**PPOL502-01, Spring 2016**
**Course Notes #19: Causality; Internal & External Validity; Basic Research Designs**

---

## I. STARTING POINTS

- Social sciences research attempts to understand and explain human behavior; we study social phenomena:

  -- Formulate and test suppositions about relationships among individual, organizational, societal, and political processes

  -- We can always further think of these as relationships between a *dependent variable* (the effects or result to be explained) and *independent variables* (the factors that influence the dependent variable or cause it to change somehow)

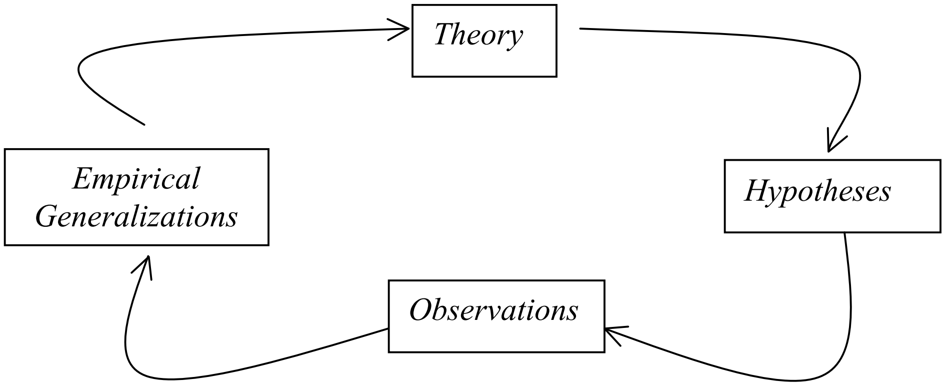
- Core elements of the scientific method

  *Theory:* explanations of the relationships between phenomena

  *Hypotheses:* statements about relationships between dependent and independent variables. More specifically tied to a particular research question or situation than a theory would be

  *Observations / Data:* information we collect or use to test the hypotheses. We collect these for a particular unit of analysis (e.g., a person, organization, period of time, country, etc.)

  *Empirical generalizations:* assess the theory and hypotheses that guided the analysis; perhaps develop other generalizations based on the patterns observed.



## II. CAUSALITY

- In public policy and evaluation settings, often a research goal is to identify the ***causal*** relationships among phenomena we study – the *what if* question.

  -- There are many different views of exactly what constitutes a causal condition, which is at its root a philosophical concept (e.g., Bradley & Schaefer reading,).

  -- In some views (i.e., essentialists), *A* must be both a necessary and sufficient condition in order for *B* to occur.

  -- More realistically in the social sciences, we talk about causality in terms of sufficient but not necessary (i.e., *A* alone can lead to *B,* but other factors alone may also cause B); or necessary but not sufficient (i.e., multiple factors are required, often in a specific order in time).

  Let's define these terms further:

  *Necessary condition*: a condition that must be present for an event to occur. Although the presence of a necessary condition does not guarantee occurrence of the event, the absence of the condition guarantees nonoccurrence.

  *Sufficient condition*: a condition that guarantees that the event will occur whenever it is present. However, the event may still occur in its absence.

  -- Scholars and commentators in the policy world use these terms, and you should have a sense of what they mean.

-- For our purposes in this course and throughout your time at GPPI, the three factors necessary for causal inference identified by Paul Lazarsfeld probably provide the most useful guide:

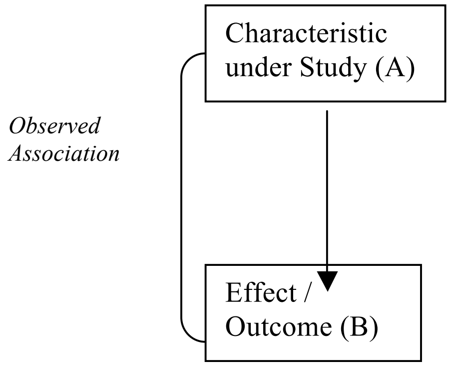
(1) The cause (*A)* must precede the effect *(B)*.

(2) The cause and effect must be related (i.e., correlated).

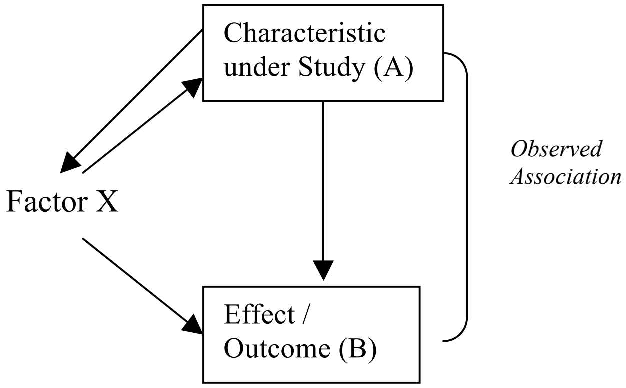
(3) Other explanations of the cause-effect relationship must be eliminated (i.e., rule out spurious or confounding factors)

-- A picture might help:

**A causes B:**



**A appears to cause B, but the association is spurious:**



- A key concept for thinking about whether a causal relationship has been established is the *counterfactual*. The counterfactual state is <u>what would have happened in the absence of the characteristic or program</u>. The counterfactual outcome is what would have resulted as a consequence of this path (whatever the outcome of interest is).

