Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies

Teppei Yamamoto
Massachusetts Institute of Technology / Stanford University

Tutorial at the 19th Labor Economics Conference

September 17, 2016

Joint work with
Kosuke Imai   Luke Keele   Dustin Tingley

Project References (click the article titles)

- **General:**
  - Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies. *American Political Science Review*
  - Identifying Mechanisms behind Policy Interventions via Causal Mediation Analysis. *Journal of Policy Analysis and Management*
- **Theory:**
  - Identification, Inference, and Sensitivity Analysis for Causal Mediation Effects. *Statistical Science*
- **Extensions:**
  - Experimental Designs for Identifying Causal Mechanisms. *Journal of the Royal Statistical Society, Series A (with discussions)*
  - Identification and Sensitivity Analysis for Multiple Causal Mechanisms: Revisiting Evidence from Framing Experiments. *Political Analysis*
- **Software:**
  - mediation: R Package for Causal Mediation Analysis. *Journal of Statistical Software*

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?

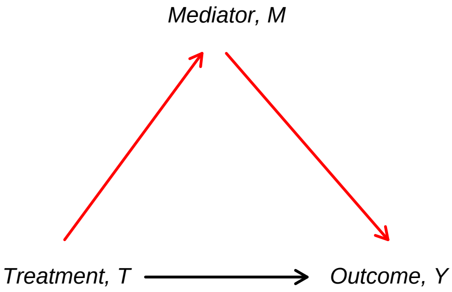
Overview of the Talk

Present a general framework for statistical analysis and research design strategies to understand causal mechanisms

1. Show that the sequential ignorability assumption is required to identify mechanisms even in experiments
2. Offer a flexible estimation strategy under this assumption
3. Introduce a sensitivity analysis to probe this assumption
4. Illustrate how to use statistical software `mediation`
5. Consider research designs that relax sequential ignorability

Causal Mediation Analysis

- Graphical representation



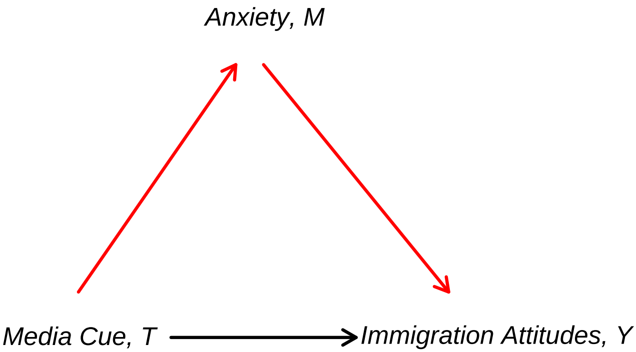
- Goal is to decompose total effect into direct and indirect effects
- Alternative approach: decompose the treatment into different components
- Causal mediation analysis as quantitative process tracing

Example: Psychological Study of Media Effects

- Large literature on how media influences public opinion
- A media framing experiment of Brader *et al.*:
  1. (White) Subjects read a mock news story about immigration:
     - Treatment: Hispanic immigrant in the story
     - Control: European immigrant in the story
  2. Measure attitudinal and behavioral outcome variables:
     - Opinions about increasing or decrease immigration
     - Contact legislator about the issue
     - Send anti-immigration message to legislator
- Why is group-based media framing effective?: role of emotion
- Hypothesis: Hispanic immigrant increases anxiety, leading to greater opposition to immigration

- The primary goal is to examine how, not whether, media framing shapes public opinion

Causal Mediation Analysis in Brader *et al.*



- Does the media framing shape public opinion by making people anxious?
- An alternative causal mechanism: change in beliefs
- Can we identify mediation effects from randomized experiments?

The Standard Estimation Method

- Linear models for mediator and outcome:

$$
\begin{aligned}
Y_i &= \alpha_1 + \beta_1 T_i + \xi_1^\top X_i + \epsilon_{1i} \\
M_i &= \alpha_2 + \beta_2 T_i + \xi_2^\top X_i + \epsilon_{2i} \\
Y_i &= \alpha_3 + \beta_3 T_i + \gamma M_i + \xi_3^\top X_i + \epsilon_{3i}
\end{aligned}
$$

  where X_i is a set of pre-treatment or control variables
  1. Total effect (ATE) is β_1
  2. Direct effect is β_3
  3. Indirect or mediation effect is $\beta_2\gamma$
  4. Effect decomposition: $\beta_1 = \beta_3 + \beta_2\gamma$.

