# SOCI 40258

Causal Mediation Analysis

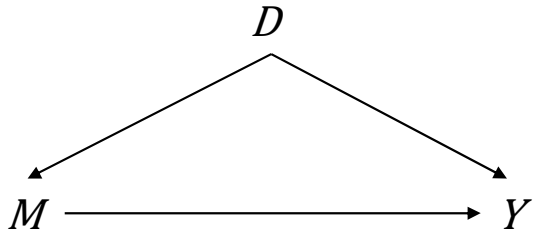
Week 3: The Natural Effects Decomposition

# Outline

- Graphical mediation models
- Joint, nested, and cross-world potential outcomes
- Natural direct and indirect effects
- Nonparametric identification
- Nonparametric estimation

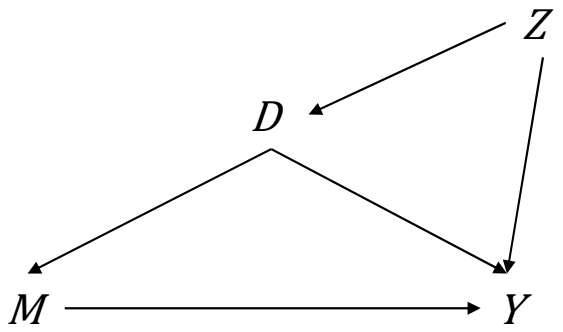
# A simple mediation model

- $D$ is an exposure of interest, $Y$ is an outcome, and $M$ is a putative mediator
- $D$ directly affects $M$, which in turn directly affects $Y$, as indicated by the causal chain $D \rightarrow M \rightarrow Y$
- $D$ also directly affects $Y$, as indicated by the $D \rightarrow Y$ path



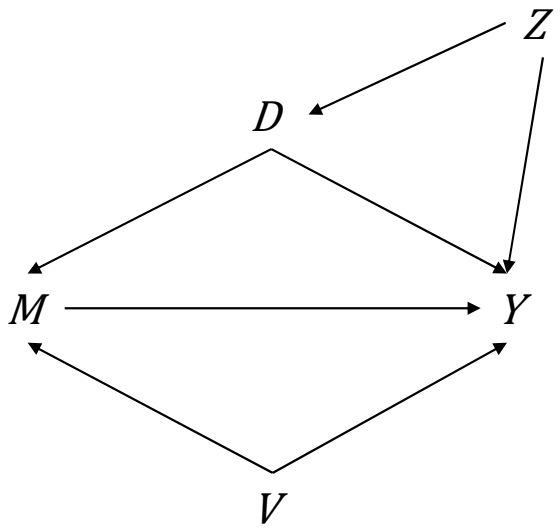
# A model with $D \leftarrow Z \rightarrow Y$ confounding

- Confounding occurs when two variables share a common cause, known as a confounder
- In this model, $D$ affects $Y$ directly and indirectly through $M$, as before
- However, the effect of $D$ on $Y$ is now confounded by $Z$
- $Z$ is known as an exposure-outcome confounder



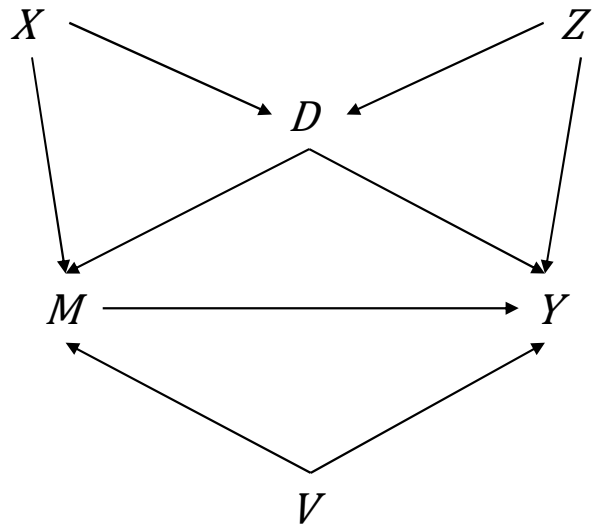
# A model with $M \leftarrow V \rightarrow Y$ confounding

