Identification and Sensitivity Analysis of Contagion Effects in Randomized Placebo-Controlled Trials

Kosuke Imai

Harvard University

Keynote Talk, Causal Mediation Analysis Training
Mailman School of Public Health, Columbia University
August 14, 2020

Joint work with Zhichao Jiang (UMass. Amherst)

Causal Mediation Analysis in Scientific Research

- Causal inference is a central goal of scientific research
- Scientists care about mechanisms, not just effects $\rightsquigarrow$ external validity
- Policy makers want to devise better policies $\rightsquigarrow$ target key mechanism

- 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:

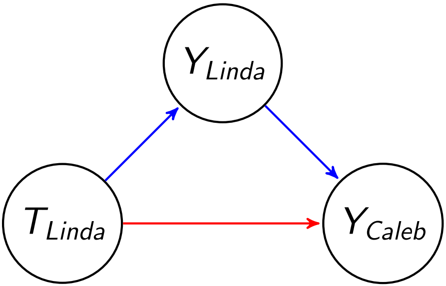
  black box view of causality
- Qualitative research $\rightsquigarrow$ process tracing

- Question: How can we learn about causal mechanisms from experimental and observational studies? $\rightsquigarrow$ causal mediation analysis

Causal Mechanisms of Spillover Effects

- Common assumption: no interference between units
- One's outcome is affected only by their own treatment

- A growing methodological literature on spillover effects
- What are causal mechanisms of spillover effects?
  1. contagion effects
  2. direct effects



Randomized Placebo-controlled Trial of GOTV Campaign

- How does the effect of GOTV canvassing spread within a household?
- Placebo-controlled RCTs in Denver and Minneapolis during the 2002 Congressional primary election (Nickerson, 2008)

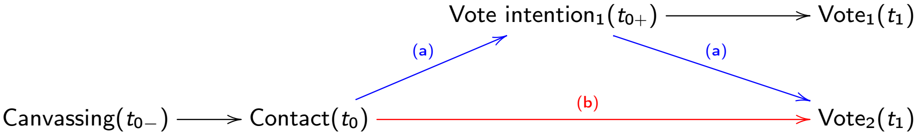
- 956 households with two registered voters
  1. treatment: canvasser visits to encourage turnout
  2. placebo: canvasser visits to encourage recycling
  3. control: no canvasser

- Placebo condition is useful:
  - nobody answered the door for more than a half of households
  - contacted voters in the treatment and placebo groups are comparable
- Control condition is only used for estimating placebo effects

Spillover Effects within Households

	Denver		Minneapolis		Pooled	
	Contact	Non-contact	Contact	Non-contact	Contact	Non-contact
Turnout rates						
GOTV	0.477	0.424	0.272	0.238	0.392	0.346
	(0.030)	(0.029)	(0.031)	(0.030)	(0.022)	(0.022)
Recycling	0.391	0.369	0.162	0.173	0.298	0.289
	(0.029)	(0.029)	(0.027)	(0.027)	(0.021)	(0.021)
Control	0.384		0.172		0.312	
	(0.012)		(0.013)		(0.009)	
Causal effects						
Treatment effect	0.086	0.055	0.110	0.065	0.094	0.057
	(0.042)	(0.041)	(0.041)	(0.041)	(0.031)	(0.030)
Placebo effect	0.006		0.005		0.005	
	(0.017)		(0.018)		(0.113)	
Number of households	1124		786		1910	

A Causal Model of Within-Household Spillover Effects



- Mediator = vote intention immediately after contact
- Two mechanisms:
  - a. Contagion effect: a canvasser convinces a contacted voter who then convinces a non-contacted voter
  - b. Direct effect: canvassing has no effect on a contacted voter's turnout but still influences a non-contacted voter
- Methodological challenges:
  1. unmeasured (pre-treatment) confounding: e.g., interests in politics
  2. missing mediator: we do not observe vote intention
  3. measurement error: contacted voter may change their mind

