Experimental Identification of Causal Mechanisms

Kosuke Imai

Princeton University

Joint work with
Luke Keele, Dustin Tingley, and Teppei Yamamoto

May 21, 2010

Experiments, Statistics, and Causal Mechanisms

- Causal inference is a central goal of most scientific research
- Experiments as **gold standard** for estimating *causal effects*
- A major criticism of experimentation:

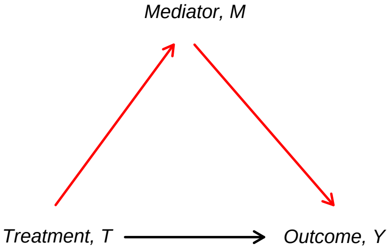
  it can only determine whether the treatment causes changes in the outcome, but not how and why

- Experiments merely provide a **black box** view of causality
- But, scientific theories are all about causal mechanisms
- Knowledge about causal mechanisms can also improve policies

- Key Challenge: How can we *design* and analyze experiments to identify causal mechanisms?

Causal Mechanisms as Indirect Effects

- What is a causal mechanism?
- Cochran (1957)'s example:
  soil fumigants increase farm crops by reducing eel-worms
- Political science example: incumbency advantage
- Causal mediation analysis



- Quantities of interest: Direct and indirect effects
- Fast growing methodological literature

Formal Statistical Framework of Causal Inference

- Binary treatment: $T_i \in \{0, 1\}$
- Mediator: $M_i \in \mathcal{M}$
- Outcome: $Y_i \in \mathcal{Y}$
- Observed covariates: $X_i \in \mathcal{X}$

- Potential mediators: $M_i(t)$ where $M_i = M_i(T_i)$
- Potential outcomes: $Y_i(t, m)$ where $Y_i = Y_i(T_i, M_i(T_i))$
- Total causal effect:

$$\tau_i \equiv Y_i(1, M_i(1)) - Y_i(0, M_i(0))$$

Defining and Interpreting Indirect and Direct Effects

- Robins and Greenland, Pearl, Petersen et al., and many others
- Indirect effects (a specific causal mechanism):

$$\delta_i(t) \equiv Y_i(t, M_i(1)) - Y_i(t, M_i(0))$$

- Effect of a change in M_i on Y_i that would be induced by treatment
- Direct effects (other causal mechanisms):

$$\zeta_i(t) \equiv Y_i(1, M_i(t)) - Y_i(0, M_i(t))$$

- Causal effect of T_i on Y_i, holding mediator constant at its potential value that would be realized when $T_i = t$
- Decomposition: Total effect $=$ indirect effect $+$ direct effect:

$$\tau_i = \frac{1}{2}\{\delta_i(0) + \delta_i(1) + \zeta_i(0) + \zeta_i(1)\}$$

Mechanisms and Manipulations

- **Mechanisms**: Direct and Indirect effects:
  - Counterfactuals about treatment-induced mediator values
- **Manipulations**
  - Controlled direct effects:

$$\xi_i(t, m, m') \equiv Y_i(t, m) - Y_i(t, m')$$

  - Causal effect of directly manipulating the mediator under $T_i = t$
- Fallacy of the "Causal Chain" approach:

Prop.	$M_i(1)$	$M_i(0)$	$Y_i(t,1)$	$Y_i(t,0)$	$\delta_i(t)$
0.3	1	0	0	1	-1
0.3	0	0	1	0	0
0.1	0	1	0	1	1
0.3	1	1	1	0	0

$\mathbb{E}(M_i(1) - M_i(0)) = \mathbb{E}(Y_i(t,1) - Y_i(t,0)) = 0.2$, but $\bar{\delta}(t) = -0.2$

Single Experiment Design

1) Randomize treatment

2) Measure mediator

3) Measure outcome

Assumption Satisfied

- Randomization of treatment

$$\{Y_i(t, m), M_i(t')\} \perp\!\!\!\perp T_i, \mid X_i = x$$

Key Identifying Assumption

- Sequential Ignorability:

$$Y_i(t', m) \perp\!\!\!\perp M_i \mid T_i = t, X_i = x$$

- Selection on observables
- Violated if there are unobservables that affect mediator and outcome

Identification under the Single Experiment Design

- Sequential ignorability yields nonparametric identification
- Many alternative assumptions exist

- Sequential ignorability is an untestable assumption
- Without it, the identification power is weak
- The sign is not identified in the binary case
- Back to an obervational study

- Sensitivity analysis: How large a departure from sequential ignorability must occur for the conclusions to no longer hold?
- Possible pre-treatment unobserved confounders

Alternative Experimental Designs

- Can we design experiments to better identify causal mechanisms?

