Analysis & interpretation of evidence

Hypothesis testing: a simple scenario

SCENARIO 1

- DO measured upstream & downstream over 9 months
  - Upstream: $\bar{x}$ = 9.3 mg/L
  - Downstream: $\bar{x}$ = 8.4 mg/L
- Difference significant at P<0.05

SCENARIO 2

- DO measured upstream & downstream over 3 months
  - Upstream: $\bar{x}$ = 7.9 mg/L
  - Downstream: $\bar{x}$ = 4.2 mg/L
- Difference **not** significant at P<0.05

Which scenario presents a stronger case for DO causing impairment?

Use statistics responsibly

Use caution in interpreting differences

- Look at magnitude & consistency of differences, rather than statistical significance
- Statistical significance detects differences exceeding natural variance
  - Does not detect stressor effects
  - Does not equal biological significance
- Can use statistics, but think about what you are doing
  - Consider the relationship between minimum detectable difference (power) & the biologically relevant difference

Analyzing Data

This page provides access to descriptions of graphical and statistical techniques that are useful for analyzing data for a causal assessment.

Understand before applying any statistical method that:

Traditional statistical approaches to data interpretation are often not appropriate for causal analysis.

- Field data rarely meet the assumptions and requirements of statistical tests, which were designed for the analysis of experimental results, and
- Causal analysis requires multiple lines of evidence indicating stressor influence on biological responses. Traditional hypothesis testing does not indicate whether or not the stressor influenced the response, it only indicates whether variability in the response is greater than one would expect from random variation.

Therefore, prior to selecting methods for your data analysis, understanding the fundamentals of data analysis for causal assessment will help you plan a rigorous analysis and produce a successful assessment.

The table below lists a selection of graphical and statistical methods along with links to advice on how to use them when applying data to different steps of a causal analysis.

Step	Useful Methods
Step 2: List Candidate Causes	• Scatter plots • Correlation analysis • Box plots • Conditional probability analysis

Causal Analysis/Diagnosis Decision Information System (CADDIS)

Analyzing Data

Data Analysis Methods | Fundamentals of Data Analysis | **Get Data Analysis Tools**

This page provides access to tools that you can download and use to analyze your data.

The tools provided on this page were developed in different platforms and include a spreadsheet template, a menu-driven statistical package, and programming code. Additional tools will be developed for this page, so please share input on your preferred platform type and suggestions for additional methods.

The tools you select depend on the methods you need and your comfort level with programming. Are you:

- **Inexperienced with programming?** We recommend that you begin by using the tools that do not require programming expertise. CADStat for, example, will allow you to conduct several types of statistical analyses using a menu-driven interface, and the Species Sensitivity Distribution (SSD) Generator provides detailed instructions and macros that may be used to generate an SSD.
- **Familiar with the R Statistical package?** Some of our more complex tools require that you be familiar with command-line statistical programming. Command-line programming provides more flexibility for analyses, including batch processing of data. We have provided programs in R, a freely available statistical program. More information on R and on downloading our programs is provided below.

