Experimental Designs for Identifying Causal Mechanisms

Kosuke Imai

Princeton University

October 18, 2011
PSMG Meeting

Project Reference

My talk is based on the collaborative project with L. Keele (Penn State), D. Tingley (Harvard), and T. Yamamoto (MIT)

- "Experimental Designs for Identifying Causal Mechanisms." *Journal of Royal Statistical Society, Series A* (with discussions)

Some other related papers:

- "Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies." *American Political Science Review*
- "Identification, Inference, and Sensitivity Analysis for Causal Mediation Effects." *Statistical Science*
- "A General Approach to Causal Mediation Analysis." *Psychological Methods*
- "Causal Mediation Analysis Using R." *Advances in Social Science Research Using R*

Software `mediation` is freely available in R and Stata

Identification of Causal Mechanisms

- Causal inference is a central goal of scientific research
- Scientists care about causal mechanisms, not just about causal effects

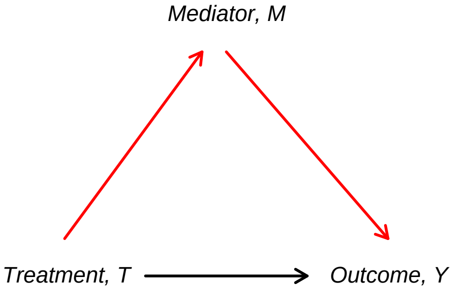
- Randomized experiments often only determine whether the treatment causes changes in the outcome
- Not how and why the treatment affects the outcome
- Common criticism of experiments and statistics:

black box view of causality

- Question: How can we learn about causal mechanisms from experimental and observational studies?

What Is a Causal Mechanism?

- Mechanisms as alternative causal pathways
- Cochran (1957)'s example:
  soil fumigants increase farm crops by reducing eel-worms
- Causal mediation analysis



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

Potential Outcomes Framework

Framework: Potential outcomes model of causal inference

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

- Potential mediators: $M_i(t)$, where $M_i = M_i(T_i)$ observed
- Potential outcomes: $Y_i(t, m)$, where $Y_i = Y_i(T_i, M_i(T_i))$ observed
- In a standard experiment, **only one potential outcome** can be observed for each i

Causal Mediation Effects

- Total causal effect:

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

- Causal mediation (Indirect) effects:

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

- Causal effect of the change in M_i on Y_i that would be induced by treatment
- Change the mediator from $M_i(0)$ to $M_i(1)$ while holding the treatment constant at t
- Represents the mechanism through M_i

Total Effect $=$ Indirect Effect $+$ Direct Effect

- Direct effects:

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

- Causal effect of T_i on Y_i, holding mediator constant at its potential value that would realize when $T_i = t$
- Change the treatment from 0 to 1 while holding the mediator constant at $M_i(t)$
- Represents all mechanisms other than through M_i

- Total effect $=$ mediation (indirect) effect $+$ direct effect:

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

Mechanisms, Manipulations, and Interactions

Mechanisms

- Indirect effects: $\delta_i(t) \equiv Y_i(t, M_i(1)) - Y_i(t, M_i(0))$
- 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 $T_i = t$

Interactions

- Interaction effects: $\xi(1, m, m') - \xi(0, m, m') \neq 0$
- Doesn't imply the existence of a mechanism

What Does the Observed Data Tell Us?

- Quantity of Interest: Average causal mediation effects

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

- Average direct effects ($\bar{\zeta}(t)$) are defined similarly
- Problem: $Y_i(t, M_i(t))$ is observed but $Y_i(t, M_i(t'))$ can never be observed
- We have an identification problem

$\Longrightarrow$ Need additional assumptions to make progress

Single Experiment Design

1) Randomize treatment

2) Measure mediator

3) Measure outcome

Assumption Satisfied

- Randomization of treatment

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

Key Identifying Assumption

- Sequential Ignorability:

$$Y_i(t,m) \perp\!\!\!\perp M_i \mid T_i, X_i$$

- Selection on (pre-treatment) observables
- Violated if there are unobservables that affect mediator and outcome
- Can't condition on post-treatment confounders

A Typical Psychological Experiment

- Brader *et al.*: media framing experiment
- Treatment: Ethnicity (Latino vs. Caucasian) of an immigrant
- Mediator: anxiety
- Outcome: preferences over immigration policy

- Single experiment design with statistical mediation analysis
- Emotion: difficult to directly manipulate

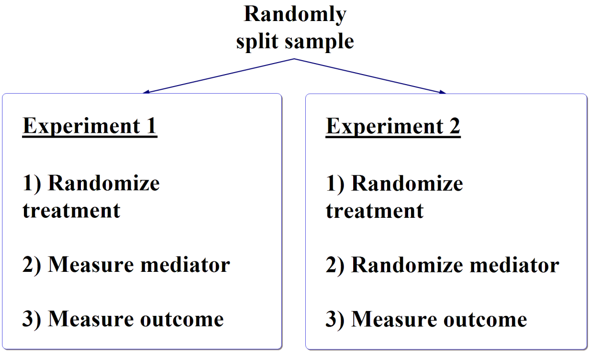
- Sequential ignorability assumption is not credible
- Possible confounding

Beyond Sequential Ignorability

- Without sequential ignorability, standard experimental design lacks identification power
- Even the sign of ACME is not identified

- Need to develop alternative experimental designs for more credible inference
- Possible when the mediator can be directly or indirectly manipulated

Parallel Design



- Must assume no direct effect of manipulation on outcome
- More informative than standard single experiment
- If we assume no T–M interaction, ACME is point identified

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)

Limitations

- Difference between manipulation and mechanism

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

- Here, $\mathbb{E}(M_i(1) - M_i(0)) = \mathbb{E}(Y_i(t,1) - Y_i(t,0)) = 0.2$, but $\bar{\delta}(t) = -0.2$
- Limitations:
  - Direct manipulation of the mediator is often impossible
  - Even if possible, manipulation can directly affect outcome
- Need to allow for subtle and indirect manipulations

Encouragement Design

- Randomly encourage subjects to take particular values of the mediator M_i
- Standard instrumental variable assumptions (Angrist et al.)

