Lecture 16: Moderators, Mediators, and Causal Explanation

POL-GA 1251
Quantitative Political Analysis II
Prof. Cyrus Samii
NYU Politics

April 3, 2018

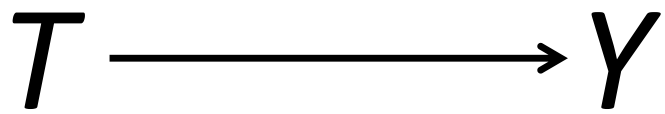
Motivation

- A single effect may give rise to different interpretations.
  - E.g., economic conditions and civil conflict: "greed"? "grievance"? state capacity?
- Different interpretations have different normative implications.
- How can we sort between different interpretations?

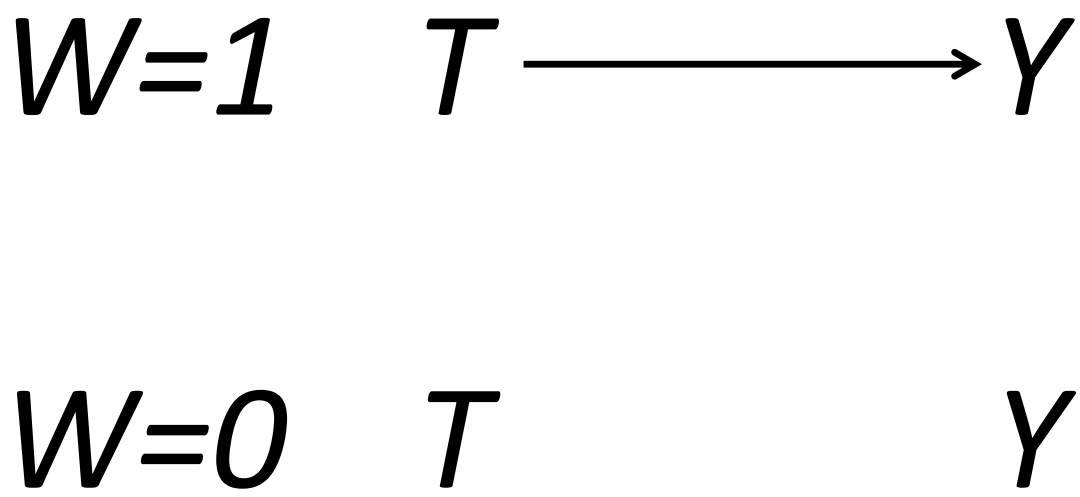
Motivation

- Different interpretations have different observable implications beyond the reduced form cause-effect relationship.
- "$\Rightarrow$ effects should be *stronger* for certain types."
- "$\Rightarrow$ effects should be *transmitted through* certain pathways."

