Mediation analysis

Xinzhou Guo

HKUST

(Credited to Zhichao Jiang)

May 1, 2024

$ACE = E(Y(1) - Y(0))$

① Definition of causal mechanism: mediator.
NDE, NIE, $ACE = NDE + NIE$

② Identifiability

③ Estimation and Inference

Causal mechanisms

- Scientists care about causal mechanisms, not just causal effects
- Randomized experiments often only determine whether the treatment causes changes in the outcome; Not how and why the treatment effects the outcome
- Common criticism of experiments and statistics: black box view of causality
- Mediation analysis studies the extent to which an effect is mediated through a particular pathway and to which the effect of a treatment on the outcome operates directly

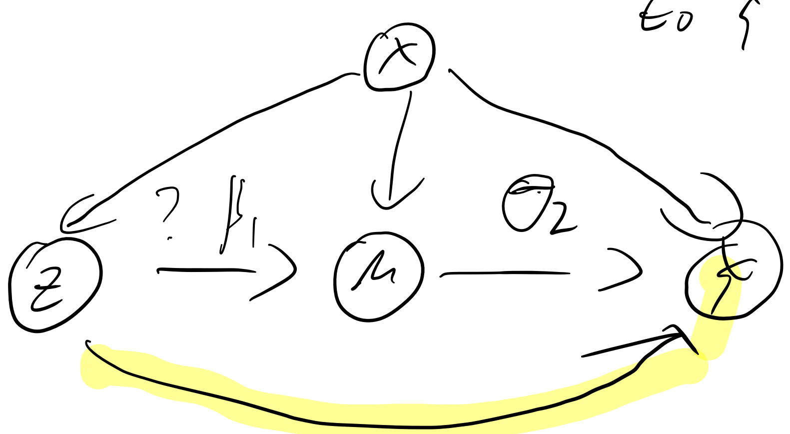
Examples

Can causal effect tell us the following (indirect) mechanism?

- Variants on chromosome 15q25.1 → smoking → lung cancer
- Neighborhood poverty → school and peer environment → adolescent substance use
- Job training → job-search self-efficacy → mental health

How do we define and identify the indirect effect?

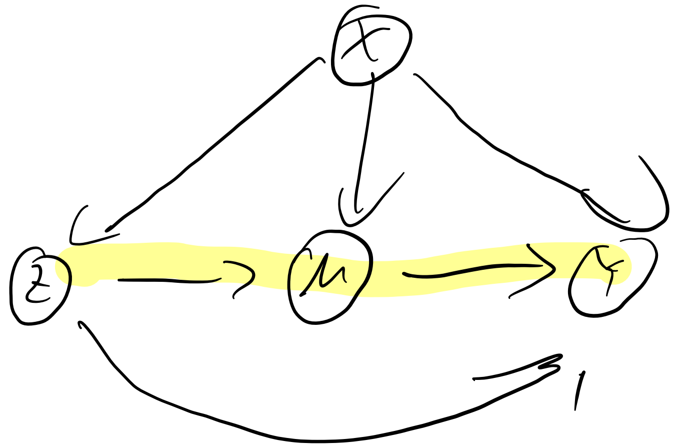
Causal Mechanism → Indirect Effect → magnitude of path from Z to M to Y



Difference and product methods

- Two methods commonly used in practice
- Product method for indirect effect:
  - regress Y on Z, M and $\mathbf{X} \rightarrow \widehat{\theta}_2$ (coef. for M)
  - regress M on Z and $\mathbf{X} \rightarrow \widehat{\beta}_1$ (coef. for Z)
  - estimator: $\widehat{\theta}_2 \widehat{\beta}_1$
- Difference method for indirect effect:
  - regress Y on Z and $\mathbf{X} \rightarrow \widehat{\tau}_1$ (coef. for Z)
  - regress Y on Z, M and $\mathbf{X} \rightarrow \widehat{\theta}_1$ (coef. for Z)
  - estimator: $\widehat{\tau}_1 - \widehat{\theta}_1$
- How is the effect defined?
- What assumptions are needed to justify these methods?

