



# 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



# Causal Analysis/Diagnosis Decision Information System (CADDIS)

Recent Additions | Contact Us Search:  GO  
 You are here: [EPA Home](#) » [CADDIS](#) » [Analyzing Data](#) » Data Analysis Methods

- CADDIS Home
- Basic Information
- Frequently Asked Questions
- Step-by-Step Guide
  - Step 1: Define the Case
  - Step 2: List Candidate Causes
  - Step 3: Evaluate Data from the Case
  - Step 4: Evaluate Data from Elsewhere
  - Step 5: Identify Probable Cause
  - Summary Table of Scores
  - Summary Tables of Types of Evidence
- Examples
- Candidate Causes
- Analyzing Data**
- Information Sources
  - Related Links
  - Databases
  - Glossary
  - References
- Site Map

## Analyzing Data



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

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.

- Links to Methods**

  - [M.1. Scatter Plots](#)
  - [M.2. Correlation Analysis](#)
  - [M.3. Box Plots](#)
  - [M.4. Conditional Probability Analysis](#)
  - [M.5. Regression Analysis](#)
  - [M.6. Predicting Environmental Conditions From Biological Observations](#)
  - [M.7. Quantile Regression](#)
  - [M.8. Classification and Regression Trees](#)
  - [M.9. Species Sensitivity Distributions](#)

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
<b>Step 2: List Candidate Causes</b>	<ul style="list-style-type: none"> <li>• <a href="#">Scatter plots</a></li> <li>• <a href="#">Correlation analysis</a></li> <li>• <a href="#">Box plots</a></li> <li>• <a href="#">Conditional probability analysis</a></li> </ul>



# Causal Analysis/Diagnosis Decision Information System (CADDIS)

[Recent Additions](#) | [Contact Us](#)Search:  [GO](#)You are here: [EPA Home](#) » [CADDIS](#) » [Analyzing Data](#) » Get Data Analysis Tools

## 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.

CADDIS Home  
Causal Information  
Frequently Asked Questions  
Step-by-Step Guide  
Step 1: Define the Case  
Step 2: List Candidate Causes  
Step 3: Evaluate Data from the Case  
Step 4: Evaluate Data from Elsewhere  
Step 5: Identify Probable Cause  
Summary Table Scores  
Summary Tables  
Types of Evidence  
Examples  
Candidate Causes  
Analyzing Data  
Information  
Resources  
Related Links  
Databases  
Glossary