- In this model, $D$ affects $Y$ directly and indirectly through $M$, as before
- In addition to exposure-outcome confounding by $Z$, the effect of $M$ on $Y$ is also now confounded by $V$
- $V$ is known as a mediator-outcome confounder



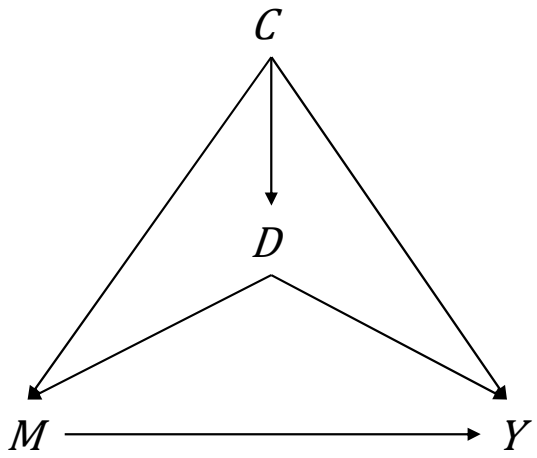
# A model with $D \leftarrow X \rightarrow M$ confounding

- In this model, $D$ affects $Y$ directly and indirectly through $M$, as before
- There is both exposure-outcome and mediator-outcome confounding
- In addition, the effect of $D$ on $M$ is now confounded by $X$
- $X$ is known as an exposure-mediator confounder



# A model with baseline confounding

- In this model, $D$ affects $Y$ directly and indirectly through $M$, as before
- The confounder, denoted by $C$, jointly affects the exposure, mediator, and outcome of interest
- Thus, $C$ confounds the exposure-outcome, exposure-mediator, and mediator-outcome relationships simultaneously
- We will refer to variables like $C$ as <u>baseline confounders</u>



# Graphical mediation models

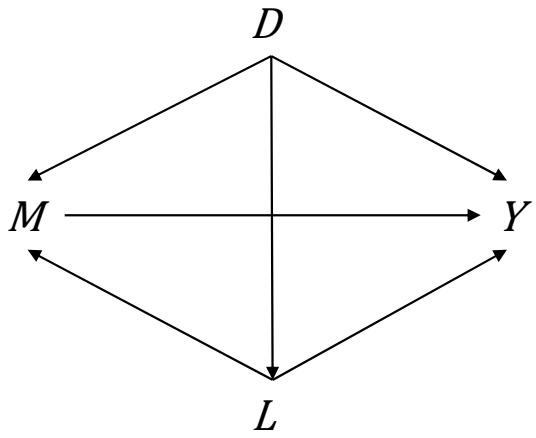
- The methods covered this week and next are appropriate for data arising from a causal process resembling the graphical models depicted previously

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

- But these methods are also appropriate for settings without any confounding or for applications with exposure-outcome, mediator-outcome, or exposure-mediator confounding only

# Exposure-induced confounding

- In this model, $D$ affects $L$, which in turn affects both $M$ and $Y$
- Thus, $L$ is a mediator-outcome confounder that is affected by the exposure
- Variables like $L$ are known as exposure-induced confounders
- The methods we cover this week and next are not appropriate if exposure-induced confounders are present



# Conventional potential outcomes

- Potential outcomes provide a notation for defining measures of causation
- A potential outcome, denoted by $Y(d)$, is the value of the outcome that would occur if the exposure were equal to $d$, possibly contrary to fact
  - For example, with a binary exposure:
    - $Y(1)$ is the value for the outcome that would occur if an individual were exposed
    - $Y(0)$ is the value for the outcome that would occur if an individual were not exposed
- These are the same as the potential outcomes we previously used to define total effects; we will henceforth refer to them as <u>conventional potential outcomes</u>

# 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 equal to $d$ and the mediator were equal to $m$, possibly contrary to fact
  - For example, with a binary exposure and mediator:
    - $Y(1,1)$ 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

# Nested potential outcomes

- Nested potential outcomes are a special type of joint potential outcome
- A nested potential outcome, here denoted by $Y\big(d, M(d)\big)$, is the outcome that would occur if the exposure were equal to $d$ and, by extension, the mediator were equal to its value that would arise naturally under exposure $d$
- This quantity is known as a nested potential outcome because the potential value for the mediator $M$, denoted by $M(d)$, is nested within the potential outcome for $Y$