  -- This is a very important concept that you should always have in mind whenever a researcher (or other person) claims to have discovered a causal relationship.

  -- The counterfactual is a *concept* that can never actually happen (e.g., the very same persons cannot simultaneously live two different lives)

  -- The ideal counterfactual would be the outcome for a group ***exactly*** like the treatment group in ***all*** ways ***except for*** the intervention being studied.

  -- A research goal is often to "construct a counterfactual" that is as close as possible to the "true" counterfactual. We often refer to this constructed counterfactual as the "comparison group."

  -- The best way to approximate this ideal counterfactual is through randomly assigning subjects to receive or not receive the treatment. While the individuals randomly assigned are not identical, if the groups are large enough, the groups are similar *on average*

Example: Does enrollment of pregnant women in pre-natal classes cause healthier babies to be born (i.e., babies who are not low birthweight)? (Note: a baby is considered low birthweight if it weighs less than 2,500 grams, or 5 pounds 8 ounces)

Suppose I studied all pregnant women in DC during the last year, and found the following:

**Table 1: Participation in prenatal courses and LBW**

"Healthy" babies"

| | **Mother Participated in Pre-natal course** | | |
|---|---|---|---|
| **Low Birthweight (LBW) baby** | No | Yes | *Number of observations* |
| No | 100 | 500 | 600 |
| Yes | 300 | 0 | 300 |
| *Number of observations* | 400 | 500 | 900 |

a. What is the dependent variable?

What is the independent variable?

b. Is participation in prenatal courses a necessary condition for "healthy" babies?

c. Is participation in prenatal courses a sufficient condition for "healthy" babies?

d. Did the independent variable precede the dependent variable?

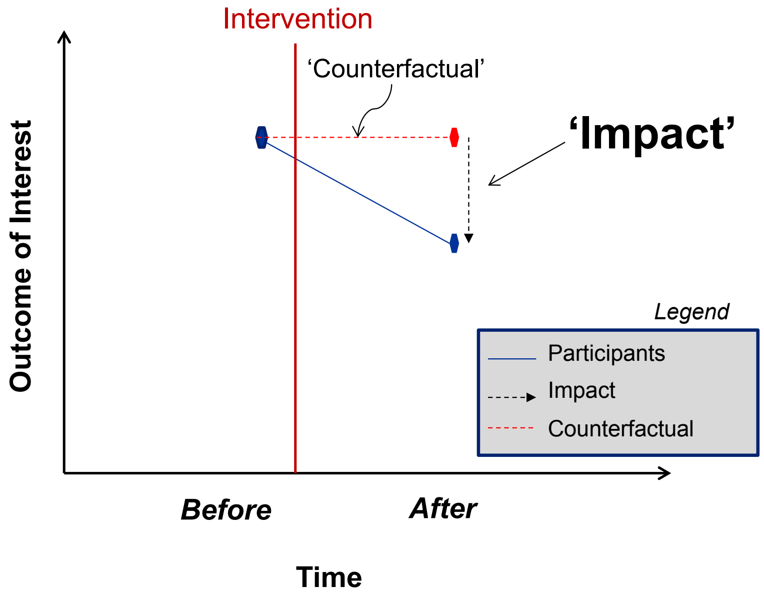
e. Are the independent and dependent variables associated?

f. Are spurious or confounding factors present or likely?

g. How would you describe the counterfactual condition?

h. Do you think mothers' participation in a prenatal program causes healthy babies (defined by birthweight) to be born?

i. Based on these data, what would be your advice to a policymaker about whether to support a program that provided prenatal care to teenage mothers?



## III. VALIDITY

- The notion of causality is directly tied to the idea of *internal validity*. In fact, there are four general kinds of validity, defined and explicated by Cook & Campbell (1979) (most recently by Shadish, Cook, and Campbell 2002): statistical conclusion; internal; construct; and external.

- We have been discussing *statistical conclusion validity* throughout the course (and through PPOL 508 and 512).

- Later in this course, we'll talk about *construct validity*.

