Causal Mediation Analysis with Time Varying Exposures, Mediators and Confounders - Application to Longitudinal Settings

Supervisor: Bo Fu
Reporter: Yiqiao Jin
Date: 2023/06/07

DATASET

- **Data Type:** Annual financial panel data
- **Source:** Shanghai Health Administrative Database (SHAD)
- **Time:** from 2014 to 2020 (1 Jan to 31 Dec)
- **Description:** 93 secondary and 50 tertiary hospitals with a total of 1001 records
- **Variables:**

  (a) basic characteristics of hospitals

  (b) annual revenue

  (c) categorical attributes of hospitals

SETTING

- **Zero Markup Drug Policy (ZMDP)**
  - **Method:** Government Financial Subsidy
  - **Purpose:**

    (a) Promote the progress towards Universal Health Coverage (UHC)

    (b) Control the unreasonable increase in medical expenses growth

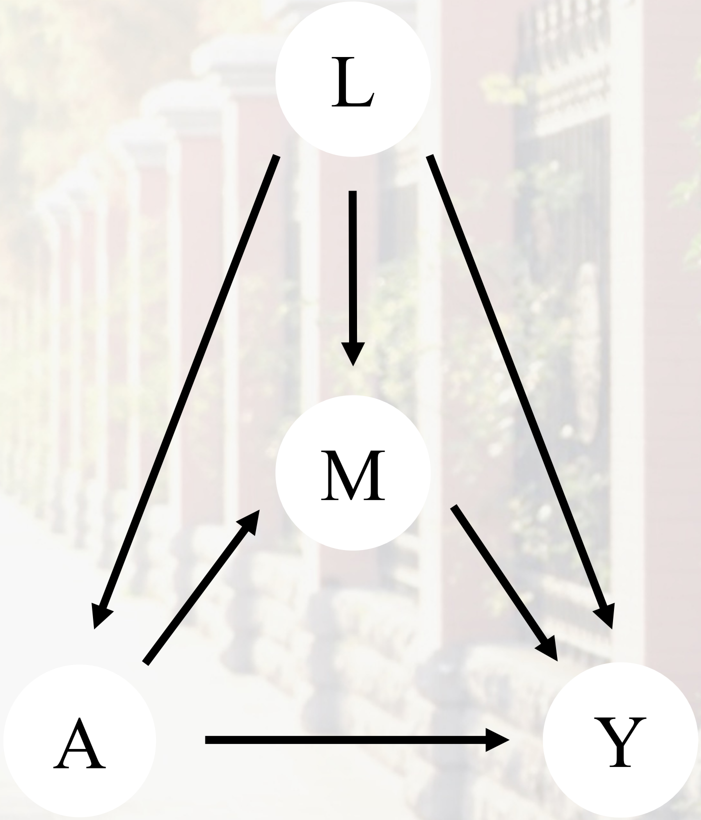
    (c) Optimize the medical expense structure
  - **Evaluation:**

    Effectiveness (Annual ratio of the service expense to the drug expense)

