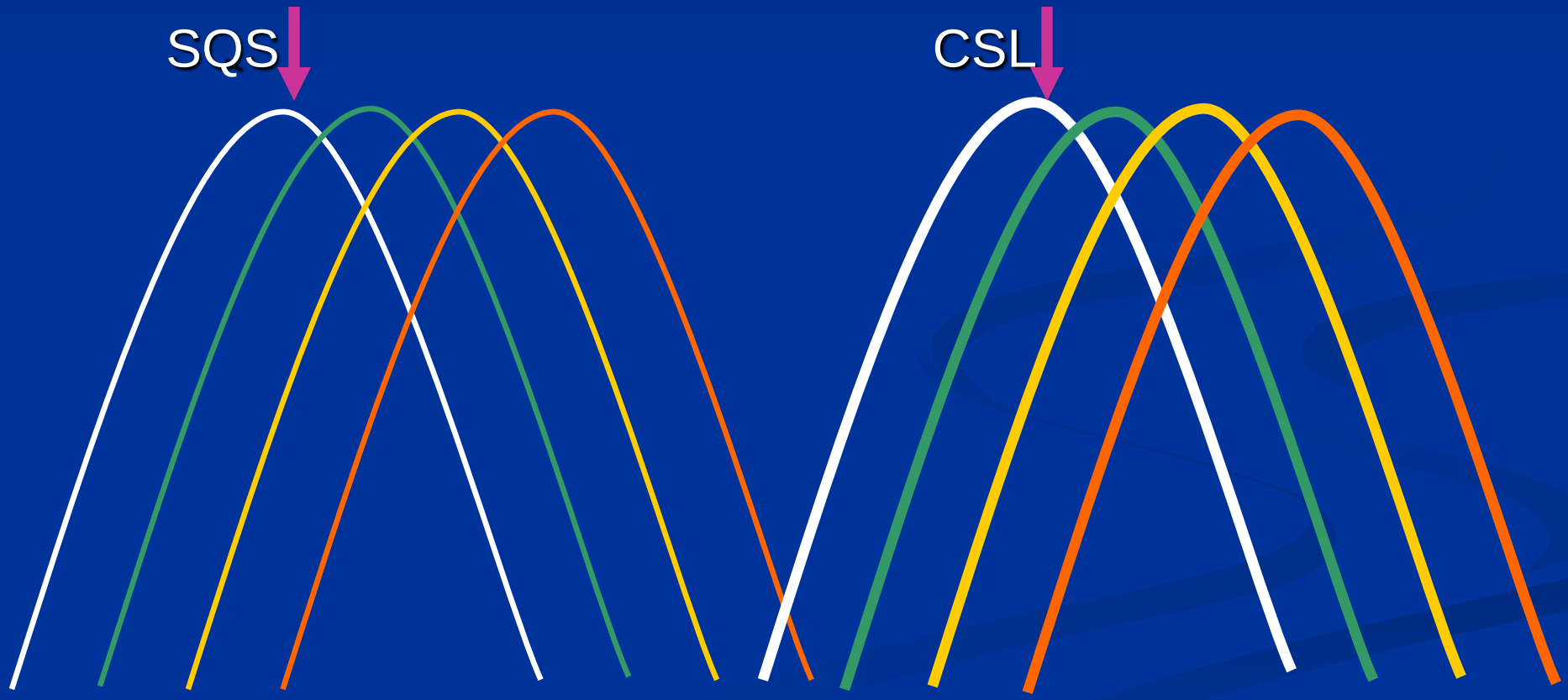
Data Interpretation and Regulatory Decision-Making:

- Is the overall framework and selection of final SQVs consistent with the SMS and marine SQVs?

- Is the approach used to select final SQVs scientifically defensible?

Challenges – Criteria Selection

Expectation: SQS values clustered below CSL values



Challenges – Criteria Selection

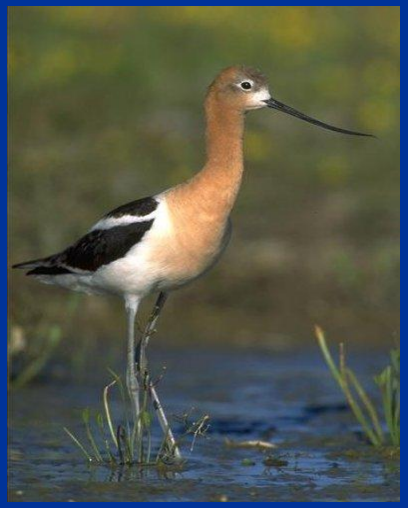
Reality: Differences between bioassays were much greater than differences between endpoints → try species sensitivity distribution approach