- Some motivating questions:
  1. What should we do when we have interaction or nonlinear terms?
  2. What about other models such as logit?
  3. In general, under what conditions can we interpret β_1 and $\beta_2\gamma$ as causal effects?
  4. What do we really mean by causal mediation effect anyway?

Potential Outcomes Framework of Causal Inference

- Observed data:
  - 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 outcomes model (Neyman, Rubin):
  - Potential mediators: $M_i(t)$ where $M_i = M_i(T_i)$
  - Potential outcomes: $Y_i(t, m)$ where $Y_i = Y_i(T_i, M_i(T_i))$
- Total causal effect:

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

- **Fundamental problem of causal inference**: only one potential outcome can be observed for each i

Back to the Example

- $M_i(1)$:
  - Level of anxiety individual i would report if he reads the story with Hispanic immigrant
- $Y_i(1, M_i(1))$:
  - Immigration attitude individual i would report if he reads the story with Hispanic immigrant and reports the anxiety level $M_i(1)$
- $M_i(0)$ and $Y_i(0, M_i(0))$ are the converse

Causal Mediation Effects

- 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
- Zero treatment effect on mediator $\Longrightarrow$ Zero mediation effect

- Example:
  - Difference in immigration attitudes that is due to the change in anxiety induced by the treatment news story

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:

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

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: $\xi(1, m, m') - \xi(0, m, m')$
- The extent to which controlled direct effects vary by the treatment

What Does the Observed Data Tell Us?

- Recall the Brader *et al.* experimental design:
  1. randomize T_i
  2. measure M_i and then Y_i

- Among observations with $T_i = t$, we observe $Y_i(t, M_i(t))$ but not $Y_i(t, M_i(1-t))$ unless $M_i(t) = M_i(1-t)$
- But we want to estimate

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

- For $t = 1$, we observe $Y_i(1, M_i(1))$ but not $Y_i(1, M_i(0))$
- Similarly, for $t = 0$, we observe $Y_i(0, M_i(0))$ but not $Y_i(0, M_i(1))$

- We have an identification problem $\Longrightarrow$ Need assumptions or better research designs

Counterfactuals in the Example

- A subject viewed the news story with Hispanic immigrant ($T_i = 1$)
- For this person, $Y_i(1, M_i(1))$ is the observed immigration opinion
- $Y_i(1, M_i(0))$ is his immigration opinion in the counterfactual world where he still views the story with Hispanic immigrant but his anxiety is at the same level as if he viewed the control news story
- $Y_i(1, M_i(0))$ **cannot be observed** because $M_i(0)$ is not realized when $T_i = 1$

Sequential Ignorability Assumption

- A sufficient condition for identification: Sequential Ignorability (SI)

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

$$Y_i(t', m) \perp\!\!\!\perp M_i(t) \mid T_i = t, X_i = x \qquad (2)$$

- In words,
  1. T_i is (as-if) randomized conditional on $X_i = x$
  2. $M_i(t)$ is (as-if) randomized conditional on $X_i = x$ and $T_i = t$
- Important limitations:
  1. In a standard experiment, (1) holds but (2) may not
  2. X_i needs to include all confounders
  3. X_i must be pre-treatment confounders $\Longrightarrow$ post-treatment confounder is not allowed
  4. Randomizing M_i via manipulation is not the same as assuming $M_i(t)$ is as-if randomized

Sequential Ignorability in the Standard Experiment

Back to Brader *et al.*:

- Treatment is randomized $\Longrightarrow$ (1) is satisfied
- But (2) may not hold:
  1. Pre-treatment confounder or X_i: state of residence
     those who live in AZ tend to have higher levels of perceived harm and be opposed to immigration
  2. Post-treatment confounder: alternative mechanism
     beliefs about the likely negative impact of immigration makes people anxious

- Pre-treatment confounders $\Longrightarrow$ measure and adjust for them
- Post-treatment confounders $\Longrightarrow$ adjusting is not sufficient

Nonparametric Identification

Under SI, both ACME and average direct effects are nonparametrically identified (can be consistently estimated without modeling assumption)