Tool 1: Species Sensitivity Distribution Generator

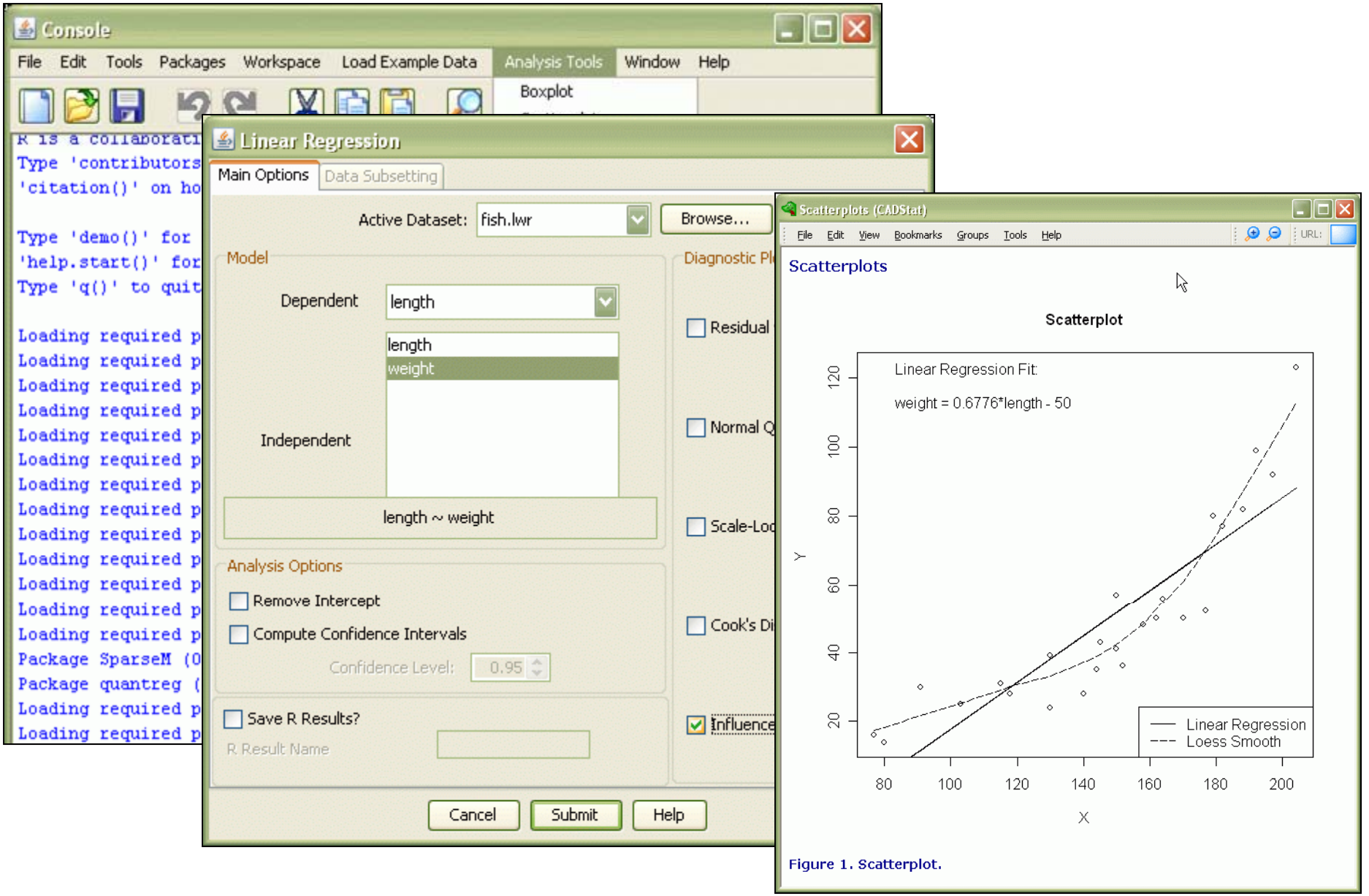
Download the SSD_Generator_V1.xlt (2.8MB, .xlt) to calculate and plot the proportion of species affected at different levels of exposure in laboratory toxicity tests. See the SSD methods page for further details on the use of SSD plots in causal analysis. EPA (2005) provides more detail on selecting data for SSDs and the method used in generating them.

This is a Microsoft Excel template that depends on macros for operation, so you must select **"enable macros"** when you open the template.

