National Biological Assessment and Criteria Workshop

Advancing State and Tribal Programs

Coeur d'Alene, Idaho
31 March – 4 April, 2003

SI 201

Appropriate Use of Statistics

Presented by

Glenn Suter II, USEPA,
Office of Research & Development

What is the Problem?

- Correlation does not prove causation
  - Except in experiments
- Practitioners often attempt to test causal hypotheses with observational data
  - t tests, ANOVA, etc. are experimental stats.
- Practitioners often mistake statistical significance for biological significance
  - Which can produce misleading results
- Practitioners often believe that regression, ordination, etc. demonstrate causation

Lessons

- Statistics cannot derive causation from observational data
  - The correlation of *C* and *E* is not the probability that *C* causes *E*
  - You can not test the hypothesis that *C* causes *E*
- Statistics can help to identify and quantify
  - Associations
  - Magnitudes of differences
  - Gradients and other patterns

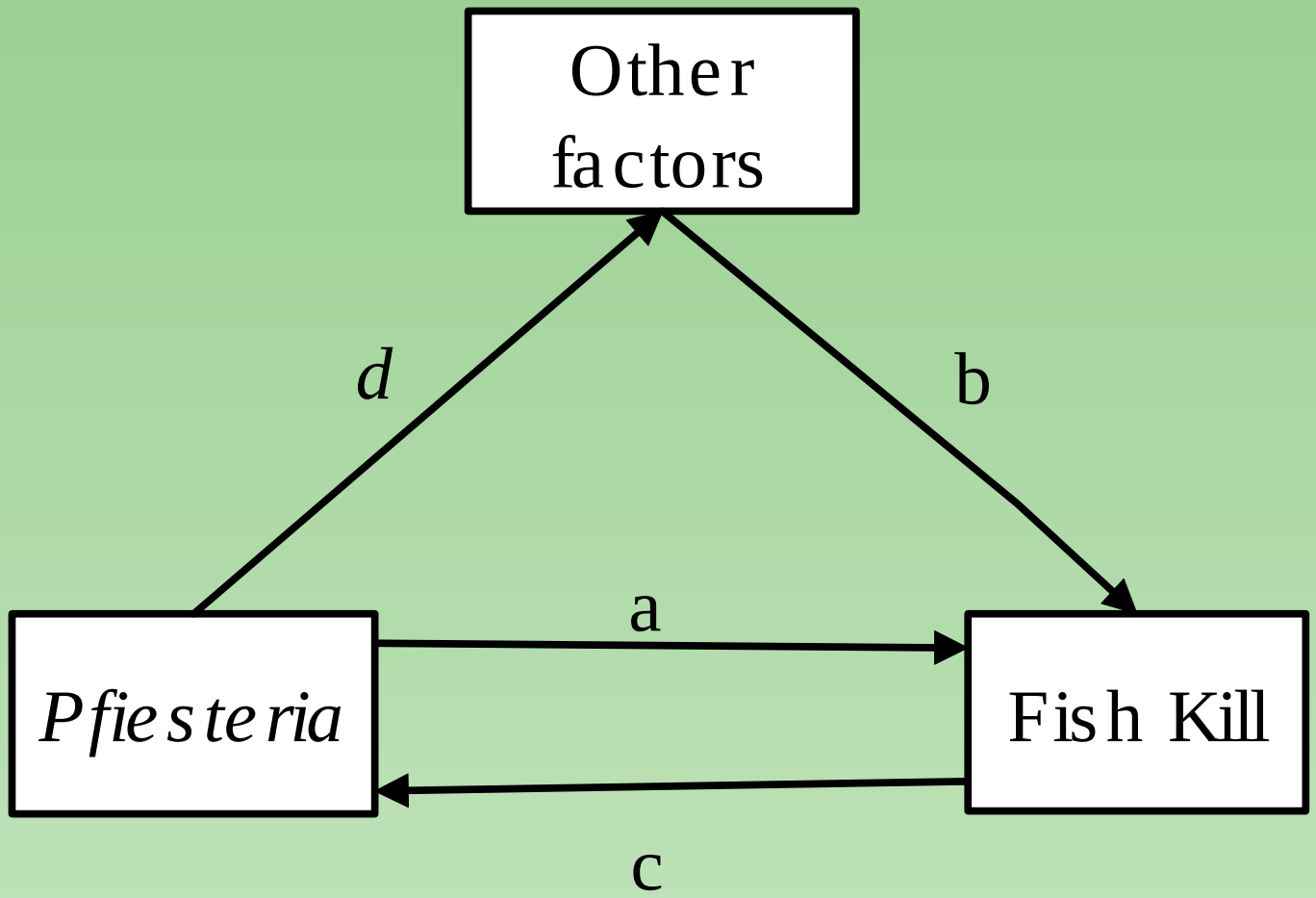
Uses for Statistics in SI

- Summarize data using descriptive statistics
  - Mean, range, variance, recurrence rates, etc.
- Explore data sets
  - Multivariate correlations
- Quantification of associations
  - Correlation, regression, etc.
- Statistical modeling
  - Exposure-response relationships
- Comparison of models
  - Goodness of fit, maximum likelihood

Pitfalls for Statistics

- Few samples at few times and locations
- No replication
- No randomization
- No control of treatments
- Many confounding variables
  - Flow augmentation
  - Stream gradient
  - Julian day (seasonality)
  - Habitat structural attributes

Pfiesteria Conceptual Model



Conclusions

- Statistics can not determine the cause
  - but it can help a lot
- Define needed inference first
  - then statistics
- Use statistics like a drunk uses a lamp post
  - more for support than enlightenment