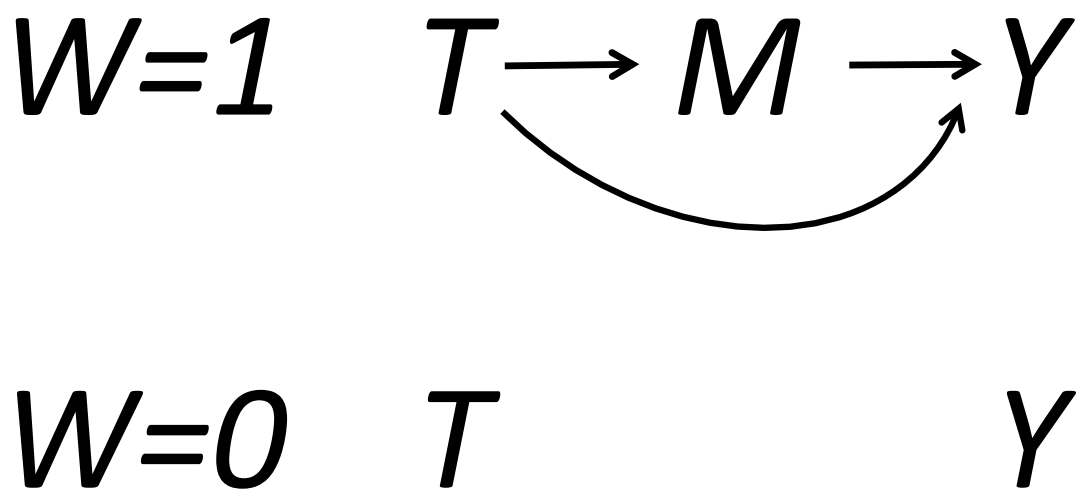
Motivation



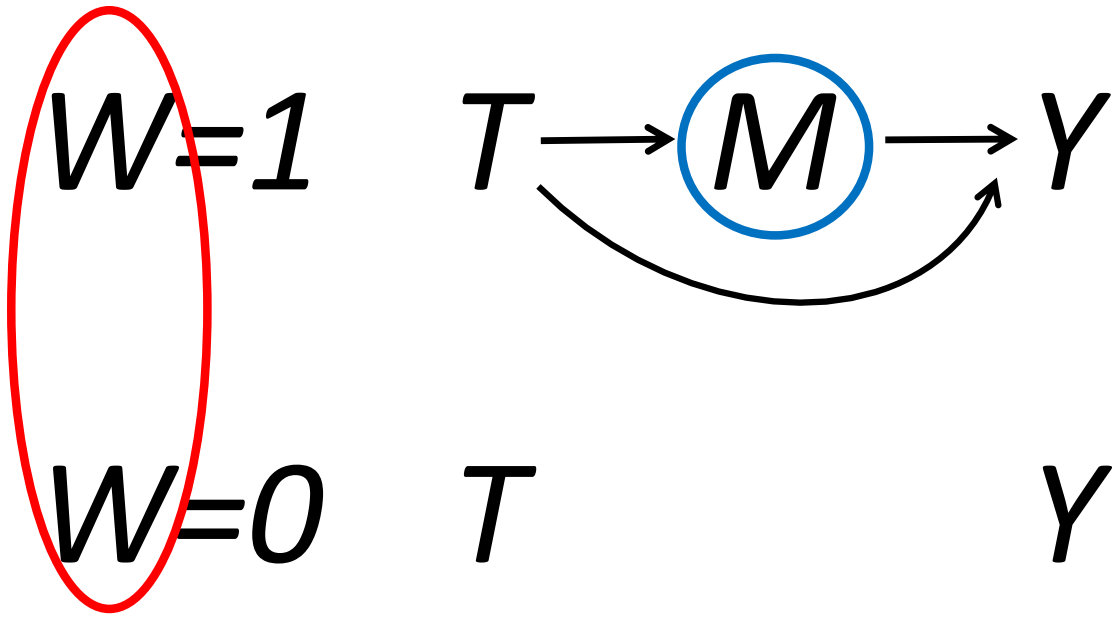
Motivation



Motivation



Motivation



- M =“Mediator”
- W =“Moderator”

Motivation

- Today: mediation, mechanisms, direct effects.
- Next time: moderators and effect heterogeneity.

Mediation

- Following Imai et al. (2010, 2011) notation.

Mediation

- Following Imai et al. (2010, 2011) notation.
- Random sample of size N.

Mediation

- Following Imai et al. (2010, 2011) notation.
- Random sample of size N.
- Treatment T_i assigned randomly conditional on covariates, X_i.

Mediation

- Following Imai et al. (2010, 2011) notation.
- Random sample of size N.
- Treatment T_i assigned randomly conditional on covariates, X_i.
- Mediator M_i with potential values $M_i(t)$ for $T_i = t$.

Mediation

- Following Imai et al. (2010, 2011) notation.
- Random sample of size N.
- Treatment T_i assigned randomly conditional on covariates, X_i.
- Mediator M_i with potential values $M_i(t)$ for $T_i = t$.
- Potential outcomes, $Y_i(t, m)$ for $T_i = t$ and $M_i = m$.

Mediation

- Following Imai et al. (2010, 2011) notation.
- Random sample of size N.
- Treatment T_i assigned randomly conditional on covariates, X_i.
- Mediator M_i with potential values $M_i(t)$ for $T_i = t$.
- Potential outcomes, $Y_i(t,m)$ for $T_i = t$ and $M_i = m$.
- We observe as data $X_i, T_i, M(T_i), Y_i(T_i, M(T_i))$.

Mediation

- Following Imai et al. (2010, 2011) notation.
- Random sample of size N.
- Treatment T_i assigned randomly conditional on covariates, X_i.
- Mediator M_i with potential values $M_i(t)$ for $T_i = t$.
- Potential outcomes, $Y_i(t,m)$ for $T_i = t$ and $M_i = m$.
- We observe as data $X_i, T_i, M(T_i), Y_i(T_i, M(T_i))$.
- For exposition, suppose $T_i = 0, 1$ (everything generalizes).

Mediation

Different types of effects:

- "Total effect."
- "Mediation effect,"
- "Direct effect."

Mediation

Different types of effects:

- "Total effect":

  Gross effect of T on Y.

- Define unit level "total effect" (CME):

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

- Average total effect (ATE):

$$\bar{\tau}(t) = \mathrm{E}\left[\tau_i(t)\right] = \mathrm{E}\left[Y_i(1, M_i(1)) - Y_i(0, M_i(0))\right].$$

Mediation

Different types of effects:

- "Mediation effect":

  Extent that effect of T on Y is transmitted via M.

- Define unit level "causal mediation effect" (CME):

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

- Average casual mediation effect (ACME):

$$\bar{\delta}(t) = \mathrm{E}\left[\delta_i(t)\right] = \mathrm{E}\left[Y_i(t, M_i(1)) - Y_i(t, M_i(0))\right].$$

Mediation

- "Direct effect":

  Extent that effect of T on Y travels through channels that bypass M.

- Define unit level (natural) "direct effect" (DE):

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

- Average direct effect (ADE):

$$\bar{\zeta}(t) = \mathrm{E}\left[\zeta_i(t)\right] = \mathrm{E}\left[Y_i(1, M_i(t)) - Y_i(0, M_i(t))\right].$$

Mediation

- Now,

$$\begin{aligned}
\bar{\tau} &= \mathrm{E}\left[Y_i(1, M_i(1)) - Y_i(0, M_i(0))\right] \\
&= \mathrm{E}\Big[\underbrace{Y_i(1, M_i(1)) - Y_i(1, M_i(0))}_{\delta_i(1)} + \underbrace{Y_i(1, M_i(0)) - Y_i(0, M_i(0))}_{\zeta_i(0)}\Big] \\
&= \mathrm{E}\Big[\underbrace{Y_i(1, M_i(1)) - Y_i(0, M_i(1))}_{\zeta_i(1)} + \underbrace{Y_i(0, M_i(1)) - Y_i(0, M_i(0))}_{\delta_i(0)}\Big]
\end{aligned}$$

$$\Rightarrow \bar{\tau} = \bar{\delta}(t) + \bar{\zeta}(1-t).$$

- Therefore, identification of any two identifies the third.
- We know identifying conditions for $\bar{\tau}$. What about for the others?