NOTATION 1

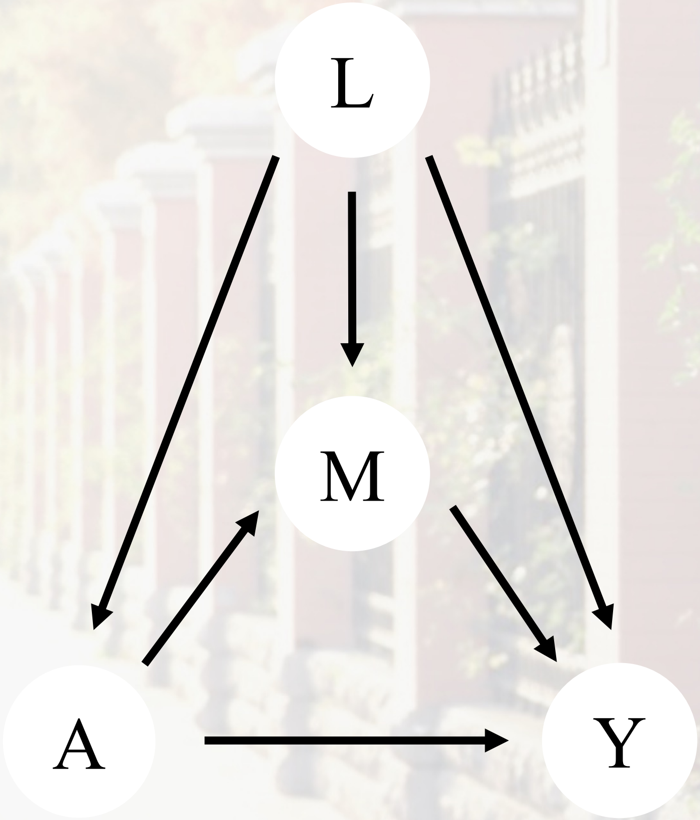
Take on a single value

- $\boldsymbol{A}$: Exposure (Index $a = 1$, Reference $a^* = 0$)
- $\boldsymbol{Y}$: Outcome
- $\boldsymbol{M}$: Mediator
- $\boldsymbol{L}$: Covariates for potential confounding control
- $\boldsymbol{Y_a}$: Potential outcome Y had A set to a
- $\boldsymbol{M_a}$: Mediator M had A set to a
- $\boldsymbol{Y_{am}}$: Potential outcome Y had A set to a and M set to m
- $\boldsymbol{Y_{aM_{a^*}}}$: Potential outcome Y had A set to a and M set to $\boldsymbol{M_{a^*}}$ Natural potential M when settting A set to a^*)



Robins, J.M. and Greenland, S. (1992) Identifiability and Exchangeability for Direct and Indirect Effects. Epidemiology, 3, 143-155.

SCENARIO

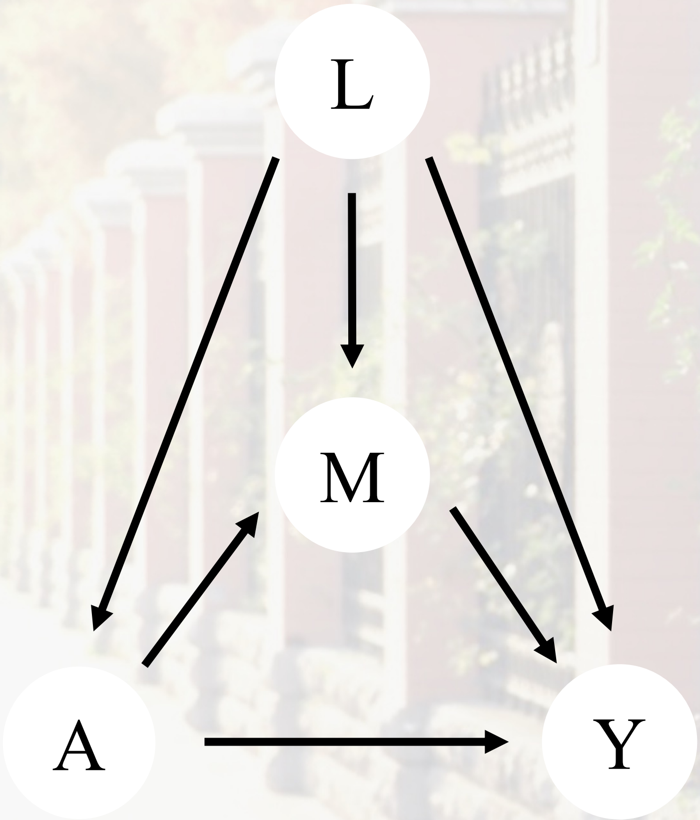


Ps: Sample size limitation may lead to algorithmic unreliability of causal discovery.