Potential Outcomes Framework

- Households: $i = 1, 2, \ldots, N$
- Pre-treatment covariates: $\boldsymbol{X}_i$
- Treatment conditions: $Z_i = \begin{cases} 0 & \text{control} \\ 1 & \text{treatment} \\ 2 & \text{placebo} \end{cases}$
- Contact:
  - potential values: $D_i(z) \in \{0, 1\}$ for $z = 0, 1, 2$
  - observed value: $D_i = D_i(Z_i)$
- Complier status: $G_i = \begin{cases} c & \text{if } D_i(1) = D_i(2) = 1 \\ n & \text{if } D_i(1) = D_i(2) = 0 \end{cases}$
- Vote intention for contacted voter in a complier household $G_i = c$:
  - potential values: $Y_{i1}^*(z) \in \{0, 1\}$
  - realized (but unobserved) value: $Y_{i1}^* = Y_{i1}^*(Z_i)$
- Turnout for contacted and non-contacted voters:
  - potential values: $Y_{i1}(z, y_1^*), Y_{i2}(z, y_1^*) \in \{0, 1\}$
  - observed values: $Y_{i1} = Y_{i1}(Z_i, Y_{i1}^*(Z_i))$ and $Y_{i2} = Y_{i2}(Z_i, Y_{i1}^*(Z_i))$

Causal Quantities of Interest

1. Average Spillover Effect (e.g., Halloran and Struchiner 1995)

$$\theta \ = \ \mathbb{E}\{Y_{i2}(1, Y^*_{i1}(1)) - Y_{i2}(0, Y^*_{i1}(0)) \mid G_i = c\}.$$

2. Average Contagion Effects (e.g., Robins and Greenland 1992; Pearl 2001)

$$\tau(z) \ = \ \mathbb{E}\{Y_{i2}(z, Y^*_{i1}(1)) - Y_{i2}(z, Y^*_{i1}(0)) \mid G_i = c\} \quad \text{for } z = 0, 1.$$

3. Average Direct Effects

$$\eta(z) \ = \ \mathbb{E}\{Y_{i2}(1, Y^*_{i1}(z)) - Y_{i2}(0, Y^*_{i1}(z)) \mid G_i = c\} \quad \text{for } z = 0, 1.$$

Decomposition of the Average Spillover Effects:

$$\theta \ = \ \tau(1) + \eta(0) \ = \ \tau(0) + \eta(1).$$

Identification Assumptions

1. Randomization of treatment assignment

$$Z_i \perp\!\!\!\perp \{D_i(z), Y^*_{i1}(z), Y_{i2}(z, y^*_1), \boldsymbol{X}_i\}$$

for $z = 0, 1, 2$ and $y^*_1 = 0, 1$.

2. Latent ignorability of mediator among compliers

$$Y_{i2}(z', y^*_1) \perp\!\!\!\perp Y^*_{i1}(z) \mid Z_i = z, G_i = c, \boldsymbol{X}_i$$

for $z, z' = 0, 1$, and $y^*_1 = 0, 1$.

3. Zero average placebo effect among compliers

$$\begin{aligned} \mathbb{E}\{Y^*_{i1}(2) \mid G_i = c, \boldsymbol{X}_i\} &= \mathbb{E}\{Y^*_{i1}(0) \mid G_i = c, \boldsymbol{X}_i\} \\ \mathbb{E}\{Y_{i2}(2, y^*_1) \mid G_i = c, \boldsymbol{X}_i\} &= \mathbb{E}\{Y_{i2}(0, y^*_1) \mid G_i = c, \boldsymbol{X}_i\} \end{aligned}$$

for $y^*_1 = 0, 1$.

Assumptions about Potential Measurement Error

1. Perfect proxy for mediator among compliers

$$Y_{i1}^* \;=\; Y_{i1}, \quad \text{for all } i \text{ with } G_i = c.$$

2. Non-differential measurement error of the mediator among compliers

$$\begin{aligned} & \Pr(Y_{i1} = y_1^* \mid Y_{i1}^* = y_1^*, G_i = c, Y_{i2}, Z_i, \boldsymbol{X}_i) \\ = \;& \Pr(Y_{i1} = y_1^* \mid Y_{i1}^* = y_1^*, G_i = c, \boldsymbol{X}_i) \end{aligned}$$

for $y_1^* = 0, 1$.