- ACME $\bar{\delta}(t)$

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

- Average direct effects $\bar{\zeta}(t)$

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

Implies the general mediation formula under any statistical model

Traditional Estimation Methods: LSEM

- Linear structural equation model (LSEM):

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

- Fit two least squares regressions separately
- Use product of coefficients ($\hat{\beta}_2 \hat{\gamma}$) to estimate ACME
- Use asymptotic variance to test significance (Sobel test)

- Under SI and the no-interaction assumption ($\bar{\delta}(1) \neq \bar{\delta}(0)$), $\hat{\beta}_2 \hat{\gamma}$ consistently estimates ACME
- Can be extended to LSEM with interaction terms

- Problem: Only valid for the simplest LSEM

Popular Baron-Kenny Procedure

- The procedure:
  1. Regress Y on T and show a significant relationship
  2. Regress M on T and show a significant relationship
  3. Regress Y on M and T, and show a significant relationship between Y and M

- Problems:
  1. First step can lead to false negatives especially if indirect and direct effects in opposite directions
  2. The procedure only anticipates simplest linear models
  3. Only about statistical significance; effect sizes are more informative

A General Estimation Algorithm

1. Model outcome and mediator
   - Outcome model: $p(Y_i \mid T_i, M_i, X_i)$
   - Mediator model: $p(M_i \mid T_i, X_i)$
   - These models can be of any form (linear or nonlinear, semi- or nonparametric, with or without interactions)
2. Predict mediator for both treatment values ($M_i(1)$, $M_i(0)$)
3. Predict outcome by first setting $T_i = 1$ and $M_i = M_i(0)$, and then $T_i = 1$ and $M_i = M_i(1)$
4. Compute the average difference between two outcomes to obtain a consistent estimate of ACME
5. Simulation-based methods (e.g. bootstrap) to estimate uncertainty

Example: Binary Mediator and Outcome

- Two logistic regression models:

$$\begin{aligned} \Pr(M_i = 1 \mid T_i, X_i) &= \text{logit}^{-1}(\alpha_2 + \beta_2 T_i + \xi_2^\top X_i) \\ \Pr(Y_i = 1 \mid T_i, M_i, X_i) &= \text{logit}^{-1}(\alpha_3 + \beta_3 T_i + \gamma M_i + \xi_3^\top X_i) \end{aligned}$$

- Can't multiply β_2 by γ
- Difference of coefficients $\beta_1 - \beta_3$ doesn't work either

$$\Pr(Y_i = 1 \mid T_i, X_i) = \text{logit}^{-1}(\alpha_1 + \beta_1 T_i + \xi_1^\top X_i)$$

- Can use our algorithm (example: $\mathbb{E}\{Y_i(1, M_i(0))\}$)
  1. Predict $M_i(0)$ given $T_i = 0$ using the first model
  2. Compute $\Pr(Y_i(1, M_i(0)) = 1 \mid T_i = 1, M_i = \widehat{M}_i(0), X_i)$ using the second model

Sensitivity Analysis

- Standard experiments require sequential ignorability to identify mechanisms
- The sequential ignorability assumption is often too strong
- Need to assess the robustness of findings via sensitivity analysis
- Question: How large a departure from the key assumption must occur for the conclusions to no longer hold?
- Parametric sensitivity analysis by assuming

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

  but not

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

- Possible existence of unobserved *pre-treatment* confounder

Parametric Sensitivity Analysis

- Sensitivity parameter: $\rho \equiv \mathrm{Corr}(\epsilon_{i2}, \epsilon_{i3})$
- Sequential ignorability implies $\rho = 0$
- Set ρ to different values and see how ACME changes

- **Result:**

$$\bar{\delta}(0) = \bar{\delta}(1) = \frac{\beta_2 \sigma_1}{\sigma_2}\left\{\tilde{\rho} - \rho\sqrt{(1-\tilde{\rho}^2)/(1-\rho^2)}\right\},$$

where $\sigma_j^2 \equiv \mathrm{var}(\epsilon_{ij})$ for $j = 1, 2$ and $\tilde{\rho} \equiv \mathrm{Corr}(\epsilon_{i1}, \epsilon_{i2})$.