Table 3-7. Selection of Recommended Sediment Quality Guidelines

Analyte	Distribution of Floating Percentile Model Values[a]										SL1/SQS[b]	SL2/CSL[c]
Conventional Pollutants (mg/kg)												
Ammonia	230	300	> 780	> 780	> 780	> 780					230	300
Total sulfides	39	61	340	340	360	540	920	920			39	61
Metals (mg/kg)												
Antimony	0.3	0.3	12	42	42	42	42	> 63			0.3	12
Arsenic	14	14	16	16	120	120	120	180	200	200	14	120
Cadmium	2.1	2.1	5.4	6.3	6.3	13	13	> 23	> 23	> 23	2.1	5.4
Chromium	72	72	82	88	220	220	220	> 350	> 350		72	82
Copper	400	1200	1600	1600	1900	1900	> 1900	> 1900	> 11000		400	1200
Lead	360	360	> 1300	> 1300	> 1400	> 1400	> 1400	> 1400	> 1400	> 1400	360	> 1300
Mercury	0.66	0.66	0.8	0.8	0.8	0.87	0.87	> 0.87	3.04		0.66	0.8
Nickel	26	> 27	> 100	> 100	110	110	360	360	> 590	> 590	26	110
Selenium	11	11	> 20	> 20	> 20	> 20					11	> 20
Silver	0.58	0.64	1.7	1.7	4.1	4.1					0.58	1.7
Zinc	3200	3200	3200	> 4200	> 4200	> 14400	> 14400	> 14400			3200	> 4200
Organic Chemicals (µg/kg)												
4-Methylphenol	260	260	2000	2000	2400	2400	> 6300	> 6300			260	2000
Benzoic acid	2900	3800	3800	4100							2900	3800

- “>” values- no toxicity observed for that endpoint up to the listed concentration. Sample concentrations at or above this level should undergo toxicity testing.
- Approach for selection of CSL: “next significantly different value”

Questions?



Petroleum Toxicity

PAHs contribute to the greatest number of errors:

- PAHs do not sufficiently capture petroleum toxicity
- No form of normalizing or summing solves the problem
- Addressed through side-by-side PAH/TPH & combined model runs
- Best results when both were included
- May be legacy issues with not enough TPH data