# 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



# Begin with data exploration

---

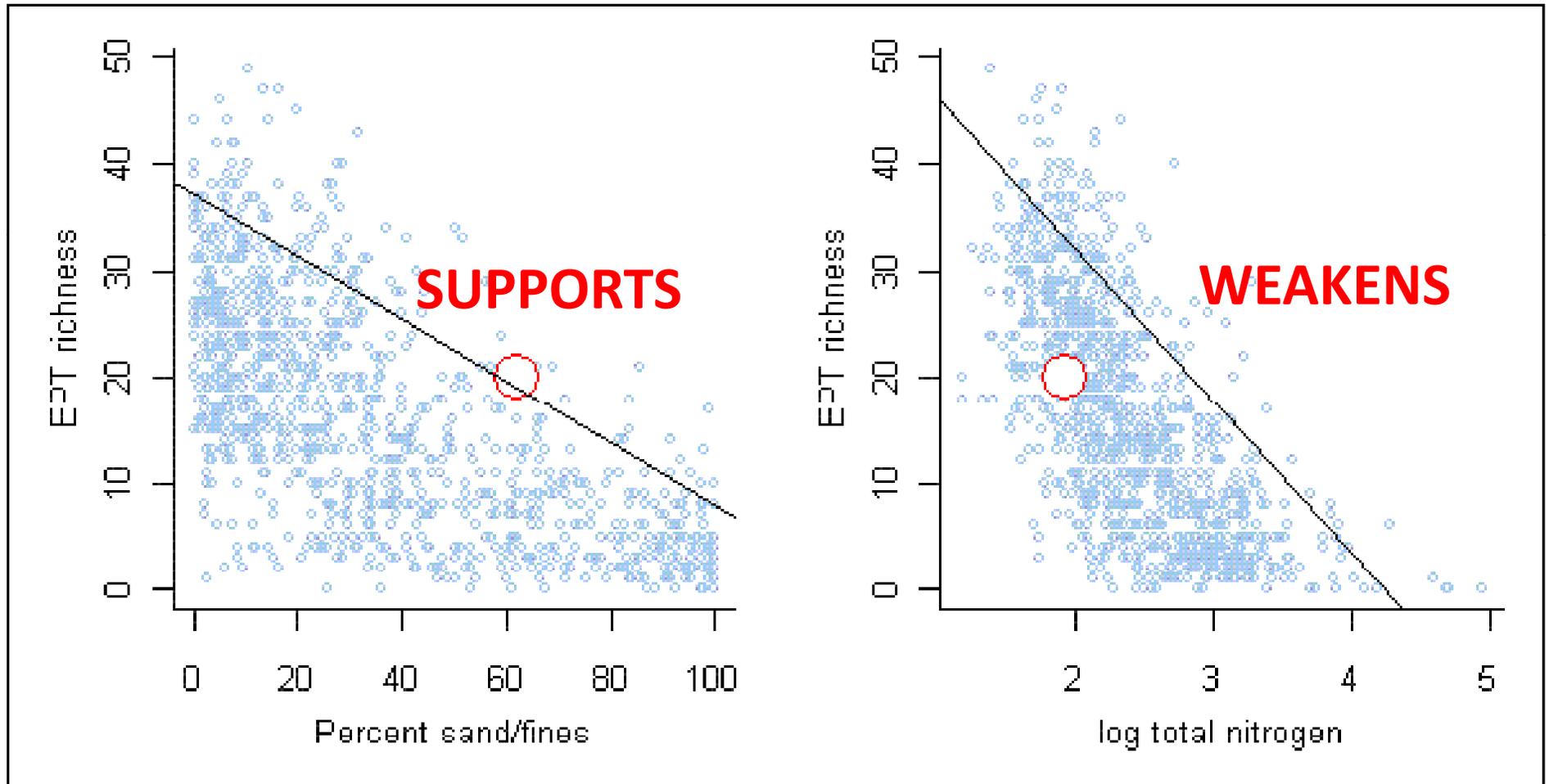
- Want to examine relationships between different variables
  - Linear & more complex
  - Expected & unexpected
- 1<sup>st</sup> step is visualizing data
  - Scatterplots & boxplots
  - No statistics, hypotheses, or p-values

# Quantile regression

---

- Models relationship between specified quantile of response variable and explanatory variable (stressor)
  - 50<sup>th</sup> quantile gives median line, under which 50% of observed responses occur
  - 90<sup>th</sup> 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

Microsoft Excel - SSD\_Generator V11

File Edit View Insert Format Tools Data Window Help Adobe PDF Type a question for help

A1

Introduction How to use the SSD Generator Calculations behind an SSD

Step 1) Select Data Step 2) Calculate Step 3) Fit Distribution [Fit to Screen](#)

### Introduction: What is the 'Species Sensitivity Distribution Generator'?

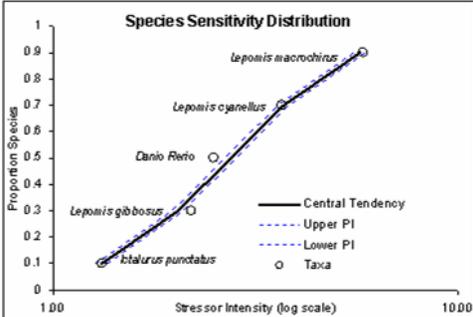
The Species Sensitivity Distribution Generator is a tool to create custom Species Sensitivity Distributions (SSDs). It fits a commonly applied distribution, the log-probit (i.e., linearized log-normal) to data for concentrations at which different species exhibit a standard response to a stressor. Using this tool, you may place stressor concentration data from your site in context of concentration-response data you selected for inclusion in an SSD. You may use your own data, data from the ECOTOX database or data from the CADDIS Metals SSD Library.