$Y \sim M$?



X : confounder

M : mediator

ACE : $Z \to M \to Y \; + \; Z \to Y$

$ACE - Z \to Y$

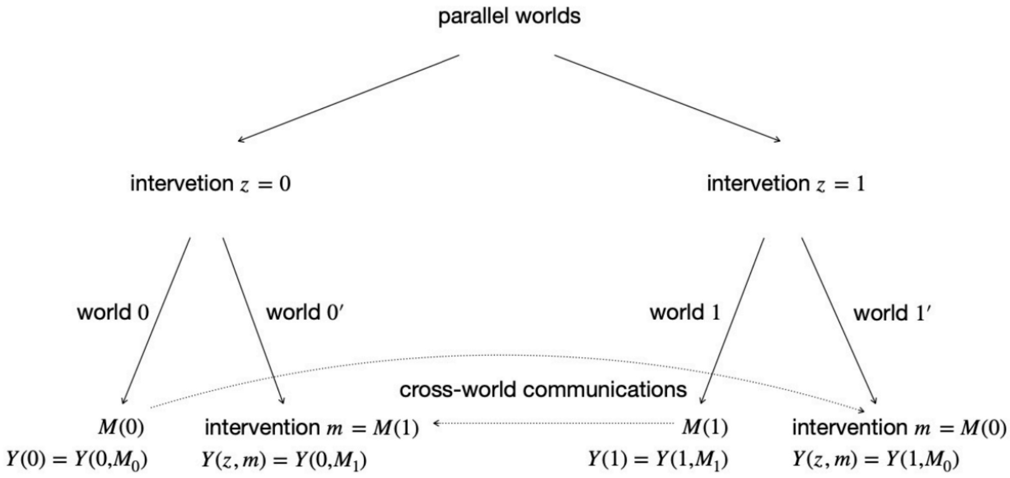
$Z \to M \quad \times \quad M \to Y$

Robins-Greenland-Pearl nested potential outcomes

- Binary treatment Z_i
- Potential mediators and outcomes: $M_i(z)$ and $Y_i(z)$
- Potential outcomes under both z and m : $Y_i(z, m)$
- Robins and Greenland (1992) and Pearl (2001) further consider the nested potential outcomes corresponding to intervention on z and $m = M\left(z'\right) : Y_i\left(z, M\left(z'\right)\right)$
- For example, $Y(1, M(0))$ is the hypothetical outcome if the unit received treatment 1 but its mediator were set at its natural value $M(0)$ without the treatment.
- Observed outcome: $Y_i = Y_i\left(Z_i, M_i\left(Z_i\right)\right)$. If $Z_i = 1, M_i(1) = 0$, then $Y_i = Y_i(1, 0)$

Cross-world potential outcomes

- $Y(1, M_i(0))$: interventions $Z = 1$ and $M = M(0)$ cannot simultaneously happen in any realized experiment – why?
- We need to imagine the parallel worlds



Metaphysics or science?

- Difference between $\{Y(1), Y(0)\}$ and $\{Y(1, M(0)), Y(0, M(1))\}$
- Frangakis and Rubin (2002) called $Y(1, M(0))$ and $Y(0, M(1))$ a priori counterfactuals because we cannot observe them in any physical experiments. In this sense, they do not exist a priori.
- According to Popper (1963), a way to distinguish science and metaphysics is the falsifiability of the statements. That is, if a statement is not falsifiable based on any physical experiments or observations, then it is not a scientific but rather a metaphysical statement.
- A strict Popperian statistician would view mediation analysis as metaphysics

$$
\begin{aligned}
ACE &= E\left(Y(1) - Y(0) \right) \\
&= E\left\{ Y(1, M(1)) - Y(0, M(0)) \right\} \\
&= E\left\{ Y(1, M(1)) - Y(1, M(0)) \right\} \\
&\quad + E\left\{ Y(1, M(0)) - Y(0, M(0)) \right\} \\
&= NIE + NDE
\end{aligned}
$$

Causal effects

We can depomose the ACE to natural direct effect and natural indirect effect.

