Use of Receiver Operating Characteristic (ROC) Curves to Select and Evaluate Sediment Quality Guidelines

Jim Shine, Crista Trapp: Dept. Env. Health

Brent Coull: Dept. Biostatistics

Harvard School of Public Health

Outline:

1. Description of ROC Curves

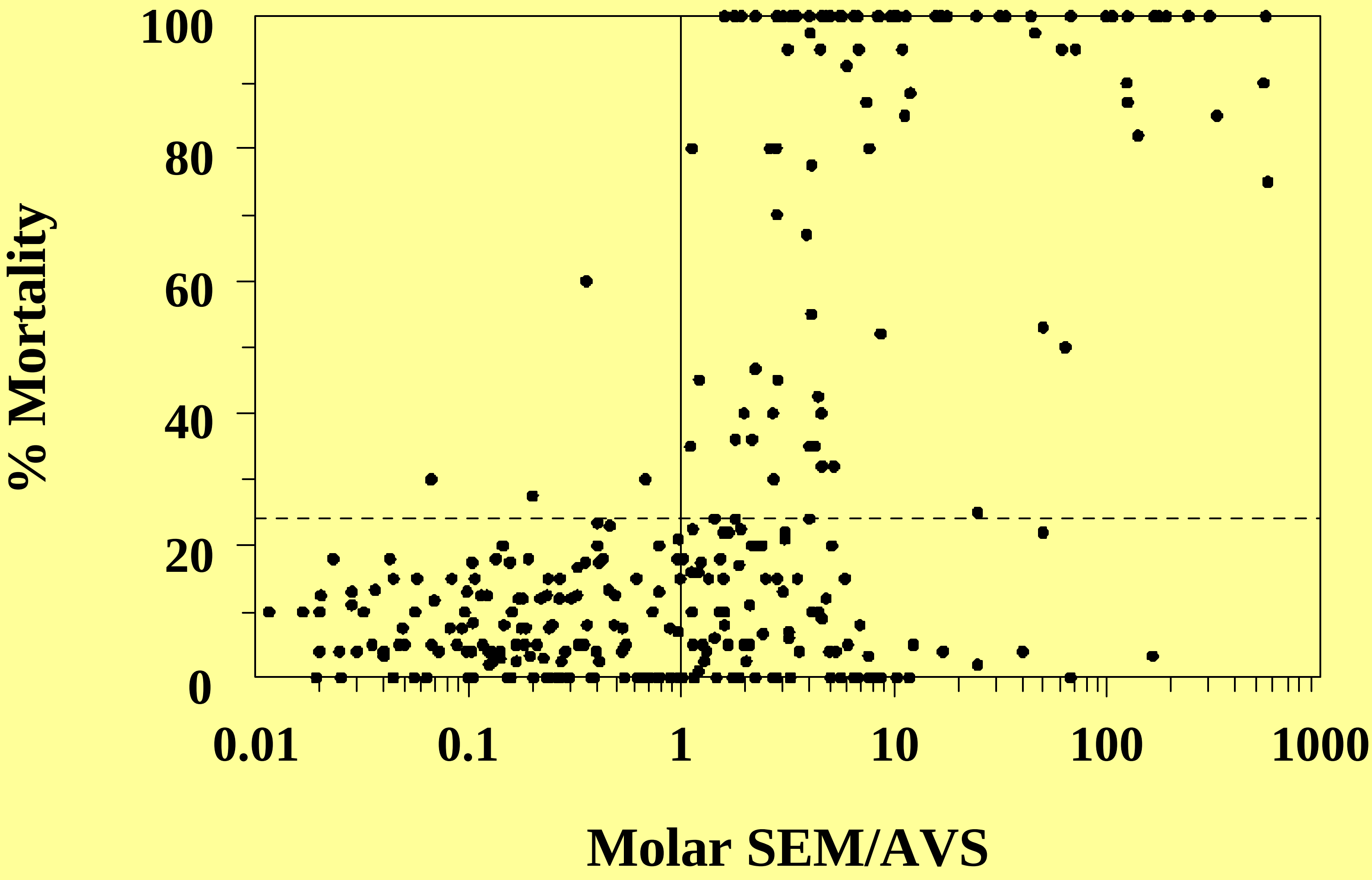
2. Application to Sediment Quality Guidelines for Metals

- Value of a metric that optimizes true positive and false positive classification rates