# Cross-world potential outcomes

- Cross-world potential outcomes are a special type of nested potential outcome
- A cross-world potential outcome, such as $Y\big(d, M(d^*)\big)$, is the outcome that would occur if the exposure were equal to $d$ but the mediator were equal to its value that would arise naturally under another exposure $d^*$
- This quantity is known as a <u>cross-world potential outcome</u>
  - It involves setting the exposure at one value $d$ and then setting the mediator at its value from an alternative world in which the exposure was set to $d^*$ instead

# The natural effects decomposition

- Nested and cross-world potential outcomes allow us to formalize the concept of causal mediation

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

$$
\begin{aligned}
ATE(d, d^*) &= E\big(Y(d) - Y(d^*)\big) \\
&= E\left(Y\big(d, M(d)\big) - Y\big(d^*, M(d^*)\big)\right) \\
&= E\left(Y\big(d, M(d^*)\big) - Y\big(d^*, M(d^*)\big)\right) + E\left(Y\big(d, M(d)\big) - Y\big(d, M(d^*)\big)\right)
\end{aligned}
$$

# The natural effects decomposition

- Nested and cross-world potential outcomes allow us to formalize the concept of causal mediation

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

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

$$= E\left(Y\big(d, M(d)\big) - Y\big(d^*, M(d^*)\big)\right)$$

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

# The natural direct effect

- The natural direct effect:

$$NDE(d, d^*) = E\left(Y\left(d, M(d^*)\right) - Y\left(d^*, M(d^*)\right)\right)$$

- The $NDE(d, d^*)$ is the expected difference in the outcome if individuals had been exposed to $d$ rather than $d^*$ and if they had experienced the level of the mediator that would have arisen naturally for them under exposure $d^*$

- It captures an effect of the exposure $D$ on the outcome $Y$ that operates through all mechanisms other than those involving the mediator $M$

# The natural direct effect

- The natural direct effect:

$$NDE(d, d^*) = E\left(Y\left(d, M(d^*)\right) - Y\left(d^*, M(d^*)\right)\right)$$

- The $NDE(d, d^*)$ 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 its value under only one level of the exposure $M(d^*)$
- This comparison deactivates the component of the total effect that is transmitted through a causal chain from the exposure to the mediator to the outcome, denoted by $D \rightarrow M \rightarrow Y$

# The natural indirect effect

- The natural indirect effect:

$$NIE(d, d^*) = E\left(Y\big(d, M(d)\big) - Y\big(d, M(d^*)\big)\right)$$

- The $NIE(d, d^*)$ is the expected difference in the outcome if individuals had been exposed to $d$ and then…
  - experienced the level of the mediator that would have arisen naturally for them under exposure $d$ rather than the level of the mediator that would have arisen under exposure $d^*$

- It captures an effect of the exposure $D$ on the outcome $Y$ that operates specifically through a causal mechanism involving the mediator $M$