- Total effect: ACE $= \mathbb{E}\{Y_i(1) - Y_i(0)\}$
- Natural direct effect:

$$\text{NDE} = Y_i\,(1, M_i(0)) - Y_i\,(0, M_i(0)) = Y_i\,(1, M_i(0)) - Y_i(0)$$

- Natural indirect effect – what is the interpretation?:

$$\text{NIE} = Y_i\,(1, M_i(1)) - Y_i\,(1, M_i(0)) = Y_i(1) - Y_i\,(1, M_i(0))$$

- ACE = NDE + NIE
- $\widehat{NIE} = \widehat{ACE} - \widehat{NDE}$: difference method

$$Z \to \hat{M} \to Y \Rightarrow NIE$$

$$NIE = E\big(Y(Z=1, M=M(1)) - Y(Z=1, M=M(0)) \big)$$

$$= 0 \quad (\text{because } M(1) = M(0))$$

$$CIE = E\{ Y(Z=1, M=1) - Y(Z=1, M=0) \}$$

$$\neq 0$$

$$\begin{tikzcd} & M \arrow[dr] & \\ Z \arrow[rr, "/" marking] & & Y \end{tikzcd}$$

$$E(Y(1) - Y(0))$$

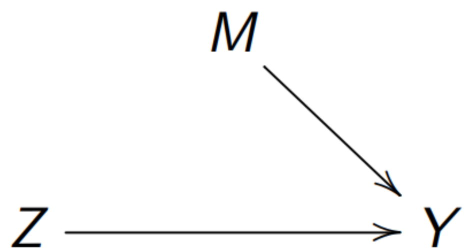
$$= E(Y(1,1) - Y(1,0))$$

$$+ E(Y(1,0) - Y(0,0))$$

$$= CIE + CDE$$

Controlled direct and indirect effects

- Controlled direct effect: $Y_i(1,0) - Y_i(0,0)$
- Controlled indirect effect: $Y_i(1,1) - Y_i(1,0)$
- Controlled indirect effect is not the effect of the treatment – why?



$$
\begin{aligned}
\text{NIE} &= Y_i(1, M_i(1)) - Y_i(1, M_i(0)) = 0 \\
\text{CIE} &= Y_i(1,1) - Y_i(1,0) \neq 0
\end{aligned}
$$

Identification assumptions

We need several confounding assumptions to identify the causal quantity.

- (A) No treatment-outcome confounding: $Z_i \perp Y_i(z, m) \mid \mathbf{X}_i$
- (B) No mediator-outcome confounding: $M_i \perp Y_i(z, m) \mid (Z_i, \mathbf{X}_i)$
- (C) No treatment-mediator confounding: $Z_i \perp M_i(z) \mid \mathbf{X}_i$
- (D) Cross-world independence between the potential outcomes and potential mediators: $Y_i(z, m) \perp M_i(z') \mid \mathbf{X}_i$
- (A)+(B): $(Z_i, M_i) \perp Y_i(z, m) \mid \mathbf{X}_i \rightsquigarrow \mathbb{E}\{Y_i(z, m) \mid \mathbf{X}_i\}$ is identifiable
- $(A) + (B) + (C)$ hold under experiments with sequentially randomized treatment and mediator
- (D) is fundamentally meta-physical because no physical experiment can ensure it.