- When do my results go away completely?
- $\bar{\delta}(t) = 0$ if and only if $\rho = \tilde{\rho}$
- Easy to estimate from the regression of Y_i on T_i:

$$Y_i = \alpha_1 + \beta_1 T_i + \epsilon_{i1}$$

Interpreting Sensitivity Analysis with R squares

- Interpreting ρ: how small is too small?
- An unobserved (pre-treatment) confounder formulation:

$$\epsilon_{i2} = \lambda_2 U_i + \epsilon'_{i2} \quad \text{and} \quad \epsilon_{i3} = \lambda_3 U_i + \epsilon'_{i3}$$

- How much does U_i have to explain for our results to go away?
- Sensitivity parameters: R squares
  1. Proportion of previously unexplained variance explained by U_i

$$R_M^{2*} \equiv 1 - \frac{\text{var}(\epsilon'_{i2})}{\text{var}(\epsilon_{i2})} \quad \text{and} \quad R_Y^{2*} \equiv 1 - \frac{\text{var}(\epsilon'_{i3})}{\text{var}(\epsilon_{i3})}$$

  2. Proportion of original variance explained by U_i

$$\widetilde{R}_M^2 \equiv \frac{\text{var}(\epsilon_{i2}) - \text{var}(\epsilon'_{i2})}{\text{var}(M_i)} \quad \text{and} \quad \widetilde{R}_Y^2 \equiv \frac{\text{var}(\epsilon_{i3}) - \text{var}(\epsilon'_{i3})}{\text{var}(Y_i)}$$

- Then reparameterize ρ using (R_M^{2*}, R_Y^{2*}) (or $(\widetilde{R}_M^2, \widetilde{R}_Y^2)$):

$$\rho \quad = \quad \text{sgn}(\lambda_2\lambda_3) R_M^* R_Y^* \; = \; \frac{\text{sgn}(\lambda_2\lambda_3)\widetilde{R}_M \widetilde{R}_Y}{\sqrt{(1-R_M^2)(1-R_Y^2)}},$$

  where R_M^2 and R_Y^2 are from the original mediator and outcome models

- $\text{sgn}(\lambda_2\lambda_3)$ indicates the direction of the effects of U_i on Y_i and M_i

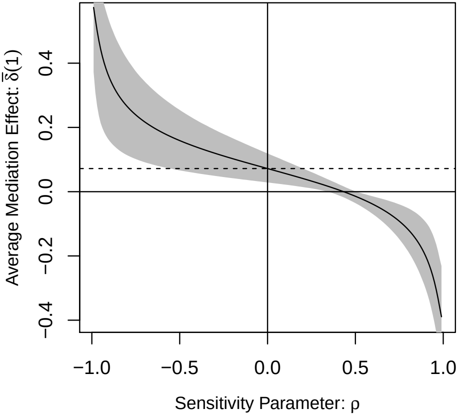
- Set (R_M^{2*}, R_Y^{2*}) (or $(\widetilde{R}_M^2, \widetilde{R}_Y^2)$) to different values and see how mediation effects change

Reanalysis: Estimates under Sequential Ignorability

- Original method: Product of coefficients with the Sobel test
  — Valid only when both models are linear w/o T–M interaction (which they are not)
- Our method: Calculate ACME using our general algorithm

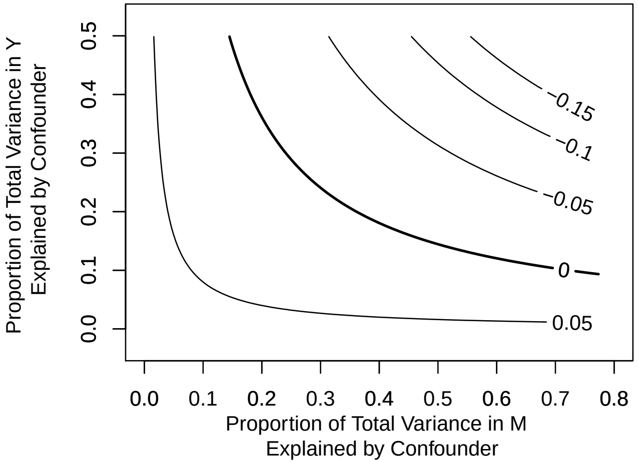
Outcome variables	Product of Coefficients	Average Causal Mediation Effect (δ)
Decrease Immigration	.347	.105
$\bar{\delta}(1)$	[0.146, 0.548]	[0.048, 0.170]
Support English Only Laws	.204	.074
$\bar{\delta}(1)$	[0.069, 0.339]	[0.027, 0.132]
Request Anti-Immigration Information	.277	.029
$\bar{\delta}(1)$	[0.084, 0.469]	[0.007, 0.063]
Send Anti-Immigration Message	.276	.086
$\bar{\delta}(1)$	[0.102, 0.450]	[0.035, 0.144]