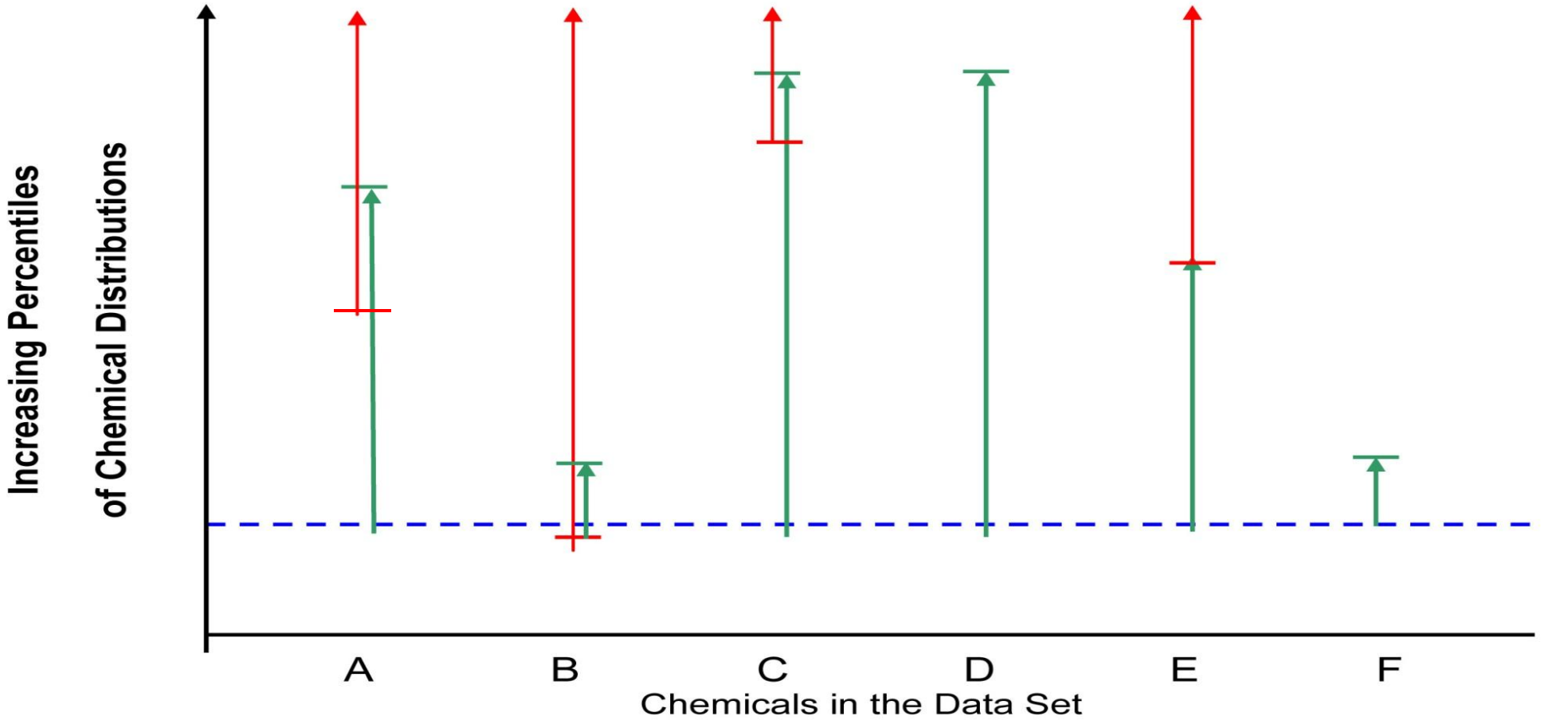
Covariance

- All chemicals are addressed simultaneously – avoids inappropriate assignment of toxicity
- Covariance analysis can be run ahead of time
- Model results allow visual identification of covariance patterns
- Appropriate chemical classes can be summed
- Those that can't be summed but often covary are subjected to multiple runs with different starting points to find the low concentration for each chemical, then selection of the combination with the highest reliability

