# SOCI 40258

Causal Mediation Analysis

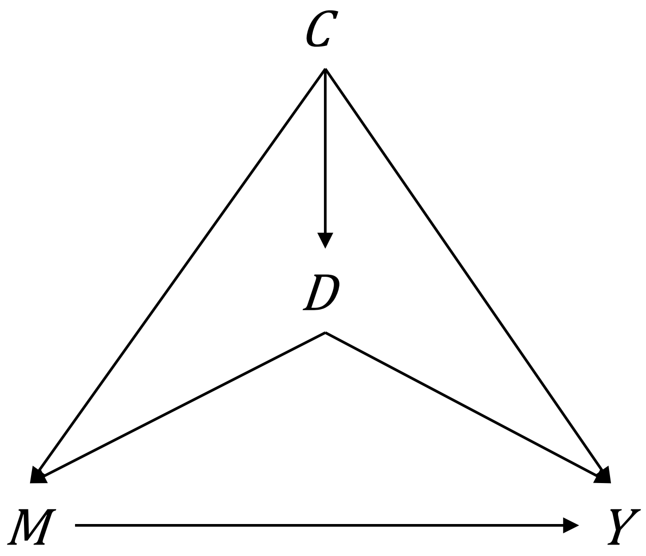
Week 5: Interventional Effects

# Outline

- Graphical mediation models
- The problem of exposure-induced confounding
- Controlled direct effects
- Interventional direct and indirect effects
- Nonparametric identification and estimation

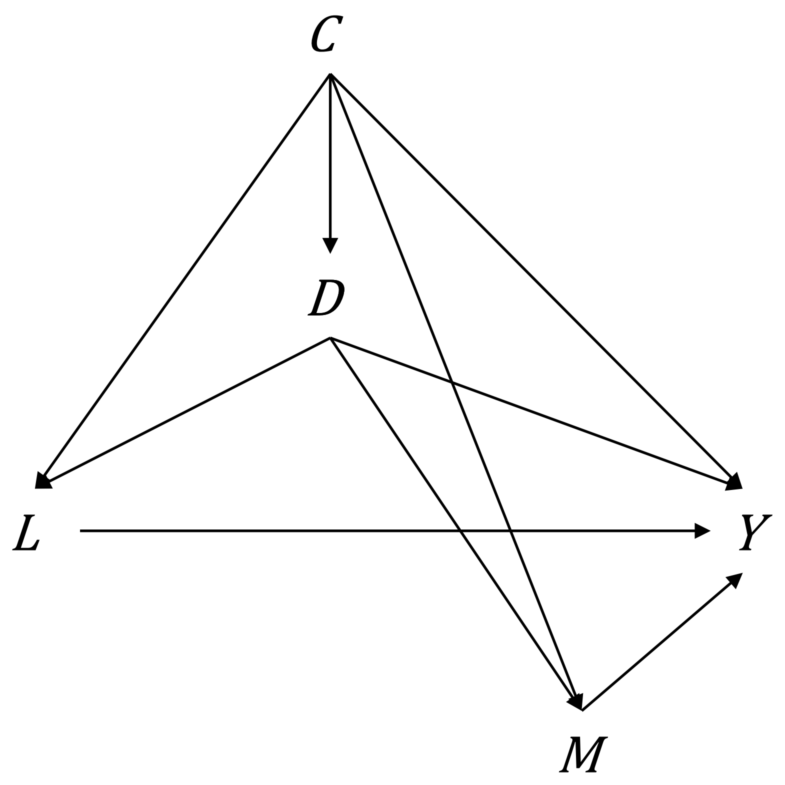
# A model with baseline confounding

- Thus far, we have been focusing on models with baseline confounding only
- All the methods we discuss this week and next can be used in applications with baseline confounding
- In addition, they can also be used in settings with more complex forms of confounding



# Models with multiple mediators

- In this model, the exposure $D$ affects two mediators, $L$ and $M$, which both affect the outcome $Y$
- $L$ does not affect $M$, nor does $M$ affect $L$
  - In other words, the two mediators are causally independent of each other
- In this setting, we can safely use the methods covered previously to analyze how $M$ mediates the effect of $D$ on $Y$
  - Likewise, we could also use methods covered previously to analyze how $L$ mediates the effect of $D$ on $Y$

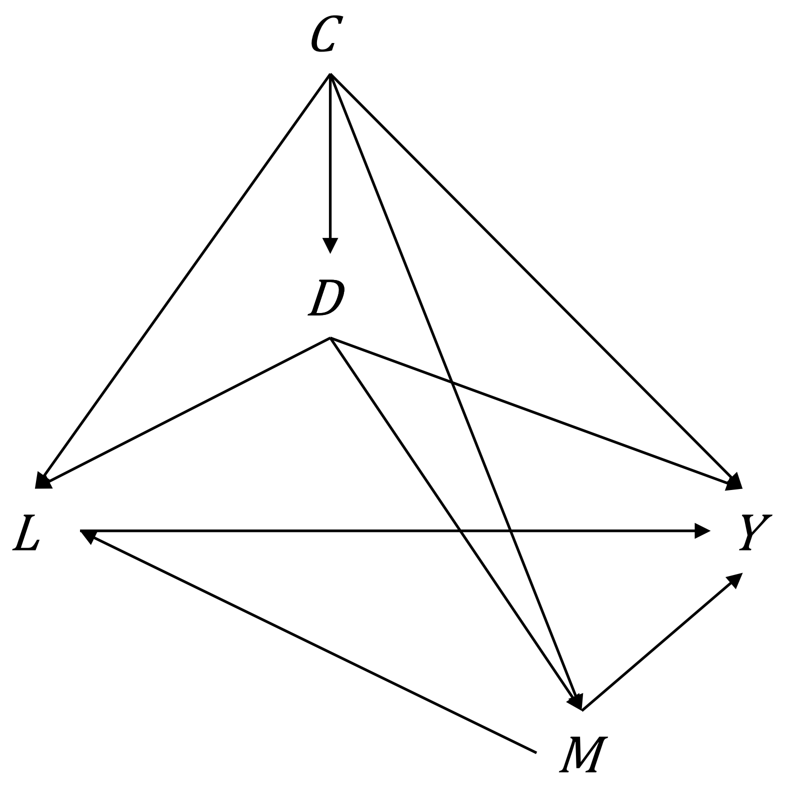


# Models with multiple mediators

- In this model, the exposure $D$ affects two mediators, $L$ and $M$, which both affect the outcome $Y$