#### What are species sensitivity distributions?

Species sensitivity distributions (SSDs, example at right) model the variation in the sensitivity of different species to a stressor. SSDs assist in the interpretation of site data for stressor identification and risk assessment by relating them to the proportion of species expected to be affected at prescribed concentrations. SSDs are usually created using data from laboratory toxicity tests.

#### Creation of an SSD involves three basic steps:

- 1) **Select data** for the exposure intensities at which different species exhibit a standard response to the stressor.
- 2) **Calculate** proportions by first ranking these data from lowest to highest, then converting ranks to proportions:  $\text{Proportion} = (\text{Rank} - 0.5) / \text{Number of Species}$ .
- 3) **Fit a statistical or empirical distribution** to Proportion (y axis) vs Stressor Intensity (x axis)



Taxa	Stressor Intensity (log scale)	Proportion Species
Labeo punctatus	1.00	0.10
Lepomis gibbosus	~1.5	0.30
Danio Retio	~2.5	0.50
Lepomis cyanellus	~4.0	0.70
Lepomis macrochirus	10.00	0.90

#### Key Concept

**Standard Response:** An equivalent response such as the concentration at which half of the individuals exposed die (LC50) or the concentration at which reproduction is reduced by 20% (EC20).

#### Citations

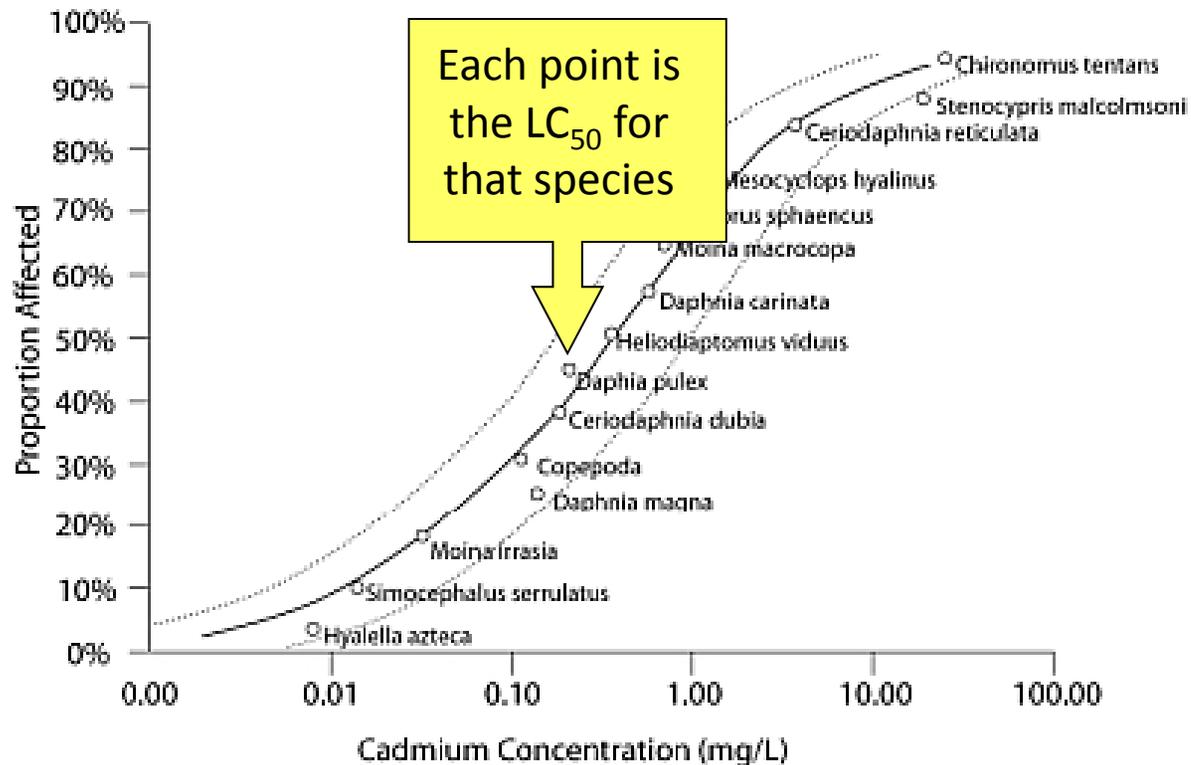
Neter, J., W. Wasserman and M.H. Kutner. 1990. Applied Linear Statistical Models, 3<sup>rd</sup> ed. Irwin, Boston, MA. 1184 pp.