$D \implies Y(1,1) \perp M(0) \mid X$

Mediation formula (Pearl, 2001)

Theorem

Under (A) to (D), $\mathbb{E}\left\{Y_i\left(z, M\left(z'\right)\right)\right\} = \mathbb{E}\left[\mathbb{E}\left\{Y_i\left(z, M\left(z'\right)\right) \mid \mathbf{X}_i\right\}\right]$, where

$$\mathbb{E}\left\{Y_i\left(z, M\left(z'\right)\right) \mid \mathbf{X}\right\} = \sum \mathbb{E}(Y \mid Z = z, M = m, \mathbf{X}) \operatorname{pr}\left(M = m \mid Z = z', \mathbf{X}\right)$$

$$\begin{aligned}
& \mathbb{E}\left\{Y_i\left(z, M\left(z'\right)\right) \mid \mathbf{X}\right\} \\
= & \sum_m \mathbb{E}\left\{Y_i\left(z, M\left(z'\right)\right) \mid M\left(z'\right) = m, \mathbf{X}\right\} \operatorname{pr}\left(M\left(z'\right) = m \mid \mathbf{X}\right) \\
= & \sum_m \mathbb{E}\left\{Y_i(z, m) \mid M\left(z'\right) = m, \mathbf{X}\right\} \operatorname{pr}\left(M\left(z'\right) = m \mid \mathbf{X}\right) \\
= & \sum_m \mathbb{E}\left\{Y_i(z, m) \mid \mathbf{X}\right\} \operatorname{pr}\left(M\left(z'\right) = m \mid \mathbf{X}\right) \\
= & \sum_m \mathbb{E}\left\{Y_i \mid Z = z, M = m, \mathbf{X}\right\} \operatorname{pr}\left(M = m \mid Z = z', \mathbf{X}\right)
\end{aligned}$$

$$E(Y(z, M(z')) \mid X)$$

$$= \sum_m E(Y \mid Z=z, M=m, X)\, P(M=m \mid Z=z', X)$$

$$NIE(x) = E(Y(1, M(1)) \mid X) - \underbrace{E(Y(1, M(0)) \mid X)}$$

$$= \sum_m E(Y \mid Z=1, M=m, X)\, P(M=m \mid \underline{Z=1}, X)$$

$$- \sum_m E(Y \mid Z=1, M=m, X)\, P(M=m \mid Z=0, X)$$

$Y \sim Z + X$

$Y \sim Z + M + X$

Mediation formula

- Mediation formula for NDE($\mathbf{X}$) and NIE($\mathbf{X}$)

$$\begin{aligned}
\text{NDE}(\mathbf{X}) =& \mathbb{E}\{Y_i(1, M(0)) \mid \mathbf{X}\} - \mathbb{E}\{Y_i(0, M(0)) \mid \mathbf{X}\} \\
=& \sum_m \{\mathbb{E}(Y_i \mid Z = 1, M = m, \mathbf{X}) - \mathbb{E}(Y_i \mid Z = 0, M = m, \mathbf{X})\} \\
& \cdot \text{pr}(M = m \mid Z = 0, \mathbf{X}) \\
\text{NIE}(\mathbf{X}) =& \mathbb{E}\{Y_i(1, M(1)) \mid \mathbf{X}\} - \mathbb{E}\{Y_i(1, M(0)) \mid \mathbf{X}\} \\
=& \sum_m \mathbb{E}(Y_i \mid Z = 1, M = m, \mathbf{X}) \\
& \cdot \{\text{pr}(M = m \mid Z = 1, \mathbf{X}) - \text{pr}(M = m \mid Z = 0, \mathbf{X})\}
\end{aligned}$$

$z=1$

$z'=1$

$z=1$

$z'=0$

- Average over $\mathbf{X}$ to obtain the NDE and NIE

$Y \sim Z + M + X$

$M \sim Z + X$

Baron-Kenny method (Baron and Kenny, 1986)