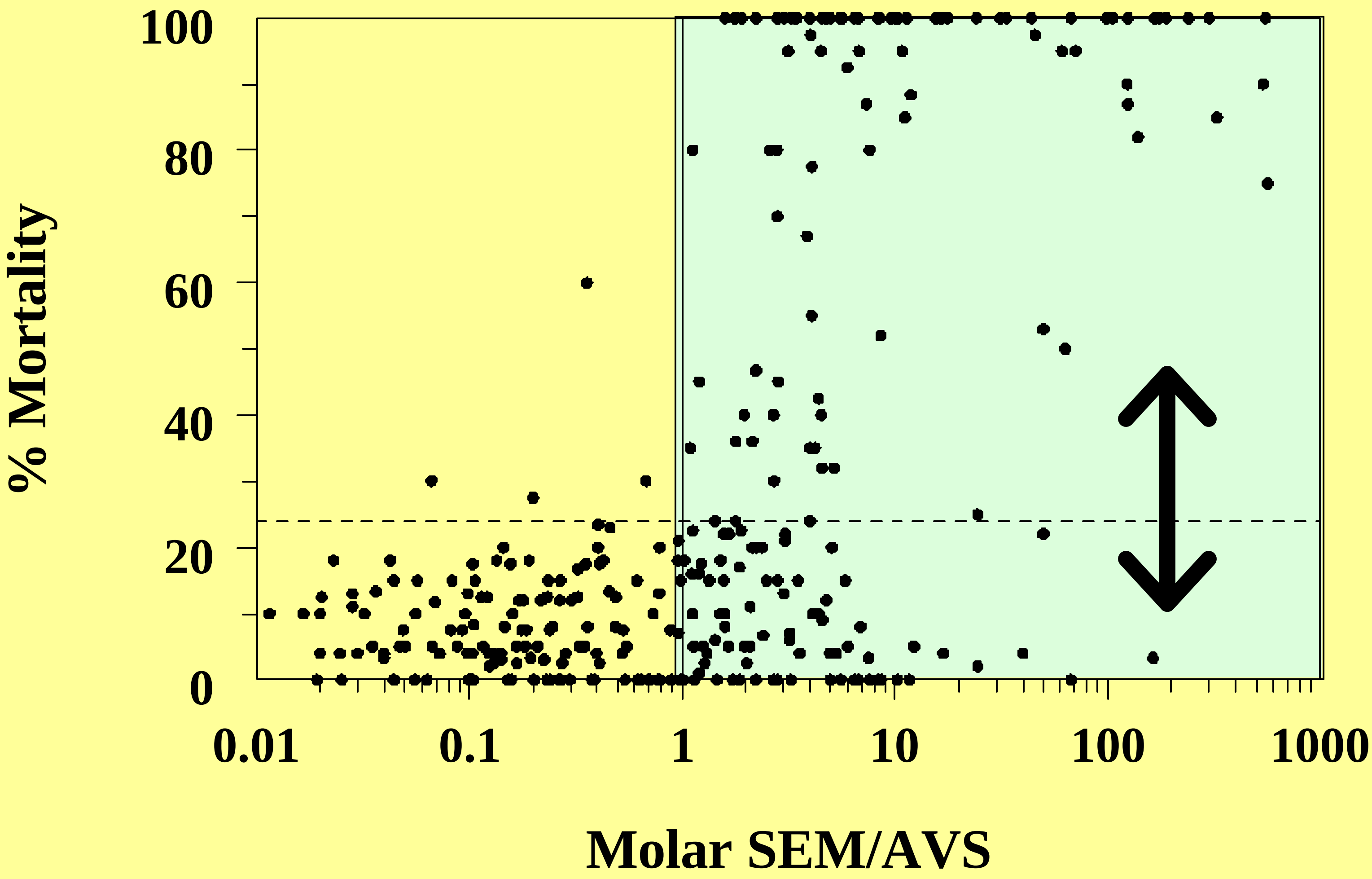
- Non ad-hoc comparisons of different metrics

