Experimental Identification of Causal Mechanisms

Kosuke Imai

Princeton University

Joint work with Dustin Tingley and Teppei Yamamoto

March 26, 2010

Experiments, Statistics, and Causal Mechanisms

- Causal inference is a central goal of most scientific research
- Experiments as **gold standard** for estimating *causal effects*
- But, scientists actually care about *causal mechanisms*
- Knowledge about causal mechanisms can also improve policies

- 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

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

Some Papers

- Imai, Keele, and Yamamoto. "Identification, Inference, and Sensitivity Analysis for Causal Mediation Effects." *Statistical Science*, in-press.
- Imai, Tingley, and Yamamoto. "Experimental Identification of Causal Mechanisms." Working paper.

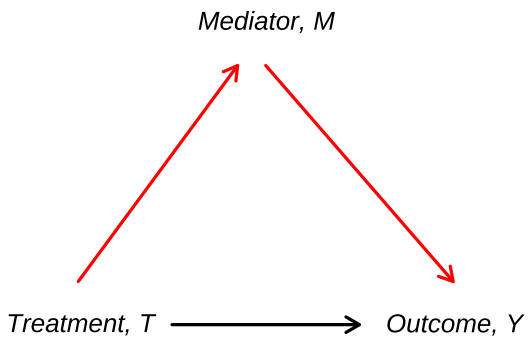
available at **http://imai.princeton.edu**

Overview of the Talk

- Show the limitation of a common approach
- Consider alternative experimental designs
- What is a minimum set of assumptions required for identification under each design?
- How much can we learn without the key identification assumptions under each design?
- Identification of causal mechanisms is possible but difficult
- Distinction between design and statistical assumptions
- Roles of creativity and technological developments

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 examples: resource curse, habitual voting
- 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 $Y_i = Y_i(T_i, M_i(T_i))$

- Fundamental problem of causal inference (Holland):
  Only one potential value is observed

Defining and Interpreting Indirect Effects

- Total causal effect:

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

- Indirect (causal mediation) effects (Robins and Greenland; Pearl):

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

- Change $M_i(0)$ to $M_i(1)$ while holding the treatment constant at t
- Effect of a change in M_i on Y_i that would be induced by treatment

- Fundamental problem of causal mechanisms:

  For each unit i, $Y_i(t, M_i(t))$ is observable but
  $Y_i(t, M_i(1-t))$ is not even observable

Defining and Interpreting Direct Effects

- Direct effects:

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

- Change T_i from 0 to 1 while holding the mediator constant at $M_i(t)$
- Causal effect of T_i on Y_i, holding mediator constant at its potential value that would be realized when $T_i = t$

- Total effect = indirect effect + direct effect:

$$\begin{aligned}\tau_i &= \frac{1}{2}\{\delta_i(0) + \delta_i(1) + \zeta_i(0) + \zeta_i(1)\} \\ &= \delta_i + \zeta_i \quad \text{if } \delta_i = \delta_i(0) = \delta_i(1) \text{ and } \zeta_i = \zeta_i(0) = \zeta_i(1)\end{aligned}$$

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

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

Identification under the Single Experiment Design

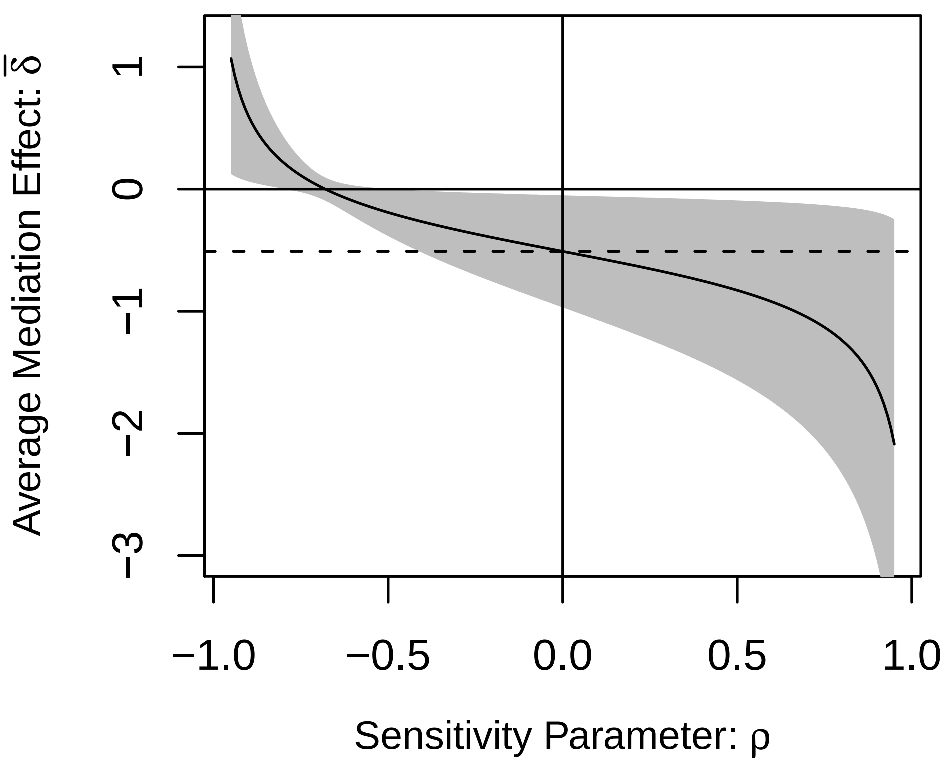
- Sequential ignorability yields nonparametric identification
- Under the single experiment design and sequential ignorability,