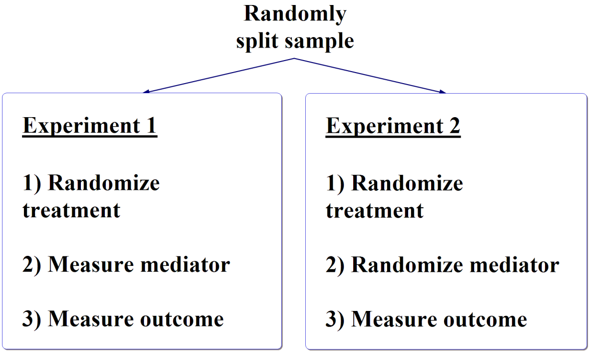
- Perfect manipulation of the mediator:
  1. Parallel Design
  2. Crossover Design

- Imperfect manipulation of the mediator:
  1. Parallel Encouragement Design
  2. Crossover Encouragement Design

- Implications for desining observational studies

The Parallel Design

- No manipulation effect assumption: The manipulation has no direct effect on outcome other than through the mediator value
- Running two experiments in parallel:



Identification under the Parallel Design

- Is the randomization of mediator sufficient? No!

- Sharp bounds: Binary mediator and outcome
- Use of linear programming (Balke and Pearl):
  - Objective function:

$$\mathbb{E}\{Y_i(1, M_i(0))\} = \sum_{y=0}^{1}\sum_{m=0}^{1}(\pi_{1ym1} + \pi_{y1m1})$$

    where $\pi_{y_1 y_0 m_1 m_0} = \Pr(Y_i(1,1) = y_1, Y_i(1,0) = y_0, M_i(1) = m_1, M_i(0) = m_0)$
  - Constraints implied by $\Pr(Y_i = y, M_i = m \mid T_i = t, D_i = 0)$, $\Pr(Y_i = y \mid M_i = m, T_i = t, D_i = 1)$, and the summation constraint

- More informative than those under the single experiment design
- Can sometimes identify the sign of average direct/indirect effects

An Example from Behavioral Neuroscience

Why study brain?: Social scientists' search for causal mechanisms underlying human behavior

- Psychologists, economists, and even political scientists

Question: What mechanism links low offers in an ultimatum game with "irrational" rejections?

- A brain region known to be related to fairness becomes more active when unfair offer received (single experiment design)

Design solution: manipulate mechanisms with TMS

- Knoch et al. use TMS to manipulate — turn off — one of these regions, and then observes choices (parallel design)

The Parallel Encouragement Design

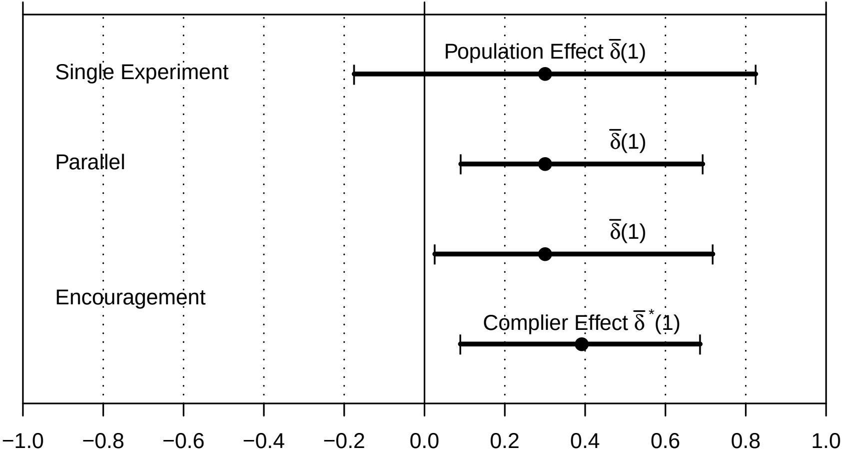
- Direct manipulation of mediator is often difficult
- Even if possible, the violation of no manipulation effect can occur
- Need for indirect and subtle manipulation