- In addition, $M$ now affects $L$ as well

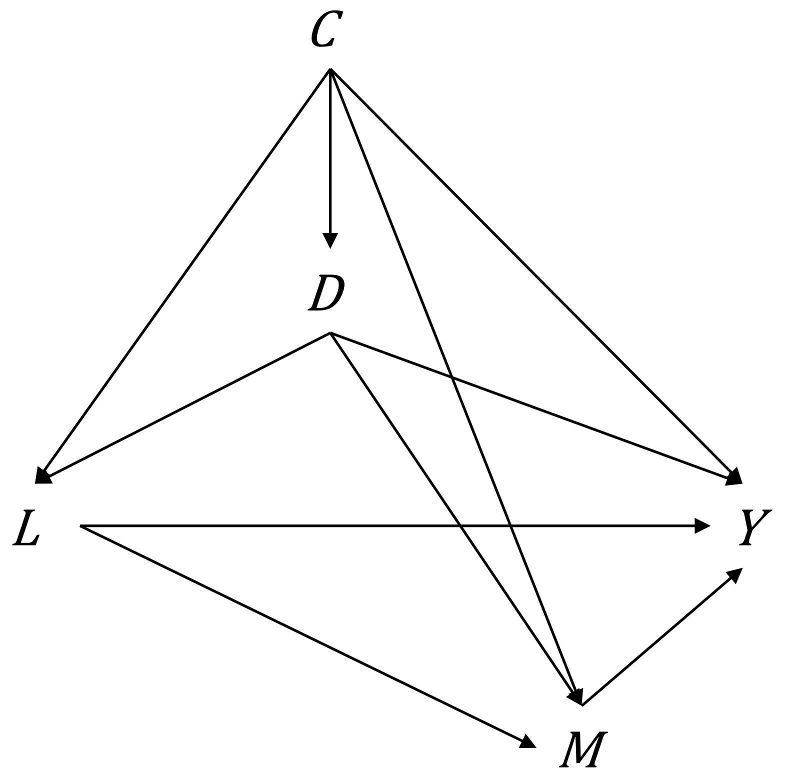
- In this setting, we can still safely use methods covered previously to analyze how $M$ mediates the effect of $D$ on $Y$
  - However, we cannot use methods covered previously to analyze how $L$ mediates the effect of $D$ on $Y$ because of exposure-induced confounding



# Models with multiple mediators

- In this model, the exposure $D$ affects two mediators, $L$ and $M$, which both affect the outcome $Y$, and $L$ now affects $M$

- In this setting, we can cannot use methods covered previously to analyze how $M$ mediates the effect of $D$ on $Y$
  - Natural effects cannot be nonparametrically identified in the presence of exposure-induced confounding
  - However, we could use the methods covered previously to analyze how $L$ mediates the effect of $D$ on $Y$

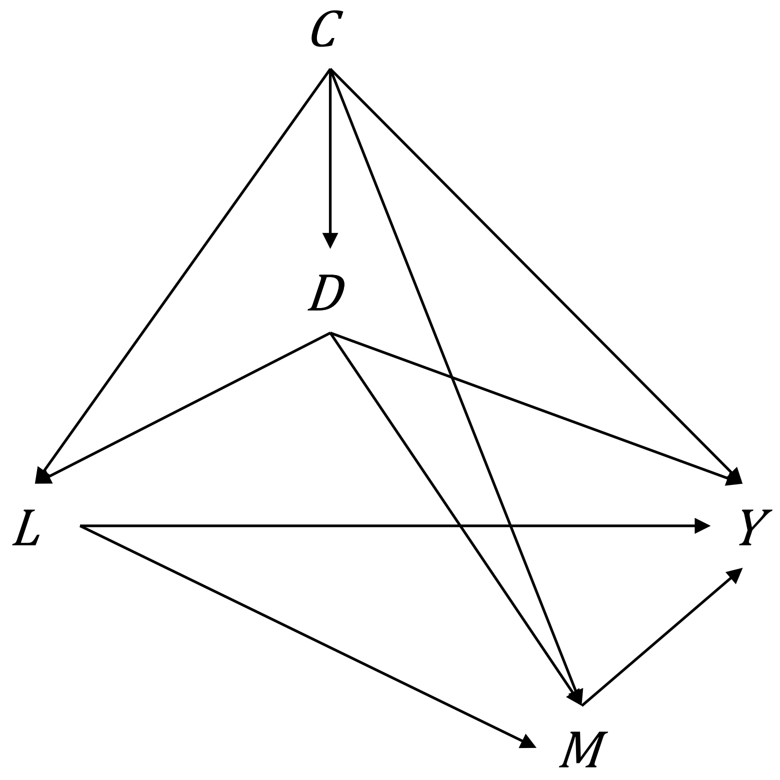


# A model with exposure-induced confounding

- In this model, $D$ affects $L$, which in turn affects both $M$ and $Y$

- Variables like $L$ are known as *exposure-induced confounders*

- The methods we cover this week and next are appropriate for evaluating…
  - how a single mediator $M$ transmits the effect of an exposure $D$ on an outcome $Y$…
  - …when exposure-induced confounders like $L$ are present



# Summary

- The methods covered this week and next are appropriate for data arising from a causal process resembling any of the graphical models depicted previously

- My presentation of these methods is tailored for models that allow general patterns of both baseline and exposure-induced confounding

- These methods are also appropriate, however, for settings without any baseline confounding or without any exposure-induced confounding

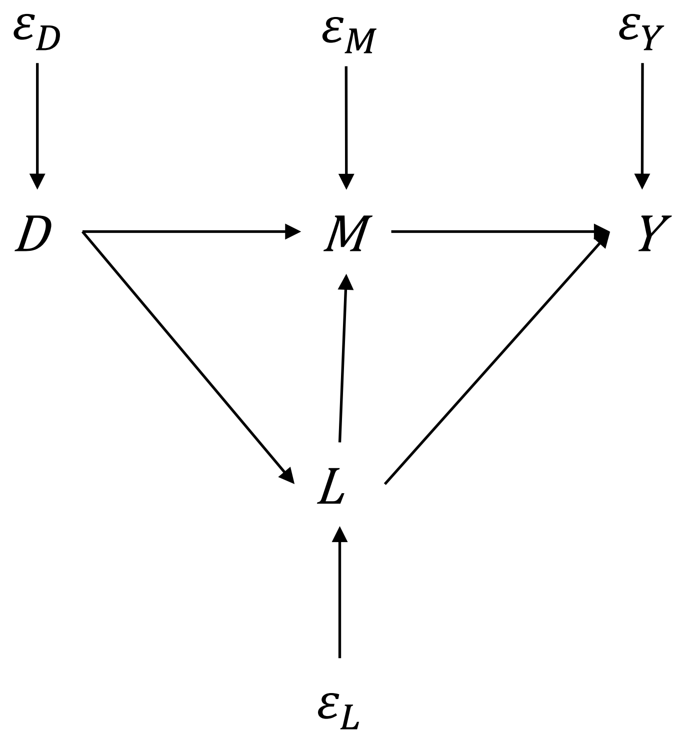
# Limitations of the natural effects decomposition

- Identifying natural direct and indirect effects hinges on the cross-world independence assumption:

$$Y(d,m) \perp M(d^*)|C$$

- This assumption is violated anytime there exists an exposure-induced confounder