$$\bar{\delta}(t) = \int\int \mathbb{E}(Y_i \mid M_i, T_i = t, X_i)\{dP(M_i \mid T_i = 1, X_i) - dP(M_i \mid T_i = 0, X_i)\}\, dP(X_i)$$

- Linear structural equation modeling (a.k.a. Baron-Kenny)

- Sequential ignorability is an untestable assumption
- Sensitivity analysis: How large a departure from sequential ignorability must occur for the conclusions to no longer hold?

Sensitivity Analysis



Identification Power of the Single Experiment Design

- How much can we learn without sequential ignorability?
- Sharp bounds on indirect effects (Sjölander):

$$\max\left\{\begin{array}{l}-P_{001}-P_{011}\\-P_{011}-P_{010}-P_{110}\\-P_{000}-P_{001}-P_{100}\end{array}\right\} \le \bar{\delta}(1) \le \min\left\{\begin{array}{l}P_{101}+P_{111}\\P_{010}+P_{110}+P_{111}\\P_{000}+P_{100}+P_{101}\end{array}\right\}$$

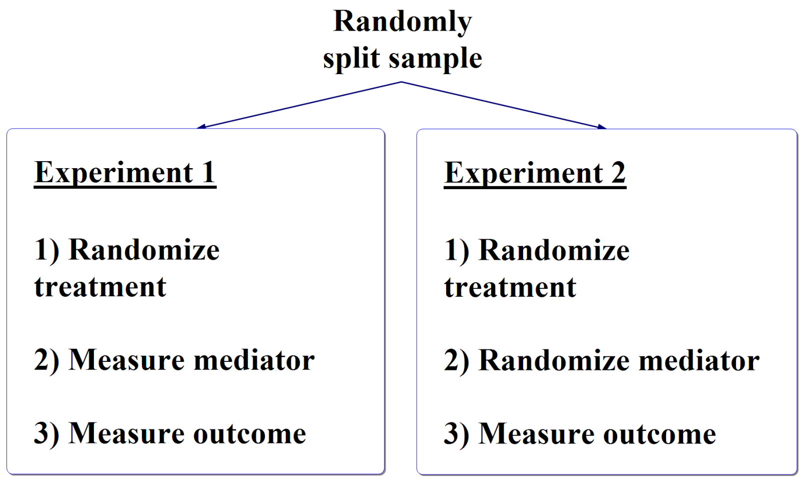
$$\max\left\{\begin{array}{l}-P_{100}-P_{110}\\-P_{011}-P_{111}-P_{110}\\-P_{001}-P_{101}-P_{100}\end{array}\right\} \le \bar{\delta}(0) \le \min\left\{\begin{array}{l}P_{000}+P_{010}\\P_{011}+P_{111}+P_{010}\\P_{000}+P_{001}+P_{101}\end{array}\right\}$$

where $P_{ymt} = \Pr(Y_i = y, M_i = m \mid T_i = t)$

- The sign is not identified
- Can we design experiments to better identify causal mechanisms?

The Parallel Design

- Suppose we can directly manipulate the mediator without directly affecting the outcome
- 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

- 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

- $\mathbb{E}(M_i(1) - M_i(0)) = \mathbb{E}(Y_i(t,1) - Y_i(t,0)) = 0.2$, but $\bar{\delta}(t) = -0.2$
- Is the randomization of mediator sufficient? No
- The no interaction assumption (Robins) yields point identification

$$Y_i(1,m) - Y_i(1,m') \;=\; Y_i(0,m) - Y_i(0,m')$$

- Must hold at the unit level
- Not directly testable but indirect tests are possible

