Causal Mediation Analysis: Concepts and Identification

Li Ge, lge7@wisc.edu
Biostatistics & Medical Informatics
University of Wisconsin-Madison
November 13, 2020

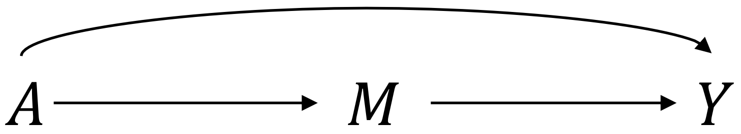
Disclaimer & Reference

- I don't take the credit other than re-interpreting the topic from my own perspective. The slides are only for pedagogical use.
- [1] VanderWeele, T. (2015). *Explanation in causal inference: methods for mediation and interaction*. Oxford University Press.
    - Note: Read Chapter 2 for a more detailed introduction on this topic. Read the Appendix for definitions and proofs.
- [2] VanderWeele's video lectures: https://youtu.be/EI5y6pV87-Q.
- I thank Felix Elwert for giving a lecture on causal mediation analysis in his seminar course SOC 952, from which I learned a lot.

Motivation

Background

- In many research contexts, we might be interested in the extent to which the effect of some exposure A on some outcome Y is mediated by an intermediate variable M and to what extent it is direct.
- Stated another way, we are interested in the direct and indirect effects of the exposure

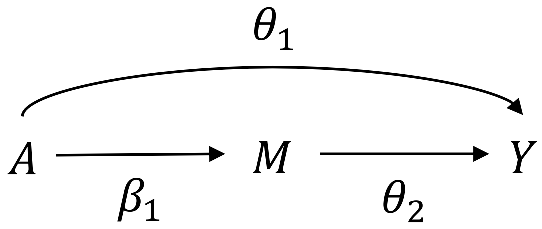


- For example, A represents some genetic variants, M represent smoking behavior, and Y represents whether someone gets lung cancer.
- We know some genetic variants are associated with lung cancer. We also know these genetic variants are associated with smoking behavior.
- **Question**: Are the genetic effects on lung direct or operate through pathways related to smoking behavior? Is there an interaction between A and M?

Standard Approaches

- Difference method: fit two linear model for outcome Y:
  - $E[Y|A=a, C=c] = \phi_0 + \phi_1 a + \phi_2' c$
  - $E[Y|A=a, m=m, C=c] = \theta_0 + \theta_1 a + \theta_2 m + \theta_4' c$
  - $\text{Indirect Effect} = \phi_1 - \theta_1, \text{Direct Effect} = \theta_1$

- Product method (Baron & Kenny 1986)[1] : fit one linear model for outcome Y, another linear model for mediator M:
  - $E[M|A=a, C=c] = \beta_0 + \beta_1 a + \beta_2' c$
  - $E[Y|A=a, m=m, C=c] = \theta_0 + \theta_1 a + \theta_2 m + \theta_4' c$
  - $\text{Indirect Effect} = \beta_1 \theta_2, \text{Direct Effect} = \theta_1$



- They are equivalent for continuous outcomes if the models are correctly specified (MacKinnon & Dwyer 1993, MacKinnon et al. 1995), i.e., linear homogeneous effect models.

[1] As of November 13, 2020, (Baron & Kenny 1986) has been cited 96,558 times according to Google Scholar.

Limitation of Standard Approaches

- Limitation 1: Both methods implicitly assume no unmeasured confounders between mediator(s) M and outcome Y.
  - Randomization can only eliminate confounding between exposure A and outcome Y, provided perfect compliance.

- Limitation 2: Both methods assume no interaction between mediator(s) M and exposure A.

- Limitation 3: We cannot causally interpret the estimated effects. Those are model-based quantities. What are the estimands?

Causal Mediation Analysis

Definitions

- Y: outcome
- M: some post-treatment intermediate(s)
- A: treatment
- C: a set of covariates
- Let $Y^{A=a}$ be the potential outcome for Y by setting $A = a$.
- Let $M^{A=a}$ be the potential outcome for M by setting $A = a$.
- Let $Y^{A=a,M=m}$ be the potential outcome for Y by setting $A = a$ and $M = m$.
- We assume the composition $Y^{a} = Y^{aM^{a}}$.