Posthuma, L., G.W. Suter II, and TP Traas. 2002. Species Sensitivity Distributions in Ecotoxicology. Lewis Publishers, Boca Raton, FL. 587 pp.

U.S. Environmental Protection Agency (EPA). 2005. Methods/Indicators for determining when metals are the cause of biological impairments of rivers and streams: species sensitivity distributions and chronic average response relationships from laboratory data. Cincinnati, Ohio, U.S. EPA Office of Research and Development, National Center for Environmental Assessment.

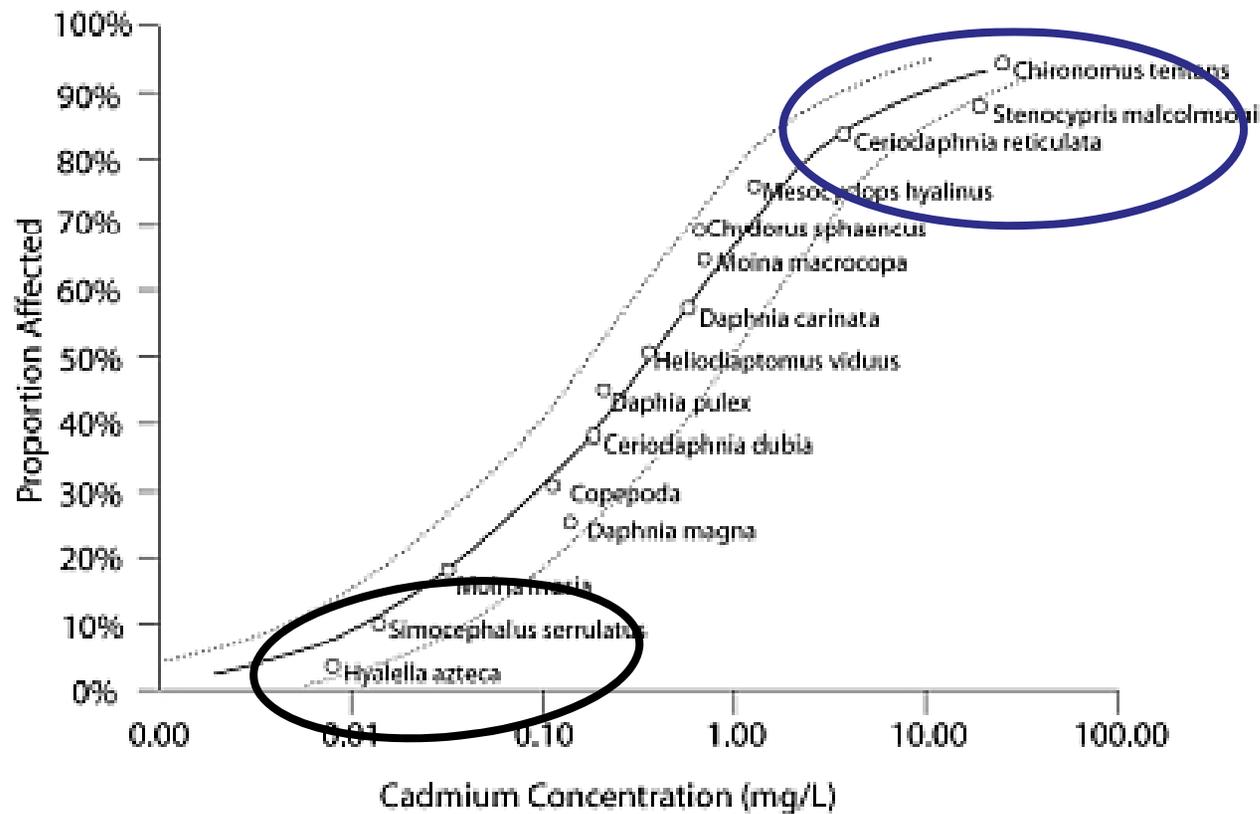
What is the SSD Generator | How to Use the SSD Generator | How are SSDs generated | Step 1) Select Data | Step 2) Calculate