- ***A***: Exposure (ZMDP Implementation)
- ***Y***: Outcome (Effectiveness of ZMDP)
- ***M***: Mediator (Government financial subsidy)
- ***L***: Covariates for potential confounding control
- **QUESTION:**
  - $TE = \mathbb{E}[Y_1 - Y_0 \mid l] = \mathbb{E}\left[Y_{1M_1} - Y_{0M_0} \mid l\right]$
  - $CDE(m) = \mathbb{E}[Y_{1m} - Y_{0m} \mid l]$
  - $NDE = \mathbb{E}\left[Y_{1M_0} - Y_{0M_0} \mid l\right]$
  - $NIE = \mathbb{E}\left[Y_{1M_1} - Y_{1M_0} \mid l\right]$

Robins, J.M. and Greenland, S. (1992) Identifiability and Exchangeability for Direct and Indirect Effects. Epidemiology, 3, 143-155.

DIFFICULTIES



- $\boldsymbol{A, Y, M, V}$ are all **time varying** variables
- **Assumption:**

$$Y_{am} \perp\!\!\!\perp A \mid V,\ \ Y_{am} \perp\!\!\!\perp M \mid \{A, V\},\ M_a \perp\!\!\!\perp A \mid V,\ Y_{am} \perp\!\!\!\perp M_{a*} \mid V$$

- NDE and NIE can not identified
  - In many settings involving time-varying exposures and mediators
  - Whenever there is a exposure-induced mediator–outcome confounding

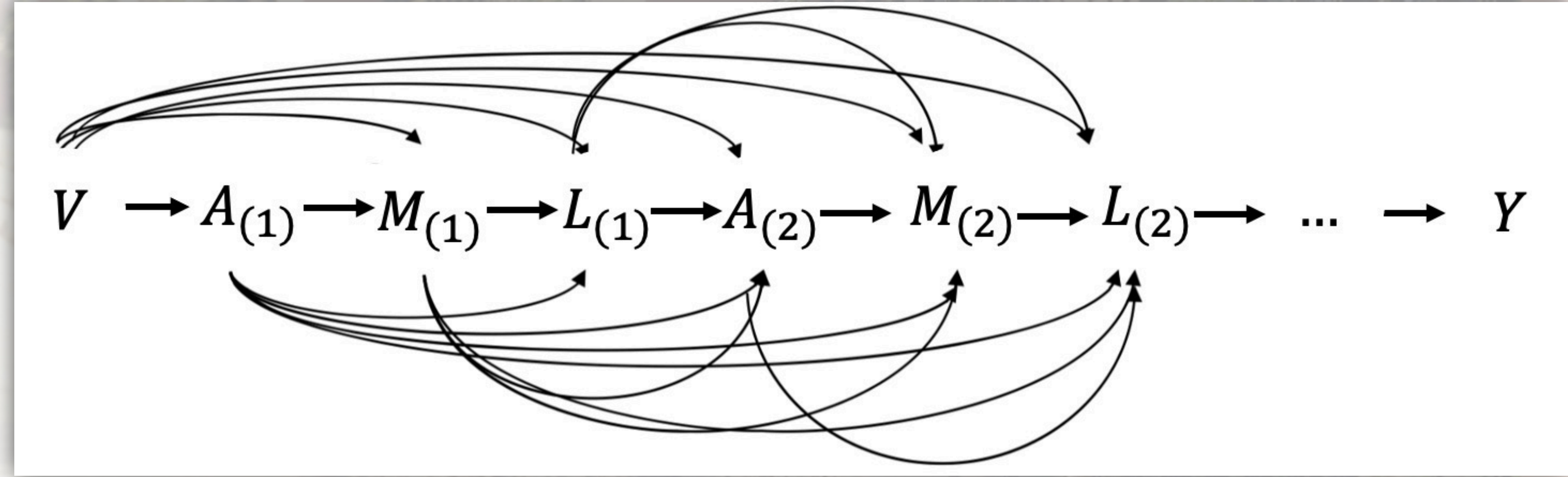
Avin, C., Shpitser, I., & Pearl, J. (2005). Identifiability of Path-Specific Effects. *International Joint Conference on Artificial Intelligence*.