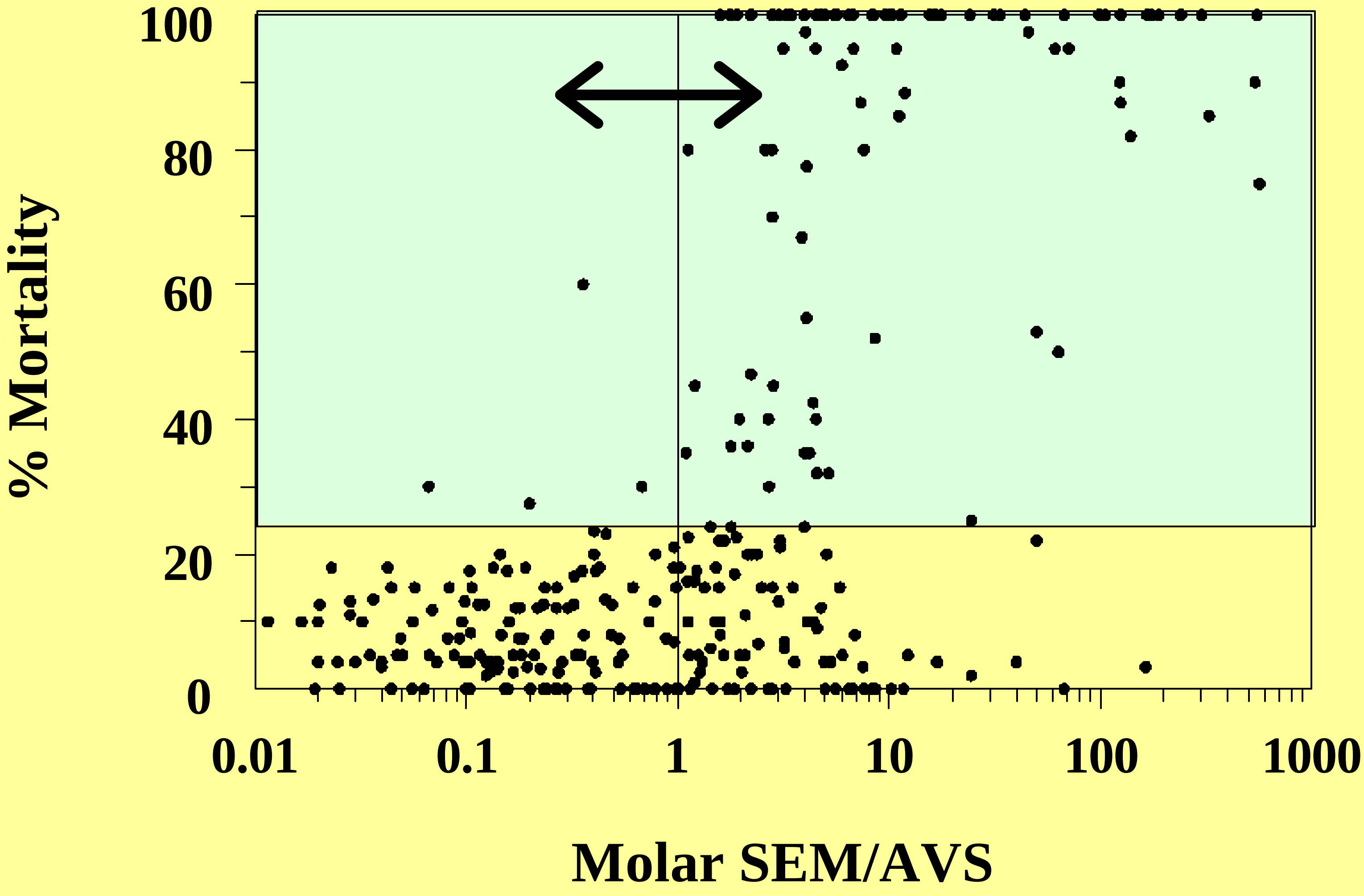
  - Speciation (i.e., SEM:AVS) vs. 'Total Metal' (i.e., NOAA ERM) based approaches

- 3. Reanalysis of NOAA/BEDS Database using ROC Curves



100
80
60
40
20
0
% Mortality
0.01
0.1
1
10
100
1000
Molar SEM/AVS


100
80
60
40
20
0
% Mortality
0.01
0.1
1
10
100
1000
Molar SEM/AVS



100
80
60
40
20
0
% Mortality
0.01
0.1
1
10
100
1000
Molar SEM/AVS

What Questions do Environmental Managers Need to Ask?

For a Given a Sediment Quality Guideline….

1. What is the likelihood that a sample above the guideline is toxic?

    - What is the probability that a sample below the guideline is non-toxic

2. What is the probability that I will correctly classify toxic samples as toxic? (Sensitivity)

    - What is the probability that I will correctly classify a non-toxic sample as non-toxic?

What is an ROC Curve?

-ROC stands for Receiver Operating Characteristic Curve

-Developed by radar operators during World War II

-Commonly used in the biomedical field to assess the discriminatory power of diagnostic tests

-A good discriminatory test has high sensitivity (correctly classifying an affected individual as affected) and high specificity (correctly classifying an unaffected individual as unaffected).

