Identification and Sensitivity Analysis for Multiple Causal Mechanisms: Revisiting Evidence from Framing Experiments

Kosuke Imai[1] **Teppei Yamamoto**[2]

[1]Princeton University

[2]Massachusetts Institute of Technology

July 25, 2012

Symposium on Causality 2012
Jena and Dornburg

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:

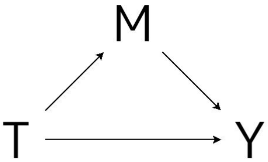
  > *it can only determine whether the treatment causes changes in the outcome, but not how and why*

- 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?

Motivation

- Use of causal mediation analysis to study causal mechanisms
- A fast-growing methodological literature on causal mediation
- Existing work tends to focus on a single mechanism:



- However, multiple mediators are common in applied settings
- Applied researchers often aim to test competing theories by comparing mediation effects

Causally Independent vs. Dependent Mechanisms

- Quantity of interest $=$ Average indirect effect with respect to M
- W represents the alternative observed mediators

- Left: Assumes independence between M and W
- Right: Allows M to be affected by W
- W represents post-treatment confounders between M and Y

- Applied researchers often implicitly assume independence

Our Contributions

1. Analyze multiple mediators that are causally dependent
2. Show that the standard path-analytic approach implicitly assumes independence between mechanisms
3. Use a semiparametric linear structural equation model to simplify analysis while not compromising too much on flexibility
4. Identification under the homogeneous interaction assumption
5. Sensitivity analysis for possible heterogeneity in the degree of treatment-mediator interaction
6. Extension to new experimental designs to avoid relying on a sequential ignorability assumption

A Review of Single Mediator Case

- 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 $M_i = M_i(T_i)$
- Potential outcomes: $Y_i(t, m)$ where $Y_i = Y_i(T_i, M_i(T_i))$

- Fundamental problem of causal inference (Rubin; Holland):

  Only one potential value is observed

  1. If $T_i = 1$, then $M_i(1)$ is observed but $M_i(0)$ is not
  2. If $T_i = 0$ and $M_i(0) = 0$, then $Y_i(0, 0)$ is observed but $Y_i(1, 0)$, $Y_i(0, m)$, and $Y_i(1, m)$ are not when $m \neq 0$

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:

$$\begin{aligned} \tau_i &= \delta_i(t) + \zeta_i(1-t) \\ &= \delta_i + \zeta_i \end{aligned}$$

where the second equality assumes $\delta_i(0) = \delta_i(1)$ and $\zeta_i(0) = \zeta_i(1)$

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

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 (Imai *et al.*, 2010):

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

- Selection on pre-treatment observables
- Unmeasured pre-treatment confounders
- Measured/unmeasured post-treatment confounders

Identification under the Single Experiment Design

- Sequential ignorability yields nonparametric identification
- Linear structural equation model (a.k.a. Baron-Kenny) as a special case
- Easy to extend to other non-linear models

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

- But, what about post-treatment confounders?

Multiple Mediator Example: A Framing Experiment

- Framing may affect how individuals perceive the issue and change attitudes and behavior (Tversky and Kahneman 1981)
- How does framing of political issues affect public opinions?

Example: Druckman and Nelson (2003) ($N = 261$)

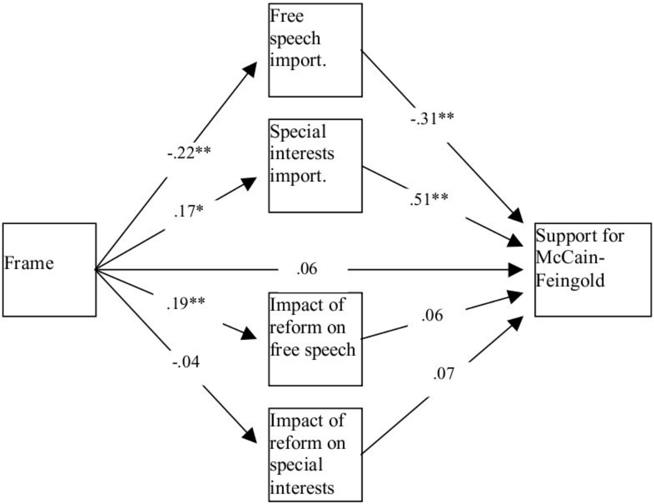
- Treatment: News paper article on a proposed election campaign finance reform, emphasizing either its positive or negative impact
- Outcome: Support for the proposed reform

- Primary mediator: Perceived importance of free speech
- Alternative (possibly confounding) mediator: Belief about the impact of the proposed reform

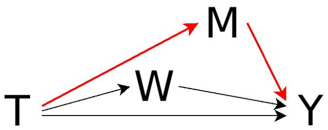
Two other examples in the paper (Slothuus 2008, Brader et al. 2008)