Sharp Bounds under the Parallel Design

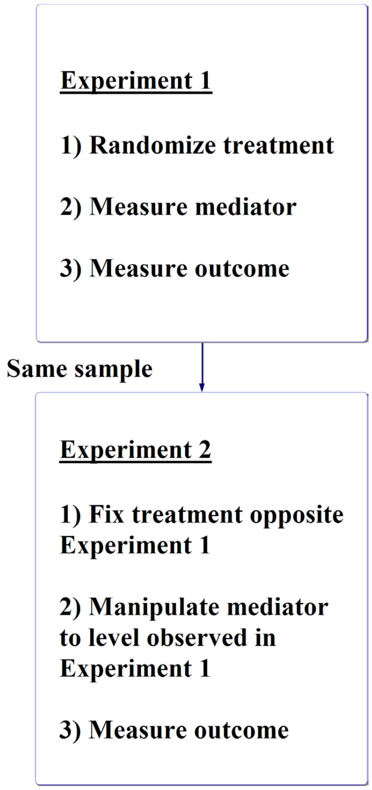
- Again, a special case of binary mediator and outcome
- Use of linear programming (Balke and Pearl)
- Objective function:

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

where $\pi_{y_1 y_0 m_1 m_0} = \Pr(Y_i(1,1) = y_1, Y_i(1,0) = y_0, M_i(1) = m_1, M_i(0) = m_0)$

- Linear 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
- Sharp bounds (expressions given in the paper) are more informative than those under the single experiment design
- Can sometimes identify the sign of average indirect effects

The Crossover Design



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

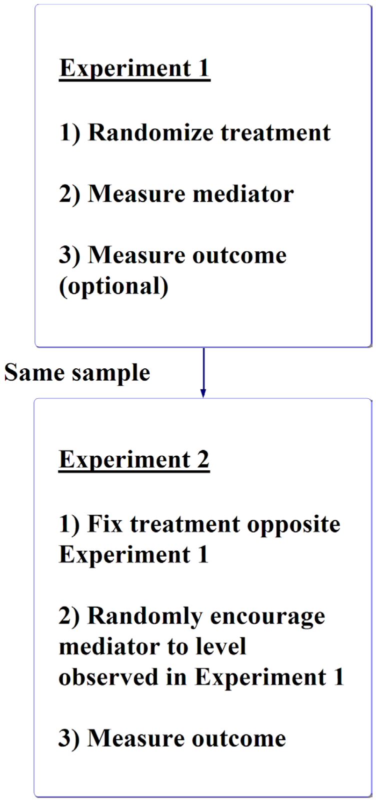
The Encouragement Design

- 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

The Crossover Encouragement Design



Key Identifying Assumptions

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

Identification Analysis

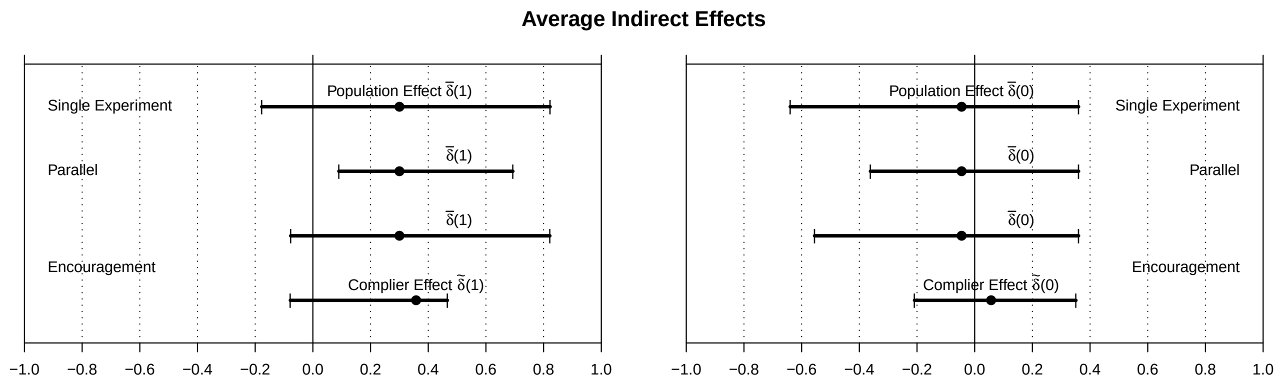
- Identify indirect effects for "compliers"
- No carryover effect assumption is indirectly testable (unlike the crossover design)

Comparing Alternative Designs

- No manipulation
  - Single experiment: sequential ignorability
- Direct manipulation
  - Parallel: no manipulation effect, no interaction
  - 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

An Example from Social Science

- Brader *et al.*: media framing experiment
  - Single experiment design with statistical mediation analysis
  - Treatment: Ethnicity (Latino vs. Caucasian) of an immigrant
  - Mediator: anxiety
  - Outcome: immigration
- Emotion: difficult to directly manipulate but indirect manipulation may be possible
- An artificial data consistent with the observed data



An Example from Behavioral Neuroscience

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)

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