Causal Inference

- A **causal inference** is a statement about why something happens.
  - A causal inference therefore states the existence of a relationship between at least two variables.
- The **dependent variable** measures that variation which we would like to explain (find a cause for).
  - Also called ***Y***, or the "outcome" or "response" or variable.
- The **independent variable** measures that variation which we think explains variation in the dependent variable.
  - Also called ***X***, or the "treatment" or "study" variable.

What is Causation?

- What does it mean to say “X causes Y” and how are we able to know this?
- This is more complicated than it seems and there are many philosophies of causation.
- We’ll use the “counterfactual” framework.
  - AKA: “potential outcomes” or “Neyman-Rubin” framework.
  - Dominant framework in the social sciences today.

The causal effect of a treatment is the difference between what happens to a unit after that treatment and what would have happened had the unit not been treated.

The Consistency Assumption

- AKA the “SUTVA”: The Stable Unit Treatment Value Assumption
- "the [potential outcome] observation on one unit should be unaffected by the particular assignment of treatments to the other units" (Cox 1958)
- $Y_i(x) = Y_i$ if $X_i = x$
- Very important/tricky in social research (hint: strategic interactions, time, etc.)

The Fundamental Problem of Causal Inference

For any unit, we only ever observe one potential outcome.

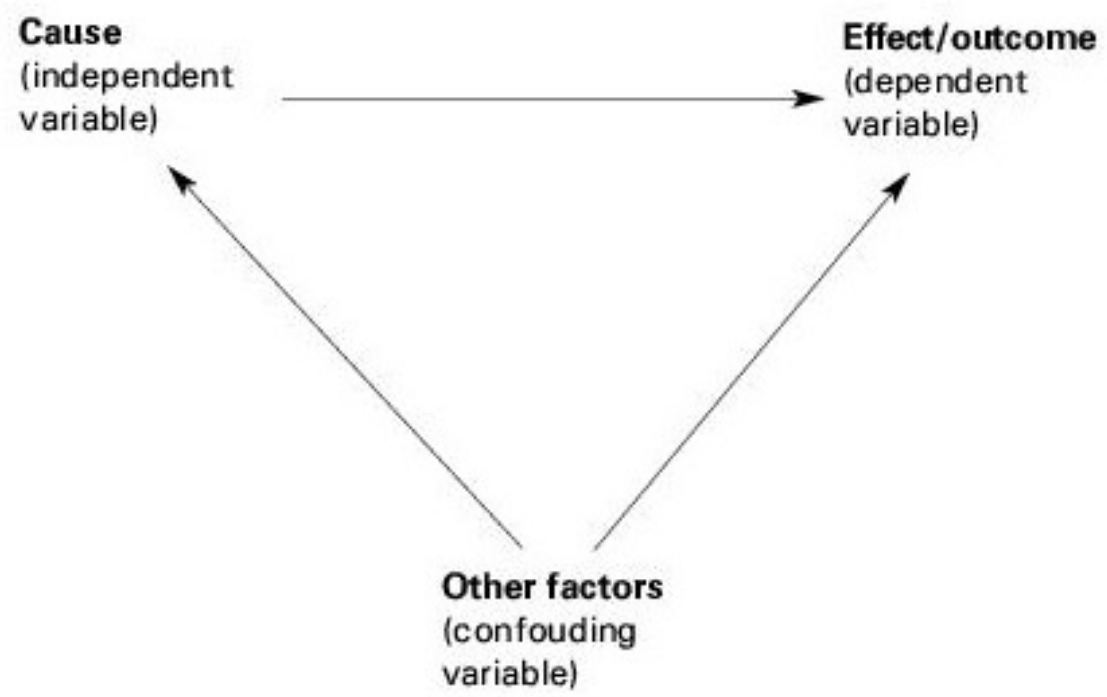
- In other words, to directly calculate a causal effect would require us to rewind the world and re-run it with a different value on the independent variable.
- In other words, causality *cannot* be directly and certainly observed.