Reanalysis: Sensitivity Analysis w.r.t. ρ



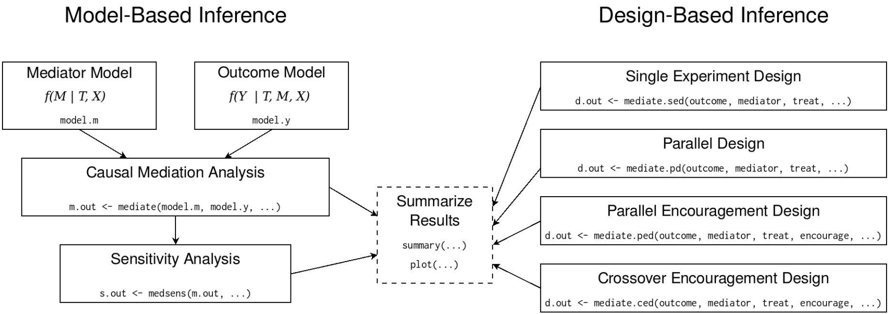
- ACME $>$ 0 as long as the error correlation is less than 0.39 (0.30 with 95% CI)

Reanalysis: Sensitivity Analysis w.r.t. $\tilde{R}^2_M$ and $\tilde{R}^2_Y$



- An unobserved confounder can account for up to 26.5% of the variation in both Y_i and M_i before ACME becomes zero

R Package: `mediation`



Implementation Examples

1. Fit models for the mediator and outcome variable and store these models

```
> m <- lm(Mediator ~ Treat + X)
> y <- lm(Y ~ Treat + Mediator + X)
```

2. Mediation analysis: Feed model objects into the `mediate()` function. Call a summary of results

```
> m.out<-mediate(m, y, treat = "Treat",
                 mediator = "Mediator")
> summary(m.out)
```

3. Sensitivity analysis: Feed the output into the `medsens()` function. Summarize and plot

```
> s.out <- medsens(m.out)
> summary(s.out)
> plot(s.out, "rho")
> plot(s.out, "R2")
```

Data Types Available via **mediation**

	Outcome Model Types						
Mediator Model Types	Linear	GLM	Ordered	Censored	Quantile	GAM	Survival
Linear (`lm`/`lmer`)	✓	✓	✓*	✓	✓	✓*	✓
GLM (`glm`/`bayesglm`/`glmer`)	✓	✓	✓*	✓	✓	✓*	✓
Ordered (`polr`/`bayespolr`)	✓	✓	✓*	✓	✓	✓*	✓
Censored (tobit via `vglm`)	-	-	-	-	-	-	-
Quantile (`rq`)	✓*	✓*	✓*	✓*	✓*	✓*	✓
GAM (`gam`)	✓*	✓*	✓*	✓*	✓*	✓*	✓*
Survival (`survreg`)	✓	✓	✓*	✓	✓	✓*	✓

Types of Models That Can be Handled by `mediate`. Stars (*) indicate the model combinations that can only be estimated using the nonparametric bootstrap (i.e. with `boot = TRUE`).

Additional Features

- Treatment/mediator interactions, with formal statistical tests
- Treatment/mediator/pre-treatment interactions and reporting of quantities by pre-treatment values
- Factoral, continuous treatment variables
- Cluster standard errors/adjustable CI reporting/p-values
- Support for multiple imputation
- Multiple mediators
- Multilevel mediation

See our JSS tutorial paper: **here.**

Stata package with limited functionalities available:

```
ssc install mediation
```

See: Hicks and Tingley, Causal Mediation Analysis. *Stata Journal*.