Use a 2×3 factorial design:

1. Randomly assign T_i
2. Also randomly decide whether to **positively encourage**, **negatively encourage**, or do nothing
3. Measure mediator and outcome

- Informative inference about the “complier” ACME
- Reduces to the parallel design if encouragement is perfect
- Application to the immigration experiment:
  Use autobiographical writing tasks to encourage anxiety

Crossover Design

- Recall ACME can be identified if we observe $Y_i(t', M_i(t))$
- Get $M_i(t)$, then switch T_i to t' while holding $M_i = M_i(t)$
- Crossover design:
  1. Round 1: Conduct a standard experiment
  2. Round 2: Change the treatment to the opposite status but fix the mediator to the value observed in the first round
- Very powerful – identifies mediation effects for each subject
- Must assume no carryover effect: Round 1 must not affect Round 2
- Can be made plausible by design

Example from Labor Economics

Bertrand & Mullainathan (2004, AER)

- Treatment: Black vs. White names on CVs
- Mediator: Perceived qualifications of applicants
- Outcome: Callback from employers

- Quantity of interest: Direct effects of (perceived) race
- Would Jamal get a callback if his name were Greg but his qualifications stayed the same?

- Round 1: Send Jamal's actual CV and record the outcome
- Round 2: Send his CV as Greg and record the outcome

- Assumptions are plausible

Crossover Encouragement Design

- Crossover encouragement design:
  1. Round 1: Conduct a standard experiment
  2. Round 2: Same as crossover, except encourage subjects to take the mediator values

EXAMPLE Hainmueller & Hiscox (2010, APSR)

- Treatment: Framing immigrants as low or high skilled
- Outcome: Preferences over immigration policy
- Possible mechanism: Low income subjects may expect higher competition from low skill immigrants

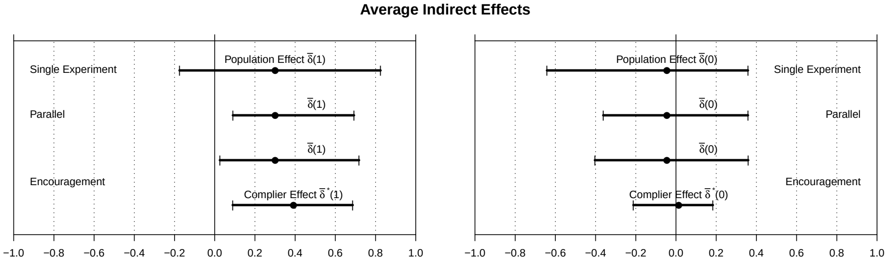
- Manipulate expectation using a news story
- Round 1: Original experiment but measure expectation
- Round 2: Flip treatment, but encourage expectation in the same direction as Round 1

Comparing Alternative Designs

- No manipulation
  - Single experiment: sequential ignorability
- Direct manipulation
  - Parallel: no manipulation effect, no interaction effect
  - Crossover: no manipulation effect, no carryover effect
- Indirect manipulation
  - Encouragement: no manipulation effect, monotonicity, no interaction (?)
  - Crossover encouragement: no manipulation effect, monotonicity, no carryover effect

Identification Power

- A numerical example based on Brader et al. (2008)
- Binary outcome, mediator, and treatment
- Sharp bounds for parallel and encouragement designs without no-interaction assumption



Designing Observational Studies

- Key difference between experimental and observational studies: treatment assignment
- Sequential ignorability:
  1. Ignorability of treatment given covariates
  2. Ignorability of mediator given treatment and covariates
- Both (1) and (2) are suspect in observational studies

- Statistical control: matching, propensity scores, etc.
- Search for quasi-randomized treatments: "natural" experiments

- How can we design observational studies?
- Experiments can serve as templates for observational studies

Example from Political Science

EXAMPLE Incumbency advantage

- Estimation of incumbency advantages goes back to 1960s
- Why incumbency advantage? Scaring off quality challenger
- Use of cross-over design (Levitt and Wolfram)
  1. 1st Round: two non-incumbents in an open seat
  2. 2nd Round: same candidates with one being an incumbent
- Assume challenger quality (mediator) stays the same
- Estimation of direct effect is possible

- Redistricting as natural experiments (Ansolabehere et al.)
  1. 1st Round: incumbent in the old part of the district
  2. 2nd Round: incumbent in the new part of the district
- Challenger quality is the same but treatment is different
- Estimation of direct effect is possible

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

The project website for papers and software:

http://imai.princeton.edu/projects/mechanisms.html

Email for comments and suggestions:

kimai@Princeton.Edu