Causal Estimands

- (Robins & Greenland 1992) and (Pearl 2001) proposed the following counterfactual definitions[2] for direct and indirect effects:
  - Controlled direct effect (CDE):
    $$CDE(m) = Y^{1m} - Y^{0m}$$
  - Natural direct effect (NDE):
    $$NDE = Y^{1M^0} - Y^{0M^0}$$
  - Natural indirect effect (NIE):
    $$NIE = Y^{1M^1} - Y^{1M^0}$$

We need more assumptions to identify this quantity.

- Remarks
  - Total Effect (TE): $TE = Y^1 - Y^0 = Y^{1M^1} - Y^{1M^0} + Y^{1M^0} - Y^{0M^0} = NIE + NDE$.
  - Alternative definition: $NDE = Y^{1M^1} - Y^{0M^1}$, $NIE = Y^{0M^1} - Y^{0M^0}$.

[2] See (Nguyen, Schmid, and Stuart 2019) for a comprehensive survey on defining causal mediation estimands.

Interpretation of CDE and NDE[3]

CDEs are “prescriptive” (more recently called “interventional”) in that they capture direct effects that result from prescribing some externally determined value for the mediator.

- CDEs do not require knowledge of what values the mediators take in nature.
- Relevant for policy evaluation, where A and M are intervened on.

NDEs are “descriptive” (or “explanatory”) in that they capture direct effects that result from fixing the mediators at the values that they would take “descriptively” in nature.

- NDEs require knowledge of the natural behavior of the mediators
- Relevant for understanding mediation processes in “nature.”
- NDEs do not usually correspond to practical interventions (because we do not usually know what value M would take in nature).

[3] This slide is from Felix Elwert.

Identification

Assumptions

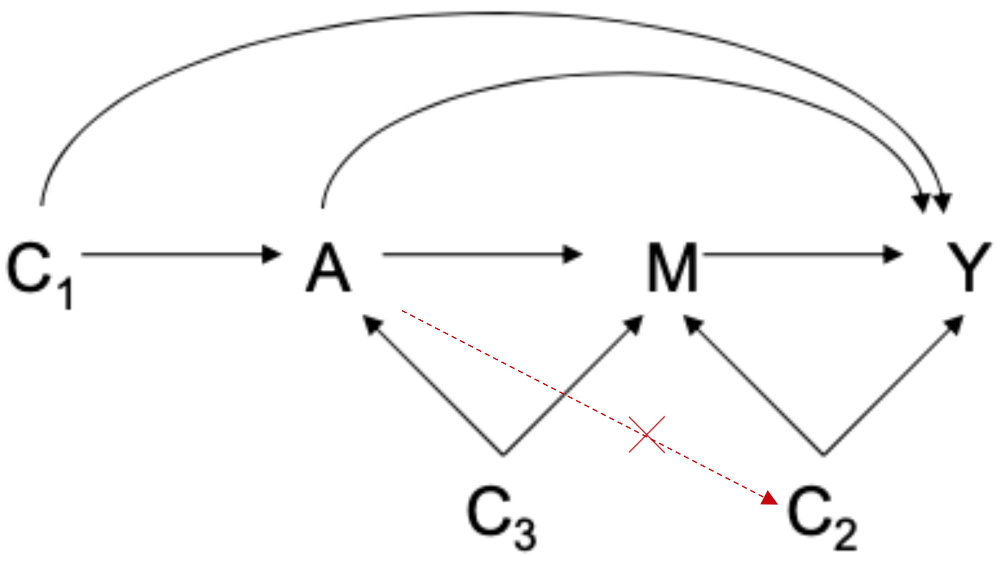
- We assume consistency, SUTVA, and positivity throughout.
- TE is identified if (A.1) holds.
- CDE is identified if both (A.1) and (A.2) hold.
- For NDE and NIE, we require two more assumptions (A.3) and (A.4).
- (A.1) $Y^{am} \perp A|C$ → TE
- (A.2) $Y^{am} \perp M|\{A, C\}$
- (A.3) $M^{a} \perp A|C$
- (A.4) $Y^{am} \perp M^{a^*}|C$