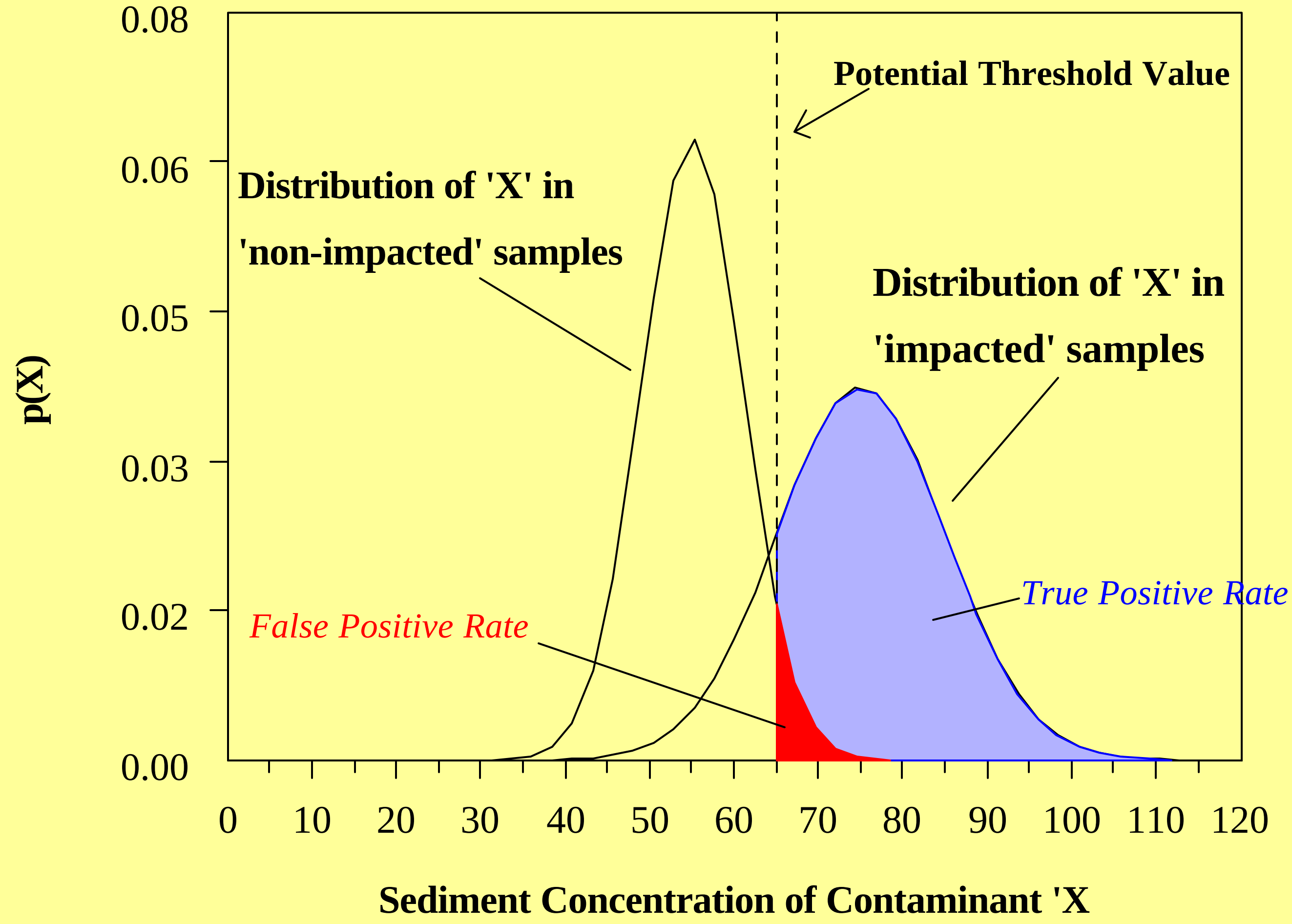
-**ROC Curves can:**

-Evaluate the overall discriminatory power of a given metric

-Make non ad-hoc comparisons amongst different tests with different units

-Elucidate the trade-offs in sensitivity and specificity when using a metric

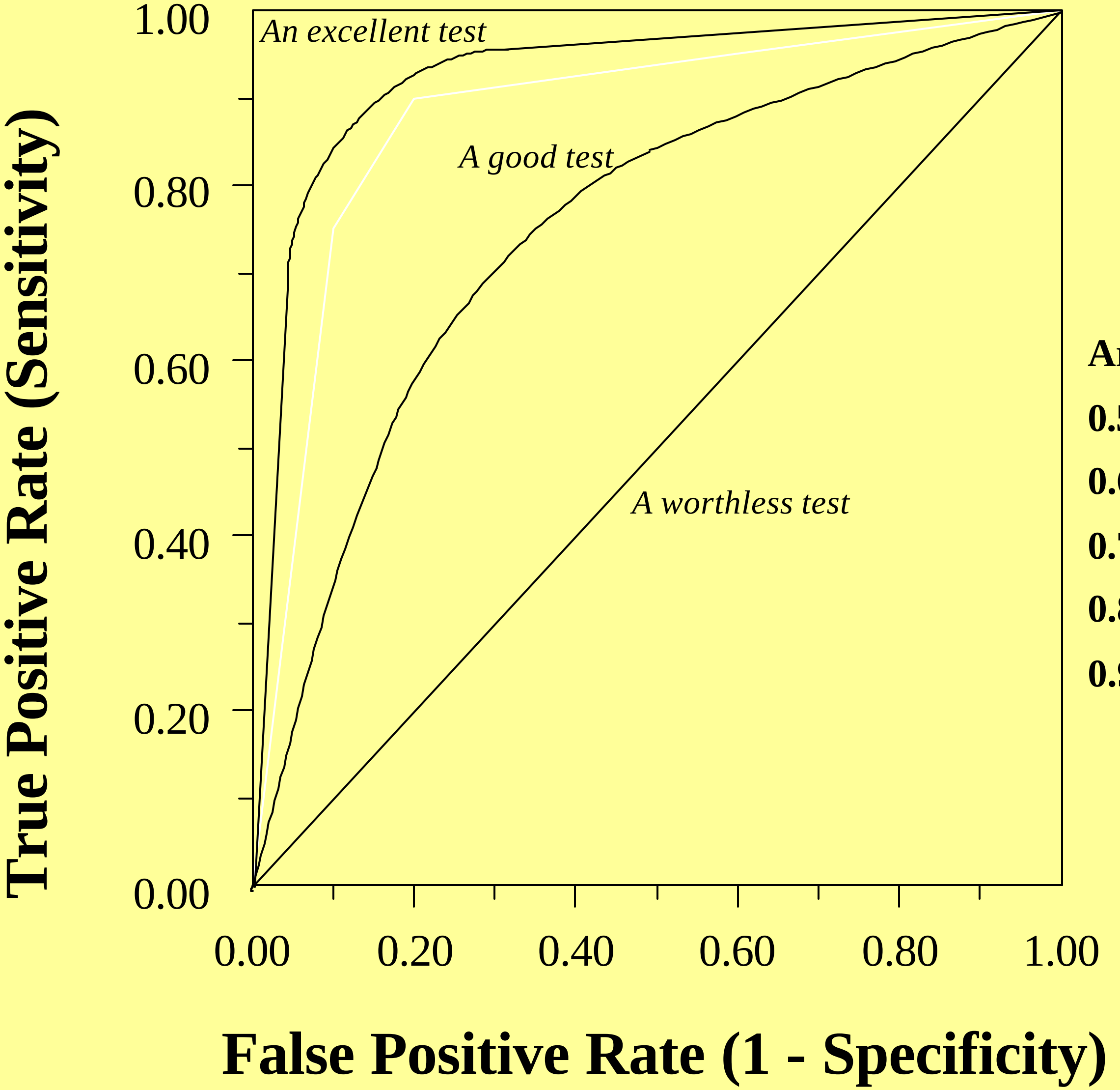
-Aid in selection of a value that best balances sensitivity and specificity.



- All values tested as potential thresholds

- True Positive Rate (Sensitivity) and False Positive Rate (1-Specificity) for each potential threshold graphed

Sample ROC Curves



Area Under the Curve:

0.5 - 0.6 = Fail (F)

0.6 - 0.7 = Poor (D)

0.7 - 0.8 = Fair (C)

0.8 - 0.9 = Good (B)

0.9 - 1.0 = Excellent (A)

Endpoints Used to Predict the Acute Toxicity of Heavy Metals in Marine Sediments:

- Speciation Based Metrics:
  - SEM/AVS ratio
  - SEM - AVS difference
  - (SEM-AVS)/f_{oc}
- Total Metal Based Metrics:
  - NOAA Effects Range Approaches
    - ERM, ERL based on distribution of 'effects' data in BEDS database
  - TEL, PEL Approaches
    - geometric means of data from effects and no-effects distributions

