Causal Mechanisms Short Course Part II:
Analyzing Mechanisms with Experimental and Observational Data

Teppei Yamamoto
Massachusetts Institute of Technology

March 24, 2012
Frontiers in the Analysis of Causal Mechanisms
Pre-Conference Short Course
Overview

Our central question:
How can we learn about causal mechanisms from data?

In this tutorial, you will learn:

- A quantitative definition of causal mechanisms
- Assumptions needed to identify a causal mechanism from data
- A general procedure to estimate a causal mechanism (given the assumptions)
- Methods for analyzing sensitivity to the violation of the assumptions
- Experimental designs to identify mechanisms with weaker assumptions
This tutorial is based on:

All methods can be implemented by the R package *mediation*

Further details are given in the (more technical) papers listed at the end.
What Is a Causal Mechanism?

- Mechanisms as causal pathways
- Example: Soil fumigants increase farm crops by reducing eel-worms (Cochran)
- Causal mediation analysis

\[
\begin{align*}
\text{Mediator, } M \\
\text{Treatment, } T \quad \rightarrow \quad \text{Outcome, } Y
\end{align*}
\]

- Quantities of interest: Direct and indirect effects
- Fast growing literature in the past 10–20 years
Potential Outcomes Framework

Framework: Potential outcomes (Neyman 1923; Rubin 1974)

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

- In a standard experiment, only one potential outcome can be observed for each \( i \)
- Moreover, some potential outcomes can never be observed: \( Y_i(t, M_i(t')) \) where \( t \neq t' \)
Causal Mediation Effects

- Total causal effect:
  \[ \tau_i \equiv Y_i(1, M_i(1)) - Y_i(0, M_i(0)) \]

- Causal mediation effects (a.k.a. natural indirect effects):
  \[ \delta_i(t) \equiv Y_i(t, M_i(1)) - Y_i(t, M_i(0)) \]

- Definition originates in Robins and Greenland (1992) and Pearl (2001)
- Causal effect of the change in \( M_i \) on \( Y_i \) that would be induced by treatment
- Represents the mechanism through \( M_i \)
Total Effect = Indirect Effect + Direct Effect

- Natural 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 \( T_i = t \)
- Represents all mechanisms other than through \( M_i \)

Cf. Controlled direct effect:

\[ \xi_i(m) \equiv Y_i(1, m) - Y_i(0, m) \]

- Causal effect of \( T_i \) on \( Y_i \) when mediator is manipulated at a fixed value \( m \) (regardless of unit \( i \)'s natural response to \( T_i \))

- Total effect = mediation (indirect) effect + direct effect:
  \[ \tau_i = \delta_i(t) + \zeta_i(1 - t) = \frac{1}{2} \{ \delta_i(0) + \delta_i(1) + \zeta_i(0) + \zeta_i(1) \} \]
What Do the Observed Data Tell Us?

- Quantity of Interest: Average causal mediation effects
  \[ \bar{\delta}(t) \equiv \mathbb{E}(\delta_i(t)) = \mathbb{E}\{Y_i(t, M_i(1)) - Y_i(t, M_i(0))\} \]

- Average direct effects (\(\bar{\zeta}(t)\)) are defined similarly

- Problem: \(Y_i(t, M_i(t))\) is observed but \(Y_i(t, M_i(t'))\) can never be observed

- We have an identification problem
  \[ \Rightarrow \text{Need additional assumptions to make progress} \]
Identification under Standard Research Design

- Standard experiment: Randomize $T_i$ and measure $M_i$ and $Y_i$
- An identification assumption: **Sequential Ignorability**

\[
\{Y_i(t', m), M_i(t)\} \perp\!
\!
\!
\!
\!
\!
\!
\!
\!
\perp T_i \mid X_i = x \quad (1)
\]

\[
Y_i(t', m) \perp\!
\!
\!
\!
\!
\!
\!
\!
\perp M_i(t) \mid T_i = t, X_i = x \quad (2)
\]

- (1) is guaranteed to hold in standard experiments
- (2) does **not** hold if there exist:
  - unobserved pre-treatment $M$–$Y$ confounders, or
  - any post-treatment $M$–$Y$ confounding, even if observed
- In observational studies, neither (1) nor (2) is guaranteed
- Alternative assumptions: Robins, Pearl, Petersen *et al.*, VanderWeele, etc.

**Theorem** (Imai et al. 2010): Under sequential ignorability, ACME and average direct effects are **nonparametrically identified** (= estimable from data)
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. Monte-Carlo or bootstrapping to estimate uncertainty

(Alternative procedures: Inverse probability weighting, conditional mean models, etc.)
The simplest example: Linear structural equation model:

\[ M_i = \alpha_2 + \beta_2 T_i + \xi_2^TX_i + \epsilon_{i2}, \]
\[ Y_i = \alpha_3 + \beta_3 T_i + \gamma M_i + \xi_3^TX_i + \epsilon_{i3}. \]

Traditional procedure (Baron & Kenny 1986):

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

Under SI and the no-interaction assumption (\(\bar{\delta}(1) \neq \bar{\delta}(0))\), this procedure consistently estimates ACME.