CDE: (A.1), (A.2)

NDE & NIE: (A.1)–(A.4)

Graphical Adjustment Criteria

- (A.1) $Y^{am} \perp A|C$: No unmeasured exposure-outcome confounders given C.
- (A.2) $Y^{am} \perp M|\{A,C\}$: No unmeasured mediator-outcome confounders given $\{C,A\}$
- (A.3) $M^{a} \perp A|C$: No unmeasured exposure-mediator confounders given C.
- (A.4) $Y^{am} \perp M^{a^*}|C$: No mediator-outcome confounder affected by exposure (i.e. no arrow from A to C_2)
- Remark: Assumptions (A.1) and (A.3) will hold if the exposure A is randomized. However, assumptions (A.3) and (A.4) may not.



- Adjusting for C_1, we can identify TE.
- Adjusting for C_1, C_2, we can identify CDE.
- Adjusting for C_1, C_2, C_3, we can identify NDE and NIE.
- In all cases, C cannot include descendent of A, i.e., no post-treatment variables should be included in C.

Nonparametric Identification of CDE

Proposition 1.1 (Robins 1986, cf. Pearl 2001):

If assumptions (A.1) and (A.2) hold, then average controlled direct effects conditional on C are identified and given by

$$E\left[Y^{am} - Y^{a^*m} \middle| c\right] = E[Y|a, m, c] - E[Y|a^*, m, c]$$

Proof:

For any a, m, we have

$$\Pr(Y^{am} \le y | C = c) = \Pr(Y^{am} \le y | a, c) \text{ by } (A.1)$$
$$= \Pr(Y^{am} \le y | a, m, c) \text{ by } (A.2)$$
$$= \Pr(Y \le y | a, m, c) \text{ by consistency}$$

In fact, we can identify the cumulative distribution function (c.d.f.) of the potential outcome using the observed data. The expected value is consequently identified.

Remark: The population average is given by $E\left[Y^{am} - Y^{a^*m}\right] = \sum_c E\left[Y^{am} - Y^{a^*m} \middle| c\right] P(c)$.

Nonparametric Identification of NDE & NIE

Proposition 1.2 (Pearl 2001):

If assumptions (A.1) – (A.4) hold, then the <u>average</u> natural direct effect conditional on C is identified and is given by

$$E\left[Y^{aM^{a^*}} - Y^{a^*M^{a^*}} \middle| c\right] = \sum_m \{E[Y|a,m,c] - E[Y|a^*,m,c]\}\, P(m|a^*,c)$$

and the <u>average</u> natural indirect effect conditional on C is identified and is given by

$$E[Y^{aM^a} - Y^{aM^{a^*}} |c] = \sum_m E[Y|a,m,c]\{P(m|a,c) - P(m|a^*,c)\}$$

Proof:

$$\Pr(Y^{aM^{a^*}} \le y|C=c) = \sum_m \Pr(Y^{am} \le y|c, M^{a^*} = m)\Pr(M^{a^*} = m|c) \text{ by iterated expectation}$$

$$= \sum_m \Pr(Y^{am} \le y|c)\Pr(M^{a^*} = m|a^*, c) \text{ by } (A.4) \text{ and } (A.3)$$