- actual turnout depends only on vote intention and pre-treatment covariates
- no confounder between non-contacted voter's turnout and contacted voter's turnout
- no causal effect of treatment on a contacted voter's turnout other than through their vote intention

Nonparametric Identification

- Average spillover effects (randomization, zero placebo effect)

$$\theta = \sum_{\boldsymbol{x}} \Pr(\boldsymbol{X}_i = \boldsymbol{x} \mid D_i = 1) \cdot \{\mathbb{E}(Y_{i2} \mid D_i = 1, Z_i = 1, \boldsymbol{X}_i = \boldsymbol{x}) - \mathbb{E}(Y_{i2} \mid D_i = 1, Z_i = 2, \boldsymbol{X}_i = \boldsymbol{x})\}$$

- Average contagion and direct effects (latent ignorability, perfect proxy)

$$\tau(z) = \sum_{\boldsymbol{x}} \Pr(\boldsymbol{X}_i = \boldsymbol{x} \mid D_i = 1) \cdot \underbrace{\{m_{\boldsymbol{x}}(1, 2-z) - m_{\boldsymbol{x}}(0, 2-z)\}}_{\text{effect of } Y_{i1} \text{ on } Y_{i2} \text{ given } Z_i} \cdot \underbrace{\{q_{\boldsymbol{x}}(1,1) - q_{\boldsymbol{x}}(1,2)\}}_{\text{effect of } Z_i \text{ on } Y_{i1}},$$

$$\eta(z) = \sum_{\boldsymbol{x}} \left[\Pr(\boldsymbol{X}_i = \boldsymbol{x} \mid D_i = 1) \sum_{y=0}^{1} \underbrace{\{m_{\boldsymbol{x}}(y,1) - m_{\boldsymbol{x}}(y,2)\}}_{\text{effect of } Z_i \text{ on } Y_{i2} \text{ given } Y_{i1}} \cdot q_{\boldsymbol{x}}(y, 2-z)\right]$$

where $m_{\boldsymbol{x}}(y,z) = \mathbb{E}(Y_{i2} \mid D_i = 1, Z_i = z, Y_{i1} = y, \boldsymbol{X}_i = \boldsymbol{x})$ and $q_{\boldsymbol{x}}(y,z) = \Pr(Y_{i1} = y \mid D_i = 1, Z_i = z, \boldsymbol{X}_i = \boldsymbol{x})$.

Identification under Nondifferential Measurement Error

- Consider the following function that controls for measurement error

$$p_{\boldsymbol{x}}(y) \;=\; \Pr(Y_{i1} = y \mid Y_{i1}^{*} = y, G_i = c, \boldsymbol{X}_i = \boldsymbol{x}) \quad \text{for } y = 0, 1$$

  where $p_{\boldsymbol{x}}(y) = 1$ for perfect proxy
- Identification

$$\begin{aligned}
\tau(z) &= \sum_{\boldsymbol{x}} \Pr(\boldsymbol{X}_i = \boldsymbol{x} \mid D_i = 1) r_{\boldsymbol{x}}(z) \\
&\qquad \cdot \{m_{\boldsymbol{x}}(1, 2-z) - m_{\boldsymbol{x}}(0, 2-z)\}\{q_{\boldsymbol{x}}(1,1) - q_{\boldsymbol{x}}(1,2)\}, \\
\eta(z) &= \sum_{\boldsymbol{x}} \Bigg[\Pr(\boldsymbol{X}_i = \boldsymbol{x} \mid D_i = 1) \cdot \sum_{y=0}^{1} \Big(\{m_{\boldsymbol{x}}(y,1) - m_{\boldsymbol{x}}(y,2)\} q_{\boldsymbol{x}}(y, 2-z) \\
&\qquad - \{1 - r_{\boldsymbol{x}}(1-z)\} m_{\boldsymbol{x}}(y, 1+z) \{q_{\boldsymbol{x}}(y,2) - q_{\boldsymbol{x}}(y,1)\} \Big) \Bigg],
\end{aligned}$$

  where $r_{\boldsymbol{x}}(z) =$
  $q_{\boldsymbol{x}}(1, 2-z)\{1 - q_{\boldsymbol{x}}(1, 2-z)\} / \{p_{\boldsymbol{x}}(1) - q_{\boldsymbol{x}}(1, 2-z)\}[p_{\boldsymbol{x}}(0) - \{1 - q_{\boldsymbol{x}}(1, 2-z)\}]$

Sensitivity Analysis

- Attenuation bias:

$$r_{\boldsymbol{x}}(z) = \frac{q_{\boldsymbol{x}}(1, 2-z)}{q_{\boldsymbol{x}}(1, 2-z) - \{1 - p_{\boldsymbol{x}}(0)\}} \cdot \frac{q_{\boldsymbol{x}}(0, 2-z)}{q_{\boldsymbol{x}}(0, 2-z) - \{1 - p_{\boldsymbol{x}}(1)\}} \geq 1.$$