Original Analysis Assumes Independent Mechanisms

Druckman and Nelson, p.738



Causally Independent Multiple Mediators



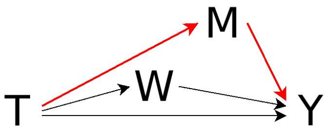
- Potential mediators: $M_i(t)$ and $W_i(t)$
- Potential outcomes: $Y_i(t, m, w)$
- The indirect and natural direct effects:

$$
\begin{aligned}
\delta_i^M(t) &\equiv Y_i(t, M_i(1), W_i(t)) - Y_i(t, M_i(0), W_i(t)) \\
\delta_i^W(t) &\equiv Y_i(t, M_i(t), W_i(1)) - Y_i(t, M_i(t), W_i(0)) \\
\zeta_i(t, t') &\equiv Y_i(1, M_i(t), W_i(t')) - Y_i(0, M_i(t), W_i(t'))
\end{aligned}
$$

- These sum up to the total effect, as expected:

$$\tau_i = \delta_i^M(t) + \delta_i^W(1-t) + \zeta_i(1-t, t)$$

Identification of Independent Multiple Mechanisms



- W is posttreatment but not a confounder between M and Y
- Independent multiple mediators can be analyzed under sequential ignorability:

$$\begin{aligned}\{Y_i(t,m,w), M_i(t'), W_i(t'')\} &\perp\!\!\!\perp T_i \mid X_i = x \\ Y_i(t', m, W_i(t')) &\perp\!\!\!\perp M_i \mid T_i = t, X_i = x \\ Y_i(t', M_i(t'), w) &\perp\!\!\!\perp W_i \mid T_i = t, X_i = x\end{aligned}$$

- SI $\Longrightarrow$ Nonparametric identification of $\bar{\delta}^M(t)$, $\bar{\delta}^W(t)$ and $\bar{\zeta}(t,t')$

Unpacking the Standard Path-Analytic Approach

- Social science applications often use structural equation models:

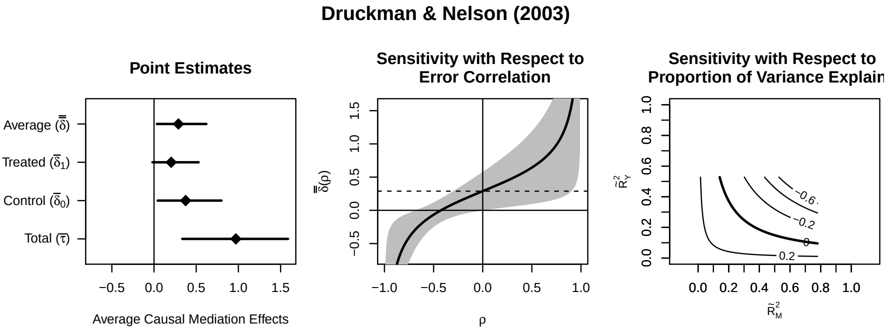
$$\begin{aligned} M_i &= \alpha_M + \beta_M T_i + \xi_M^\top X_i + \epsilon_{iM} \\ W_i &= \alpha_W + \beta_W T_i + \xi_W^\top X_i + \epsilon_{iW} \\ Y_i &= \alpha_3 + \beta_3 T_i + \gamma M_i + \theta^\top W_i + \xi_3^\top X_i + \epsilon_{i3} \end{aligned}$$

- The mediation effects are then estimated as $\hat{\beta}_M \hat{\gamma}$ for M and $\hat{\beta}_W \hat{\theta}$ for W
- Under SI, consistent for $\bar{\delta}_i^M$ and $\bar{\delta}_i^W$ (if the linear models are correct)
- However, under SI analyzing one mechanism at a time is also valid:

$$\begin{aligned} M_i &= \alpha_2 + \beta_2 T_i + \xi_2^\top X_i + \epsilon_{i2} \\ Y_i &= \alpha_3 + \beta_3 T_i + \gamma M_i + \xi_3^\top X_i + \epsilon_{i3} \end{aligned}$$

- The standard approach does not address multiple mechanisms at all!
- Correlation between M and W given $(T, X) \Longrightarrow$ potential violation of SI

Empirical Analysis under Independence Assumption



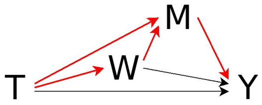
- Weakly significant average indirect effects ([0.025, 0.625]), accounting for 28.6 percent of the total effect
- Moderate degree of sensitivity to the mediator exogeneity ($\bar{\delta} = 0$ when $\rho = -0.43$ or $\tilde{R}^2_M \tilde{R}^2_Y = 0.078$)
- Potential problem (both theoretical and empirical): The importance mechanism may be affected by the belief content mechanism