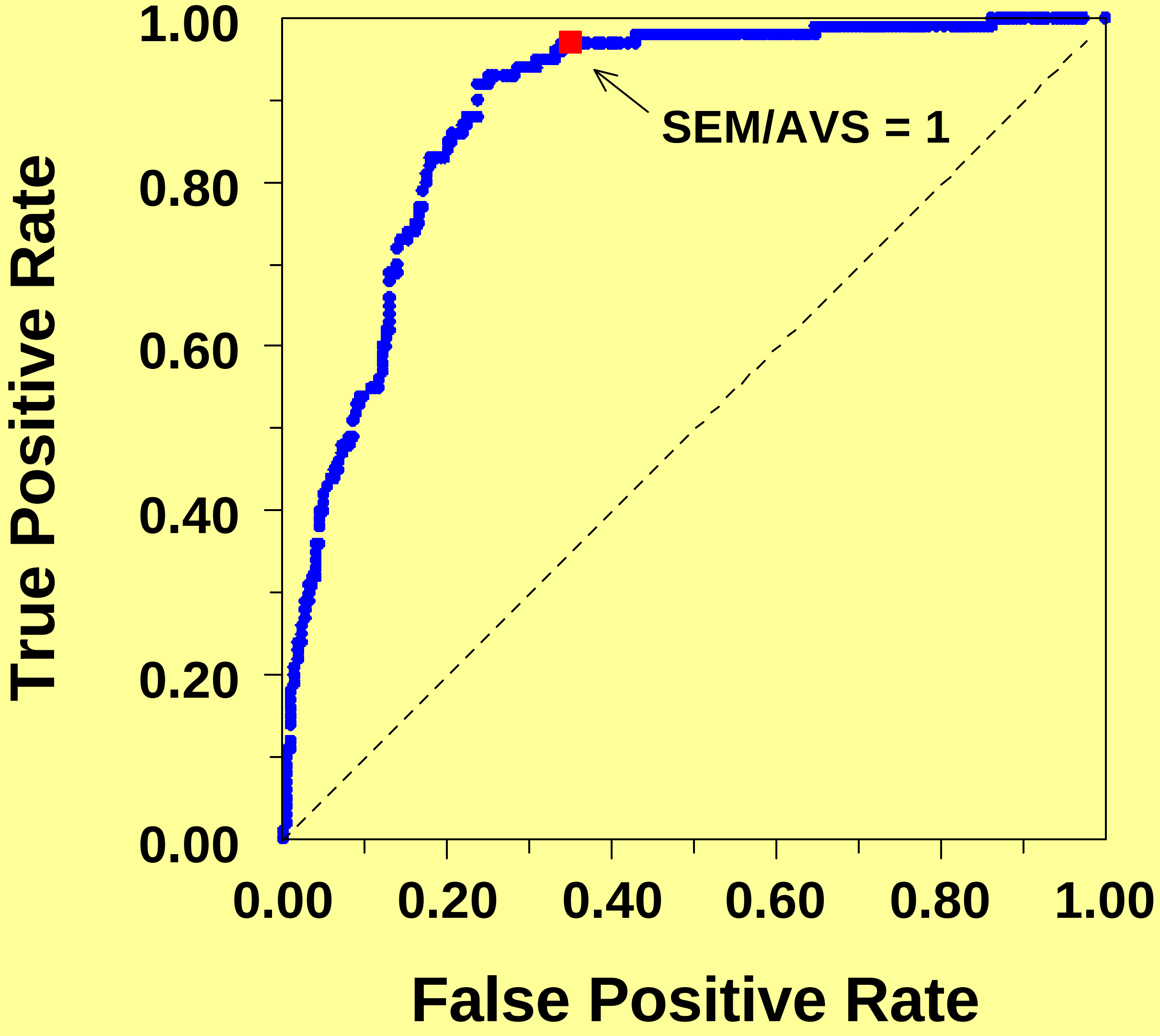
Sources of Data:

Reference	**(n)**
Berry *et.al.* (1996)	88
Berry *et.al.* (1999)	21
Call *et.al.* (1999)	2
Carlson *et.al.* (1991)	30
Casas and Crecelius (1994)	19
Hansen *et.al.* (1996)	118
Kemble *et.al.* (1994)	30
Pesch *et.al.* (1995)	49
Total Sample Size:	**357**

Test organisms include: *Hyalella azteca, Chironomus riparius, Neanthes arenaceodentata, Capitella capitata, Lumbriculus variegatus, Helisoma* spp., *Ampelisca abdita*