Data analysis tools

- CADStat
  - Menu-driven package of data visualization & statistical techniques, based on JGR (Java GUI for R)
- Command-line R scripts
  - Powerful and free statistical package [http://www.r-project.org/]
  - R scripts provided for PECBO
- Species sensitivity distribution (SSD) generator
  - Tool that calculates & plots proportion of species affected at different levels of exposure in laboratory toxicity tests

CADStat


Console
File Edit Tools Packages Workspace Load Example Data Analysis Tools Window Help
Boxplot
Type 'contributors
'citation()' on ho
Type 'demo()' for
'help.start()' for
Type 'q()' to quit
Loading required p
Package SparseM (0
Package quantreg (
Linear Regression
Main Options
Data Subsetting
Active Dataset:
fish.lwr
Browse...
Model
Dependent
length
length
weight
Independent
length ~ weight
Diagnostic Pl
Residual
Normal Q
Scale-Loc
Cook's Di
Influence
Analysis Options
Remove Intercept
Compute Confidence Intervals
Confidence Level:
0.95
Save R Results?
R Result Name
Cancel
Submit
Help
Scatterplots (CADStat)
File Edit View Bookmarks Groups Tools Help
URL:
Scatterplots
Scatterplot
Linear Regression Fit:
weight = 0.6776*length - 50
Y
X
20
40
60
80
100
120
80
100
120
140
160
180
200
Linear Regression
Loess Smooth
Figure 1. Scatterplot.

Begin with data exploration

- Want to examine relationships between different variables
  - Linear & more complex
  - Expected & unexpected

- 1st step is visualizing data
  - Scatterplots & boxplots
  - No statistics, hypotheses, or p-values