- To appreciate this, consider the graphical model shown here…



# Limitations of the natural effects decomposition

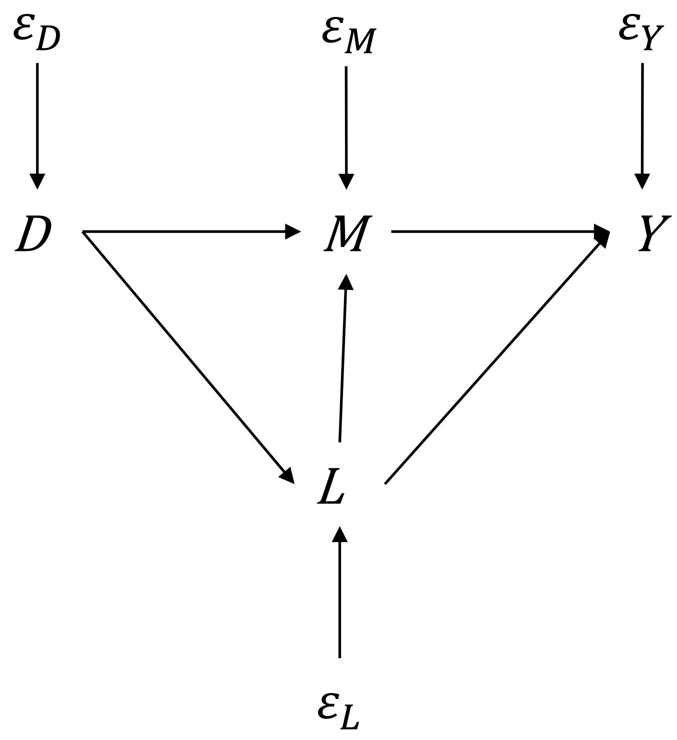
- This graphical model corresponds to the following set of nonparametric structural equations:

$$D \coloneqq f_D(\varepsilon_D)$$

$$L \coloneqq f_L(D, \varepsilon_L)$$

$$M \coloneqq f_M(D, L, \varepsilon_M)$$

$$Y \coloneqq f_Y(L, M, \varepsilon_Y)$$



# Limitations of the natural effects decomposition

- Under this model, the potential outcomes are generated as follows:

$$L(d) \coloneqq f_L(d, \varepsilon_L)$$

$$L(d^*) \coloneqq f_L(d^*, \varepsilon_L)$$

$$M(d^*) \coloneqq f_M(d^*, f_L(d^*, \varepsilon_L), \varepsilon_M)$$

$$Y(d, m) \coloneqq f_Y(f_L(d, \varepsilon_L), m, \varepsilon_Y)$$

- $M(d^*)$ is not independent of $Y(d, m)$ because both variables are affected by $\varepsilon_L$, the random component of $L$

$$\begin{tikzcd}
\varepsilon_D \arrow[d] & \varepsilon_M \arrow[d] & \varepsilon_Y \arrow[d] \\
D \arrow[r] \arrow[dr] & M \arrow[r] & Y \\
& L \arrow[u] \arrow[ur] & \\
& \varepsilon_L \arrow[u] &
\end{tikzcd}$$

# Limitations of the natural effects decomposition

- Under this model, the potential outcomes are generated as follows:

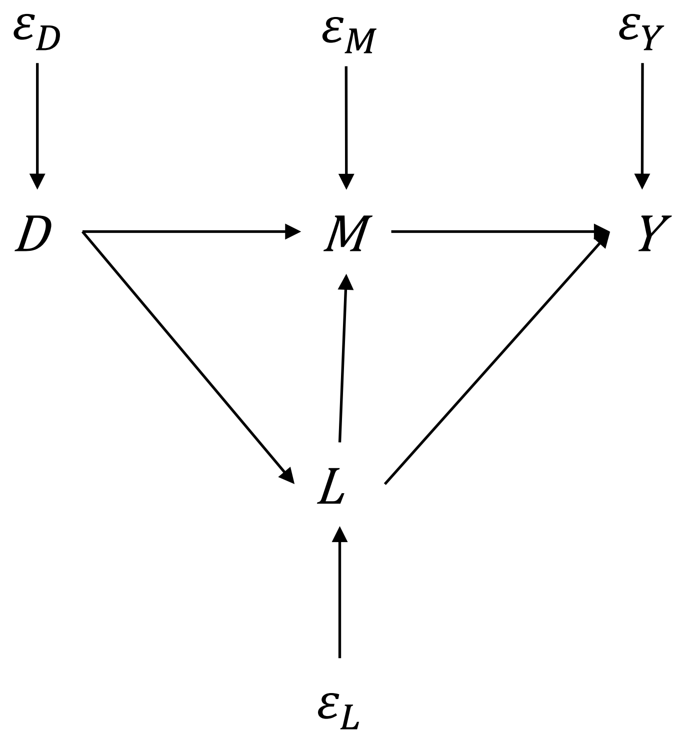
$$L(d) \coloneqq f_L(d, \varepsilon_L)$$

$$L(d^*) \coloneqq f_L(d^*, \varepsilon_L)$$

$$M(d^*) \coloneqq f_M(d^*, f_L(d^*, \varepsilon_L), \varepsilon_M)$$

$$Y(d, m) \coloneqq f_Y(f_L(d, \varepsilon_L), m, \varepsilon_Y)$$

- Moreover, $M(d^*)$ and $Y(d, m)$ are associated even conditional on the observed value of $L$



# Limitations of the natural effects decomposition

- Nonparametric identification of natural direct and indirect effects is essentially a nonstarter in the presence of exposure-induced confounders

- Parametric approaches to identification are still viable, however
    - For example, if $f_L(D, \varepsilon_L)$, $f_M(D, L, \varepsilon_M)$, and $f_Y(L, M, \varepsilon_Y)$ are all linear and additive, then the NDE and NIE can still be identified and consistently estimated

- Although parametric assumptions can facilitate identification for natural effects, these assumptions typically impose restrictions on the distribution of the data that cannot be substantiated by theory or prior knowledge

# The controlled direct effect

- Controlled direct effects are defined in terms of a joint intervention on the exposure and the mediator together

- They can be nonparametrically identified under weaker assumptions than natural effects

- Comparing the controlled direct effect with the total effect indicates whether the exposure impacts the outcome through a mechanism involving the mediator

    - The mechanism here could be either mediation (i.e., a causal chain) or interaction (i.e., the mediator modifies the effect of exposure)

    - The difference between the total effect and controlled direct effect is sometimes called the *eliminated effect*

# Joint potential outcomes

- Joint potential outcomes are defined in terms of an intervention on both the exposure and mediator together

- A joint potential outcome, denoted by $Y(d, m)$, is the value of the outcome that would occur if the exposure were set to $d$ and the mediator were set to $m$, possibly contrary to fact
  - With a binary exposure and mediator, for example…
    - $Y(1,1)$ is the outcome if an individual were exposed to treatment and the mediator
    - $Y(1,0)$ is the outcome if an individual were exposed to treatment but not the mediator
    - $Y(0,1)$ is the outcome if an individual were exposed to the mediator but not treatment
    - $Y(0,0)$ is the outcome if an individual were exposed to neither treatment nor the mediator

# The controlled direct effect

- The controlled direct effect:

$$CDE(d, d^*, m) = E\big(Y(d, m) - Y(d^*, m)\big)$$

- The $CDE(d, d^*, m)$ is the expected difference in the outcome if individuals had been exposed to $d$ rather than $d^*$ and if they had all experienced the same level of the mediator $m$

- It captures an effect of the exposure $D$ on the outcome $Y$ that persists after an intervention on the mediator $M$ that sets, or controls, its value at the same level for everyone

# The controlled direct effect

- The $CDE(d, d^*, m)$ isolates this effect by…
  - comparing outcomes across different levels of the exposure ($d$ versus $d^*$)…
  - while holding the mediator constant at a fixed value for everyone
- The $CDE(d, d^*, m)$ is not equal to the $NDE(d, d^*)$ except under special circumstances
- By extension, the difference between the $CDE(d, d^*, m)$ and the $ATE(d, d^*)$, known as the eliminated effect, does not in general have an interpretation as an indirect effect

# The controlled direct effect

- The controlled direct effect and the natural direct effect are only equivalent when there is no interaction effect between the exposure and mediator on the outcome *for all individuals*

- If the effect of contrasting exposure $d$ with $d^*$ is the same no matter the value of the mediator, then it makes no difference whether the mediator is set to $M(d^*)$, $m$, or to any other value