- Sensitivity parameter:

$$p_{\boldsymbol{x}}(1) + p_{\boldsymbol{x}}(0) \quad \geq \quad p \quad \text{for all } x \text{ and } p \in [0, 2]$$

- We examine how the bounds vary as a function of p
- The bounds of causal quantities follow from those of $r_{\boldsymbol{x}}(z)$

$$1 \leq r_{\boldsymbol{x}}(z) \leq \max\left\{\frac{q_{\boldsymbol{x}}(1, 2-z)}{q_{\boldsymbol{x}}(1, 2-z) + p - 2}, \frac{1 - q_{\boldsymbol{x}}(1, 2-z)}{1 - q_{\boldsymbol{x}}(1, 2-z) + (p-2)}\right\}$$

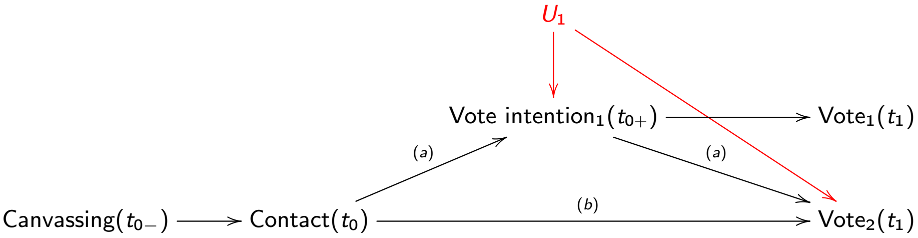
Parametric Modeling Approach

- Latent variable models for the mediator and outcome

$$\begin{aligned} Y_{i1}^*(z) &= 1\{\widetilde{Y}_{i1}(z) > 0\}, \quad \widetilde{Y}_{i1}(z) \;=\; g(z, \boldsymbol{X}_i) + \epsilon_{i1}, \\ Y_{i2}(z, y_1^*) &= 1\{\widetilde{Y}_{i2}(z, y_1^*) > 0\}, \quad \widetilde{Y}_{i2}(z, y_1^*) \;=\; f(z, y_1^*, \boldsymbol{X}_i) + \epsilon_{i2}, \end{aligned}$$

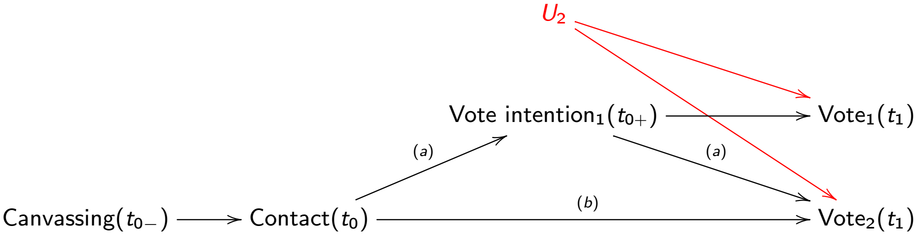
  where $\epsilon_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ for $j = 1, 2 \rightsquigarrow$ EM algorithm

- Parametric sensitivity analysis I



$\rightsquigarrow$ Correlation between ϵ_{i1} and ϵ_{i2}

- Parametric sensitivity analysis II



- Nondifferential measurement error

$$Y_{i1}(z) \;=\; 1\{\widetilde{Y}_{i1}(z) + \zeta_i > 0\} \quad \text{where} \quad \zeta_i \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2).$$

  where $\sigma^2 = 0 \rightsquigarrow$ perfect proxy
- Correlation between η_i and ϵ_{i2}

Empirical Analysis

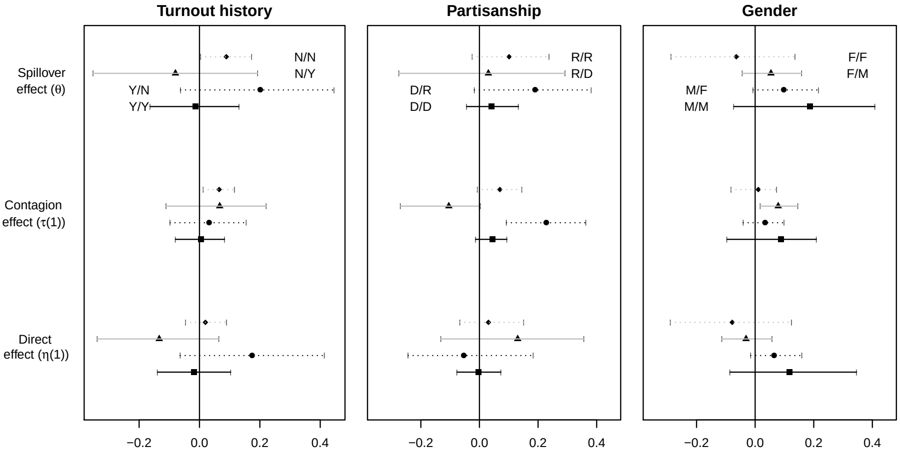
- Pre-treatment covariates: age, age^2, gender, party, prior turnout
- Parametric analysis