Identification

- Imai et al. focus on identifying $\bar{\delta}$, which by implication identifies $\bar{\zeta}$.
- "Sequential ignorability" assumption (SI):

$$\{Y_i(t', m), M_i(t)\} \perp\!\!\!\perp T_i | X_i = x \tag{1}$$

$$Y_i(t', m) \perp\!\!\!\perp M_i(t) | T_i = t, X_i = x \tag{2}$$

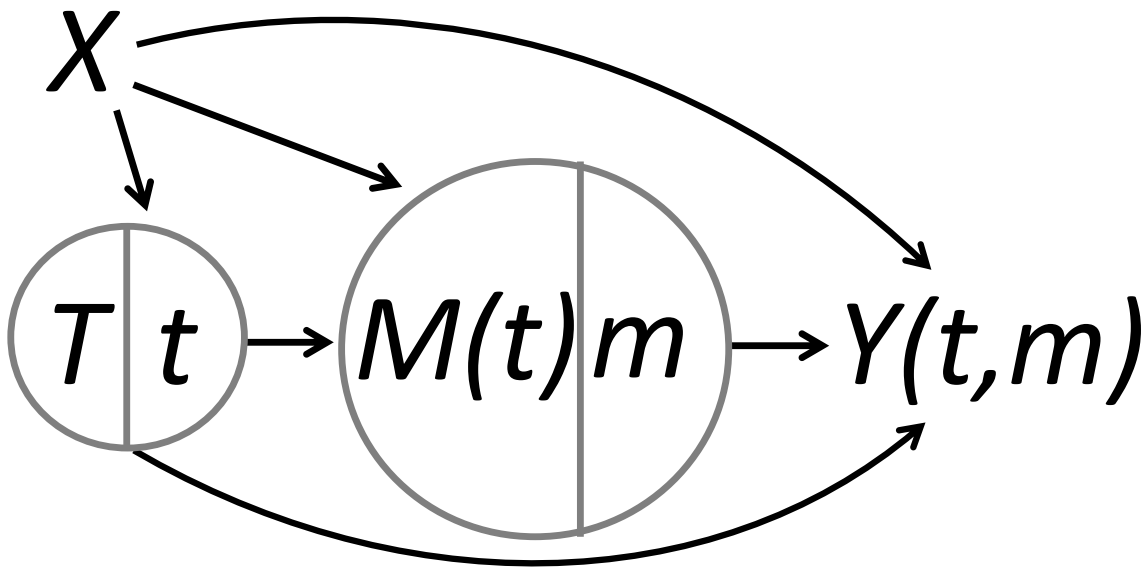
  for $t, t' = 0, 1$, all $x \in \mathscr{X}$, with $0 < \Pr[T_i = 1 | X_i = x]$ and $0 < p(M_i(t) = m | T_i = t, X_i = x)$ for $t = 0, 1$ and all $x \in \mathscr{X}$ and $m \in \mathscr{M}$.
- The various identification strategies we've studied imply (30) and get us $\bar{\tau}$.
- Second part, (30), is what's new.

Mediation

> Thus, the treatment is first assumed to be ignorable given the pre-treatment covariates, and then the mediator variable is assumed to be ignorable *given* the observed value of the treatment as well as the pre-treatment covariates. We emphasize that, unlike the standard sequential ignorability assumption in the literature (e.g., Robins, 1999), the conditional independence given in equation (5) of Assumption 1 must hold without conditioning on the observed values of post-treatment confounders. This issue is discussed further below.

(Imai et al. 2010 p. 55)

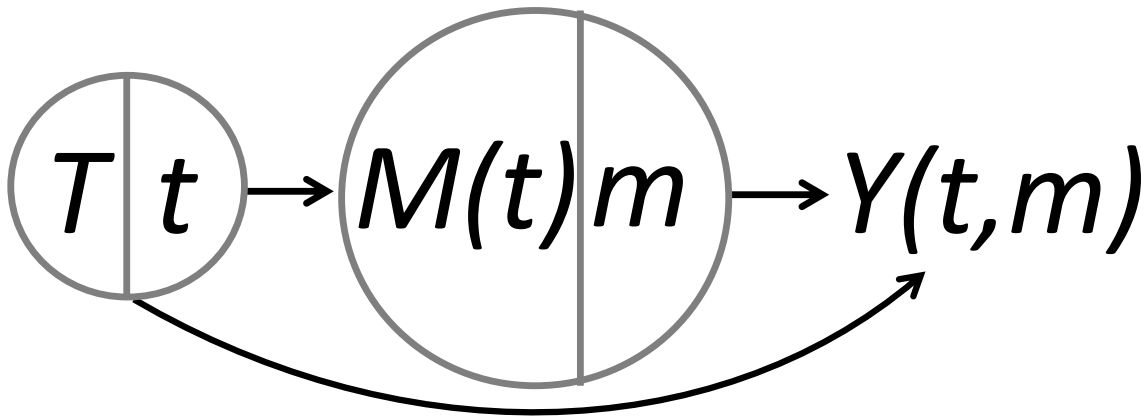
Mediation



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

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

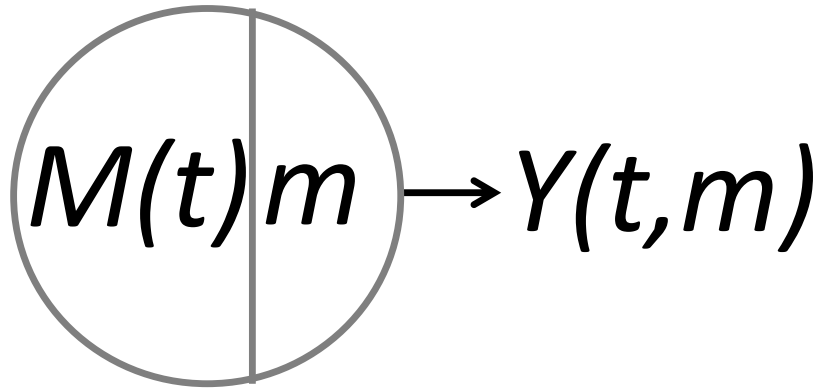
Mediation



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

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

Mediation



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

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

Identification

Lemma:

- We have

$$\{Y_i(t',m), M_i(t)\} \perp\!\!\!\perp T_i | X_i$$
$$\Rightarrow Y_i(t',m) \perp\!\!\!\perp T_i | X_i \text{ and } M_i(t) \perp\!\!\!\perp T_i | X_i$$

As such,

$$p[T_i|X_i] = p[T_i|Y_i(t',m), M_i(t), X_i]$$

and

$$p[T_i|X_i] = p[T_i|M_i(t), X_i]$$

in which case $p[T_i|Y_i(t',m), M_i(t), X_i] = p[T_i|M_i(t), X_i]$ and so

$$T_i \perp\!\!\!\perp Y_i(t',m) | M_i(t), X_i.$$

Identification

$$\text{ACME: } \bar{\delta}(t) = \mathrm{E}\left[Y_i(t, M_i(1)) - Y_i(t, M_i(0))\right].$$

- Under SI, $Y_i(t, M_i(t'))$ counterfactual obeys:

$$\begin{aligned}
\mathrm{E}\left[Y_i(t, M_i(t'))|X_i = x\right] &= \int \mathrm{E}\left[Y_i(t,m)|M_i(t') = m, X_i = x\right] dF_{M_i(t')|X_i=x}(m) \\
(lemma) &= \int \mathrm{E}\left[Y_i(t,m)|M_i(t') = m, T_i = t', X_i = x\right] dF_{M_i(t')|X_i=x}(m) \\
(30) &= \int \mathrm{E}\left[Y_i(t,m)|T_i = t', X_i = x\right] dF_{M_i(t')|T_i=t',X_i=x}(m) \\
(30), (lemma) &= \int \mathrm{E}\left[Y_i(t,m)|T_i = t, X_i = x\right] dF_{M_i(t')|T_i=t',X_i=x}(m) \\
(30) &= \int \mathrm{E}\left[Y_i(t,m)|M_i(t) = m, T_i = t, X_i = x\right] dF_{M_i(t')|T_i=t',X_i=x}(m) \\
&= \int \mathrm{E}\left[Y_i|M_i = m, T_i = t, X_i = x\right] dF_{M_i|T_i=t',X_i=x}(m)
\end{aligned}$$

- That is, weighted average of outcomes in t but with weights from dist. of m in group t'.

Identification

$$\text{ACME: } \bar{\delta}(t) = \mathrm{E}\left[Y_i(t, M_i(1)) - Y_i(t, M_i(0))\right].$$

- Plugging this in, ACME is identified as,

$$\bar{\delta}(t) = \mathrm{E}_X\{\int \mathrm{E}\left[Y_i | M_i = m, T_i = t, X_i = x\right](dF_{M_i|T_i=1,X_i=x}(m) - dF_{M_i|T_i=0,X_i=x}(m))\}$$

- Using outcomes from group t.
- Weighting by distributions of M in groups t and t'
- Taking the difference.

Identification

- When $M_i = 0, 1$:

$$\begin{aligned}
\bar{\delta}(t) &= \mathrm{E}_X\{\mathrm{E}[Y_i|M_i = 0, T_i = t, X](\Pr[M_i = 0|T_i = 1, X] - \Pr[M_i = 0|T_i = 0, X]) \\
&\quad + \mathrm{E}[Y_i|M_i = 1, T_i = t, X](\Pr[M_i = 1|T_i = 1, X] - \Pr[M_i = 1|T_i = 0, X])\} \\
&= \mathrm{E}_X\{\underbrace{(\mathrm{E}[Y_i|M_i = 1, T_i = t, X] - \mathrm{E}[Y_i|M_i = 0, T_i = t, X])}_{\gamma(t,x)} \\
&\quad \times \underbrace{(\Pr[M_i = 1|T_i = 1, X] - \Pr[M_i = 1|T_i = 0, X])}_{\beta(x)}\} \\
&= \beta\gamma(t).
\end{aligned}$$

(where X is shorthand for $X_i = x$)

Identification

$$\text{ADE: } \bar{\zeta}(t) = \mathrm{E}\left[Y_i(1, M_i(t)) - Y_i(0, M_i(t))\right].$$

- By similar arguments

$$\begin{aligned}\bar{\zeta}(t) =& \mathrm{E}_X\Big[\int \{\mathrm{E}\left[Y_i | M_i = m, T_i = 1, X_i = x\right] \\ & - \mathrm{E}\left[Y_i | M_i = m, T_i = 0, X_i = x\right]\} dF_{M_i|T_i=t, X_i=x}(m)\Big]\end{aligned}$$

- Differences across treatment and control, weighting by distribution of M in group t.

Estimation

- Classical approach uses linear structural equation models (Barron & Kenny, 1986):

$$M_i = \alpha + \beta T_i + X_i'\delta + \varepsilon_i$$
$$Y_i = \lambda + \omega T_i + \gamma M_i + X_i'\zeta + \nu_i,$$

  with $\bar{\delta}(0) = \bar{\delta}(1) = \beta\gamma$ and $\bar{\zeta} = \omega$.
- Fit via OLS. Standard errors easy to derive.
- Consistency requires homogeneity, functional form (nb: no interaction), and SI to be true.
- Modest generalization adds the interaction:

$$Y_i = \gamma_b + \omega_b T_i + \gamma_b M_i + X_i'\zeta_b + \kappa T_i M_i + \nu_{bi},$$

  with $\bar{\delta}(t) = \beta(\gamma_b + t\kappa_b)$ and $\bar{\zeta} = \omega_b + \kappa(\alpha + t\beta)$.