- By extension, the difference between the $CDE(d, d^*, m)$ and the $ATE(d, d^*)$ —that is, the eliminated effect—is equal to the natural indirect effect when there is no exposure-mediator interaction

# Nonparametric identification

- Controlled direct effects can be nonparametrically identified if the following conditions are met:

Assumption CE.1: $Y(d,m) \perp D|C$

Assumption CE.2: $Y(d,m) \perp M|C,D,L$

Assumption CE.3: $P(d|c) > 0$ and $P(m|c,d,l) > 0$

Assumption CE.4: $Y = Y(D,M)$

# No unobserved $D$-$Y$ confounding

- Assumption CE.1:

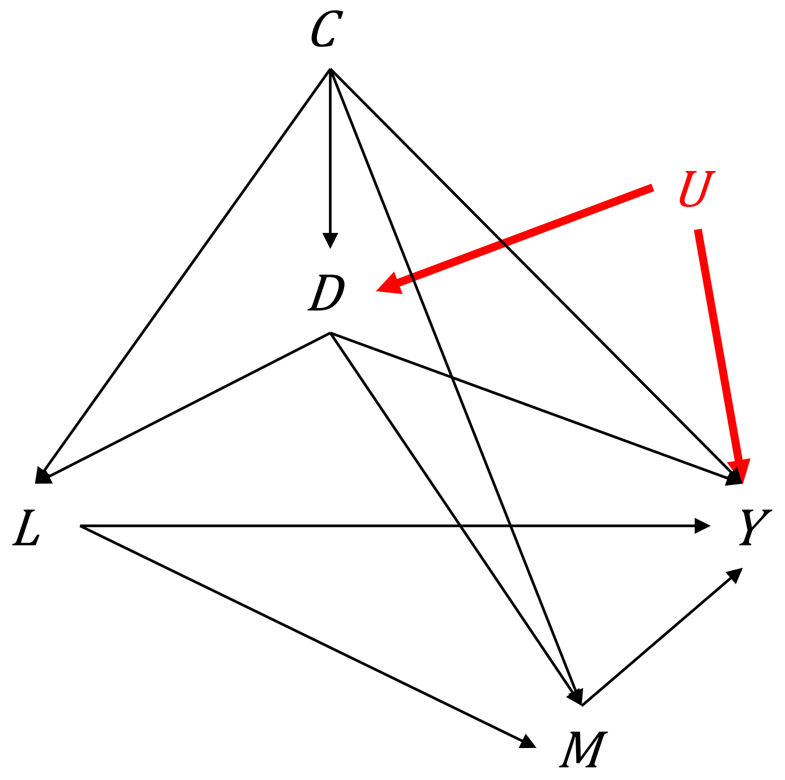
$$Y(d, m) \perp D|C$$

- This assumption requires that the exposure $D$ must be statistically independent of the joint potential outcomes $Y(d, m)$, conditional on the baseline confounders $C$

- Substantively, this assumption requires that there must not be any unobserved factors that confound the exposure-outcome relationship

# No unobserved $D$-$Y$ confounding

- Assumption CE.1 would be violated if an unobserved variable jointly affects the exposure and outcome
- In this graph, $U$ is an unobserved confounder for the $D \rightarrow Y$ relationship



# No unobserved $M$-$Y$ confounding

- Assumption CE.2:

$$Y(d,m) \perp M|C,D,L$$

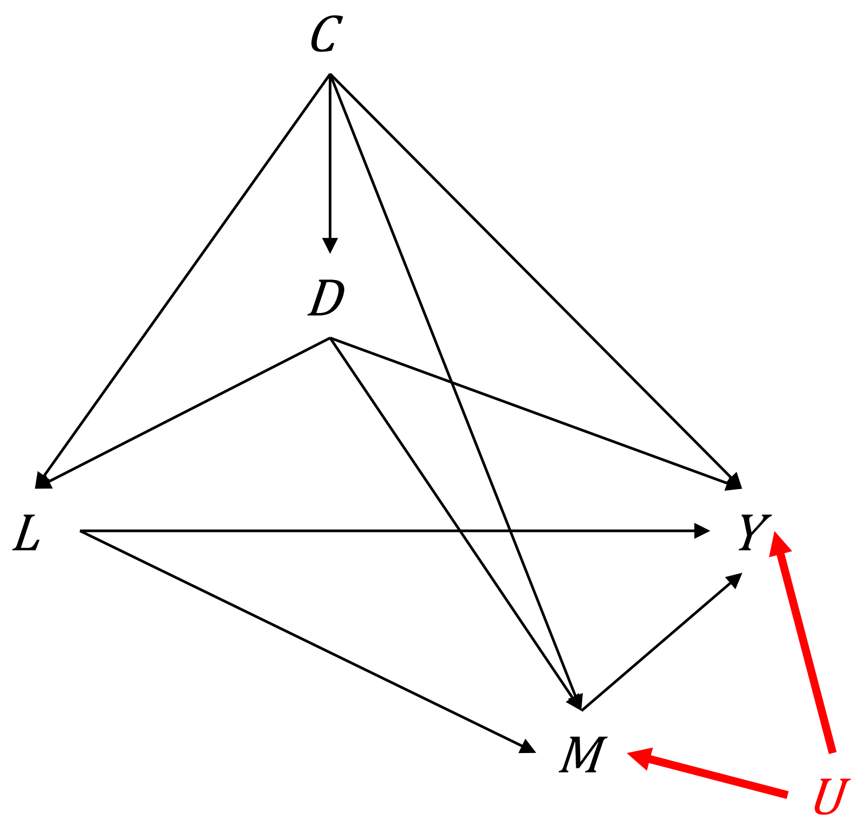
- This assumption requires that the mediator $M$ must be independent of the joint potential outcomes $Y(d,m)$, conditional on the baseline confounders $C$, prior exposure $D$, and any exposure-induced confounders $L$

- Substantively, this assumption requires that there must not be any unobserved factors that confound the mediator-outcome relationship

# No unobserved $M$-$Y$ confounding

- Assumption CE.2 would be violated if an unobserved variable jointly affects the mediator and outcome

- In this graph, $U$ is now an unobserved confounder for the $M \rightarrow Y$ relationship



# Positivity

- Assumption CE.3:

$$P(d|c) > 0 \text{ and } P(m|c,d,l) > 0$$

- This assumption requires that there must be a positive probability of all values for the exposure conditional on the baseline confounders

- It also requires that there must be a positive probability of all values for the mediator conditional on the baseline confounders, exposure, and exposure-induced confounders

# Consistency

- Assumption CE.4:

$$Y = Y(D, M)$$

- This assumption requires that the observed and joint potential outcomes are consistent with one another

# The fundamental problem of mediation revisited

- Unlike the conditions required for identifying natural effects, assumptions CE.1 to CE.4 can be met by experimental design

- In particular, the assumptions for identifying controlled direct effects can be met by design in a study where the exposure and mediator are jointly or sequentially randomized

  - <u>Joint randomization</u>: sample members are randomly assigned to different levels of both the exposure and mediator at baseline

  - <u>Sequential randomization</u>: sample members are first randomly assigned to different levels of the exposure at baseline, and then they are subsequently assigned to different levels of the mediator

# Nonparametric identification formula

- Under assumptions CE.1 to CE.4, the controlled direct effect can be equated with a function of observable data rather than the joint potential outcomes