- Now, we'll briefly define *internal* and *external validity*. Basically:

  -- A study has *internal validity* if it is able to distinguish a causal effect between an independent variable and a dependent variable. "The validity of inferences about whether observed covariation between A (the presumed treatment) and B (the presumed outcome) reflects a causal relationship from A to B as those variables were manipulated or measured." (SCC p. 38)

  -- A study has *external validity* if its results can be generalized to other people, places, and times. "The validity of inferences about whether the cause-effect relationship holds over variation in persons, settings, treatment variables, and measurement variables." (SCC p. 38)

  -- Specific factors that may lead you to question a study's internal or external validity are referred to as "threats to validity"

-- "Lists of validity threats are heuristic aids; they are not etched in stone, and they are not universally relevant across all research areas in the social sciences (Shadish, Cook, & Campbell, p. 40).

-- "A validity typology can greatly aid…design, but it does not substitute for critical analysis of the particular case or for logic" (Mark 1986, cited in SCC, p. 39).

-- "The threats to validity are heuristic devices that are intended to raise consciousness about priorities and tradeoffs, not to be a source of skepticism or despair" (SCC, p. 97)

-- When you analyze a study (including your own), you can use the validity typology to help you understand the strengths and limitations of the study. SCC suggest that you ask yourself these questions:

  --- How does a particular threat apply in the study? (i.e., it is not sufficient to say that "history" might be a threat to internal validity. Instead, you need to think through exactly what "history" means in the context of the study you're examining.)

  --- Is the threat plausible (not just possible) in the study?

  --- If the threat is plausible, would its effects results in a bias in the same direction as the effect found? Could it partly explain the findings?

## A. Threats to Internal Validity

"Internal validity confounds are forces that could have occurred in the absence of the treatment and could have caused some or all of the outcomes observed" (SCC p. 95).

1. *Ambiguous temporal precedence* (first condition of Lazarsfeld)

2. *Selection*: on the part of (a) participants; (b) program administrators. E.g., Were the mothers who participated in the program likely to be different from those who didn't participate for any reason?

3. *History*: events occurring concurrently with the treatment. Did anything else other than the prenatal program happen in these mothers' lives during their pregnancy that could have affected the birth-weight of their babies?

4. *Maturation:* Naturally-occurring changes over time. Did the mothers learn about healthy eating, or mature/grow in other ways?

5. *Regression to the mean*. When units are selected based on extreme scores, they'll tend to "settle back down" toward the mean on subsequent measures (regardless of treatment)

6. *Attrition: (aka "mortality").* Loss of participants (a type of selection). Did any of the moms drop out of the study sample? The program?

7. *Testing:* learning through testing

8. *Instrumentation*: the measurement itself may change over time (think about reliability) and this may be confounded with treatment effects.

### B. External Validity (Generalizability)

-- remember you can think about whether these results generalize to other people, places, and times. What factors in this study might affect our ability to generalize our results (assuming that we actually found a causal effect)?

- We can think about generalizing *to* specific populations of interest; and generalizing *across* populations of interest. Here's the basic idea:

  -- Generalizing *to* a specific population of interest: The idea here is whether the study subjects and conditions are indicative of those of the population the sample is drawn from (suppose pregnant women in NW DC last year were sampled).

  -- Generalizing *across* specific populations of interest: "which different populations (or subpopulations) have been affected by a treatment" (C&C p. 71).

## IV. RESEARCH DESIGNS

- We've focused on internal and external validity so far; we'll talk about construct validity later; and we'll be thinking about statistical conclusion validity throughout this course and others.
- Underlying everything we've talked about so far is the idea that we're never able to observe all elements of a population of interest. We'll talk more about sampling, and random sampling especially, in another class.

- We have a sense of what to watch out for – now, what kinds of research designs can we implement to pursue a research question? And, what are the basic advantages and disadvantages of these designs?

- 3 basic types of design: *descriptive, experimental (may include "natural experiments")*, and *nonexperimental.*

- In this class, we can't go into a full explanation of these types of designs, and their advantages and disadvantages. You'll explore them further in PPOL 512. However, it's good to know that different approaches exist to uncover the cause-effect relationships (or to simply describe, predict, etc. not even trying to assert causal relationships)

Experimental designs.

-- Also called randomized experiments, random assignment designs.

-- Participants (persons, programs, schools, etc. – whatever the unit of analysis is) are randomly assigned, through a lottery-like process, to either a *treatment* group or to a *control* group

-- The subjects receiving the intervention or program being studied are called the "treatment" or "program" group; the subjects who do not receive the intervention being studied are called the "control" group.

-- The difference in outcomes between treatment and control group members represents the *impact*, or *value-added* of the program or intervention.

-- This random assignment process assures that, with sufficient sample sizes, the participants in the two groups are not statistically different from each other; i.e., they are equivalent groups on average. The only differences between the two groups would be due to chance: Random assignment designs are considered the "gold standard" for showing cause-effect relationships. WHY?

-- Can be difficult to implement (for administrative or moral reasons)
Example: Studying the effects of different electoral systems.

-- May not be possible to generalize the results of an experiment carried out in a particular place in a particular time with particular people to other places, times, and people (i.e., external validity can be problematic)