The Experiment as an Imperfect Solution to the FPCI

- Suppose some units $i = 1, ..., N$
- A dependent variable Y_i
- An independent variable X_i
- The value of ***Y*** given some treatment $Y_i(x = 1)$
- The value of ***Y*** given no treatment is $Y_i(x = 0)$
- A basic formal statement of the inferred causal effect is

$$\frac{1}{N}\Sigma_{i=1}^{N} Y_i(x = 1) - Y_i(x = 0)$$

Identifying causal effects in observational research is very hard.



http://www.slideshare.net/tosifahmad395/research-methodology3

Designing observational research is about collecting and analysing information in a way that mimics experiments.

2. If doing case studies, we select cases strategically to maximize causal leverage.

E.g., two countries that are as similar as possible but different on the independent variable.

3. If quantitative data is available, we can use statistical models to mathematically isolate correlation between an independent and dependent variable.

E.g., regression analysis.

Evaluating a Causal Inference

1. Is there a credible causal mechanism that connects X to Y?
2. Can we rule out the possibility that Y' could cause X?
3. Is there covariation between X and Y?
4. Have we controlled for all **confounding variables** Z that might make the association between X and Y **spurious?**[5]

Kellstedt and Whitten

Roadmap for Evaluating Your Causal Inference

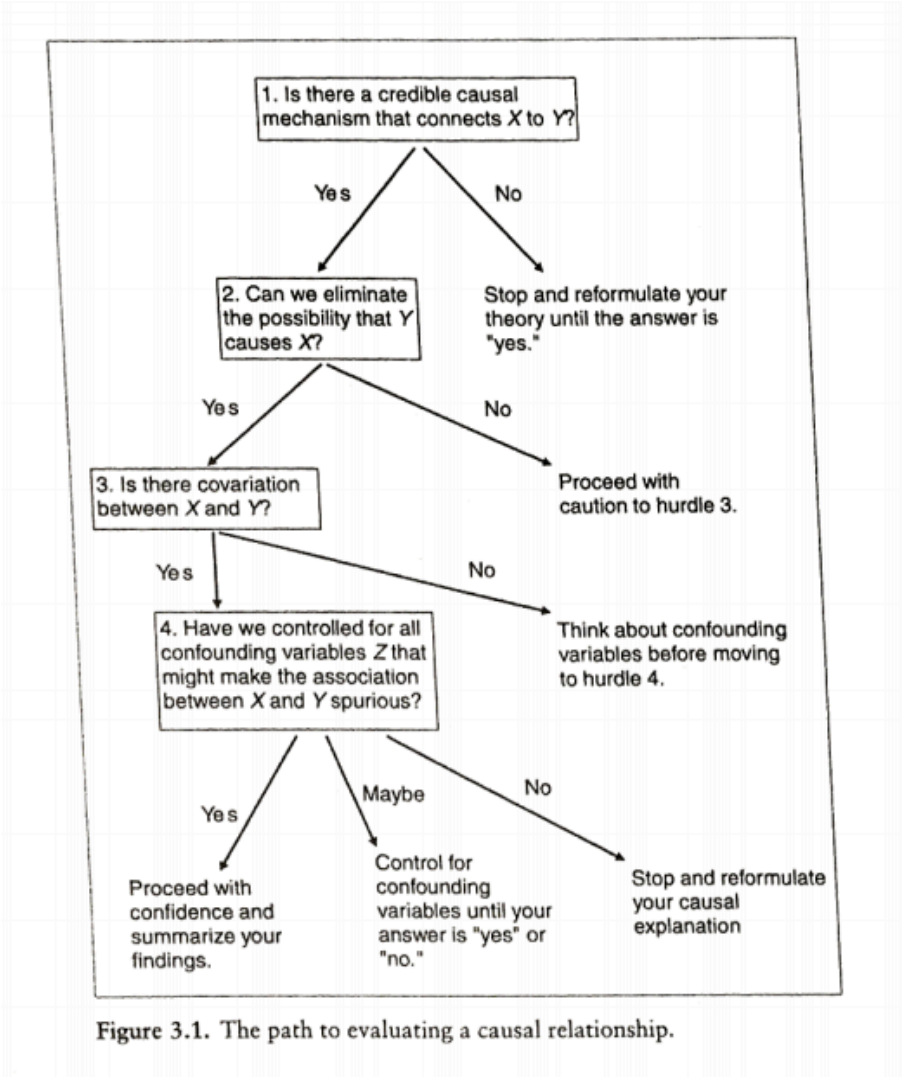


Figure 3.1. The path to evaluating a causal relationship.

Kellstedt and Whitten