- Mediation formula under linear models

$$\mathbb{E}(M \mid Z, X) = \beta_0 + \beta_1 Z + \beta_X^\top X$$

$$\mathbb{E}(Y \mid Z, M, X) = \theta_0 + \theta_1 Z + \theta_2 M + \theta_X^\top X$$

$$\begin{aligned}
\mathrm{NDE}(\mathbf{x}) &= \sum_m \left[\mathbb{E}\left\{Y_i \mid Z = 1, M = m, \mathbf{X}\right\} - \mathbb{E}\left\{Y_i \mid Z = 0, M = m, \mathbf{X}\right\}\right] \\
&\quad \cdot \operatorname{pr}(M = m \mid Z = 0, \mathbf{X}) \\
&= \sum_m \theta_1 \operatorname{pr}(M = m \mid Z = 0, \mathbf{X}) = \theta_1 \\
\mathrm{NIE}(\mathbf{x}) &= \sum_m \mathbb{E}\left\{Y_i \mid Z = 1, M = m, \mathbf{X}\right\} \\
&\quad \cdot \left\{\operatorname{pr}(M = m \mid Z = 1, \mathbf{X}) - \operatorname{pr}(M = m \mid Z = 0, \mathbf{X})\right\} \\
&= \sum_m \left(\theta_0 + \theta_1 + \theta_2 m + \theta_X^\top X\right) \\
&\quad \cdot \left\{\operatorname{pr}(M = m \mid Z = 1, \mathbf{X}) - \operatorname{pr}(M = m \mid Z = 0, \mathbf{X})\right\} \\
&= \theta_2 \left\{\mathbb{E}(M \mid Z = 1, \mathbf{X}) - \mathbb{E}(M \mid Z = 0, \mathbf{X})\right\} = \theta_2 \beta_1
\end{aligned}$$

$$NIE(X) = \sum E(Y \mid Z=1, M=m, X)$$

$$\{p(M=m \mid Z=1, X) - p(M=m \mid Z=0, X)\}$$

$$= \sum_m (\theta_0 + \theta_1 + \theta_2 m + \theta_x^T X)$$

$$(p(M=m \mid Z=1, X) - p(M=m \mid Z=0, X))$$

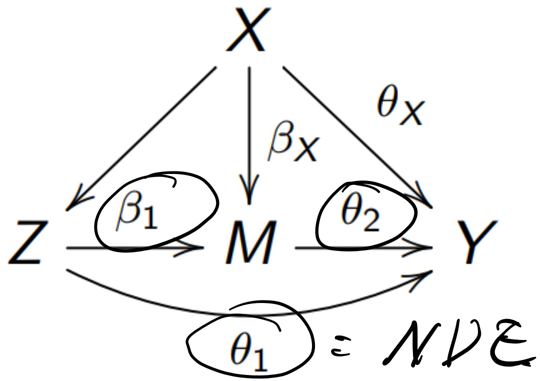
$$= (\theta_0 + \theta_1 + \theta_x^T X) \sum_m (p(M=m \mid Z=1, X) - p(M=m \mid Z=0, X))$$

$$+ \theta_2 \sum_m m (p(M=m \mid Z=1, X) - p(M=m \mid Z=0, X))$$

$$= \theta_2 (E(M \mid Z=1, X) - E(M \mid Z=0, X))$$

Baron-Kenny method

ACE: $Y \sim Z + X$ (coef of Z)



- Regress Y on Z, M and $X \to \widehat{\theta}_1$ and $\widehat{\theta}_2$
- Regress M on Z and $X \to \widehat{\beta}_1$
- Point estimates: $\widehat{\mathrm{NDE}} = \widehat{\theta}_1$ and $\widehat{\mathrm{NIE}} = \widehat{\theta}_2\widehat{\beta}_1$
- Variance of NIE: Delta method $\to$ the asymptotic variance of the NIE is $\mathrm{var}\left(\widehat{\theta}_2\right)\beta_1^2 + \mathrm{var}\left(\widehat{\beta}_1\right)\theta_2^2$
- Variance estimator $\widehat{\mathrm{var}}\left(\widehat{\theta}_2\right)\widehat{\beta}_1^2 + \widehat{\mathrm{var}}\left(\widehat{\beta}_1\right)\widehat{\theta}_2^2$