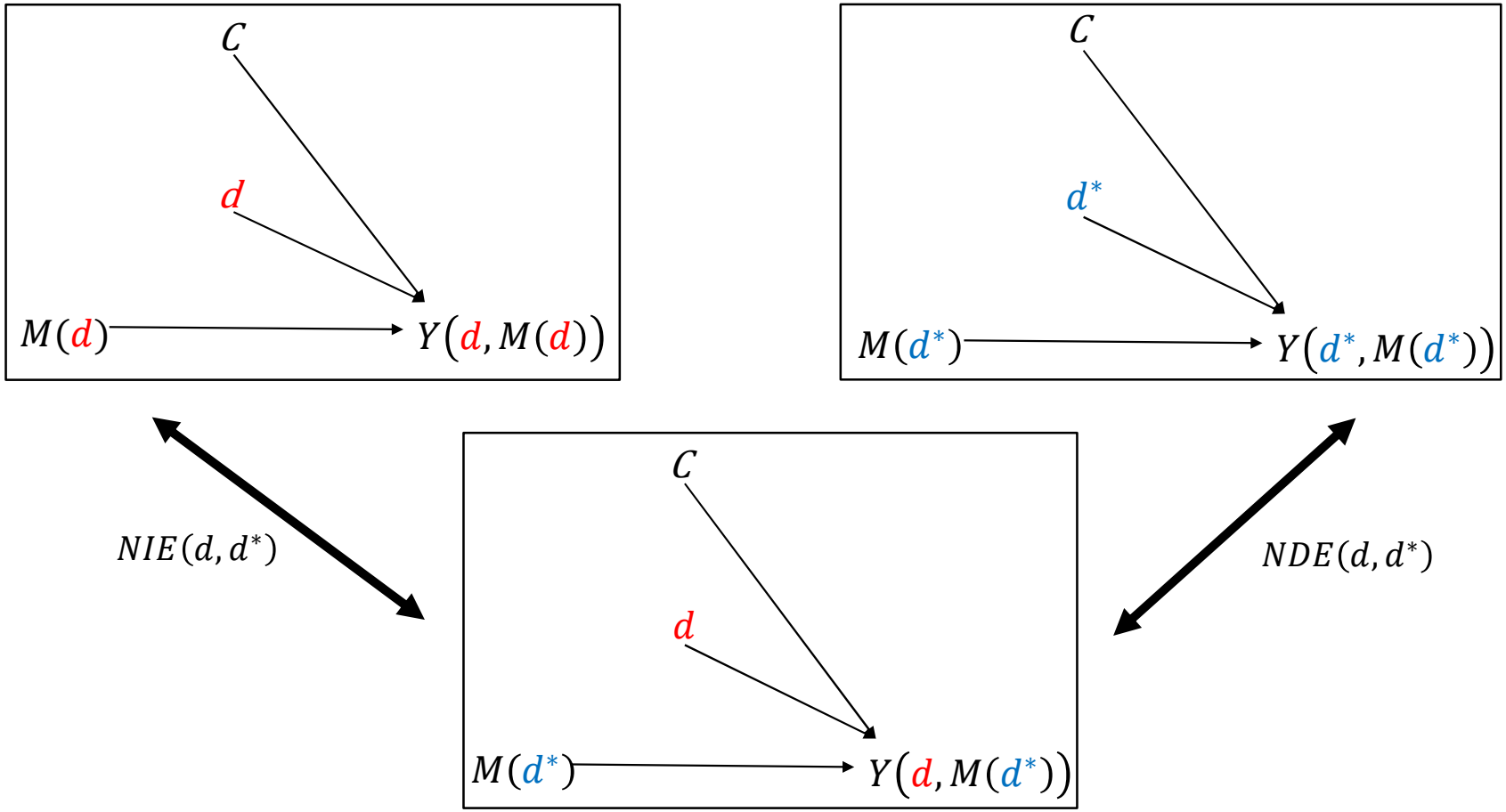
# The natural indirect effect

- The natural indirect effect:

$$NIE(d, d^*) = E\left(Y\big(d, M(d)\big) - Y\big(d, M(d^*)\big)\right)$$

- The $NIE(d, d^*)$ isolates an effect operating through the mediator by…
  - holding the exposure for each individual constant at $d$…
  - while comparing outcomes across differences in the mediator that would have arisen under different exposures, $M(d)$ versus $M(d^*)$
- This comparison deactivates all causal mechanisms connecting the exposure to the outcome except for a causal chain operating through the mediator, denoted by the $D \rightarrow M \rightarrow Y$ path

# Three counterfactual worlds



# The fundamental problem of causal inference

- We never observe the potential outcomes under exposure $d$, $Y(d)$, for individuals who are not actually exposed to $d$
- We also never observe the potential outcomes under exposure $d^*$, $Y(d^*)$, for individuals who are not actually exposed to $d^*$
- As result, it is impossible to observe or compute an individual causal effect, $Y(d) - Y(d^*)$, because we are always missing one piece of data

# The fundamental problem of causal inference

- But, at least we observe either $Y(d)$ OR $Y(d^*)$ for everyone (just not both)

# The fundamental problem of causal mediation

- The fundamental problem of causal mediation is even more challenging than the fundamental problem of causal inference
- The central challenge is that we never observe the cross-world potential outcome, $Y\big(d, M(d^*)\big)$, for anyone!
  - In other words, cross-world potential outcomes cannot ever be observed in reality
- Thus, mediation analyses based on the natural effects decomposition must contend with not one but two fundamental problems, both of which are formidable

