

Distinguishing Causal from Noncausal Associations: Issues Intrinsic to Behavioral and Other Complex Interventions

William R. Shadish

University of California, Merced



UC Merced Construction
November 12th 04

Photo By Hans Marsen

Overview of This Talk

- In general, I'm going to discuss the logical, design, and statistical tools that we use to make causal inferences, and point to problems in each of those areas that cause us to make inferential errors
- Most of the issues are not specific to complex or behavioral interventions, though I will point out some particular problems in that area.

Three traditional tools to help judge treatment effectiveness

- Logic
- Design
- Statistics

Logic

- Causal inference is always, first and foremost, a matter of logic.
 - Mill's Canons (temporal precedence, covariation, alternatives)
 - Counterfactual Models (e.g., Rubin's model)
- This means causal inference is possible with nonrandomized designs
 - Smoking and cancer in humans
 - Dietary treatment of PKU (short time series)
 - Campbell's (1975) Degrees of Freedom and the Case Study
 - But the conditions facilitating causal inference in case studies are rare

Example of Aids to Logic: Campbell's Validity Types

- Statistical Conclusion Validity (did the treatment covary with the outcome?)
- Internal Validity (did the treatment affect the outcome?)
- Construct Validity (what labels or constructs best represent what we did?)
- External Validity (to what does the effect generalize?)

Threats to Validity

- Threats to Internal Validity
 - Ambiguous Causal Precedence
 - History
 - Maturation
 - Attrition
 - Selection
 - Regression
 - Testing
 - Instrumentation
- The idea is to rule each of these out as alternative explanations for the observed effect
- Similar lists of threats to the other validities
- Similar lists in other literatures, such as Sackett (1979) for case control studies

Logical Problems

- Many scientists are surprisingly unaware of basic causal logic
- Our ability to enumerate specific contextual threats to validity is limited.
 - Need research programs to identify threats
 - Need ways to quantify qualitative threats to validity
- Scientists are just people, and people are not good at making causal inferences based on logic alone.
 - E.g., Confirmation bias.

Design

- Randomized experiments (lots of examples of their feasibility with behavioral interventions, though some special problems do arise)
- Strong quasi-experiments
 - Interrupted time series (on groups or individuals)
 - regression discontinuity (cutoff assignment)
- Weaker quasi-experiments
 - Nonequivalent comparison group designs
- Observational Studies
- Meta-Analysis
 - Of randomized experiments
 - Of mixed methods work

Design Problems

- Specious arguments against randomization
- Occasional premature randomized experiments
- Misplaced faith in one large randomized trial over many small randomized trials. Both are important, but:
 - Many small trials allows exploration of generalization
- Under-use of stronger quasi-experiments
- Failure to take full advantage of meta-analysis to explore generalization of effects
 - Response surface modeling using multimethod databases
 - Routine exploration of generalization over moderators
- Misplaced faith in single meta-analyses
 - They are only as good as the data going into them
 - E.g., most negative effects for Vitamin E occurred in studies using both Vit E and beta-carotene

Statistical Tools

- In Randomized Experiments
 - Attrition Analyses
 - Incomplete treatment implementation
 - Analysis of nested data
- In Quasi-Experiments
 - Selection Bias Modeling
 - Propensity Score and Hidden Bias Analysis
- In Meta-Analysis
 - Response Surface Modeling

Statistical Problems

- Lack of awareness of recent developments for both randomized and nonrandomized experiments.
 - E.g., continued failure to take nesting into account, leading to Type I errors
- Misplaced preference for single point answers rather than bracketing
- Lack of clear methods or software to do some of the analyses (e.g., response surface modeling in meta-analysis)

My Favorite Top 3 Problems

1. Misplaced faith in “singletons”
 - Single point estimates
 - Single studies
 - Single meta-analyses
2. Misanalyzed data (especially nesting and attrition)
3. Failure to take generalization as seriously as causation in design and analysis

But we should not overstate the problem

- We do want to minimize preventable problems, but:
- Experiments are about DISCOVERY of effects of treatments. We should not be upset that the effects are not always what we anticipated.
- We rarely know when the “last word” is in (even for HRT or beta-carotene).