Estimation

- Non-parametric or semi-parametric methods can relieve us of homogeneity and functional form assumptions.
- E.g., for $M_i = 0, 1$, $T_i = 0, 1$, within strata defined by $X_i = x$, compute:

$$\hat{\gamma}(t) = \frac{\sum_{i=1}^{N} Y_i I(M_i = 1, T_i = t)}{\sum_{i=1}^{N} I(M_i = 1, T_i = t)} - \frac{\sum_{i=1}^{N} Y_i I(M_i = 0, T_i = t)}{\sum_{i=1}^{N} I(M_i = 0, T_i = t)}$$

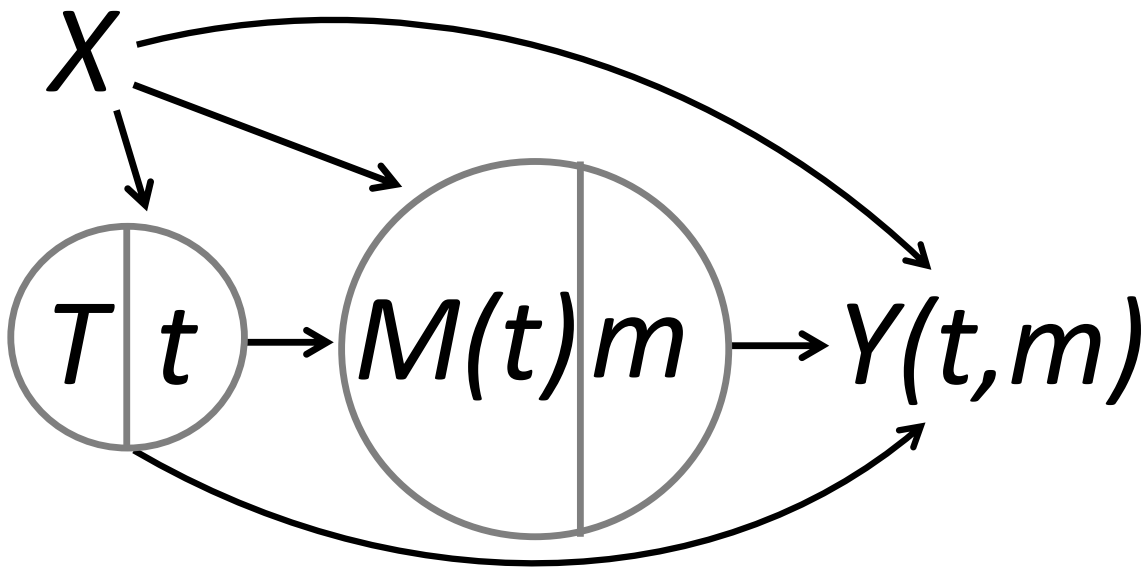
$$\hat{\beta} = \frac{\sum_{i=1}^{N} M_i I(T_i = 1)}{\sum_{i=1}^{N} I(T_i = 1)} - \frac{\sum_{i=1}^{N} M_i I(T_i = 0)}{\sum_{i=1}^{N} I(T_i = 0)}$$

  (could be done with series of simple regressions or a single interacted regression)
- Then $\hat{\bar{\delta}}(t) = \hat{\beta}\hat{\gamma}(t)$. Similar for $\bar{\zeta}(t)$.
- Standard errors from delta method or bootstrap.
- Imai et al. demonstrate approaches for general T_i and M_i.

Discussion

- All seems great. What's the problem?