VanderWeele, T. J., & Tchetgen Tchetgen, E. J. (2017). Mediation analysis with time varying exposures and mediators. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, *79*(3), 917–938.

VanderWeele. T.J. (2015). *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford University Press.

NOTATION 2

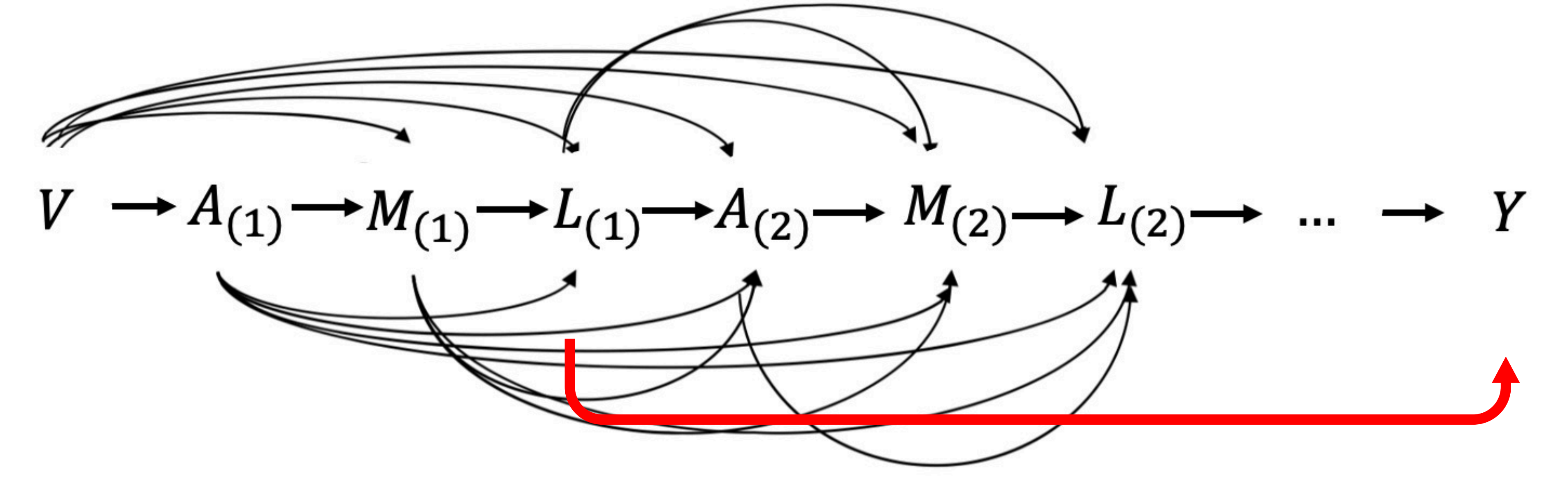
- Let $(\boldsymbol{A}_{(1)},\dots,\boldsymbol{A}_{(T)})$, $(\boldsymbol{M}_{(1)},\dots,\boldsymbol{M}_{(T)})$ and $(\boldsymbol{L}_{(1)},\dots,\boldsymbol{L}_{(T)})$ denote values of time varying exposure, mediator and confounders at periods $\mathbf{0},\dots,\boldsymbol{T}$.
- Let $\boldsymbol{V}$ denote baseline covariates that occur before the first exposure period.
- We assume a subsequent temporal ordering $\boldsymbol{A}_{(t)},\boldsymbol{M}_{(t)},\boldsymbol{L}_{(t)}$ as below.



VanderWeele. T.J. (2015). *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford University Press.

