Causal Analysis and Probability Surveys: Examples

John F. Paul ¹, Henry A. Walker ², Michael E. McDonald ¹
National Health and Environmental Effects Research Laboratory
U.S. Environmental Protection Agency

¹ Research Triangle Park, North Carolina
² Narragansett, Rhode Island

11 April 2007
Acknowledgements

Steve Hedtke
Susan Cormier
Sue Norton
Glenn Suter
EMAP-Surface Waters
EMAP-Coastal
Presentation Outline

- What is causal analysis?
- Hill’s elements of causality
- Stressor Identification Guidance
- Examples
- Take home message
What is causal analysis?

Identify what is causing an effect.

Difficult, if not impossible, to prove cause-effect for natural resources.

What most people consider “proof” is really convincing evidence (for them).

Easier to disprove cause-effect
Sir Bradford Hill's Elements of Causality


1. Strength of Association
2. Consistency
3. Specificity of Association
4. Temporality
5. Biological Gradient
6. Plausibility
7. Coherence
8. Experiment
9. Analogy
Stressor Identification Guidance Document

- Identify alternative candidate causes
- Logically eliminate what you can
- Diagnose when you can
- Use strength of evidence when you can
- Do not claim proof of causation
- Identify the most likely cause
- Use a consistent process
- Document the evidence and inferences
Criteria to Judge Evidence

- Time order SDI
- Strength SDI
- Consistency SDI – replication
- Coherence SDI – agrees with fact & theory
- Specificity SDI – occurrence of one with other

** Beware of publication bias **
Aquatic Examples

EMAP data
Eutrophication in Northeast Lakes

EMAP sampling of lakes in northeast in 1991-94

Response variable – chlorophyll a

Define eutrophy as chla > 30 ug/L

Threshold parameter – phosphorus

333 lakes sampled

Data available at www.epa.gov/emap
$r^2 = 0.52$
Reverse CDF and CCDFs for Total P

Total P (µg/L)

Proportion > x

Eutrophic Lakes

All lakes

Non-eutrophic lakes

All lakes

Eutrophic Lakes

Non-eutrophic lakes
Conditional Probability for Exceeding Total Phosphorus Level

Nutrient Ecoregion VIII -
Level III Ecoregions 58 & 82
Mid-Atlantic EMAP Streams

Status of biological resources – EMAP indicators

Sampling with EMAP probability design

380 stream segments sampled in 1993-96

1st to 3rd order wadable streams

Data available at www.epa.gov/emap
Scatter plots (EPT taxa richness vs. stressors) and correlation coefficients

$r = -0.341959039872149$

$r = -0.304275836453176$

$r = -0.318362941528538$

$r = -0.535333858269259$

$r = -0.330471827455699$

$r = -0.322870692745639$

$r = -0.350966838017365$

$r = 0.518934335569322$

$r = -0.312632572254351$

How about log x-scale plots?
Distribution of Turbidity Values Per Stream Mile

1993-96 data

Cumulative Distribution Functions (CDFs)

Conditional Cumulative Distribution Functions (CCDFs)

Proportion of Stream Miles <= x

log_{10} ( turbidity (NTUs) )

Data categories:
- All
- Reference
- Degraded Benthos
- Good Benthos
Conditional Probability Analysis for Total Nitrogen

Degraded Benthic Condition for EPT Taxa < 9

\[ \log_{10} (\text{Total Nitrogen (ug/L)}) \]

- Probability of Degraded Benthic Condition
  - 69%
  - 54%
- 933 ug/L
Conditional Probability Analysis for Total Phosphorus
Degraded Benthic Condition for EPT Taxa < 9

- 82% at 91.2 μg/L
- 54%
Fraction of Stream Miles with DO Levels $\leq x_c$

- dashed line: degraded benthic conditions
- solid line: all
- dotted line: good benthic conditions

$x_c$, Bottom Dissolved Oxygen (mg/L)
Estuarine example
All sites
Narragansett Bay
Probability of Benthic Impairment for not exceeding $x_c$.

- $x_c = 0.875$
- $x_c = 0.44$
- $x_c = 0.28$

DO = 2.3 mg/L

DO = 4.75 mg/L
But ..........

What value do probability surveys add?

Instead of results being applicable only to sites of observations, now they can be extrapolated to entire statistical population
Things to Remember

- Use a consistent process
- Identify possible candidate causes
- Document, document, document
- Stressor Identification Guidance
- WWW. EPA.GOV/CADDIS
Take Home Message

Probability surveys
  • Unbiased estimates
  • Candidate criteria thresholds
    • Paul & McDonald, 2005, JAWRA
    • SABS WQC Document EPA-822-R-06-001
  • Quantitative probabilistic ERA
    • Poster
    • Paul & Munns in review.
  • Diagnostics

Cost-effectiveness of probability survey as backbone of monitoring program