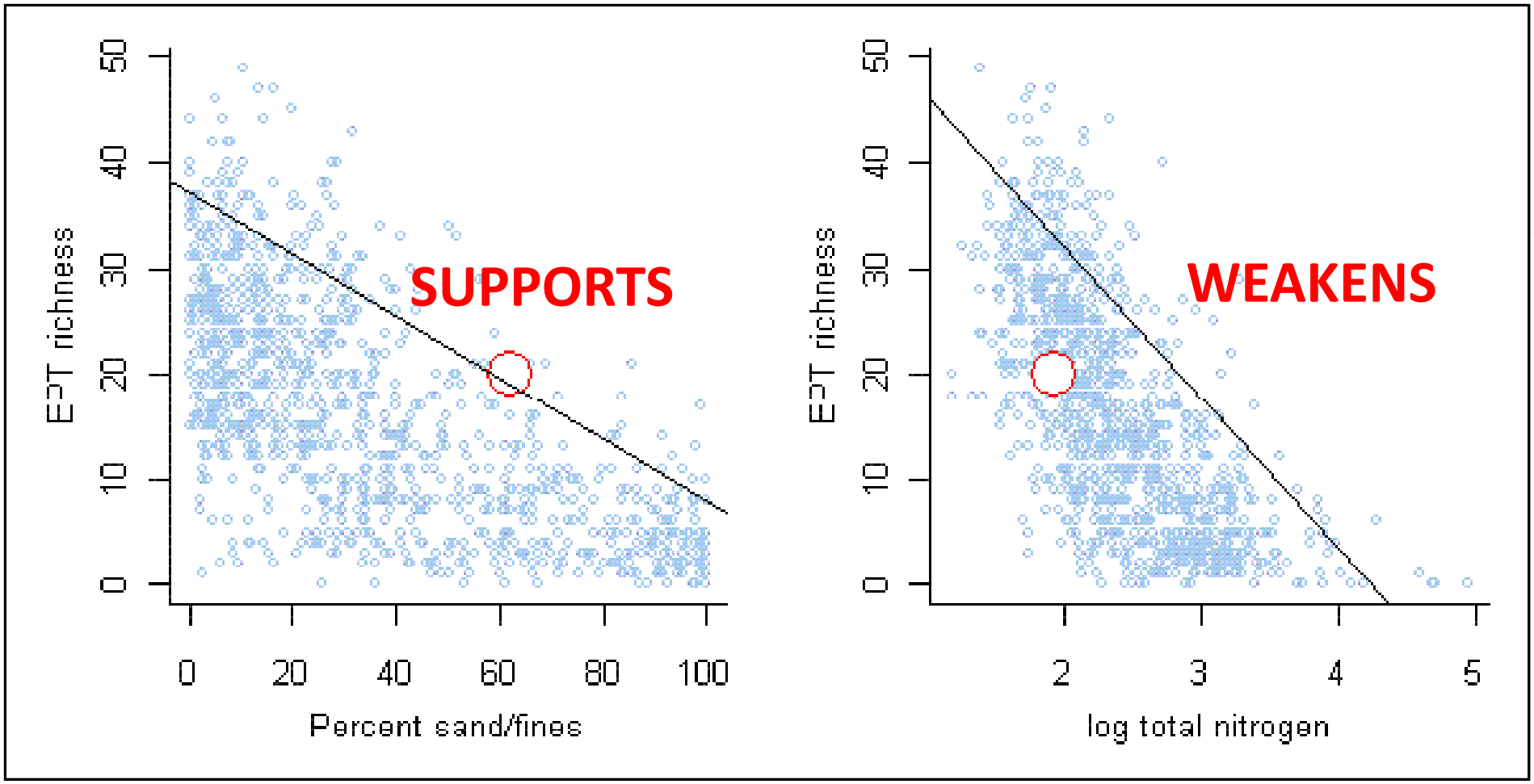
Quantile regression

- Models relationship between specified quantile of response variable and explanatory variable (stressor)
  - 50^{th} quantile gives median line, under which 50% of observed responses occur