With interaction

$$\mathbb{E}(M \mid Z, X) = \beta_0 + \beta_1 Z + \beta_X^\top X$$

$$\mathbb{E}(Y \mid Z, M, X) = \theta_0 + \theta_1 Z + \theta_2 M + \theta_3 Z M + \theta_X^\top X$$

$$\mathrm{NDE}(\mathbf{x}) = \sum_m [\mathbb{E}\{Y_i \mid Z = 1, M = m, \mathbf{X}\} - \mathbb{E}\{Y_i \mid Z = 0, M = m, \mathbf{X}\}]$$

$$\cdot \operatorname{pr}(M = m \mid Z = 0, \mathbf{X})$$

$$= \sum_m (\theta_1 + \theta_3 m) \operatorname{pr}(M = m \mid Z = 0, \mathbf{X}) = \theta_1 + \theta_3 \left(\beta_0 + \beta_X^\top X\right)$$

$$\mathrm{NIE}(\mathbf{x}) = \sum_m \mathbb{E}\{Y_i \mid Z = 1, M = m, \mathbf{X}\}$$

$$\cdot \{\operatorname{pr}(M = m \mid Z = 1, \mathbf{X}) - \operatorname{pr}(M = m \mid Z = 0, \mathbf{X})\}$$

$$= \sum_m \left(\theta_0 + \theta_1 + \theta_2 m + \theta_3 m + \theta_X^\top X\right)$$

$$\cdot \{\operatorname{pr}(M = m \mid Z = 1, \mathbf{X}) - \operatorname{pr}(M = m \mid Z = 0, \mathbf{X})\}$$

$$= (\theta_2 + \theta_3) \{\mathbb{E}(M \mid Z = 1, \mathbf{X}) - \mathbb{E}(M \mid Z = 0, \mathbf{X})\} = (\theta_2 + \theta_3) \beta_1$$

Logistic model for binary mediator

$$\text{logit}\{\text{pr}(M = 1 \mid Z, X)\} = \beta_0 + \beta_1 Z + \beta_X^\top X$$

$$\mathbb{E}(Y \mid Z, M, X) = \theta_0 + \theta_1 Z + \theta_2 M + \theta_X^\top X$$

$$\text{NDE}(\mathbf{x}) = \sum_m \theta_1 \,\text{pr}(M = m \mid Z = 0, \mathbf{X}) = \theta_1$$

$$\text{NIE}(\mathbf{x}) = \sum_m \mathbb{E}\{Y \mid Z = 1, M = m, \mathbf{X}\}$$

$$\text{-}\ \{\text{pr}(M = m \mid Z = 1, \mathbf{X}) - \text{pr}(M = m \mid Z = 0, \mathbf{X})\}$$

$$= \theta_2\{\mathbb{E}(M \mid Z = 1, \mathbf{X}) - \mathbb{E}(M \mid Z = 0, \mathbf{X})\}$$

$$= \theta_2 \left(\frac{e^{\beta_0 + \beta_1 + \beta_X^\top \mathrm{x}}}{1 + e^{\beta_0 + \beta_1 + \beta_X^\top x}} - \frac{e^{\beta_0 + \beta_X^\top \mathrm{x}}}{1 + e^{\beta_0 + \beta_X^\top x}} \right)$$

$$\text{NIE} = \theta_2 \mathbb{E} \left(\frac{e^{\beta_0 + \beta_1 + \beta_x^\top X}}{1 + e^{\beta_0 + \beta_1 + \beta_X^\top X}} - \frac{e^{\beta_0 + \beta_x^\top X}}{1 + e^{\beta_0 + \beta_X^\top X}} \right)$$