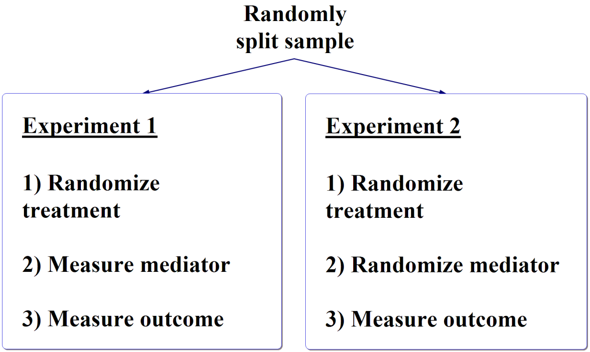
Beyond Sequential Ignorability

- 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
- All proposed designs preserve the ability to estimate the ACME under the SI assumption
- Trade-off: statistical power

- These experimental designs can then be extended to natural experiments in observational studies

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

Why Do We Need No-Interaction Assumption?

- Numerical Example:

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 $\bar{\delta}(t) = -0.2$
- The Problem: Causal effect heterogeneity
  - T increases M only *on average*
  - M increases Y only *on average*
  - $T - M$ interaction: Many of those who have a positive effect of T on M have a negative effect of M on Y (first row)
- A solution: sensitivity analysis (see Imai and Yamamoto, 2013)
- Pitfall of “mechanism experiments” or “causal chain approach”

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?

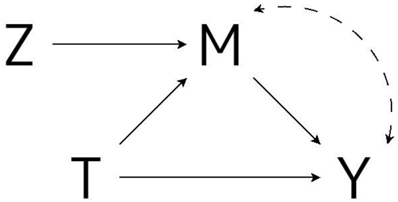
- A brain region known to be related to fairness becomes more active when unfair offer received (single experiment design)

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)

Extension: Parallel Encouragement Design

- Direct manipulation of mediator is difficult in most situations
- Use an instrumental variable approach:



- Advantage: allows for unobserved confounder between M and Y
- Key Assumptions:
  1. Z is randomized or as-if random
  2. No direct effect of Z on Y (a.k.a. exclusion restriction)

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

- The no carryover effect assumption: Round 1 must not affect Round 2
- Strong, but can be made plausible by design

Example: Labor Market Discrimination

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

- Assumption: their different names do not change the perceived qualifications of applicants
- Under this assumption, the direct effect can be interpreted as blunt racial discrimination

Multiple Mediators

- Quantity of interest = The average indirect effect with respect to M
- W represents the alternative observed mediators
- Left: Assumes independence between the two mechanisms
- Right: Allows M to be affected by the other mediators W
- Applied work often assumes the independence of mechanisms
- Under this independence assumption, one can apply the same analysis as in the single mediator case
- For causally dependent mediators, we must deal with the heterogeneity in the $T \times M$ interaction as done under the parallel design $\Longrightarrow$ sensitivity analysis

Unpacking the Standard Path-Analytic Approach

- Applied social scientists often use the following model:

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

- The mediation effects are then estimated as $\hat{\beta}_M \hat{\gamma}$ for M and $\hat{\beta}_W \hat{\theta}$ for W
- We can show that these are consistent for $\bar{\delta}_i^M$ and $\bar{\delta}_i^W$ under the above assumption and linearity
- However, because of the assumed independence between mechanisms, analyzing one mechanism at a time will also be valid, e.g.,

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

Identification of Causally Related Mechanisms

- Consider the (weak) sequential ignorability assumption:

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

  for any t, m, w, x.

- Unconfundedness of M_i conditional on both pre-treatment (X_i) and observed post-treatment (W_i) confounders
- Corresponds to sequential randomization unlike Assumption 1
- The no $T \times M$ interaction assumption required for the identification of $\bar{\delta}(t)$ under Assumption 2:

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

The Proposed Framework

- Problem: The no interaction assumption is often too strong (e.g. Does the effect of perceived issue importance invariant across frames?)
- We use a varying-coefficient linear structural equations model to:
  1. Allow for homogeneous interaction for point identification
  2. Develop a sensitivity analysis in terms of the degree of heterogeneity in the interaction effect
- Consider the following 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 are allowed to vary arbitrarily across units