  - 90^{th} quantile gives line under which 90% of observed responses occur

- Provides means of estimating location of upper boundary on scatterplot

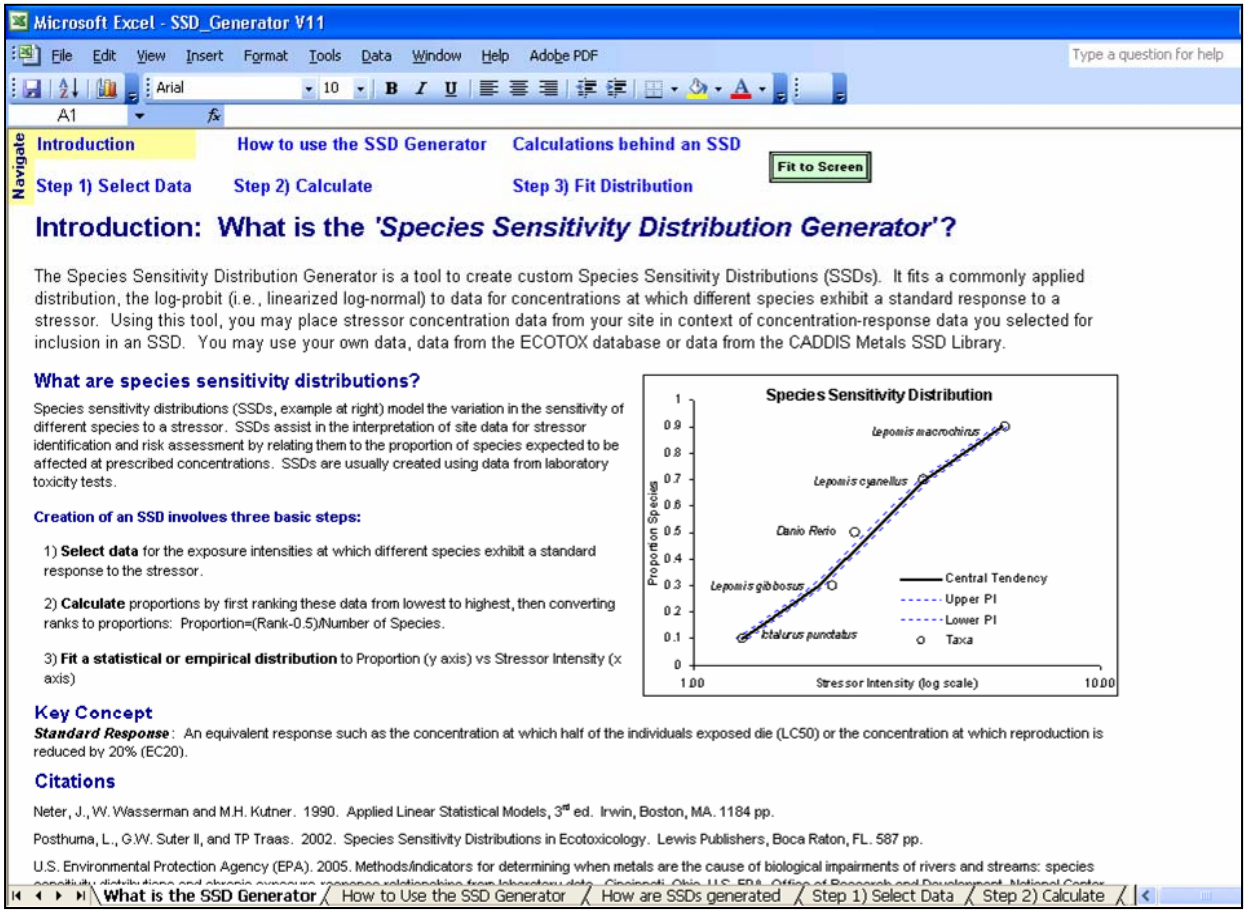
- Used to evaluate stressor-response relationships from other field studies (Step 4: Evaluating data from elsewhere)