# Nonparametric identification

- Natural direct and indirect effects can be nonparametrically identified if the following conditions are met:

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

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

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

Assumption NE.4: $Y(d,m) \perp M(d^*)|C$

Assumption NE.5: $P(d,m|c) > 0$

Assumption NE.6: $Y = Y(D) = Y\big(D, M(D)\big) = Y(D,M)$

# No unobserved $D$-$Y$ confounding

- Assumption NE.1:

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

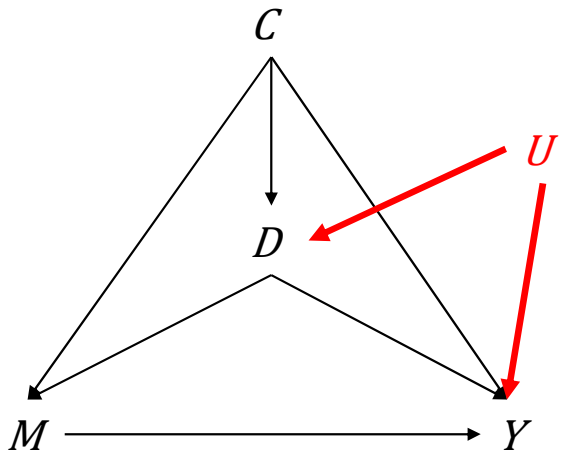
- This assumption requires that the exposure $D$ must be statistically independent of the joint potential outcomes $Y(d, m)$, conditional on the baseline confounders $C$

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

# No unobserved $D$-$Y$ confounding

- Assumption NE.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 NE.2:

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

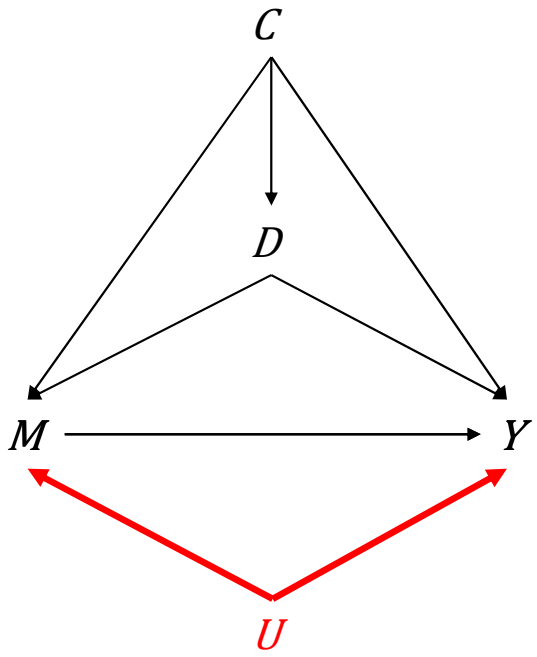
- This assumption requires that the mediator $M$ must be statistically independent of the joint potential outcomes $Y(d,m)$, conditional on the baseline confounders $C$ and exposure $D$

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

# No unobserved $M$-$Y$ confounding

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

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



# No unobserved $D$-$M$ confounding

- Assumption NE.3:

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

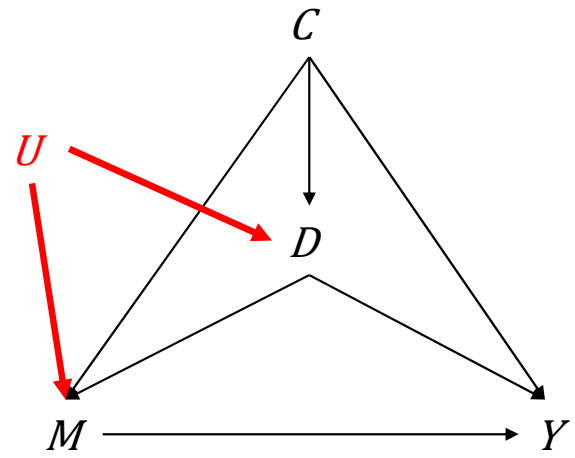
- This assumption requires that the exposure $D$ must be statistically independent of the potential values of the mediator $M(d)$, conditional on the baseline confounders $C$

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

# No unobserved $D$-$M$ confounding

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

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



# No exposure-induced confounding

- Assumption NE.4:

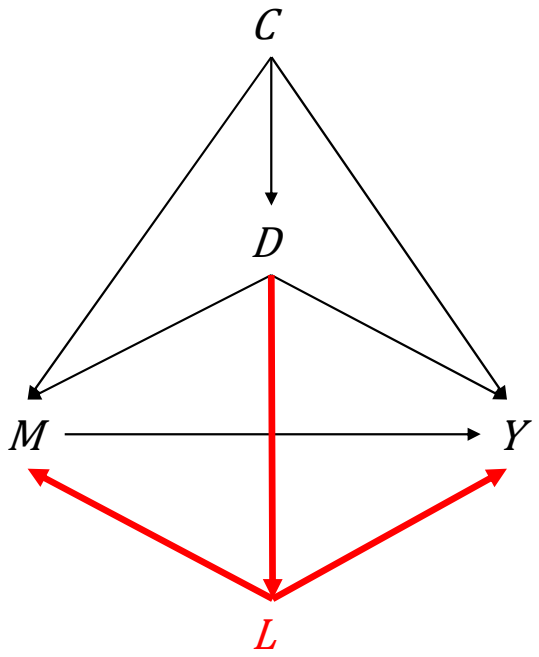
$$Y(d, m) \perp M(d^*)|C$$

- It requires that the potential values of the mediator under exposure $d$ must be independent of the joint potential outcomes under exposure $d^*$, conditional on the baseline confounders $C$

- Known as a cross-world independence assumption, it requires that there must not be any exposure-induced confounders, whether they are observed or not

# No exposure-induced confounding

- Assumption NE.4 would be violated if any variable, observed or not, jointly affects the mediator and outcome, and is also affected by the exposure
- In this graph, $L$ is a confounder for the $M \rightarrow Y$ relationship that is affected by $D$
  - That is, $L$ is an exposure-induced confounder



# Positivity

- Assumption NE.5:

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

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

- Substantively, it stipulates that there must be at least some chance that individuals experience all possible levels of the exposure and mediator within every subpopulation defined by the confounders

# Consistency

- Assumption NE.6:

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

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

# Consistency

- Assumption NE.6:

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

- This assumption requires that the observed, potential, nested potential, and joint potential outcomes are all consistent with one another
  - The first equality in this expression requires that an individual's observed outcome must be equal to their potential outcome under the level of the exposure they did in fact experience

# Consistency

- Assumption NE.6:

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

- This assumption requires that the observed, potential, nested potential, and joint potential outcomes are all consistent with one another
  - The second equality requires that an individual's potential outcome under their observed exposure must be equal to their nested potential outcome under this same exposure and…
  - …by extension, under the value of the mediator that would arise naturally from this exposure

# Consistency

- Assumption NE.6:

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

- This assumption requires that the observed, potential, nested potential, and joint potential outcomes are all consistent with one another
  - The last equality requires that the nested potential outcome must be equal to an individual's joint potential outcome under the levels of both the exposure and the mediator that they did in fact experience
  - This last equality subsumes another consistency assumption about the observed and potential values of the mediator—namely, that $M = M(D)$

# The fundamental problem revisited

- Unlike the assumptions for identifying total effects, the assumptions required to nonparametrically identify natural direct and indirect effects cannot all be met by experimental design
- In a conventional experiment where the exposure is randomly assigned, only assumptions NE.1 and NE.3 would be met by design
- Even in an experiment where the exposure and mediator are jointly randomized, only assumptions NE.1 to NE.5 would be met by design
  - Randomization of both the exposure and mediator would violate assumption NE.6, since $M \neq M(D)$ under this design

# Identification formula for NDE

- Under assumptions NE.1 to NE.6, the natural direct effect can be equated with a function of observable data

- The nonparametric identification formula for the natural direct effect:

$$
\begin{aligned}
NDE(d, d^*) &= E\left(Y\left(d, M(d^*)\right) - Y\left(d^*, M(d^*)\right)\right) \\
&= \sum_c \sum_m [E(Y|c, d, m) - E(Y|c, d^*, m)] P(m|c, d^*) P(c) \\
&= E_C\left(E_{M|c,d^*}\left(E(Y|C, d, M) - E(Y|C, d^*, M)\right)\right)
\end{aligned}
$$