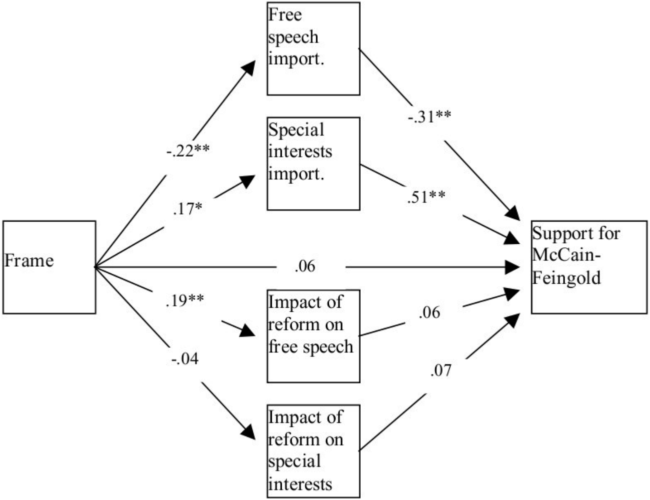
An Example: Framing Experiment

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

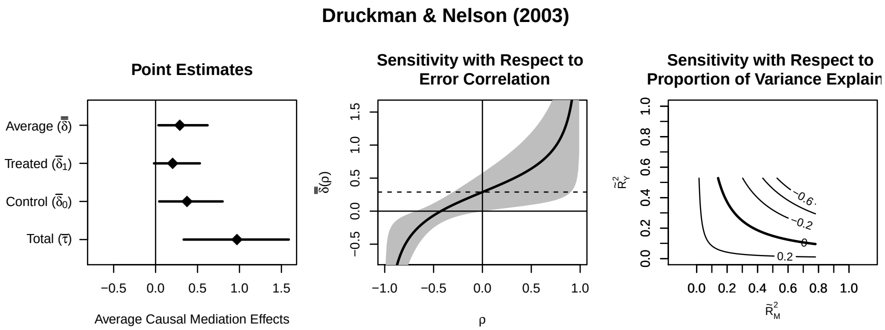
- Treatment: News paper article on a proposed election campaign finance reform, emphasizing either its positive or negative aspect
- Outcome: Support for the proposed reform
- Primary mediator: Perceived importance of free speech
- Alternative (confounding) mediator: Belief about the impact of the proposed reform
- Original analysis finds the importance mechanism to be significant, implicitly assuming its independence from beliefs

Original Analysis Assumes Independent Mechanisms

Druckman and Nelson, p.738

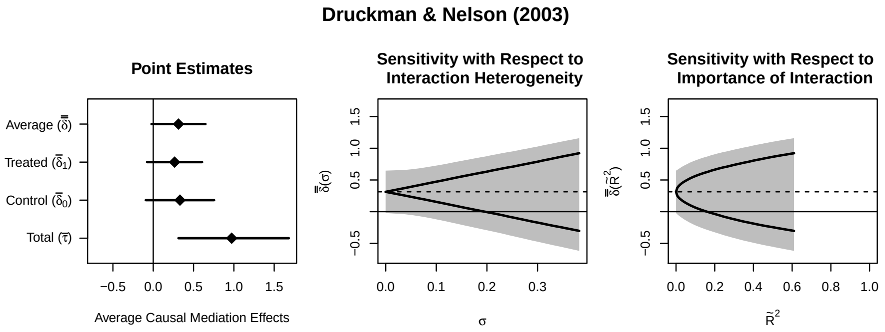


Analysis with the Independence Assumption



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

Analysis without the Independence Assumption



- Similar results with slightly wider CI ($[-0.021, 0.648]$)
- Lower bound on $\bar{\delta}$ is zero when $\sigma = 0.195$, or 51% of its upper bound
- This translates to the interaction heterogeneity explaining 15.9% of the variance of the outcome variable

Concluding Remarks

- Even in a randomized experiment, a strong assumption is needed to identify causal mechanisms
- However, progress can be made toward this fundamental goal of scientific research with modern statistical tools
- A general, flexible estimation method is available once we assume sequential ignorability
- Sequential ignorability can be probed via sensitivity analysis
- More credible inferences are possible using clever experimental designs
- Insights from new experimental designs can be directly applied when designing observational studies
- Multiple mediators require additional care when they are causally dependent

Thank You!

Email your questions and suggestions to: teppei@mit.edu