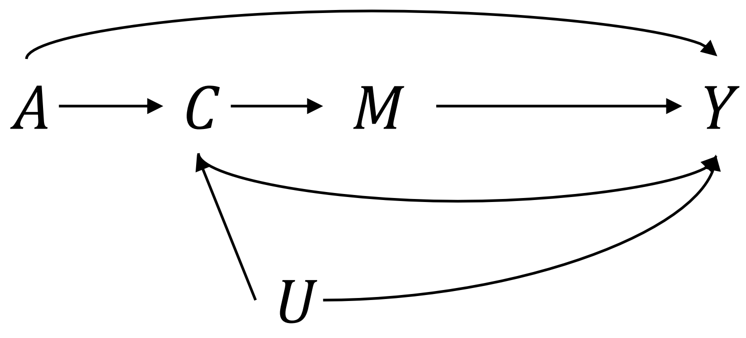
$$= \sum_m \Pr(Y^{am} \le y|a,c)\Pr(M = m|a^*,c) \text{ by } (A.1) \text{ and consistency}$$

$$= \sum_m \Pr(Y^{am} \le y|a,m,c)\Pr(m|a^*,c) \text{ by } (A.2)$$

$$= \sum_m \Pr(Y \le y|a,m,c)\Pr(m|a^*,c) \text{ by } (A.2) \text{ by consistency}$$

Sequential Adjustment Criteria for CDE

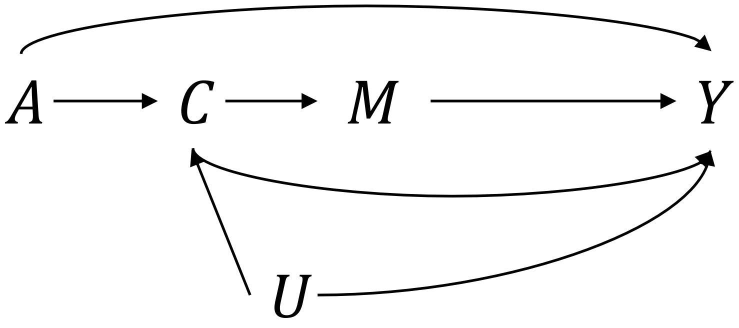
- Remember that the adjustment set C cannot include post-treatment variables.
- e.g., C is a mediator-outcome confounder, but is also a post-treatment variable, is identifying CDE possible?
- CDE can be identified using some more relaxed sequential adjustment criteria, see (Pearl and Robins 1995). The theory is also known as sequential g-estimation.



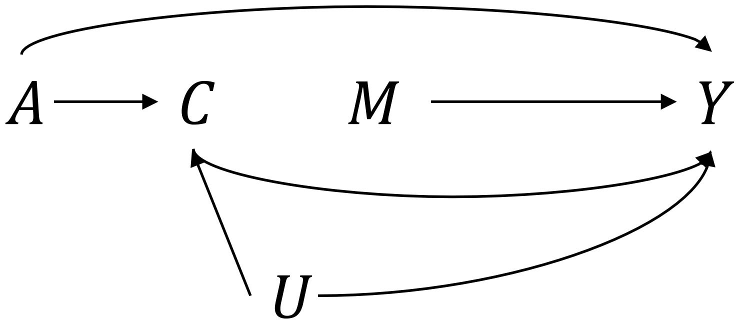
Sequential Adjustment Criteria for CDE

Proposition 1.3 [Sequential backdoor criterion for CDE] (Pearl and Robins 1995): The CDEs of A on Y relative to the (set of) mediator(s) M is identifiable if

- The total effect of M on Y can be identified by adjustment for some set of variables V, and
- The total effect of A on Y in a mutilated DAG *where all arrows into M are deleted* can be identified via adjustment for some set of variables W.



First adjust for $V = \{A, C\}$



Then adjust for $W = \emptyset$. Or $W = \{M\}$ to increase statistical power.