# Identification formula for NIE

- Under assumptions NE.1 to NE.6, the natural indirect effect can also be equated with a function of observable data

- The nonparametric identification formula for the natural indirect effect:

$$NIE(d, d^*) = E\left(Y\big(d, M(d)\big) - Y\big(d, M(d^*)\big)\right)$$

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

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

# 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
- In practice, however, we rarely have data from an entire target population and thus cannot simply ignore sampling variability
- Nonparametric estimation just involves plugging in sample analogs for the population quantities that comprise the nonparametric identification formulae

# Nonparametric estimation of NDE

- A nonparametric estimator for the natural direct effect:

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

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

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

# Nonparametric estimation of NIE

- A nonparametric estimator for the natural indirect effect:

$$\widehat{NIE}(d, d^*)^{np} = \sum_c \sum_m \hat{E}(Y|c, d, m)\left[\hat{P}(m|c, d) - \hat{P}(m|c, d^*)\right]\hat{P}(c)$$
$$= \sum_c \sum_m \bar{Y}_{c,d,m}\left[\hat{\pi}_{m|c,d} - \hat{\pi}_{m|c,d^*}\right]\hat{\pi}_c$$

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

# Limitations of nonparametric estimation

- **Sparsity:** perfect stratification of the data may produce strata that are empty, leading to violations of the positivity assumption

- **The curse of dimensionality:** the problem of sparsity increases with the dimension of the data—that is, with many confounders, or with confounders, an exposure, or a mediator with many different levels

- **Excessive sampling variability:** even when nonparametric estimation is not precluded by sparsity, perfect stratification of the data may produce strata with very few observations, leading to imprecision

# Summary

- The natural effects decomposition permits a separation of total effects into components operating through a single mediator of interest versus other mechanisms
- Natural direct and indirect effects can be nonparametrically identified under a series of assumptions, but these assumptions cannot all be satisfied by experimental design
- Nevertheless, if these assumptions are met, natural direct and indirect effects can be identified and estimated from observed data, without the need for any restrictions on their probability distribution

# 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 ($M$)
    - 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 protecting its recipients from financially strenuous and mentally taxing spells of unemployment
- Does unemployment mediate the effect of college attendance on depression?

# Example: NLSY79

- Draw a DAG that best represents the causal relations among variables in the NLSY79, based on theory and prior knowledge

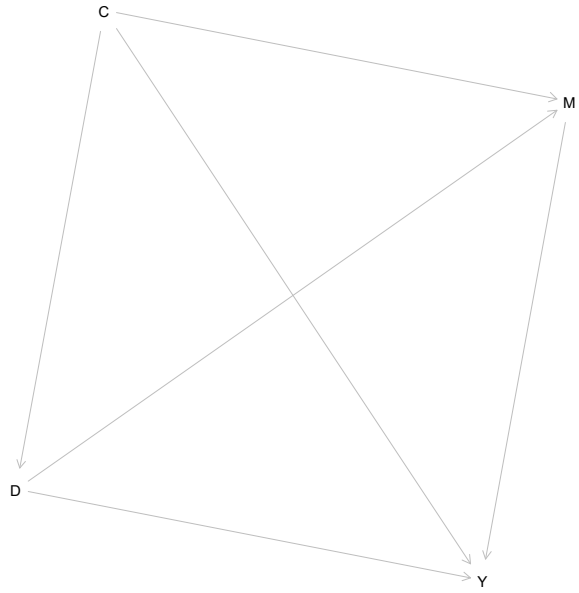
```
### wk 3 nlsy tutorial ###
rm(list=ls())

# 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 -> M -> Y
    C -> M
    C -> Y
    D -> Y
}")

plot(graphLayout(nlsyDAG))
```



# Example: NLSY79

- Formally define the natural direct and indirect effects of college attendance on depression, as mediated by unemployment
  - How are these estimands defined in terms of potential outcomes?
  - How are these estimands interpreted substantively?

# Example: NLSY79

- Can the natural direct and indirect effects of college attendance on depression be nonparametrically identified from the observed data?
  - Why or why not?
  - What assumptions are necessary, and which are violated?