Applying quantile regression

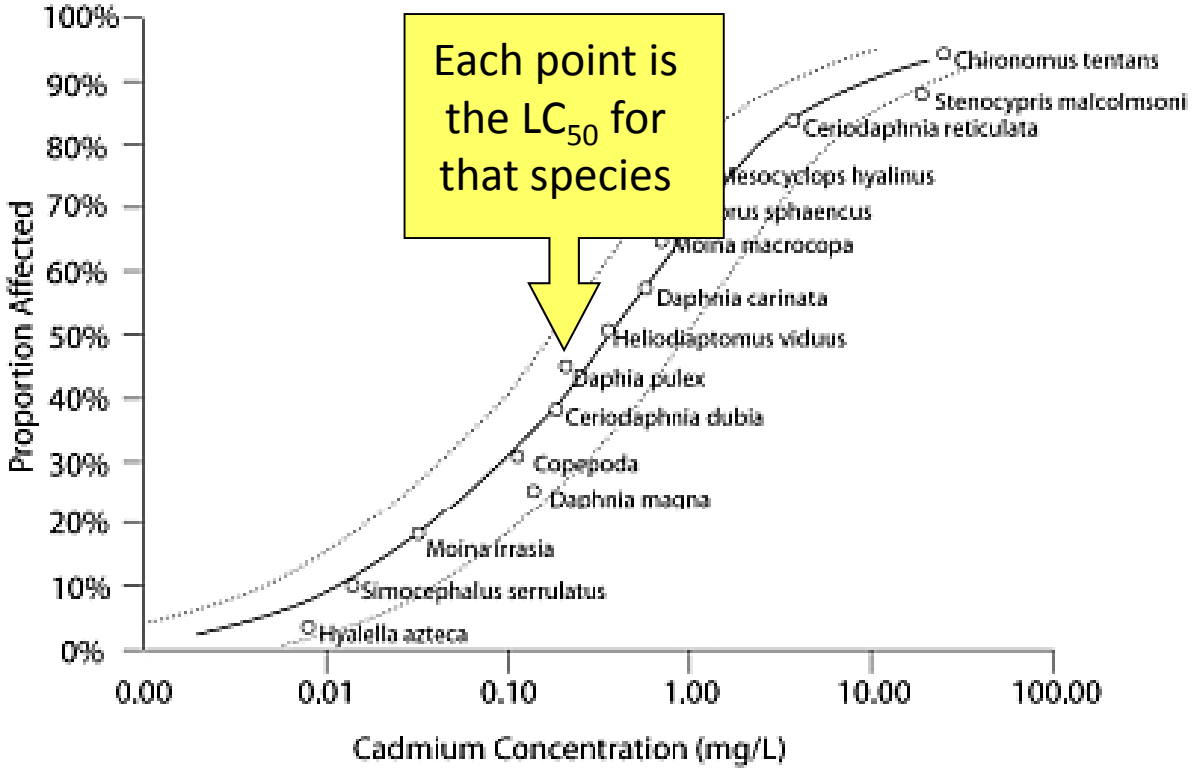


Data analysis tools

- Species sensitivity distribution (SSD) generator
  - Tool that calculates & plots proportion of species affected at different levels of exposure in laboratory toxicity tests