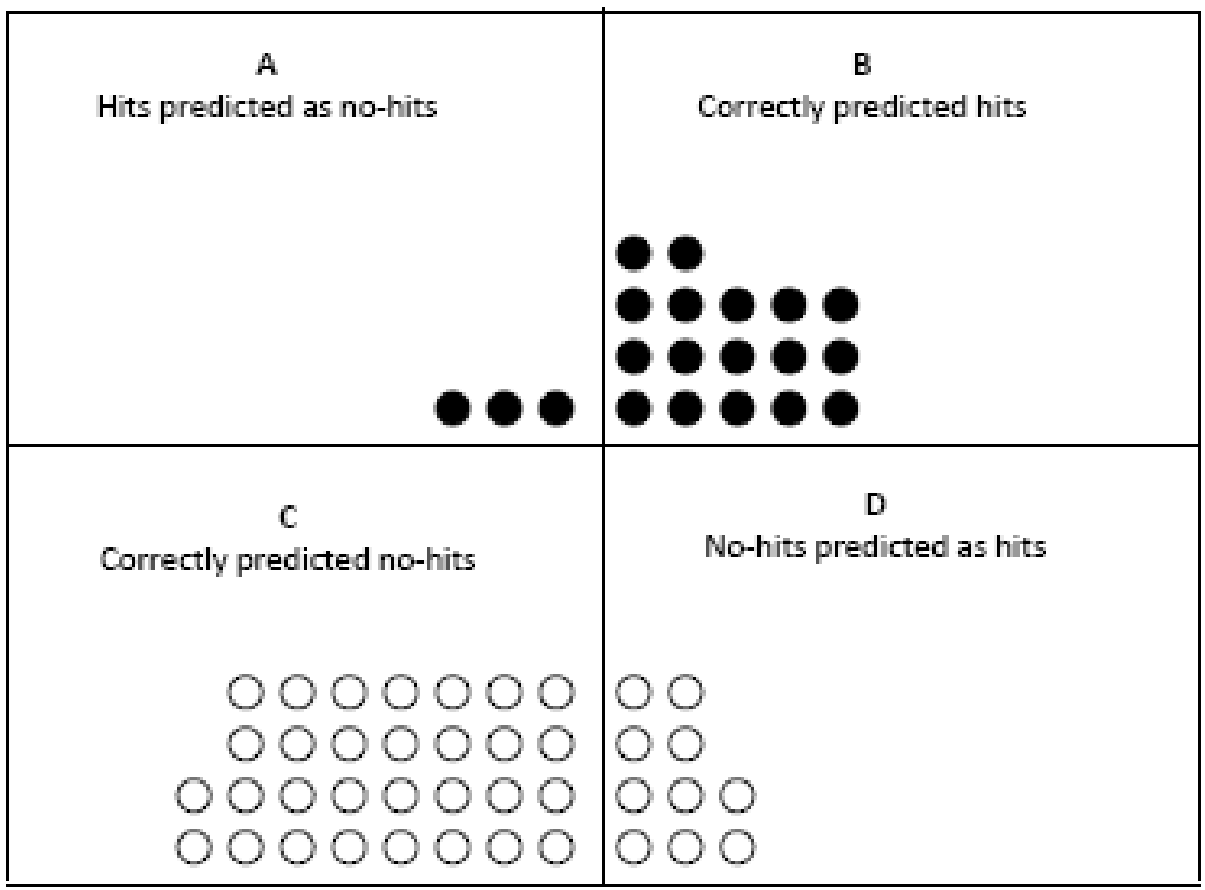
Summation of Chemical Classes

- PAHs/TPH classes
- PCB Aroclors
- Dioxins/Furans
- Chlordanes
- DDT, DDE, DDD isomers
- Heptachlors

Figure 2-1. Floating Percentile Method



Reliability Measures Diagram



Sensitivity = B / (A + B)
False Negatives = A / (A + B)

Predicted-Hit Reliability = B / (B + D)
Predicted-No-Hit Reliability = C / (A + C)

Efficiency = C / (C + D)
False Positives = D / (C + D)

Overall Reliability = (B + C) / (A + B + C + D)

Reliability Goals

The RSET Workgroup set the following goals before beginning SQV development:

	SQS (%)	CSL (%)
Sensitivity	80 – 90	75 – 85
Efficiency	70 – 80	75 – 85
Predicted Hit Reliability	70 – 80	75 – 85
Predicted No-Hit Reliability	80 – 90	75 – 85
Overall Reliability	80 – 90	80 – 90

Table 3-3. Reliability of the FPM Results and Existing SQG Sets at the SQS/SL1 Level

a. Chironomus 10-day growth

FPM FN Percentiles	% False Negatives	% False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
5	4.6	44.8	95.4	55.2	23.1	98.8	60.2
10	9.2	35.9	90.8	64.1	26.3	98.0	67.4
15	13.8	31.7	86.2	68.3	27.7	97.2	70.5
20	20.0	17.0	80.0	83.0	40.0	96.7	82.7
25	24.6	19.6	75.4	80.4	35.3	95.9	79.8
30	29.2	13.5	70.8	86.5	42.6	95.4	84.6

SQG	% False Negatives	% False Positives	Hit Reliability	NoHit Reliability	PredHit Reliability	PredNoHit Reliability	Overall Reliability
ERL	6.2	85.9	93.8	14.1	13.4	94.2	24.0
TEL	4.6	91.3	95.4	8.7	12.9	93.0	19.4
TEC	7.7	79.6	92.3	20.4	14.1	94.9	29.3
LEL	9.2	88.3	90.8	11.7	12.7	90.0	21.5

Proposed Biological Sediment Standards

Regulatory Framework

- Confirmatory bioassays override chemistry
- Two tier structure: SQS and CSL
- Bioassay suite – Multiple species/sensitive life-history stages
- Minimum of 3 endpoints
- Both acute and chronic tests

Bioassay Interpretation: Comparison to Control

Test	QA limits Control	QA limits Reference	SQS	CSL
Hyalella azteca				
10-day mortality	C ≤ 20%	R ≤ 25%	T – C > 15%	T – C > 25%
28-day mortality	C ≤ 20%	R ≤ 30%	T – C > 10%	T – C > 25%
28-day growth	CF ≥ 0.15 mg/ind	RF ≥ 0.15 mg/ind	T/C < 0.75	T/C < 0.6
Chironomus dilutus				
10-day mortality	C ≤ 30%	R ≤ 30%	T – C > 20%	T – C > 30%
10-day growth	CF ≥ 0.48 mg/ind	RF/CF ≥ 0.8	T/C < 0.8	T/C < 0.7
20-day mortality	C ≤ 32%	R ≤ 35%	T – C > 15%	T – C > 25%
20-day growth	CF ≥ 0.48 mg/ind	RF/CF ≥ 0.8	T/C < 0.75	T/C < 0.6

C=Control, R=Reference, T=Test, F=Final, (SQS &CSL hits statistically sign. diff.)