$$g(z, \boldsymbol{x}) = \alpha_0 + \alpha_Z z + \boldsymbol{x}^\top \boldsymbol{\alpha}_X + z\boldsymbol{x}^\top \boldsymbol{\alpha}_{ZX},$$

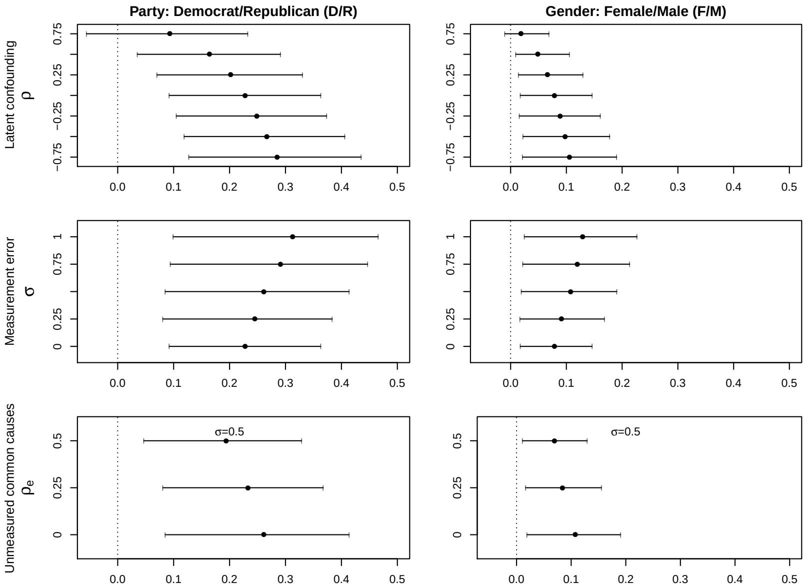
$$f(z, y_1, \boldsymbol{x}) = \beta_0 + \beta_Z z + \beta_Y y_1 + \beta_{ZY} z y_1 + \boldsymbol{x}^\top \boldsymbol{\beta}_X + z\boldsymbol{x}^\top \boldsymbol{\beta}_{ZX} + y_1 \boldsymbol{x}^\top \boldsymbol{\beta}_{YX}.$$

		Denver	Minneapolis	Pooled
Spillover effect	θ	0.069 $(-0.004, 0.139)$	0.073 $(0.007, 0.141)$	0.070 $(0.001, 0.140)$
Contagion effects	$\tau(1)$	0.052 $(0.010, 0.093)$	0.068 $(0.027, 0.115)$	0.059 $(0.014, 0.103)$
	$\tau(0)$	0.054 $(0.011, 0.098)$	0.057 $(0.019, 0.106)$	0.055 $(0.013, 0.098)$
Direct effects	$\eta(1)$	0.016 $(-0.042, 0.073)$	0.016 $(-0.042, 0.073)$	0.015 $(-0.045, 0.075)$
	$\eta(0)$	0.017 $(-0.036, 0.070)$	0.005 $(-0.057, 0.066)$	0.012 $(-0.046, 0, 070)$

Heterogeneous Effects



Sensitivity Analysis



Concluding Remarks

- Application of mediation analysis to spillover effects
- Contagion vs. direct effects
  1. vaccine trials
     - reducing the probability of infection
     - raising the awareness of contagious diseases
  2. moving-to-the-opportunity (MTO) experiment
     - one household's decision to move encourages another household to move
     - voucher offer prompts discussions about the pros and cons of moving
- Better policy recommendation
- Mediator that arises immediately after the administration of treatment $\rightsquigarrow$ no post-treatment confounding

https://imai.fas.harvard.edu