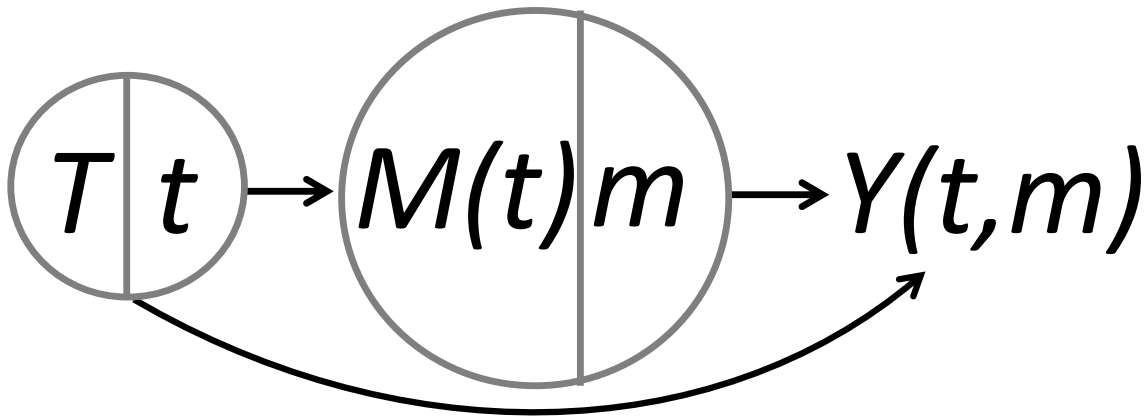
Mediation



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

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

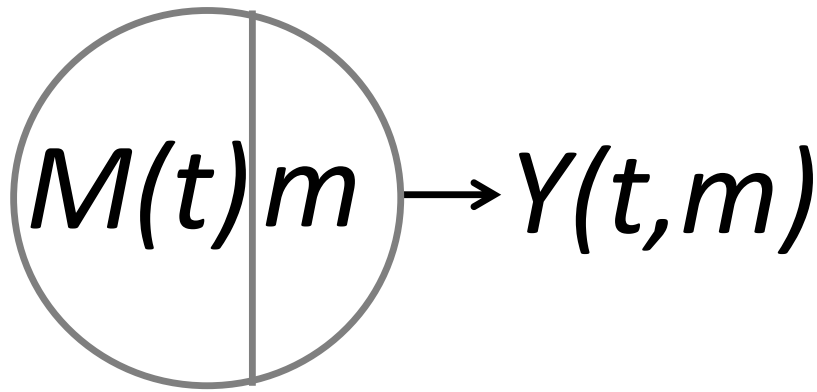
Mediation



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

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

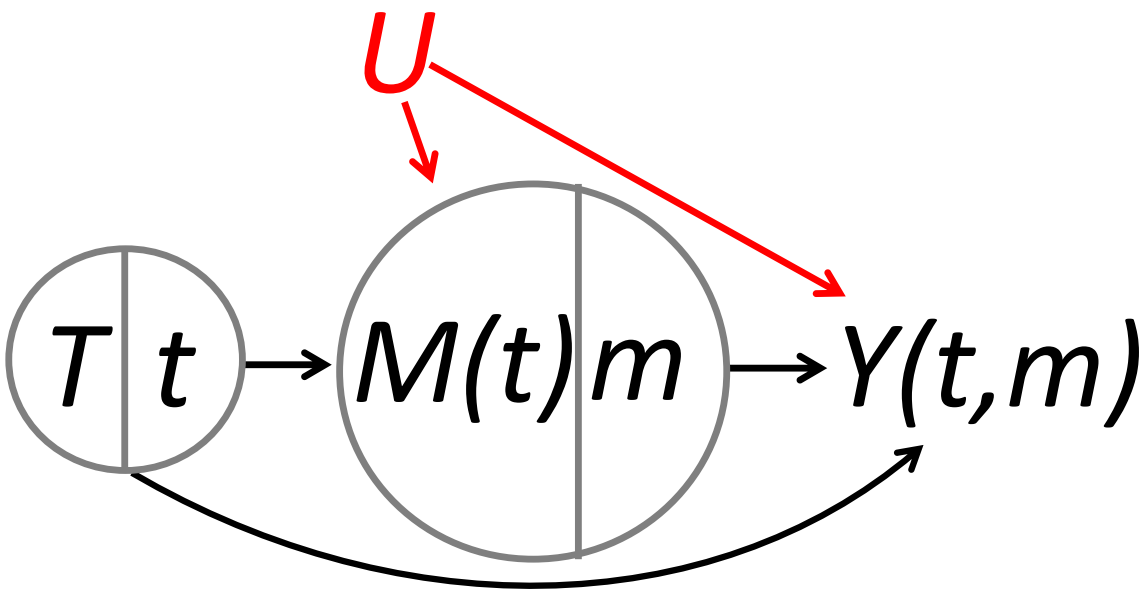
Mediation



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

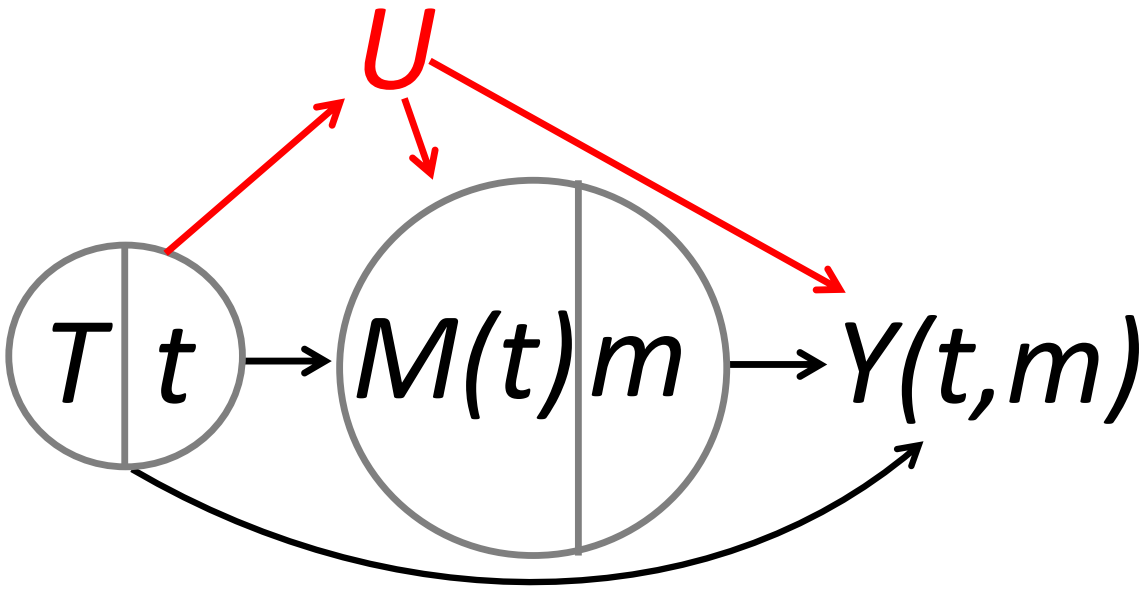
$$Y_i(t', m) \perp\!\!\!\perp M_i(t) | T_i = t, X_i = x$$