- Randomly encourage units to take a certain value of the mediator
- Instrumental variables assumptions (Angrist *et al.*):
  1. Encouragement does not discourage anyone
  2. Encouragement does not directly affects the outcome

- Not as informative as the parallel design
- Sharp bounds on the average "complier" indirect effects can be informative

A Numerical Example

- Based on the marginal distribution of a real experiment



The Crossover Design

Experiment 1

1) Randomize treatment

2) Measure mediator

3) Measure outcome

Same sample

Experiment 2

1) Fix treatment opposite Experiment 1

2) Manipulate mediator to level observed in Experiment 1

3) Measure outcome

Basic Idea

- Want to observe $Y_i(1-t, M_i(t))$
- Figure out $M_i(t)$ and then switch T_i while holding the mediator at this value
- Subtract direct effect from total effect

Key Identifying Assumptions

- No Manipulation Effect
- No Carryover Effect: First experiment doesn't affect second experiment
- Not testable, longer "wash-out" period

A Labor Market Discrimination Experiment

- Bertland and Mullainathan: manipulation of names on resumes
- Treatment: Black vs. White and Male vs. Female sounding names
- Mediator: perceived qualifications of applicants
- Outcome: callback rates

- (Natural) direct effects of applicants' race may be of interest
- Would Jamal get a callback if we send his resume as Greg?
- $\mathbb{E}(Y_i(1, M_i(1)) - Y_i(0, M_i(1)))$ vs. $\mathbb{E}(Y_i(1, m) - Y_i(0, m))$
- Key difference: use of actual resumes rather than fictitious ones

- First, send Jamal's resume as it is and record the outcome
- Then, send his resume as Greg and record the outcome

- No manipulation effect: potential employers are unaware
- Carryover effect: can be avoided if we send resumes to different (randomly matched) employers at the same time

The Crossover Encouragement Design

Experiment 1

1) Randomize treatment

2) Measure mediator

3) Measure outcome (optional)

Same sample ↓

Experiment 2

1) Fix treatment opposite Experiment 1

2) Randomly encourage mediator to level observed in Experiment 1

3) Measure outcome

Key Identifying Assumptions

- Encouragement doesn't discourage anyone
- No Manipulation Effect
- No Carryover Effect

Identification Analysis

- Identify indirect effects for "compliers"
- No carryover effect assumption is indirectly testable (unlike the crossover design)

Implications for the Design of Observational Studies

- Use of "natural experiments" in the social sciences
- Attempts to "replicate" experiments in observational studies

- Political science literature on incumbency advantage
- During 70s and 80s, the focus is on estimation of causal effects
- Positive effects, growing over time
- Last 20 years, search for causal mechanisms
- How large is the "scare-off/quality effect"?
- Estimation of direct effects using the crossover design:
  1. Use of repeat match-ups over two elections (Levitt) identifies $\mathbb{E}(Y_i(1, M_i(0)) - Y_i(0, M_i(0)))$ for some districts
  2. Use of redistricting (Ansolabehere et al, Sekhon & Titiunik) identifies $\mathbb{E}(Y_i(1, M_i(1)) - Y_i(0, M_i(1)))$ for parts of some districts

Concluding Remarks

- Identification of causal mechanisms is difficult but is possible
- Additional assumptions are required
- Five strategies:
  1. Single experiment design
  2. Parallel design
  3. Crossover design
  4. Encouragement design
  5. Crossover encouragement design
- Statistical assumptions: sequential ignorability, no interaction
- Design assumptions: no manipulation, no carryover effect
- Experimenters' creativity and technological development to improve the validity of these design assumptions

Project Reference

- Imai, Keele, and Yamamoto. (2010). "Identification, Inference, and Sensitivity Analysis for Causal Mediation Effects." *Statistical Science*, Forthcoming.
- Imai, Keele, and Tingley. "A General Approach to Causal Mediation Analysis." Working paper.
- Imai, Tingley, and Yamamoto. "Experimental Identification of Causal Mechanisms." Working paper.

available at **http://imai.princeton.edu**