Causally Dependent Multiple Mechanisms

- Binary treatment: $T_i \in \{0, 1\}$
- We allow W to influence both M and Y:

Potential mediators: $W_i(t)$ and $M_i(t, w)$

Potential outcomes: $Y_i(t, m, w)$



- Causal mediation effect (natural indirect effect):

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

- Natural direct effect:

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

- These sum up to the total effect (again):

$$\begin{aligned} \tau_i &\equiv Y_i(1, M_i(1, W_i(1)), W_i(1)) - Y_i(0, M_i(0, W_i(0)), W_i(0)) \\ &= \delta_i(t) + \zeta_i(1 - t) \end{aligned}$$

Identification of Causally Related Mechainsms

- Consider the (weak) sequential ignorability (SI) assumption:

$$\begin{aligned} \{Y_i(t,m,w), M_i(t,w), W_i(t)\} &\perp\!\!\!\perp T_i \mid X_i = x \\ \{Y_i(t,m,w), M_i(t,w)\} &\perp\!\!\!\perp W_i \mid T_i = t,\ X_i = x \\ \{Y_i(t,m,w)\} &\perp\!\!\!\perp M_i \mid W_i(t) = w,\ T_i = t,\ X_i = x \end{aligned}$$

- A special case of Robins' FRCISTG (1986)
- Observed posttreatment confounding (W) is allowed
- Empirically verifiable, at least in theory

- Robins (2003): Under FRCISTG, the no interaction assumption (between T and M) nonparametrically identifies $\bar{\delta}(t)$:

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

Why Do We Need the No-Interaction Assumption?

- A hypothetical population:

Prop.	$M_i(1,w)$	$M_i(0,w)$	$Y_i(t,1,w)$	$Y_i(t,0,w)$	$\delta_i(t)$
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

- Suppose there is no confounding. We can identify

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

- But the product of coefficients method fails miserably:

$$\bar{\delta}(t) = -0.2 \neq 0.2 \times 0.2 = 0.04$$

- Why? Interaction between T and M

- Implications:
  1. Is the randomization of mediator sufficient? No
  2. Test the assumption indirectly at the mean level
  3. Analyze a group of homogeneous units

The Proposed Framework

- Problem: The no interaction assumption is too strong in most cases (e.g. Is the effect of issue importance invariant across frames?)
- Solution: Assume a flexible (semi-parametric) linear model

$$\begin{aligned} M_i(t, w) &= \alpha_2 + \beta_{2i} t + \xi_{2i}^\top w + \mu_{2i}^\top tw + \lambda_{2i}^\top x + \epsilon_{2i}, \\ Y_i(t, m, w) &= \alpha_3 + \beta_{3i} t + \gamma_i m + \kappa_i tm + \xi_{3i}^\top w + \mu_{3i}^\top tw + \lambda_{3i}^\top x + \epsilon_{3i}, \end{aligned}$$

where $\mathbb{E}(\epsilon_{2i}) = \mathbb{E}(\epsilon_{3i}) = 0$

- Allows for dependence of M on W
- Coefficients can vary arbitrarily across units

Sensitivity Analysis w.r.t. Interaction Heterogeneity

- The model can be rewritten as:

$$\begin{aligned} M_i(t,w) &= \alpha_2 + \beta_2 t + \xi_2^\top w + \mu_2^\top tw + \lambda_2^\top x + \eta_{2i}(t,w), \\ Y_i(t,m,w) &= \alpha_3 + \beta_3 t + \gamma m + \kappa tm + \xi_3^\top w + \mu_3^\top tw + \lambda_3^\top x + \eta_{3i}(t,m,w), \end{aligned}$$

  where $\beta_2 = \mathbb{E}(\beta_{2i})$, etc.

- FRCISTG implies

$$\mathbb{E}(\eta_{2i}(T_i, W_i) \mid X_i, T_i, W_i) = \mathbb{E}(\eta_{3i}(T_i, M_i, W_i) \mid X_i, T_i, W_i, M_i) = 0$$

  The mean coefficients β_2, etc. can thus be estimated without bias

- We show that $\bar{\delta}(t)$ and $\bar{\zeta}(t)$ can be written as

$$\begin{aligned} \bar{\delta}(t) &= \bar{\tau} - \bar{\zeta}(1-t) \\ \bar{\zeta}(t) &= \beta_3 + \kappa \mathbb{E}(M_i \mid T_i = t) + \rho_t \sigma \sqrt{\mathbb{V}(M_i \mid T_i = t)} \\ &\quad + (\xi_3 + \mu_3)^\top \mathbb{E}(W_i \mid T_i = 1) - \xi_3^\top \mathbb{E}(W_i \mid T_i = 0) \end{aligned}$$

  where $\rho_t = \mathrm{Corr}(M_i(t, W_i(t)), \kappa_i)$ and $\sigma = \sqrt{\mathbb{V}(\kappa_i)}$ are the only unidentified quantities

