Causality and Research Design

POLI 205 Doing Research in Politics

Causality and Research Design

POLI 205 Doing Research in Politics

Fall 2015
Causal Theories

- The goal of political science (and all science) is to create and then evaluate causal theories
- Cause (X) and effect (Y)
- Basis of causality
  - Time Ordering: The cause precedes the effect
  - Co-Variation: Changes in X are associated with changes in Y
  - Non-Spuriousness: There is not a variable Z that causes both X and Y
Determinism and Probabilities

- **Deterministic**: If $X$ then $Y$ with certainty
  \[
  Y_i = \alpha + \beta X_i
  \]

- **Probabilistic**: If $X$ then $Y$ with uncertainty
  \[
  Y_i = \alpha + \beta X_i + \epsilon_i
  \]
Multiple Causes

- Theories are *bivariate*, $X$ causes $Y$, but reality is *multivariate*
- $X$ causes $Y$ but some other variables, $Z$, might also (or instead) cause $Y$
- If we don’t control for $Z$, the other possible causes of $Y$, then our conclusions about whether $X$ causes $Y$ might very well be mistaken
- How do we control for $Z$?
  - Research design
Four Causal Hurdles

- Is there a credible *causal mechanism* that connects $X$ to $Y$?
- Can we rule out *reverse causation* the possibility that $Y$ could cause $X$?
- Is there *covariation* between $X$ and $Y$?
- Have we controlled for all *confounding variables* $Z$ that might make the association between $X$ and $Y$ spurious?
Causal Mechanism

- Answer the “how” and “why” questions
- What is the process or mechanism that, logically speaking, suggests that $X$ might be a cause of $Y$?
- What is it specifically about having more(less) $X$ that causes more(less) $Y$?
Reverse Causation

- Endogeneity
- A country’s level of economic development causes it to be more or less democratic
- Variations in consumer confidence cause a president’s approval rating to change
- Ethnic conflict causes civil wars.
Covariation

- Measured empirically
- *Correlation is not causation*, but it’s normally a key component of causation
Confounding Variables

- When one scholar is evaluating another’s work, perhaps the most frequent objection is that the researcher “failed to control for” some potentially important cause of the dependent variable.
- So long as a credible case can be made that some uncontrolled-for $Z$ might be related to both $X$ and $Y$, we cannot conclude with full confidence that $X$ indeed causes $Y$.
- The importance of *research design*
Causal Checklist

1. Is there a credible causal mechanism that connects $X$ to $Y$?
   - Yes
   - No

2. Can we eliminate the possibility that $Y$ causes $X$?
   - Yes
   - No

3. Is there covariation between $X$ and $Y$?
   - Yes
   - No

4. Have we controlled for all confounding variables $Z$ that might make the association between $X$ and $Y$ spurious?
   - Yes
   - Maybe
   - No

   - Proceed with confidence and summarize your findings.
   - Control for confounding variables until your answer is "yes" or "no."
   - Stop and reformulate your causal explanation.
Use of the Scientific Method

- Scientific method
- Scientific research:
  - The goal is *inference*
  - Procedures are public
  - Conclusions are uncertain
  - *Observe* and *explain*
  - Can be replicated
Inference and Validity

- **Inference**: the process of using *what we know* to learn about *what we do not know*
  - What we do not know: is our theory correct?
  - What we do know: data

- **Types of inference**:
  - *Descriptive* inference: using observations (data) to learn about unobserved facts
  - *Causal* inference: using observations (data) to learn about causal effects
Inference and Validity

- **Validity**: Are we making *valid* inferences about the relationship between $X$ and $Y$?

- Types of validity:
  - External
  - Internal
Types of Validity

- **External validity**: Degree to which we can be confident that our results apply to *other contexts*
  - Are the results generalizable?
- **Internal validity**: Degree to which we can be confident that $X$ causes $Y$
- Threats to validity
  - What could reduce our confidence about our results?
  - Causal hurdles
Causality and Research Design

POLI 205
Doing Research in Politics

Causality
Research Design

Causal Hurdles

1. Is there a credible causal mechanism that connects X to Y?
   - Yes
   - No
     2. Can we eliminate the possibility that Y causes X?
       - Yes
       - No
         3. Is there covariation between X and Y?
           - Yes
           - No
             Proceed with caution to hurdle 3.
             Think about confounding variables before moving to hurdle 4.
           4. Have we controlled for all confounding variables Z that might make the association between X and Y spurious?
             - Yes
             - Maybe
             - No
               Control for confounding variables until your answer is "yes" or "no."
               Stop and reformulate your causal explanation
               Proceed with confidence and summarize your findings.
Other Threats to Validity

- Poor measures of variables
  - are the measures valid and reliable?
- Selection bias
  - is the sample representative?
- Setting
  - *Hawthorne* effect
Global warming isn't real because I was cold today! Also great news: World hunger is over because I just ate.
So much for “drug-addict” welfare recipients...