Mediation analysis for a job training program

- A randomized field experiment that investigates the efficacy of a job training intervention on unemployed workers. The program is designed to not only increase reemployment among the unemployed but also enhance the mental health of the job seekers.
- Treatment Z : indicator of encouragement; mediator M : job-search self-efficacy; Y : measure of depressive symptoms
- Covariates X : age, gender, baseline depressive measure, etc.
- Baron-Kenny method
  - NDE: point est. $= -0.035$, s.e. $= 0.011$
  - NIE: point est. $= -0.011$, s.e. $= 0.009$
- Noncompliance is present in the study

given $M(1) = M(0)$, $Y(1, M(1)) = Y(1, M(0))$

$$\Downarrow$$

$$= 0$$

$$E(Y(1) - Y(0) \mid M(1) = M(0))$$

$$= E(Y(1, M(1)) - Y(0, M(0)) \mid M(1) = M(0))$$

$$= E(Y(1, M(1)) - Y(1, M(0)) \mid M(1) = M(0))$$

$$+ E(Y(1, M(0)) - Y(0, M(0)) \mid M(1) = M(0))$$

Connection between principal stratification and mediation analysis

Principal stratification is to stratify the population by post-treatment variable and is different from stratified experiment.

- In strata with $M(1) = M(0)$, the indirect effect is zero – why?

$$
\begin{aligned}
&\mathbb{E}\{Y(1) - Y(0) \mid M(1) = M(0)\} \\
=&\mathbb{E}\{Y(1, M(1)) - Y(0, M(0)) \mid M(1) = M(0)\} \\
=&\mathbb{E}\{Y(1, M(1)) - Y(1, M(0)) \mid M(1) = M(0)\} \\
&+ \mathbb{E}\{Y(1, M(0)) - Y(0, M(0)) \mid M(1) = M(0)\} \\
=&\mathbb{E}\{Y(1, M(0)) - Y(0, M(0)) \mid M(1) = M(0)\}
\end{aligned}
$$

- Principal strata direct effect: $\mathbb{E}\{Y(1) - Y(0) \mid M(1) = M(0) = m\}$
- VanderWeele (2008) studies the relations between the principal causal effects and natural direct and indirect effects
- Forastiere et al. (2018) discuss the connections between the assumptions

Connection between principal stratification and mediation analysis

- Principal strata indirect effect:

$$\mathbb{E}\{Y(1) - Y(0) \mid M(1) = 1, M(0) = 0\}?$$

$$\begin{aligned} &\mathbb{E}\{Y(1) - Y(0) \mid M(1) = 1, M(0) = 0\} \\ =&\mathbb{E}\{Y(1, M(1)) - Y(0, M(0)) \mid M(1) = 1, M(0) = 0\} \\ =&\mathbb{E}\{Y(1,1) - Y(0,0) \mid M(1) = 1, M(0) = 0\} \end{aligned}$$

- $\mathbb{E}\{Y(1) - Y(0) \mid M(1) = 1, M(0) = 0\}$ consists of both direct and indirect effects

Summary

$Z \longrightarrow M \longrightarrow Y$

- Mediation analysis studies the extent to which an effect is mediated through a particular pathway and to which the effect of a treatment on the outcome operates directly
- Natural direct and indirect effects → definitions rely on nested potential outcomes
- Identification assumptions
  - no treatment-outcome confounding
  - no mediator-outcome confounding
  - no treatment-mediator confounding
  - cross-world independence
- Different mediation formula under different models

Extensions

- Sensitivity analysis for mediation analysis (Imai et al., 2010)
- Partial identification without cross-world independence
- Multiple mediator
  - generalization of the mediation analysis to more than one mediators $\rightsquigarrow$ path analysis
  - focus on one mediator $M_i \rightsquigarrow$ NIE is the effect through M_i and NDE is the sum of the direct effect and the effect through other mediators

Suggested readings

- Natural direct and indirect effects
  - Pearl. 2001. "Direct and indirect effects"
- Connection between principal stratification and mediation analysis
  - Forastiere et al. 2018. "Principal ignorability in mediation analysis: through and beyond sequential ignorability'
- Sensitivity analysis for mediation analysis
  - Imai et al. 2010. "Identification, inference and sensitivity analysis for causal mediation effects"