# Example: NLSY79

- Suppose, for the sake of illustration, that the natural effects of interest can be identified by adjusting only for a single confounder: whether a respondent's mother attended college themselves
  - Under this supposition, compute all the components of nonparametric estimators for the natural direct and indirect effects

```
# load data #
datadir <- "C:/Users/Geoff/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("cesd_age40", "ever_unemp_age3539", "att22", "momedu")]),]

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

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

# components of np estimator #
nlsy %>%
    group_by(momcol, att22, ever_unemp_age3539) %>%
            dplyr::summarize(
                mean = mean(std_cesd_age40),
                n = n(),
                .groups = "drop")

```

# Example: NLSY79

- Suppose, for the sake of illustration, that the natural effects of interest can be identified by adjusting only for a single confounder: whether a respondent's mother attended college themselves
  - Under this supposition, compute all the components of nonparametric estimators for the natural direct and indirect effects

```
> # components of np estimator #
> nlsy %>%
+ group_by(momcol, att22, ever_unemp_age3539) %>%
+ dplyr::summarize(
+ mean = mean(std_cesd_age40),
+ n = n(),
+ .groups = "drop")
# A tibble: 8 × 5
  momcol att22 ever_unemp_age3539      mean     n
   <dbl> <dbl>              <dbl>     <dbl> <int>
1      0     0                  0 -0.000218  1836
2      0     0                  1  0.319      548
3      0     1                  0 -0.219      594
4      0     1                  1  0.0786      96
5      1     0                  0 -0.0239     178
6      1     0                  1  0.367       40
7      1     1                  0 -0.171      347
8      1     1                  1 -0.0665      43
```

# Example: NLSY79

- Under the same supposition, compute a nonparametric estimate of average total effect

```
40  # nonparametric estimate of ATE #
41  EhatY_CD <- lm(std_cesd_age40~att22*momcol, data=nlsy)
42
43  ATEhat <- mean(marginal_effects(EhatY_CD, nlsy, variables=c("att22"))$dydx_att22)
44
```

```
> # nonparametric estimate of ATE #
> EhatY_CD <- lm(std_cesd_age40~att22*momcol, data=nlsy)
>
> ATEhat <- mean(marginal_effects(EhatY_CD, nlsy, variables=c("att22"))$dydx_att22)
>
> print(ATEhat)
[1] -0.2434601
>
```

# Example: NLSY79

- Under the same supposition, compute a nonparametric estimate of the natural direct effect

```
47  # nonparametric estimate of NDE #
48  PhatM_CD <- lm(ever_unemp_age3539~att22*momcol, data=nlsy)
49
50  EhatY_CDM <- lm(std_cesd_age40~att22*ever_unemp_age3539*momcol, data=nlsy)
51
52  gdata <- nlsy
53
54  gdata$att22 <- 0
55  gdata$ever_unemp_age3539 <- 0
56  EhatY_D0M0C <- predict(EhatY_CDM, gdata)
57  PhatM_D0C <- predict(PhatM_CD, gdata)
58
59  gdata$att22 <- 1
60  EhatY_D1M0C <- predict(EhatY_CDM, gdata)
61  PhatM_D1C <- predict(PhatM_CD, gdata)
```

```
63  gdata$ever_unemp_age3539 <- 1
64  EhatY_D1M1C <- predict(EhatY_CDM, gdata)
65
66  gdata$att22 <- 0
67  EhatY_D0M1C <- predict(EhatY_CDM, gdata)
68
69  NDEhat <- mean((EhatY_D1M0C-EhatY_D0M0C)*(1-PhatM_D0C)
70      +(EhatY_D1M1C-EhatY_D0M1C)*PhatM_D0C)
71
72  print(NDEhat)
```

```
> print(NDEhat)
[1] -0.2196634
```

# Example: NLSY79

- Under the same supposition, compute a nonparametric estimate of the natural indirect effect

```
74  # nonparametric estimate of NIE #
75  NIEhat <- mean((EhatY_D1M0C)*((1-PhatM_D1C)-(1-PhatM_D0C))
76      +(EhatY_D1M1C)*((PhatM_D1C)-(PhatM_D0C)))
77
78  print(NIEhat)
79
```

```
> print(NIEhat)
[1] -0.02379668
```

# Example: NLSY79

- What should we conclude from this analysis? Is nonparametric estimation a sensible approach with these data? Why or why not?