# What is a species sensitivity distribution

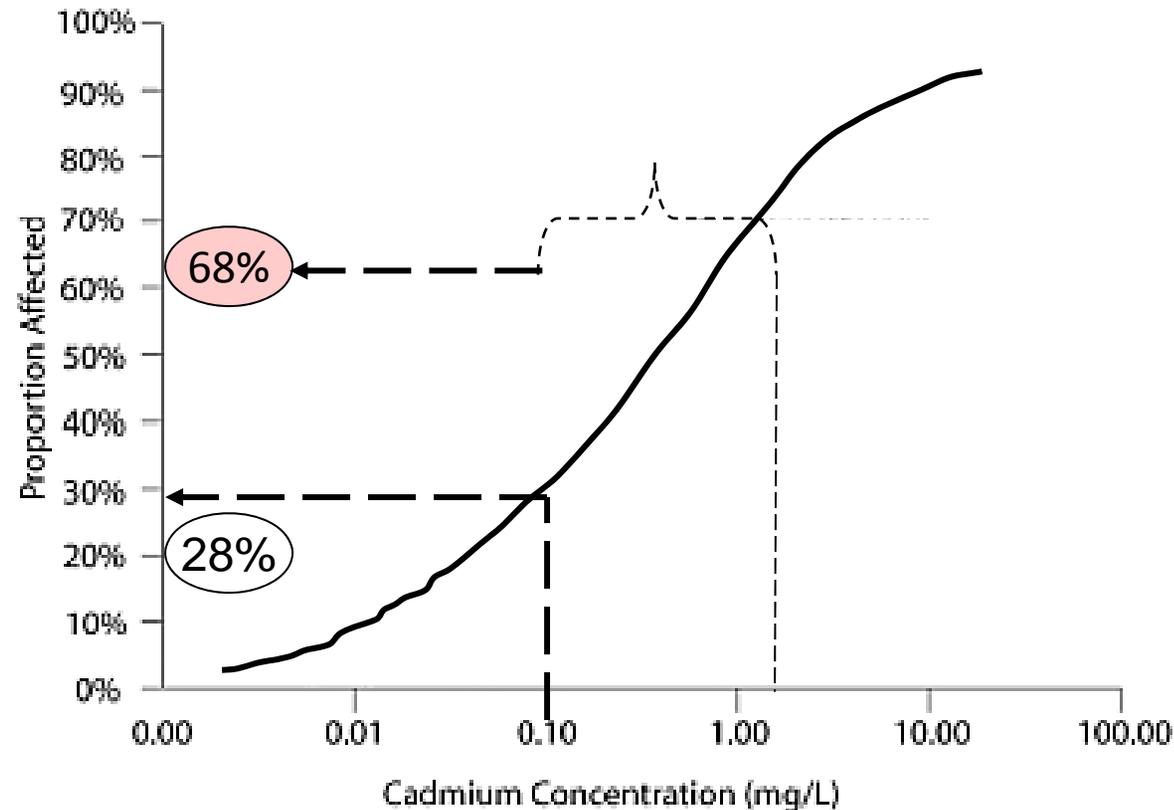


- An exposure-response model for a single stressor
- Basis for NAWQC
- Represents relative sensitivities to the stressor of interest

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