Discussion



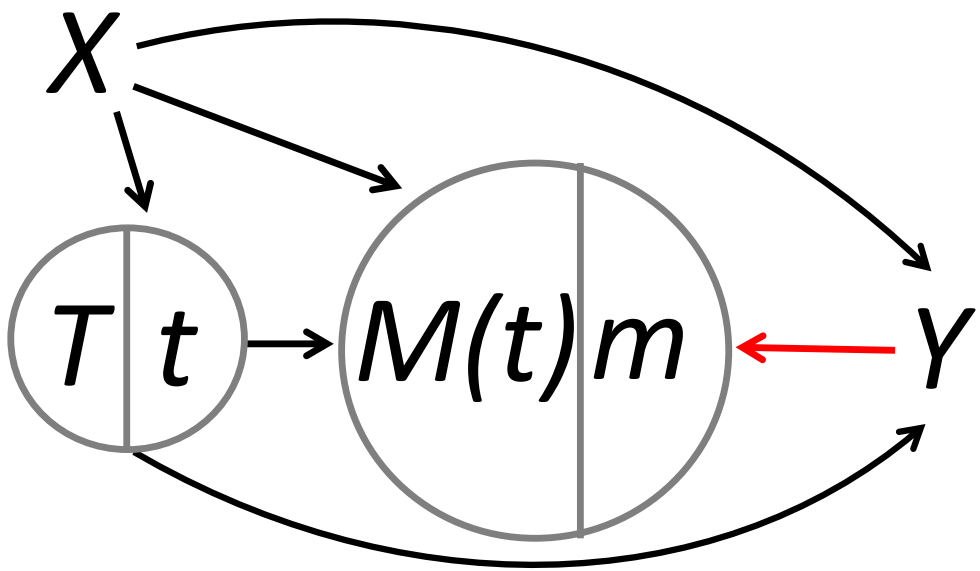
- Randomization or CIA wrt T does not identify mediation effect.
- Need a distinct control strategy to address mediation confounds.

Discussion



- Mediation confounding could be *endogenous* to T.
  - "Effect of ethnic diversity on conflict mediated through economic growth?"
  - Diversity lowers growth directly through communication barriers but also indirectly through mistrust, but the latter also affects conflict in a more direct way...
- If so, controlling only for pre-treatment variables is inadequate.

Discussion



- Do we have the causal ordering correct?
- Cross-sectional outcome data cannot rule out that we have misspecified this.

Discussion

- Typically, neither the data nor design provide immediate ways to judge the plausibility of SI, absence of post-treatment confounding, or causal ordering.
- Has to come from other substantive information.

Sensitivity Analysis

- Imai et al. develop methods for sensitivity analysis.
- They consider sensitivity to pre-treatment confounders that we fail to include in the X_i vector.
- Helpful to a certain extent, but leaves open post-treatment confounding and causal ordering questions.

Experimental Designs for ACME?

- Suppose you randomize T_i on a population and estimate effect on M_i, and then randomize M_i on that population and estimate effects on Y_i.
- Does this identify ACME?

Experimental Designs for ACME?

- Suppose you randomize T_i on a population and estimate effect on M_i, and then randomize M_i on that population and estimate effects on Y_i.
- Does this identify ACME?
- No. E.g., ACME accounts for the fact that...
  - ...for some people, T has no effect on M. For such people, the effect of M on Y is not part of the ACME.
  - ...or, T may have a negative effect on M for some, and positive effect on others.
  - We would need to match up such heterogeneous effects on M to corresponding effects of M on Y to identify the ACME. *This is not straightforward.*
- Experimental designs and associated assumptions are quite subtle (Imai et al. 2011).

Other causal quantities

- Imai et al. focused on the ACME and an average "natural" direct effect (ADE) that allowed the mediator value to vary by i:

$$\bar{\zeta}(t) = \mathrm{E}\left[Y_i(1, M_i(t)) - Y_i(0, M_i(t))\right]$$

- Acharya et al. (2015) consider an average "controlled" direct effect:

$$ACDE(m) = \mathrm{E}\left[Y_i(1, m) - Y_i(0, m)\right].$$

- ACDE considers interventions on both treatment and mediator.
- "ACDE [asks] 'what would the average effect of treatment be if we were to force the mediator be m for all units in the population?' while the [ADE asks] 'what would the average effect of treatment be if we forced every unit to take the value of the mediator it would have taken with no treatment?"'

Other causal quantities

- Acharya et al. (2015) work with a weaker sequential ignorability assumption:

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

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

for $t, t' = 0, 1$, all $x \in \mathscr{X}$, with $0 < \Pr[T_i = 1 | X_i = x]$ and $0 < p(M_i(t) = m | T_i = t, X_i = x, Z_i = z)$ for $t = 0, 1$ and all $x \in \mathscr{X}$ and $m \in \mathscr{M}$.

Other causal quantities

- "In short, if the assumptions hold to separately estimate the effects of T_i and M_i on Y_i, then sequential unconfoundedness holds."
- To make estimation simpler, they also invoke a "no interactions" assumption:

$$\begin{aligned}&\mathrm{E}\left[Y_i(t,m) - Y_i(t,m') | X_i = x, T_i = t, Z_i = z\right] \\ &\qquad = \mathrm{E}\left[Y_i(t,m) - Y_i(t,m') | X_i = x, T_i = t\right] \end{aligned} \tag{3}$$

  for $t = 0, 1$, all $x \in \mathscr{X}$, $m \in \mathscr{M}$, and $z \in \mathscr{Z}$.
- Then, procedure is then to (i) estimate effect of M_i versus $M_i = 0$ conditional on (T_i, X_i, Z_i), (ii) "demediate" Y_i by subtracting off the relevant mediator effect, and then (iii) estimating effect of T_i on the demediated Y_i. This yields the ACDE.
- Without the no interactions assumption, ACDE is still identified, but estimation is more complicated.