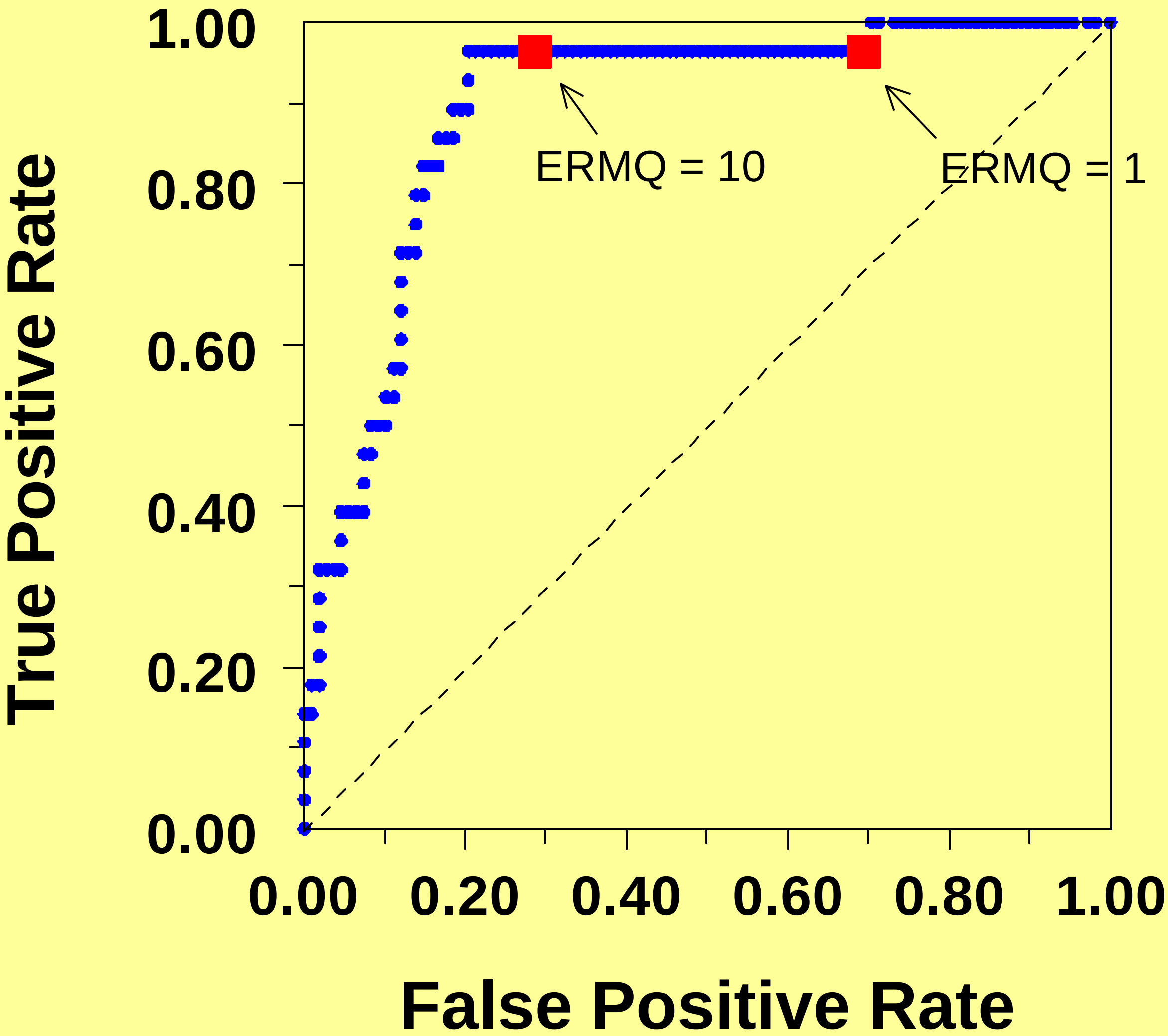
SEM/AVS ROC Curve

AUC = 0.88 ± 0.03



Mean ERMQ ROC Curve

AUC = 0.88 ± 0.03



Results: Areas Under ROC Curves

- Index of overall discriminatory power

Approach	Area Under ROC Curve (± 1 s.d.)
SEM/AVS	0.88 ± 0.03
SEM-AVS	0.84 ± 0.02
(SEM-AVS)/f_{oc}	0.87 ± 0.02
ERLQ	0.87 ± 0.03
ERMQ	0.88 ± 0.03
TELQ	0.86 ± 0.03
PELQ	0.87 ± 0.03

True Positive and False Positive Rates of Common Endpoints

Metric	Value	True Positive Rate	False Positive Rate
SEM/AVS	1	0.97	0.35
SEM-AVS	0	0.97	0.34
(SEM-AVS)/f_{oc}	0	0.95	0.26
ERLQ	1	1.00	0.85
	10	0.96	0.44
ERMQ	1	0.96	0.68
	10	0.96	0.28
TELQ	1	1.00	0.92
	10	0.96	0.53
PELQ	1	0.96	0.73
	10	0.96	0.33