Estimation: Regression Methods

Estimation of CDE, NDE & NIE

Proposition 2.3 (VanderWeele & Vansteelandt 2009):

If assumptions (A.1) – (A.4) hold and if Y and M are continuous and the following regression models for Y and M are correctly specified:

$$E[Y|a,m,c] = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4' c$$
$$E[M|a,c] = \beta_0 + \beta_1 a + \beta_2' c$$

Then the average controlled direct effect and the average natural direct and indirect effects, conditional on $C = c$, are given by

$$E\left[Y^{am} - Y^{a^*m}|c\right] = (\theta_1 + \theta_3 m)(a - a^*)$$
$$E\left[Y^{aM^{a^*}} - Y^{a^*M^{a^*}}\middle|c\right] = [\theta_1 + \theta_3(\beta_0 + \beta_1 a^* + \beta_2' c)](a - a^*)$$
$$E[Y^{aM^a} - Y^{aM^{a^*}}|c] = (\theta_2\beta_1 + \theta_3\beta_1 a)(a - a^*)$$

with standard errors given by delta method.

Proof:

$$\begin{aligned} CDE(m) &= E\left[Y^{am} - Y^{a^*m}|c\right] \\ &= E[Y|a,c,m] - E[Y|a^*,c,m] \\ &= (\theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4' c) - (\theta_0 + \theta_1 a^* + \theta_2 m + \theta_3 am + \theta_4' c) \\ &= \cdots \\ &= (\theta_1 + \theta_3 m)(a - a^*) \end{aligned}$$

Estimation of CDE, NDE & NIE

cont'd:

$$NDE = E\left[Y^{aM^{a^*}} - Y^{a^*M^{a^*}} \middle| c\right]$$
$$= \sum_m \{E[Y|a,m,c] - E[Y|a^*,m,c]\} P(m|a^*,c)$$
$$= \sum_m \{(\theta_1 + \theta_3 m)(a - a^*)\} P(m|a^*,c)$$
$$= (\theta_1 + \theta_3 E[M|a^*,c])(a - a^*)$$
$$= [\theta_1 + \theta_3(\beta_0 + \beta_1 a^* + \beta_2' c)](a - a^*)$$

$$NIE = E[Y^{aM^a} - Y^{aM^{a^*}} |c]$$
$$= \sum_m E[Y|a,m,c]\{P(m|a,c) - P(m|a^*,c)\}$$
$$= \sum_m (\theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4' c)\{P(m|a,c) - P(m|a^*,c)\}$$
$$= \theta_2(E[M|a,c] - E[M|a^*,c]) + \theta_3 a(E[M|a,c] - E[M|a^*,c])$$
$$= \cdots$$
$$= (\theta_2\beta_1 + \theta_3\beta_1 a)(a - a^*)$$

Estimation of CDE, NDE & NIE

Proposition 2.3 (VanderWeele & Vansteelandt 2009):

If assumptions (A.1) – (A.4) hold and if Y and M are continuous and the following regression models for Y and M are correctly specified:

$$E[Y|a,m,c] = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4' c$$
$$E[M|a,c] = \beta_0 + \beta_1 a + \beta_2' c$$

Then the average controlled direct effect and the average natural direct and indirect effects, conditional on $C = c$, are given by

$$CDE(m) = E\left[Y^{am} - Y^{a^*m}|c\right] = (\theta_1 + \theta_3 m)(a - a^*)$$
$$NDE = E\left[Y^{aM^{a^*}} - Y^{a^*M^{a^*}}\middle|c\right] = [\theta_1 + \theta_3(\beta_0 + \beta_1 a^* + \beta_2' c)](a - a^*)$$
$$NIE = E[Y^{aM^{a}} - Y^{aM^{a^*}}|c] = (\theta_2\beta_1 + \theta_3\beta_1 a)(a - a^*)$$

with standard errors given by delta method.

Remarks:

- Without interaction, i.e., $\theta_3 = 0$, $CDE(m) = NDE$, regardless of the value of M.
- For binary outcome Y, or binary mediator M, (VanderWeele & Vansteelandt 2010) derived similar results using GLM on the odds ratio scale.

Final Notes

- Causal mediation analysis is a very active area of research.
- Since randomization cannot guarantee unconfoundedness between mediator and outcome, sensitivity analysis is also of central interest in causal mediation analysis.
- (Imai et al. 2010) proposed to use a broad class of simulation-based parametric or semiparametric models for Y and M to estimate the natural direct and indirect effects, and the standard errors for these effects by bootstrapping.