- Sensitivity analysis: Examine how $\bar{\delta}(t)$ varies as a function of ρ_t and σ

Remarks on the Proposed Sensitivity Analysis

- Interpretation of ρ_t difficult
  $\longrightarrow$ Set $\rho_t \in [-1, 1]$ and examine sharp bounds on $\bar{\delta}(t)$ as functions of σ

- Point identification under the homogeneous interaction assumption:

$$Y_i(1, m, W_i(1)) - Y_i(0, m, W_i(0)) \;=\; B_i + Cm$$

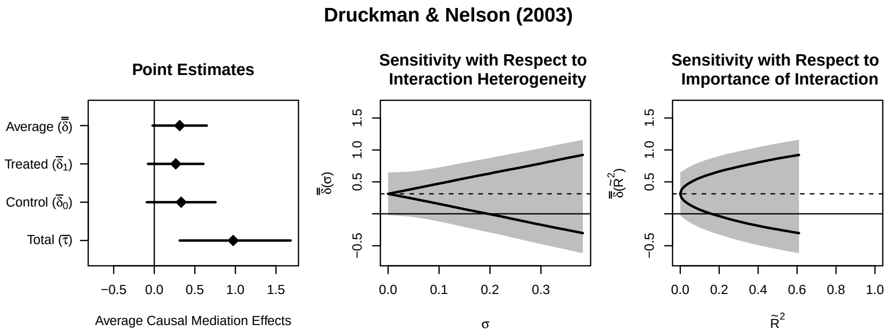
- The causal mechanism is identified as long as the degree of T–M interaction does not vary across units

- Alternative formulation using R^2 for easier interpretation:

$$R^{2*} \;=\; \frac{\mathbb{V}(\tilde{\kappa}_i T_i M_i)}{\mathbb{V}(\eta_{3i}(T_i, M_i, W_i))} \quad \text{and} \quad \widetilde{R}^2 \;=\; \frac{\mathbb{V}(\tilde{\kappa}_i T_i M_i)}{\mathbb{V}(Y_i)}$$

- How much variation in Y_i would the interaction heterogeneity have to explain for the estimate to be zero?

Reanalysis of Druckman and Nelson



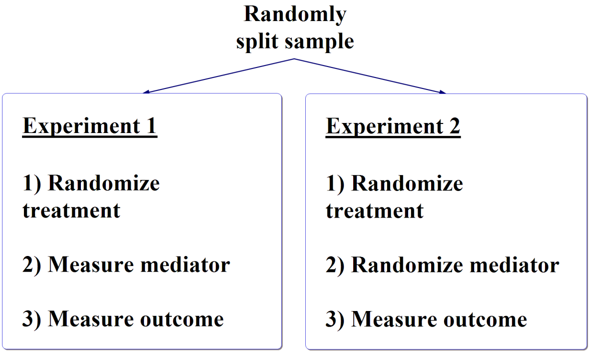
- Mediation effects insignificant at 90% ($[-0.021, 0.648]$)
- Lower bound on $\bar{\delta}$ equals zero when $\sigma = 0.195$, i.e. when σ is about half as large as its largest possible value
- Effect would go away if the interaction heterogeneity explained 15.9% of the total variance of the outcome variable

New Experimental Designs

- What about unmeasured pre and post-treatment confounding?
- Need better research designs
- New experimental designs (Imai *et al. JRSS-A* in-press):
- Manipulate mediator either directly or indirectly
  1. Parallel design
  2. Parallel encouragement design
- Proposed sensitivity analysis can be extended to these designs

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:



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)

Statistical inference:

- No interaction assumption required for point identification
- Proposed sensitivity analysis can be extended

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 affect the outcome

- Not as informative as the parallel design
- Sharp bounds on the average "complier" indirect effects can be informative
- No interaction assumption required for point identification
- Proposed sensitivity analysis can be extended to two-stage least squares

Concluding Remarks

Summary:

- We extend the causal mediation analysis framework to multiple mediators
- The framework deals with observed post-treatment confounders
- Varying coefficient linear models more flexible than traditional SEMs
- Point identification under homogeneous interaction assumption
- Sensitivity analysis with respect to the degree of interaction heterogeneity
- Extension to new experimental designs

Open-Source Software:

- All methods discussed today and much more can be implemented via:

mediation: R Package for Causal Mediation Analysis

The project website for papers and software:

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

Email for comments and suggestions:

kimai@Princeton.Edu