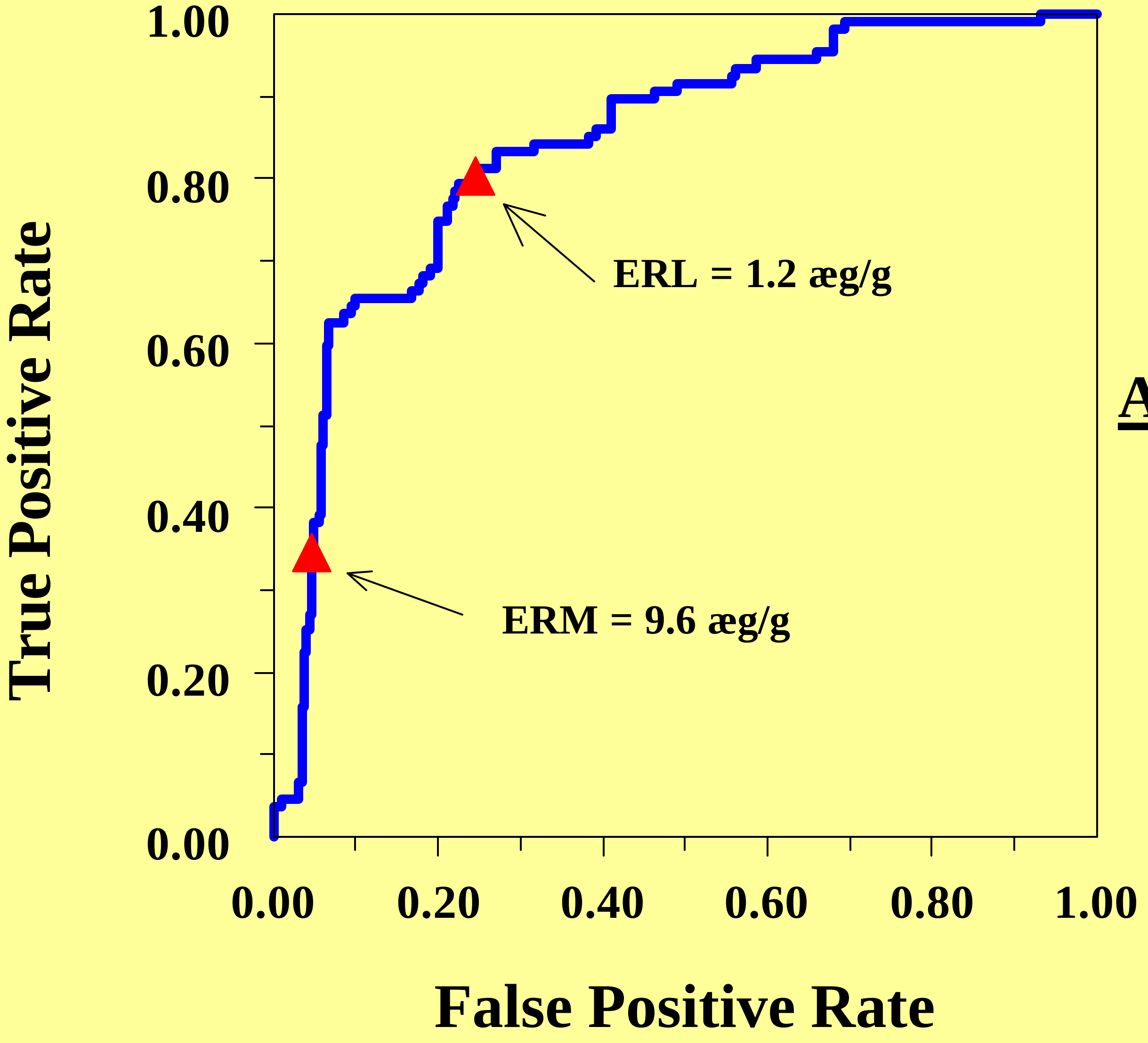
Endpoint Values Giving Desired Sensitivity, Specificity

Metric	True Positive Rate = 0.9	False Positive Rate = 0.1	Sensitivity = Specificty
SEM/AVS	1.79	5.00	2.69 [0.82]
SEM-AVS	1.54	109	5.7 [0.82]
(SEM-AVS)/f_{oc}	308	7788	367 [0.82]
ERLQ	65.8	1320	146 [0.82]
ERMQ	23.8	188	35.5 [0.82]
TELQ	79.6	1889	263 [0.82]
PELQ	33.7	351	59.2 [0.82]

Can we use ROC curves to reanalyze the BEDS database and establish more 'efficient' SQG's for individual metals?

Results: Reanalysis of BEDS Database Using ROC Curves

- Results for Cd



AUC = 0.84

- For individual metals (Conc = μg/g):

	ERM	FPR = 0.1	ERL	TPR = 0.9
Cd	9.6	3.1	1.2	0.61
Cu	270	147	34	14.2
Pb	218	123	46.7	24.2
Ni	51.6	79	20.9	8.8
Ag	3.7	1.8	1.0	0.24
Zn	410	260	150	126

In General:

ERM > FPR = 0.1 value > ERL > TPR = 0.9 value

Will these new SQG's give us more discriminatory power in the test database when the individual SQG's are combined into a 'mean quotient'?

"New Quotient" Denominator	Area Under ROC Curve
TPR = 0.9 value	0.87
FPR = 0.1 value	0.87
Sens. = Spec. value	0.87

Conclusions:

1. ROC Curves - Applicable to Ecological Studies?

- depends on the question being asked

2. Overall discriminatory power of current models to combine data for different metals do not differ

- Speciation x Total metal based approaches

3. Common values used as thresholds may not provide desired specificity or sensitivity

- Assuming results for these test organisms sufficiently correlated with ecological endpoints of concern

4. Reanalysis of BEDS database

- Interesting results for specificity, sensitivity of ERL, ERM values for individual metals?

- New Quotients combining metals do not provide better discriminatory power in test database