- The nonparametric identification formula for the controlled direct effect can be expressed as follows:

$$
\begin{aligned}
CDE(d, d^*, m) &= E\big(Y(d, m) - Y(d^*, m)\big) \\
&= \sum_c \sum_l [E(Y|c, d, l, m)P(l|c, d) - E(Y|c, d^*, l, m)P(l|c, d^*)]P(c) \\
&= E_C\left(E_{L|c,d}\big(E(Y|C, d, L, m)\big) - E_{L|c,d^*}\big(E(Y|C, d^*, L, m)\big)\right)
\end{aligned}
$$

# Nonparametric identification formula

- If there are no exposure-induced confounders $L$, then the nonparametric identification formula simplifies to the following expression:

$$
\begin{aligned}
CDE(d, d^*, m) &= E\big(Y(d, m) - Y(d^*, m)\big) \\
&= \textstyle\sum_c [E(Y|c, d, m) - E(Y|c, d^*, m)]P(c) \\
&= E_C\big(E(Y|C, d, m) - E(Y|C, d^*, m)\big)
\end{aligned}
$$

# Randomized potential outcomes

- Randomized potential outcomes are a special type of potential outcome that capture stochastic rather than deterministic interventions

- Let $\mathcal{M}(d|C)$ denote a value of the mediator randomly selected from its distribution under exposure $d$, given the baseline confounders $C$

  - To elaborate, imagine a scenario where we assigned every individual in a target population to exposure $d$ and then subsequently measured their mediator

  - If we grouped these individuals into subpopulations based on the covariates in $C$, and then pooled all their mediator values together, these sets of values would compose the distribution of the mediator under exposure $d$, conditional on $C$

  - $\mathcal{M}(d|C)$ denotes a randomly selected value from this distribution of mediator values

# Randomized potential outcomes

- A randomized potential outcome, denoted by $Y\big(d, \mathcal{M}(d|C)\big)$, resembles a nested potential outcome but is defined in terms of a stochastic rather than deterministic intervention on the mediator

- Specifically, $Y\big(d, \mathcal{M}(d|C)\big)$ is the outcome that would occur if the exposure were set to $d$ and then the mediator were set to a value randomly drawn from its distribution under exposure $d$

- In general, $\mathcal{M}(d|C)$ may not equal $M(d)$, as the former value is a random draw from a distribution, while the later value is an individual's own potential value of the mediator

# Interventional effects decomposition

- Randomized potential outcomes allow us to define interventional direct and indirect effects

- Specifically, using these quantities, an overall effect of the exposure on the outcome can be decomposed into direct and indirect components as follows:

$$OE(d, d^*) = E\left(Y\big(d, \mathcal{M}(d|C)\big) - Y\big(d^*, \mathcal{M}(d^*|C)\big)\right)$$

$$= E\left(Y\big(d, \mathcal{M}(d^*|C)\big) - Y\big(d^*, \mathcal{M}(d^*|C)\big)\right) + E\left(Y\big(d, \mathcal{M}(d|C)\big) - Y\big(d, \mathcal{M}(d^*|C)\big)\right)$$

# Interventional effects decomposition

- Randomized potential outcomes allow us to define interventional direct and indirect effects

- Specifically, using these quantities, an overall effect of the exposure on the outcome can be decomposed into direct and indirect components as follows:

$$OE(d, d^*) = E\left(Y\big(d, \mathcal{M}(d|C)\big) - Y\big(d^*, \mathcal{M}(d^*|C)\big)\right)$$

$$= \underbrace{E\left(Y\big(d, \mathcal{M}(d^*|C)\big) - Y\big(d^*, \mathcal{M}(d^*|C)\big)\right)}_{\text{interventional direct effect}} + \underbrace{E\left(Y\big(d, \mathcal{M}(d|C)\big) - Y\big(d, \mathcal{M}(d^*|C)\big)\right)}_{\text{interventional indirect effect}}$$

# The interventional direct effect

- The interventional direct effect:

$$IDE(d, d^*) = E\left(Y\big(d, \mathcal{M}(d^*|C)\big) - Y\big(d^*, \mathcal{M}(d^*|C)\big)\right)$$

- The $IDE(d, d^*)$ is the expected difference in the outcome if individuals had been exposed to $d$ rather than $d^*$ and if they had experienced a level of the mediator randomly drawn from its distribution under exposure $d^*$

- It captures an effect of the exposure $D$ on the outcome $Y$ that does not operate through its influence on the distribution of the mediator $M$

# The interventional direct effect

- The $IDE(d, d^*)$ isolates an effect not involving the mediator by…
  - comparing outcomes across different levels of the exposure ($d$ versus $d^*$)…
  - while holding the mediator constant at a value randomly selected from its distribution under only one level of the exposure, $\mathcal{M}(d^*|C)$
- This comparison functions to deactivate the component of the overall effect that is transmitted through a causal chain from the exposure to the mediator to the outcome

# The interventional direct effect

- Interventional direct effects are closely related to controlled direct effects
- Specifically, the interventional direct effect is a weighted average of controlled direct effects evaluated across values of the mediator:

$$IDE(d, d^*) = E\left(Y\big(d, \mathcal{M}(d^*|C)\big) - Y\big(d^*, \mathcal{M}(d^*|C)\big)\right)$$

$$= \sum_c \sum_m E(Y(d, m) - Y(d^*, m)|c)\, P(M(d^*) = m|c)P(c)$$

$$= \sum_c \sum_m CDE(d, d^*, m|c)\, P(M(d^*) = m|c)P(c),$$

where $CDE(d, d^*, m|c)$ is a controlled direct effect conditional on $C = c$

# The interventional indirect effect

- The interventional indirect effect:

$$IIE(d, d^*) = E\left(Y\big(d, \mathcal{M}(d|C)\big) - Y\big(d, \mathcal{M}(d^*|C)\big)\right)$$

- The $IIE(d, d^*)$ is the expected difference in the outcome if individuals had been exposed to $d$ and then…
  - experienced a level of the mediator randomly drawn from its distribution under exposure $d$ rather than from its distribution under exposure $d^*$

- It captures an effect of the exposure $D$ on the outcome $Y$ that arises from a shift in the population distribution of the mediator caused by changes in the exposure

# The interventional indirect effect

- The $IIE(d, d^*)$ isolates an effect operating through the mediator by…
  - holding the exposure for each individual constant at $d$…
  - while comparing outcomes across levels of the mediator, $\mathcal{M}(d|C)$ versus $\mathcal{M}(d^*|C)$, chosen at random from two different counterfactual distributions
- This comparison deactivates all causal mechanisms connecting the exposure to the outcome except for a causal chain operating through the mediator

# The overall effect

- The overall effect:

$$OE(d,d^*) = E\left(Y\big(d,\mathcal{M}(d|C)\big) - Y\big(d^*,\mathcal{M}(d^*|C)\big)\right)$$

$$= IDE(d,d^*) + IIE(d,d^*)$$

- Similar to an average total effect, the $OE(d,d^*)$ it represents the expected difference in the outcome if…
  - individuals had been exposed to $d$ rather than $d^*$ and…
  - they had experienced a level of the mediator randomly selected from its distribution under exposure $d$ as opposed to its distribution under exposure $d^*$
- It captures the effect of a change in the exposure and a corresponding shift in the population distribution of the mediator