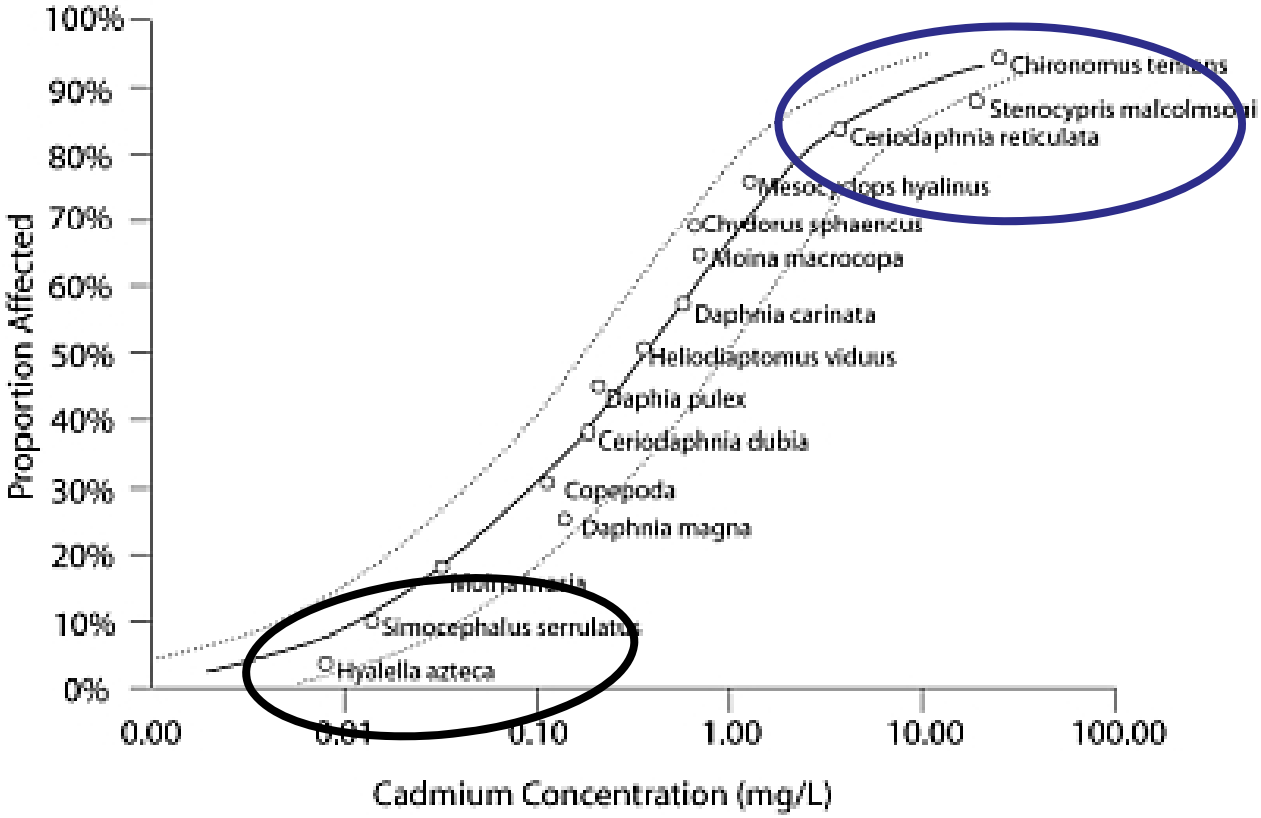


What is a species sensitivity distribution

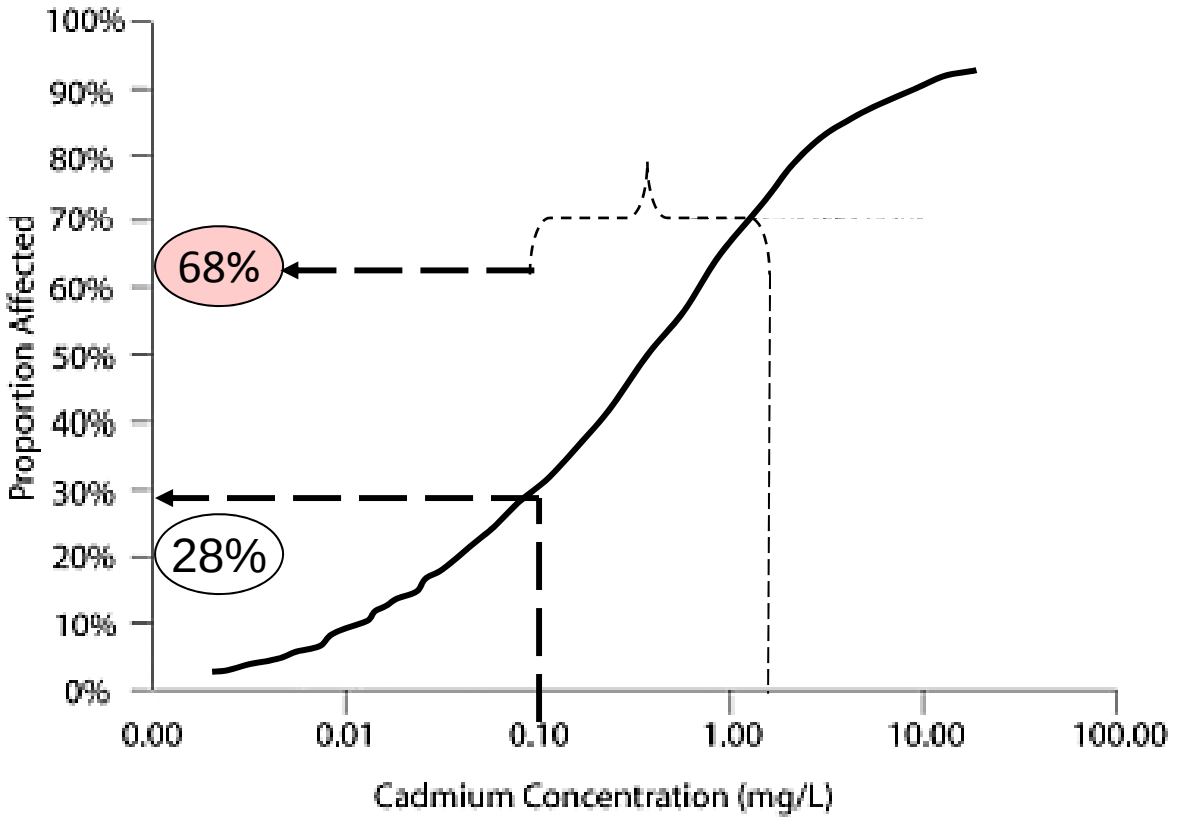


- **An exposure-response model for a single stressor**
- **Basis for NAWQC**
- **Represents relative sensitivities to the stressor of interest**

Use species at the site as evidence of specific alteration specific to the cause



Use effect and concentration at the site compared to effect and concentration of the SSD curve as evidence of sufficiency of the cause



For an SSD of LC50s, the corresponding effect would be local extirpation or occurrence in a kill event.

Quality Control Advice

- Lab conditions should be similar to field conditions
- Quality of the SSD depends on relevance of the data used
  - Life-stages of organisms
  - Duration of test
  - Responses