UNIDENTIFIABILITY

- The mediator $\boldsymbol{L}_{(1)}$ is affected by the exposure $\boldsymbol{A}_{(1)}$ and confounds the mediator-outcome relationship between $\boldsymbol{M}_{(2)}$ and $\boldsymbol{Y}$. Hence, NDE and NIE are not identified in this setting.



Avin, C., Shpitser, I., & Pearl, J. (2005). Identifiability of Path-Specific Effects. *International Joint Conference on Artificial Intelligence*.
VanderWeele. T.J. (2015). *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford University Press.

SOLUTION & NOTATION 3

- The interventional direct and indirect effects could be identified.
- $\bar{W} = \bar{W}_{(T)} = \left(W_{(1)}, \ldots, W_{(T)}\right), \bar{W}_{(t)} = \left(W_{(1)}, \ldots, W_{(t)}\right)$
- $\boldsymbol{Y_{\overline{am}}}$**:** Counterfactual outcome Y if $\bar{A}$ was set to $\bar{a}$ and $\bar{M}$ was set to $\bar{m}$
- $\boldsymbol{M_{\bar{a}}(t)}$**:** Counterfactual value of $M_{(t)}$ if $\bar{A}$ was set to $\bar{a}$
- $\boldsymbol{\bar{G}_{\bar{a}|v}(t)}$**:** a random draw from the distribution of the mediator $\bar{M}_{(t)}$ that would have been observed in the population with baseline covariates $V = v$ if exposure status $\bar{A}$ had been fixed to $\bar{a}$

VanderWeele, T. J., & Tchetgen Tchetgen, E. J. (2017). Mediation analysis with time varying exposures and mediators. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, *79*(3), 917–938.

SOLUTION & NOTATION 3

- A decomposition of the interventional overall effect:
  - $\boldsymbol{TE^R} = \mathbb{E}\left[\boldsymbol{Y_{\bar{a}\bar{G}_{\bar{a}|v}(t)}}\middle|\boldsymbol{v}\right] - \mathbb{E}\left[\boldsymbol{Y_{\bar{a}^*\bar{G}_{\bar{a}^*|v}(t)}}\middle|\boldsymbol{v}\right]$

    $= \mathbb{E}\left[\boldsymbol{Y_{\bar{a}\bar{G}_{\bar{a}|v}}}\middle|\boldsymbol{v}\right] - \mathbb{E}\left[\boldsymbol{Y_{\bar{a}\bar{G}_{\bar{a}^*|v}}}\middle|\boldsymbol{v}\right] + \mathbb{E}\left[\boldsymbol{Y_{\bar{a}\bar{G}_{\bar{a}^*|v}}}\middle|\boldsymbol{v}\right] - \mathbb{E}\left[\boldsymbol{Y_{\bar{a}^*\bar{G}_{\bar{a}^*|v}}}\middle|\boldsymbol{v}\right]$

    $= \boldsymbol{NIE^R} + \boldsymbol{NDE^R}$

  - $\widehat{\boldsymbol{NIE^R}} = \sum_{t\le T}\theta_2\beta_1(t)[avg\{\bar{\boldsymbol{a}}(t)\} - avg\{\bar{\boldsymbol{a}}^*(t)\}]$
  - $\widehat{\boldsymbol{NDE^R}} = \theta_1\{cum(\bar{a}) - cum(\bar{a}^*)\}$

  Where $\mathbb{E}[\boldsymbol{Y_{\overline{am}}}] = \theta_0 + \theta_1\, cum(\bar{a}) + \theta_2 cum(\bar{m})$

  and $g^{-1}[\mathbb{E}\{\boldsymbol{M}_{\bar{a}}(\boldsymbol{t})\}] = \beta_0(t) + \beta_1(t)avg\{\bar{\boldsymbol{a}}(t)\}$

ACKNOWLEDGE

THANK YOU!