# Interventional versus natural effects

- When are interventional effects equivalent to natural effects?
  - When there is no interaction effect between the exposure and mediator on the outcome
    - In this situation, the impact of different exposures is invariant regardless of whether the mediator is set to its potential value, a random draw from the distribution of these potential values, or any other value
  - When there is no exposure-induced confounding
    - In applications with a single mediator of interest and baseline confounding only, natural effects can be interpreted as interventional effects, and vice versa

# The fundamental problem revisited

- Unlike the natural effects decomposition, where we never observe the cross-world potential outcome $Y\big(d, M(d^*)\big)$ for anyone, randomized potential outcomes can be observed, at least in principle

- Specifically, $Y\big(d, \mathcal{M}(d^*|C)\big)$ can be observed for a subset of individuals by assigning them to exposure $d$ together with a value of the mediator randomly drawn from its distribution under $d^*$

- This unlocks the possibility of identification via experimental design

# Nonparametric identification

- Interventional effects can be nonparametrically identified if the following conditions are met:

Assumption IE.1: $Y(d,m) \perp D|C$

Assumption IE.2: $Y(d,m) \perp M|C,D,L$

Assumption IE.3: $M(d) \perp D|C$

Assumption IE.4: $P(d|c) > 0$ and $P(m|c,d,l) > 0$

Assumption IE.5: $Y = Y(D,M)$ and $M = M(D)$

# Nonparametric identification

- Interventional effects can be nonparametrically identified if the following conditions are met:

Assumption IE.1: $Y(d,m) \perp D|C$ ← Same as CE.1

Assumption IE.2: $Y(d,m) \perp M|C,D,L$ ← Same as CE.2

Assumption IE.3: $M(d) \perp D|C$ ← New

Assumption IE.4: $P(d|c) > 0$ and $P(m|c,d,l) > 0$ ← Same as CE.3

Assumption IE.5: $Y = Y(D,M)$ and $M = M(D)$ ← Modified

# No unobserved $D$-$M$ confounding

- Assumption IE.3:

$$M(d) \perp D|C$$

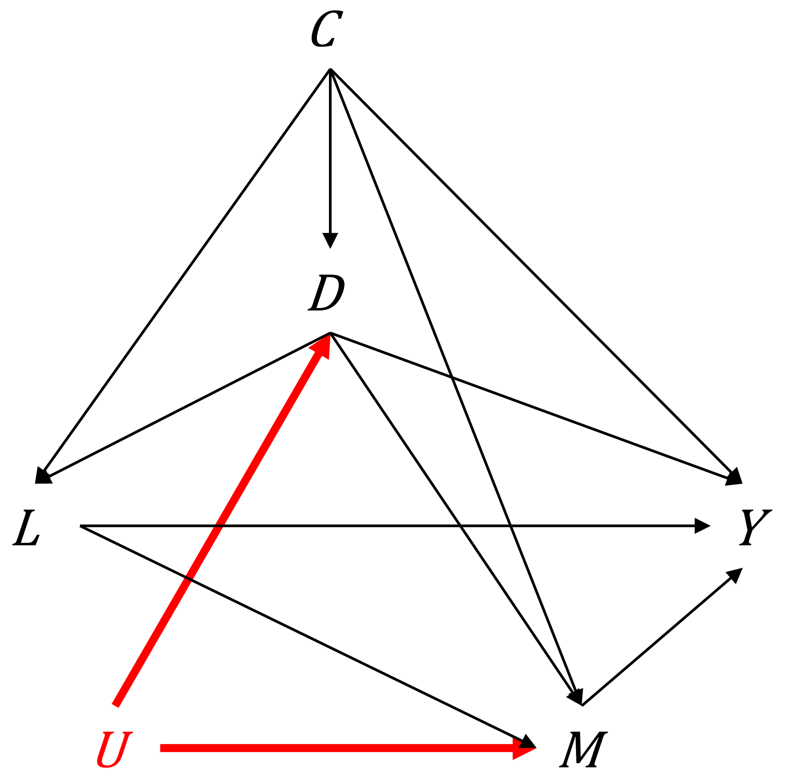
- This assumption requires that the exposure $D$ must be statistically independent of the potential values of the mediator $M(d)$, conditional on the baseline confounders $C$

- Substantively, this assumption requires that there must not be any unobserved factors that confound the exposure-mediator relationship

# No unobserved $D$-$M$ confounding

- Assumption IE.3 would be violated if an unobserved variable jointly affects the exposure and mediator

- In this graph, $U$ is now an unobserved confounder for the $D \rightarrow M$ relationship



# Consistency

- Assumption IE.5:

$$Y = Y(D, M) \text{ and } M = M(D)$$

- This assumption requires that the observed outcome must be consistent with the joint potential outcomes

- It also requires that the observed value of the mediator is consistent with its potential values

# The fundamental problem revisited

- Unlike the conditions required for identifying natural effects, assumptions IE.1 to IE.5 can be met by experimental design, at least in principle

- Consider the following multi-arm experiment, where participants are randomly assigned to each arm:
  - **Arm 1:** participants are assigned to $d$ and then their mediator is measured
  - **Arm 2:** participants are assigned to $d^*$ and then their mediator is measured
  - **Arm 3:** participants are assigned to $d$ with mediator randomly drawn from Arm 1
  - **Arm 4:** participants are assigned to $d^*$ with mediator randomly drawn from Arm 2
  - **Arm 5:** participants are assigned to $d$ with mediator randomly drawn from Arm 2

# Nonparametric identification formula

- Under assumptions IE.1 to IE.5, the interventional direct effect can be equated with a function of observable data rather than the randomized potential outcomes

- The nonparametric identification formula for the interventional direct effect can be expressed as follows:

$$IDE(d,d^*) = E\left(Y\left(d,\mathcal{M}(d^*|C)\right) - Y\left(d^*,\mathcal{M}(d^*|C)\right)\right)$$

$$= \sum_c \sum_m \sum_l [E(Y|c,d,l,m)P(l|c,d) - E(Y|c,d^*,l,m)P(l|c,d^*)]P(m|c,d^*)P(c)$$

$$= E_C\left(E_{M|C,d^*}\left(E_{L|C,d}\left(E(Y|C,d,L,M)\right) - E_{L|C,d^*}\left(E(Y|C,d^*,L,M)\right)\right)\right)$$

# Nonparametric identification formula

- Under the same set of assumptions, the interventional indirect effect can also be equated with a function of observable data rather than the randomized potential outcomes

- The nonparametric identification formula for the interventional indirect effect can be expressed as follows:

$$IIE(d, d^*) = E\left(Y\big(d, \mathcal{M}(d|C)\big) - Y\big(d, \mathcal{M}(d^*|C)\big)\right)$$

$$= \sum_c \sum_m \sum_l [P(m|c, d) - P(m|c, d^*)]\, E(Y|c, d, l, m) P(l|c, d) P(c)$$

$$= E_C\left(E_{M|C,d}\left(E_{L|C,d}\big(E(Y|C, d, M)\big)\right) - E_{M|C,d^*}\left(E_{L|C,d}\big(E(Y|C, d, M)\big)\right)\right)$$

# Nonparametric estimation

- Nonparametric identification involves equating causal effects defined in terms of counterfactuals with empirical quantities defined in terms of observable data, while ignoring random variability due to sampling

- Nonparametric estimation involves plugging in sample analogs for the population quantities in the nonparametric identification formulae outlined previously