In Tennessee, a new law requiring welfare applicants to be drug tested has turned up only 1 positive test in over 800 applications.

That’s 0.12%.

Source: The Tennessean
Research Design

- **Research Design**: the plan to collect information to address your research question
  - Two types: *Experimental* and *Observational*
  - Four Components:
    - Research question
    - Theory
    - Data
    - Data analysis

- **Data**: pieces of information
  - Can be *qualitative* or *quantitative*
Causality

• Does X (treatment) cause Y (outcome)?
• Causal (treatment) effect: $\tau_X = Y_1 - Y_0$
• $\tau_X$ : Treatment (X)
• $Y_1$ : Group that received treatment
• $Y_0$ : Group that did not receive treatment
Experimental Design

- **Experiment**: is a research design in which the researcher both *controls* and *randomly assigns* values of the independent variable (treatment) to the subjects.
- Causal (treatment) effect: $\tau_X = Y_1 - Y_0$
Experimental Design

- **Control**: the values of the independent variable that the subjects receive are not determined either by the subjects themselves, or by nature
- **Random Assignment**: All participants are equally likely to be in the control group as the treatment group
  - *Treatment* group: Group that receives treatment
  - *Control* group: Group that did not receive treatment
- Why is random assignment important for experimental designs?
• **Observational**: is a research design in which the researcher does not have control over values of the independent variable, which occur naturally
  - Variation of $X$ and $Y$ important

• Causal (treatment) effect: $\tau_X = Y_1 - Y_0$
  - more (less) $X = $ more (less) $Y$
  - statistical controls for $Z$ variables

• *Large $n$* either Cross-sectional or time-series, and case study
Large $n$ Cross-Sectional

- A *cross-sectional observational study* examines variation across a cross-section of individual spatial units
  - Same variable(s) across units
- Example: the connection between the preferences of the voters from a district ($X$) and a representative’s voting behavior ($Y$)
Large $n$ Times-Series

- A *time-series observational study* examines variation within one spatial unit over time
  - Same variable(s) and same unit(s) over time
- Example: the connection between GDP growth and presidential approval from 1995-2005
Longitudinal

- A longitudinal observational study examines variation across spatial units over time
  - Panel data: Same variable(s) across the same units and over time
  - Times-series cross-sectional (TSCS): Same variable(s) across different units and over time

- Examples:
  - Panel: the connection between GDP growth of countries in the European Union and incumbent party support from 1995-2005
  - TSCS: the connection between ideology and concern about climate change from 1995-2005 using pooled surveys
• **Case Study**: Precise description of a single case ($n=1$ or maybe a few)
  - *Exploratory*: little is known about a phenomenon
  - *Descriptive*: discover or describe what happened in a single or select few situations
  - *Explanatory*: answer how or why questions
Case Study Approaches

- Controlled Comparison:
  - *Method of difference*:
    - One (or a few cases) $Y=1$
    - One (or a few cases) $Y=0$
    - $Z$ is the same across cases; what is different? ($X$)
  - *Method of agreement*
    - Two (or a few) cases $Y=1$
    - What is similar across cases? ($X$)
Case Study Approaches

- **Congruence Procedures**: congruence between values of $X$ and $Y$
  - **Type 1**: Compare to typical values
    - Are values of $X$ higher or lower than average matched by higher or lower values of $Y$?
    - Look for extreme values of $X$ and $Y$
  - **Type 2**: Multiple within-case comparisons
    - Does $X$ and $Y$ covary across a range of circumstances within the case?
    - Need multiple occurrences of $X$ and $Y$
    - Approaches a large-$n$ study as number of observations increase
Case Study Approaches

- Process Tracing: Explores the chain of events through which $X$ causes $Y$

$$X \rightarrow p \rightarrow q \rightarrow r \rightarrow Y$$