But only if the model is correct! (linearity, unit homogeneity)

The proposed algorithm generalizes this to any types of models.
Example: Anxiety, Group Cues and Immigration

Brader, Valentino & Suhat (2008, AJPS)

- How and why do ethnic cues affect immigration attitudes?
- Theory: Anxiety transmits the effect of cues on attitudes

\[ \text{ACME} = \text{Average difference in immigration attitudes due to the change in anxiety induced by the media cue treatment} \]

\[ \text{Sequential ignorability} = \text{No unobserved covariate affecting both anxiety and immigration attitudes} \]
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

<table>
<thead>
<tr>
<th>Outcome variables</th>
<th>Product of Coefficients</th>
<th>Average Causal Mediation Effect ($\delta$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decrease Immigration $\bar{\delta}(1)$</td>
<td>0.347</td>
<td>0.105</td>
</tr>
<tr>
<td>Support English Only Laws $\bar{\delta}(1)$</td>
<td>0.204</td>
<td>0.074</td>
</tr>
<tr>
<td>Request Anti-Immigration Information $\bar{\delta}(1)$</td>
<td>0.277</td>
<td>0.029</td>
</tr>
<tr>
<td>Send Anti-Immigration Message $\bar{\delta}(1)$</td>
<td>0.276</td>
<td>0.086</td>
</tr>
</tbody>
</table>

Note: $\bar{\delta}(1)$ refers to the average causal mediation effect.
Beyond Sequential Ignorability

- Assumption (2) is too strong in most scenarios
- Can we go beyond just making this assumption?

- **Sensitivity analysis**: Assess the robustness of the estimates to the violation of sequential ignorability
- 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
  \[
  Y_i(t', m) \perp \perp M_i(t) \mid T_i = t, X_i = x
  \]
- Addresses the possible existence of unobserved *pre-treatment* confounders
- But not post-treatment confounders
- An alternative approach: VanderWeele (2010)
Parametric Sensitivity Analysis

- Assume a linear structural equations model:
  \[
  M_i = \alpha_2 + \beta_2 T_i + \xi_2^T X_i + \epsilon_{i2},
  \]
  \[
  Y_i = \alpha_3 + \beta_3 T_i + \gamma M_i + \xi_3^T X_i + \epsilon_{i3}.
  \]

- **Sensitivity parameter**: \( \rho \equiv \text{Corr}(\epsilon_{i2}, \epsilon_{i3}) \)

**Theorem** (Imai et al. 2010): ACME is identified given \( \rho \)

- Set \( \rho \) to different values and see how ACME changes
- Sequential ignorability implies \( \rho = 0 \)
- Extension to nonlinear models (e.g. logistic regression for binary outcomes)
Anxiety Example: Sensitivity Analysis w.r.t. $\rho$

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

- Interpreting $\rho$: 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 the variances of $M_i$ and $Y_i$ for our results to go away?
- Reparameterize ACME using coefficients of determination ($R^2*$ and $\tilde{R}^2$)
An anxiety example: Sensitivity Analysis with respect to $\tilde{R}^2_M$ 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.
Experimental Designs for Identifying Mechanisms

- Sensitivity analysis may be unsatisfactory
- What if we get rid of the assumption altogether?
- Under a standard design, even the sign of ACME is unidentified (Sjölander)
- Can we do any better?

- Use alternative experimental designs for more credible yet powerful inference
- Designs feasible when the mediator can be directly or indirectly manipulated
- Experiments also serve as templates for observational studies
Parallel Design

- Must assume consistency (i.e. no direct effect of manipulation on outcome)
- More informative than standard single experiment
- If we assume no $T-M$ interaction, ACME is point identified
Parallel Encouragement Design

- Direct manipulation of the mediator is often infeasible
- Even if feasible, more subtle form of intervention may be preferred to assure consistency

**Randomly split sample**

**Experiment 1**
1) Randomize treatment
2) Measure mediator
3) Measure outcome

**Experiment 2**
1) Randomize treatment
2) Encourage mediator
3) Measure outcome

- **Parallel encouragement design**: Randomly encourage subjects to take particular values of the mediator
- **Standard instrumental variable** assumptions (Angrist et al.)
Crossover Designs

- 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

**Crossover encouragement design:**

1. Round 1: Conduct a standard experiment
2. Round 2: Same as crossover, except encourage subjects to take the mediator values

- Both must assume no carryover effect
Even in a randomized experiment, a strong assumption is needed to identify causal mechanisms

Analyzing mechanisms is, therefore, not so easy!

Under the identification assumption, a general estimation procedure is available for various types of statistical models

The violation of the assumption can be addressed by:
- Analyzing sensitivity with respect to key assumptions
- Creative research designs to avoid strong assumptions

Therefore, progress can still be made!

Several software implementations to make things easy:
- R and Stata (next session)
- SAS and SPSS (session after next)


“Causal Mediation Analysis Using R.” *Advances in Social Science Research Using R*