# Nonparametric estimation of IDE

- A nonparametric estimator for the interventional direct effect:

$$\widehat{IDE}(d,d^*)^{np} = \sum_c \sum_m \sum_l \left[\hat{E}(Y|c,d,l,m)\hat{P}(l|c,d) - \hat{E}(Y|c,d^*,l,m)P(l|c,d^*)\right]\hat{P}(m|c,d^*)\hat{P}(c)$$

$$= \sum_c \sum_m \sum_l \left[\bar{Y}_{c,d,l,m}\hat{\pi}_{l|c,d} - \bar{Y}_{c,d^*,l,m}\hat{\pi}_{l|c,d^*}\right]\hat{\pi}_{m|c,d^*}\hat{\pi}_c$$

  - $\bar{Y}_{c,d,l,m}$ denotes the mean outcome among sample members with $C=c$, $D=d$, $L=l$, and $M=m$
  - $\hat{\pi}_{l|c,d}$ is the proportion of sample members for whom $L=l$ among those with $C=c$ and $D=d$
  - $\hat{\pi}_{m|c,d}$ is the proportion of sample members for whom $M=m$ among those with $C=c$ and $D=d$
  - $\hat{\pi}_c$ is the proportion of sample members with $C=c$

# Nonparametric estimation of IIE

- A nonparametric estimator for the interventional indirect effect:

$$\widehat{IIE}(d, d^*)^{np} = \sum_c \sum_m \sum_l \left[\hat{P}(m|c, d) - \hat{P}(m|c, d^*)\right]\hat{E}(Y|c, d, l, m)\hat{P}(l|c, d)\hat{P}(c)$$

$$= \sum_c \sum_m \sum_l \left[\hat{\pi}_{m|c,d} - \hat{\pi}_{m|c,d^*}\right]\bar{Y}_{c,d,l,m}\hat{\pi}_{l|c,d}\hat{\pi}_c$$

  - $\bar{Y}_{c,d,l,m}$ denotes the mean outcome among sample members with $C = c$, $D = d$, $L = l$, and $M = m$
  - $\hat{\pi}_{l|c,d}$ is the proportion of sample members for whom $L = l$ among those with $C = c$ and $D = d$
  - $\hat{\pi}_{m|c,d}$ is the proportion of sample members for whom $M = m$ among those with $C = c$ and $D = d$
  - $\hat{\pi}_c$ is the proportion of sample members with $C = c$

# Summary

- Controlled direct effects capture an effect of the exposure that would persist under an intervention that sets the mediator to the same fixed value for everyone
- The interventional effects decomposition permits a separation of overall effects into components operating through a single mediator of interest versus other mechanisms, similar to the natural effect decomposition
- Interventional and controlled effects can be nonparametrically identified and estimated under assumptions that are…
  - weaker than those required for identifying natural effects and…
  - can in principle be met by experimental design

# Example: NLSY79

- 1979 National Longitudinal Study of Youth
  - Exposure ($D$)
    - sample member attended college before age 22
  - Outcome ($Y$):
    - standardized scores on the CES-D at age 40
  - Covariates ($C$):
    - Mother attended college
  - A potential mediator ($M$)
    - household income between age 35-40
  - A potential exposure-induced confounder ($L$)
    - unemployment between age 35-40

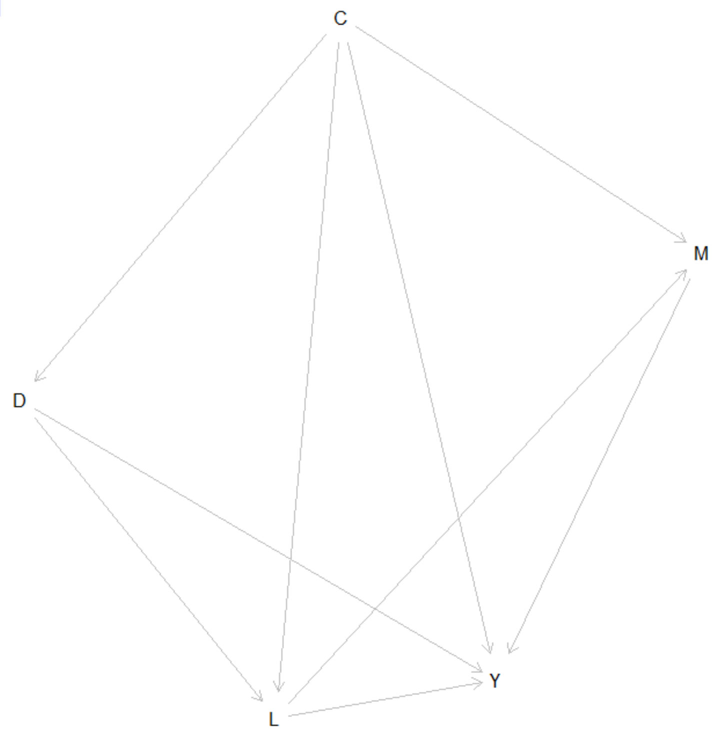
# Example: NLSY79

- Many studies have documented that going to college seems to reduce the likelihood of becoming depressed later in life—but how does this effect come about?
- One possibility is that a more advanced education reduces depression by increasing the income they have at their disposal
  - Does income mediate the effect of college attendance on depression?
- However, unemployment may independently affect both income and depression, and it is affected by college attendance
  - In other words, unemployment may be an exposure-induced confounder

# Example: NLSY79

- Draw a DAG that best represents the causal relations among these variables, based on theory and prior knowledge

```r
### wk 5 nlsy tutorial ###
rm(list=ls())

# load/install libraries #
packages<-c("dplyr", "tidyr", "foreign", "dagitty", "margins")
install.packages(packages)

for (package.i in packages) {
    suppressPackageStartupMessages(library(package.i, character.only=TRUE))
    }

# specify DAG #
nlsyDAG <- dagitty( "dag {
    C -> D -> L -> M -> Y
    C -> M
    C -> Y
    C -> L
    D -> Y
    L -> Y
}")

plot(graphLayout(nlsyDAG))

```



# Example: NLSY79

- Can the interventional direct and indirect effects of college attendance on depression, as mediated by income, be nonparametrically identified from the observed data?
  - Why or why not?
  - What assumptions are necessary?
  - Which are violated?

# Example: NLSY79

- Suppose, for the sake of illustration, that the interventional effects of interest can be identified by adjusting only for…
  - whether or not a respondent's mother attended college and…
  - whether the respondent experienced a spell of unemployment
- Suppose further that the mediating role of income can be captured by a single binary variable, coded 1 for those who earn >$50K and 0 otherwise
- Under these suppositions, compute all the components of nonparametric estimators for the interventional direct and indirect effects

# Example: NLSY79

- Under these suppositions, compute all the components of nonparametric estimators for the interventional direct and indirect effects

```r
# load data #
datadir <- "C:/Users/Geoffrey Wodtke/Dropbox/D/courses/2023-24_UOFCHICAGO/SOCI_40258_CAUSAL_MEDIATION/data/"
nlsy <- read.dta(paste(datadir, "nlsy79.dta", sep=""))

nlsy <- nlsy[complete.cases(nlsy[,
    c("ever_unemp_age3539", "momedu", "att22", "cesd_age40", "faminc_adj_age3539")]),]

nlsy$momcol <- ifelse(nlsy$momedu>12, 1, 0)

nlsy$incgt50k <- ifelse(nlsy$faminc_adj_age3539>=50000, 1, 0)

nlsy$std_cesd_age40 <- (nlsy$cesd_age40-mean(nlsy$cesd_age40))/sd(nlsy$cesd_age40)

# components of np estimator #

nptab <-nlsy %>%
        group_by(momcol, att22, incgt50k, ever_unemp_age3539) %>%
            dplyr::summarize(
                mean = mean(std_cesd_age40),
                n = n(),
                .groups = "drop")

print(nptab)
```