Discussion

Yes, But What's the Mechanism? (Don't Expect an Easy Answer)

John G. Bullock and Donald P. Green
Yale University

Shang E. Ha
Brooklyn College of the City University of New York

Enough Already about "Black Box" Experiments: Studying Mediation Is More Difficult than Most Scholars Suppose

Donald P. Green, Shang E. Ha and John G. Bullock

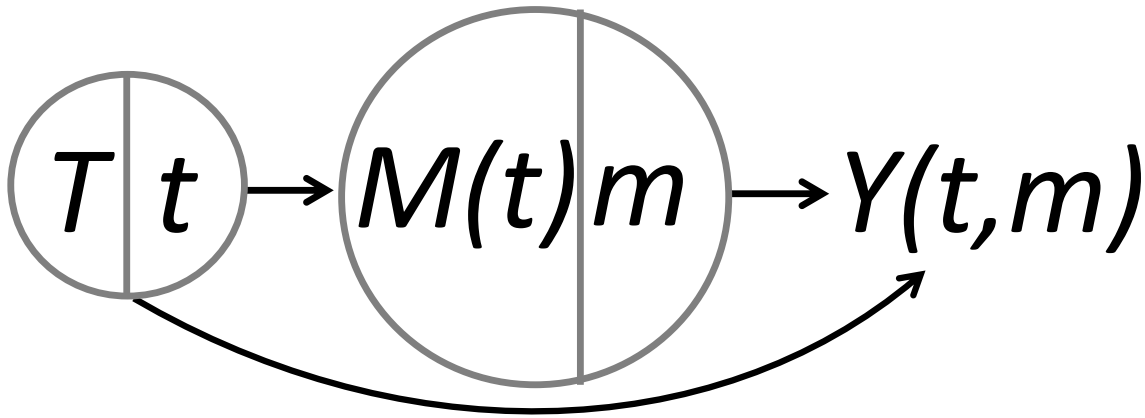
The ANNALS of the American Academy of Political and Social Science 2010 628: 200

Discussion

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

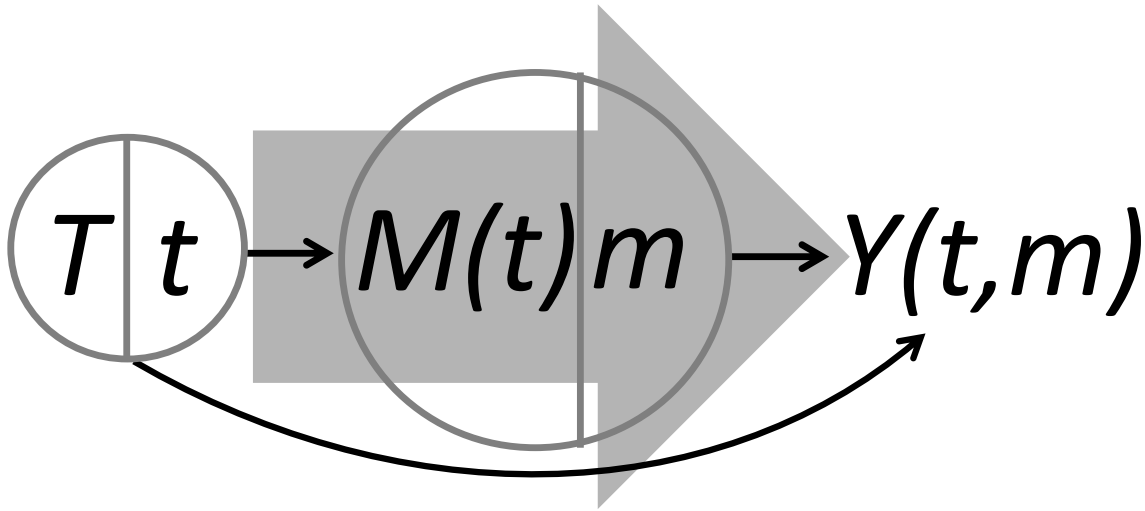
- If $M_i(1) = M_i(0)$ for all i (sharp null), then $\delta_i(t) = 0$ for all i, and so $\bar{\delta}(t) = 0$.
- If sharp null doesn't hold or even stronger, $\mathrm{E}[M_i(1) - M_i(0)] \neq 0$ (reject average null), then we cannot immediately reject $\bar{\delta}(t) \neq 0$.
- This doesn't necessarily tell us anything about the sign of the ACME.
- Nonetheless, these tests on $M_i(t)$ are identified in an experiment or when CIA holds.
- Thus, we can probe *plausibility* of mediating effects be just estimating effects of T on M.

Discussion



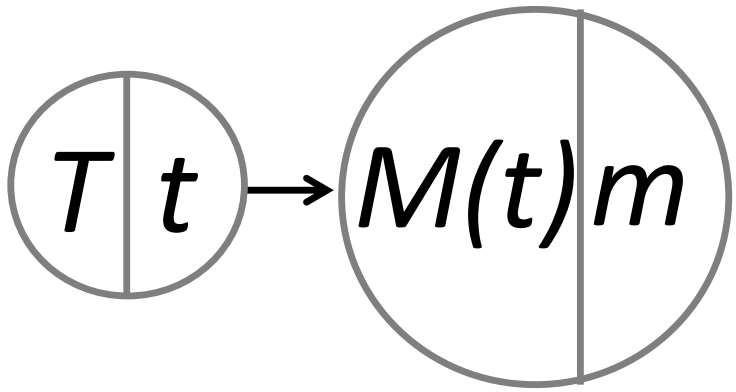
Supposing CIA or random assignment.

Discussion



We can of course estimate the effect of T on Y.

Discussion



And we can of course estimate the effect of T on M.

Discussion

- The effect of T on M allows us to assess plausibility of whether M is mediating the effect of T on Y.
- Bullock et al. and Green et al. distinguish between this kind of *well-identified but inconclusive* analysis and *assumption-laden analysis that provides an illusion of conclusiveness*.