# Example: NLSY79

- Under these suppositions, compute all the components of nonparametric estimators for the interventional direct and indirect effects

```
> print(nptab)
# A tibble: 16 × 6
   momcol att22 incgt50k ever_unemp_age3539    mean     n
    <dbl> <dbl>    <dbl>              <dbl>   <dbl> <int>
 1      0     0        0                  0  0.0717  1219
 2      0     0        0                  1  0.349    460
 3      0     0        1                  0 -0.172    528
 4      0     0        1                  1  0.0381    60
 5      0     1        0                  0 -0.0744   211
 6      0     1        0                  1  0.242     56
 7      0     1        1                  0 -0.303    358
 8      0     1        1                  1 -0.159     36
 9      1     0        0                  0  0.0510    98
10      1     0        0                  1  0.392     36
11      1     0        1                  0 -0.107     74
12      1     0        1                  1  0.344      3
13      1     1        0                  0  0.0254    85
14      1     1        0                  1  0.140     18
15      1     1        1                  0 -0.222    247
16      1     1        1                  1 -0.227     24
>
```

# Example: NLSY79

- Under the same suppositions, compute a nonparametric estimate of the interventional direct effect

```r
# compute nonparametric estimates of IDE #
YhatC0D0L0M0 <- nptab[1,5]
YhatC0D0L1M0 <- nptab[2,5]
YhatC0D0L0M1 <- nptab[3,5]
YhatC0D0L1M1 <- nptab[4,5]

YhatC0D1L0M0 <- nptab[5,5]
YhatC0D1L1M0 <- nptab[6,5]
YhatC0D1L0M1 <- nptab[7,5]
YhatC0D1L1M1 <- nptab[8,5]

YhatC1D0L0M0 <- nptab[9,5]
YhatC1D0L1M0 <- nptab[10,5]
YhatC1D0L0M1 <- nptab[11,5]
YhatC1D0L1M1 <- nptab[12,5]

YhatC1D1L0M0 <- nptab[13,5]
YhatC1D1L1M0 <- nptab[14,5]
YhatC1D1L0M1 <- nptab[15,5]
YhatC1D1L1M1 <- nptab[16,5]

phatL1_C0D0 <- mean(nlsy[which(nlsy$momcol==0 & nlsy$att22==0), "ever_unemp_age3539"])
phatL1_C0D1 <- mean(nlsy[which(nlsy$momcol==0 & nlsy$att22==1), "ever_unemp_age3539"])
phatL1_C1D0 <- mean(nlsy[which(nlsy$momcol==1 & nlsy$att22==0), "ever_unemp_age3539"])
phatL1_C1D1 <- mean(nlsy[which(nlsy$momcol==1 & nlsy$att22==1), "ever_unemp_age3539"])

phatM1_C0D0 <- mean(nlsy[which(nlsy$momcol==0 & nlsy$att22==0), "incgt50k"])
phatM1_C0D1 <- mean(nlsy[which(nlsy$momcol==0 & nlsy$att22==1), "incgt50k"])
phatM1_C1D0 <- mean(nlsy[which(nlsy$momcol==1 & nlsy$att22==0), "incgt50k"])
phatM1_C1D1 <- mean(nlsy[which(nlsy$momcol==1 & nlsy$att22==1), "incgt50k"])

phatC1 <- mean(nlsy$momcol)
```

# Example: NLSY79

- Under the same suppositions, compute a nonparametric estimate of the interventional direct effect

```r
IDEhat_npl <-
    ((YhatC0D1L0M1*(1-phatL1_C0D1) +
        YhatC0D1L1M1*phatL1_C0D1) -
        (YhatC0D0L0M1*(1-phatL1_C0D0) +
        YhatC0D0L1M1*phatL1_C0D0))*phatM1_C0D0*(1-phatC1) +
    ((YhatC1D1L0M1*(1-phatL1_C1D1) +
        YhatC1D1L1M1*phatL1_C1D1) -
        (YhatC1D0L0M1*(1-phatL1_C1D0) +
        YhatC1D0L1M1*phatL1_C1D0))*phatM1_C1D0*(phatC1) +
    ((YhatC0D1L0M0*(1-phatL1_C0D1) +
        YhatC0D1L1M0*phatL1_C0D1) -
        (YhatC0D0L0M0*(1-phatL1_C0D0) +
        YhatC0D0L1M0*phatL1_C0D0))*(1-phatM1_C0D0)*(1-phatC1) +
    ((YhatC1D1L0M0*(1-phatL1_C1D1) +
        YhatC1D1L1M0*phatL1_C1D1) -
        (YhatC1D0L0M0*(1-phatL1_C1D0) +
        YhatC1D0L1M0*phatL1_C1D0))*(1-phatM1_C1D0)*(phatC1)

print(IDEhat_npl)
```

```
> print(IDEhat_npl)
       mean
1 -0.1566999
>
```

# Example: NLSY79

- Under the same suppositions, compute a nonparametric estimate of the interventional indirect effect

```
IIEhat_npl <-
    (YhatC0D1L0M1*(1-phatL1_C0D1) +
        YhatC0D1L1M1*phatL1_C0D1)*(1-phatC1)*(phatM1_C0D1 - phatM1_C0D0) +
    (YhatC1D1L0M1*(1-phatL1_C1D1) +
        YhatC1D1L1M1*phatL1_C1D1)*phatC1*(phatM1_C1D1 - phatM1_C1D0) +
    (YhatC0D1L0M0*(1-phatL1_C0D1) +
        YhatC0D1L1M0*phatL1_C0D1)*(1-phatC1)*((1-phatM1_C0D1) - (1-phatM1_C0D0)) +
    (YhatC1D1L0M0*(1-phatL1_C1D1) +
        YhatC1D1L1M0*phatL1_C1D1)*phatC1*((1-phatM1_C1D1) - (1-phatM1_C1D0))

print(IIEhat_npl)
```

```
> print(IIEhat_npl)
         mean
1 -0.08643233
>
```

# Example: NLSY79

- Under the same suppositions, compute a nonparametric estimate of the overall effect

```
OEhat_npl <- IDEhat_npl + IIEhat_npl

print(OEhat_npl)

```

```
> print(OEhat_npl)
       mean
1 -0.2431323
>
```

# Example: NLSY79

- Under the same suppositions, compute a nonparametric estimate of the controlled direct effect

```r
CDEhat_npl <-
    ((YhatC0D1L0M1*(1-phatL1_C0D1) +
        YhatC0D1L1M1*phatL1_C0D1)*(1-phatC1) +
        (YhatC1D1L0M1*(1-phatL1_C1D1) +
        YhatC1D1L1M1*phatL1_C1D1)*(phatC1)) -
    ((YhatC0D0L0M1*(1-phatL1_C0D0) +
        YhatC0D0L1M1*phatL1_C0D0)*(1-phatC1) +
        (YhatC1D0L0M1*(1-phatL1_C1D0) +
        YhatC1D0L1M1*phatL1_C1D0)*(phatC1))

print(CDEhat_npl)
```

```
> print(CDEhat_npl)
        mean
1 -0.1658224
>
```