Mediation

Mediation

- Core idea and motivation
- Conventional approach
- Link with lifecourse models
- When conventional approach fails
- Counterfactual framework: CDE, NDE, and NIE

Mediation

$$X \longrightarrow M \longrightarrow Y$$

- If X affects M and M affects Y, then M mediates (part of) the effect of X on Y.
  - There are some weird cases where this is not true, but these are exceptions.
- M cannot mediate the effect of X on Y if X does not affect M, or if M does not affect Y.

Simple physical analogy

- In a straight row of dominoes, tipping the first one hits the second, which topples the third and so on.
- Each domino mediates the effect of the prior domino on the subsequent domino
- Removing, or holding upright, a middle domino will stop the sequence.
- Tipping domino one has no effect on domino three, conditional on domino two in the sequence.

Mediation: Direct vs Indirect Effects

$$Ed \longrightarrow \$ \longrightarrow Diet \longrightarrow BMI$$

Education increases earnings, which are used to improve diet, which results in a lower BMI.

$$\begin{tikzcd} & Info \arrow[dr, bend left] & & \\ Ed \arrow[ur, bend left] \arrow[r] & \$ \arrow[r] & Diet \arrow[r] & BMI \end{tikzcd}$$

Education also increases nutritional knowledge, which results in improved diet and lower BMI.

- Very often, we wish to decompose a “total” effect into those mediated by a hypothesized mediator (indirect effects), and those that are not mediated by that mediator (direct effects).
- Even if the effect is not mediated by the specific mediator under consideration, they are probably mediated by *something* , so it is a bit of a misnomer that we call them “direct”. It is just a convenience.
- In DAGs, we mean “direct with respect to the other variables in the DAG”.

Why do we care about effect decomposition?

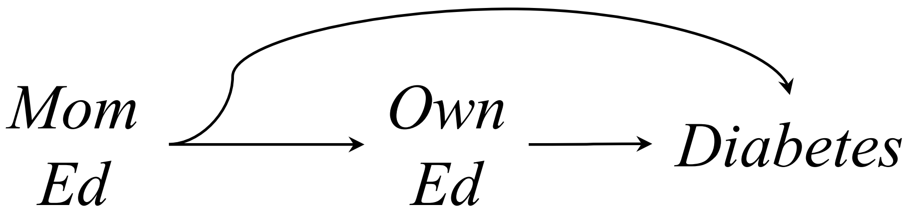
$$\begin{tikzcd} & \mathit{Info} \arrow[dr, bend left] & & \\ \mathit{Ed} \arrow[ur, bend left] \arrow[r] & \$ \arrow[r] & \mathit{Diet} \arrow[r] & \mathit{BMI} \end{tikzcd}$$

- ➢To understand etiology: what is the chain of events that leads to the disease?
- ➢To identify and prioritize intervention opportunities: should I give people money or spend the money on an educational campaign?
- ➢To assess injustice: are women paid less because their hormones make them weepy at work or because of discrimination?

Progress on Mediation Modeling

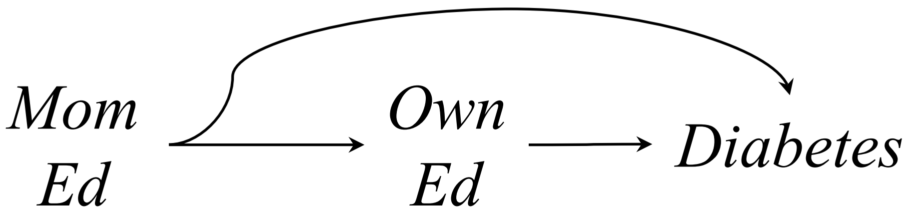
- This is an area of rapid progress in the last 20 years
- Robins et al introduced a counterfactual framework, i.e., what precisely do we mean by "direct effects"
- VanderWeele has written details on how to estimate and challenges in conventional methods
- In the simplest setting, there's nothing new from decades of prior mediation modeling.

Conventional Approach to Direct vs Indirect Effects Decomposition



- Would mother's education affect diabetes if you made everyone in the population complete college?
- Possible *direct* mechanisms:
  - Mother's education influences breastfeeding?
  - Mother's education influences diet in very early life, taste/preference formation?
- Possible *indirect* mechanisms
  - Own education influences earnings, which influences health relevant behaviors
  - Own education influences knowledge of diabetes prevention.

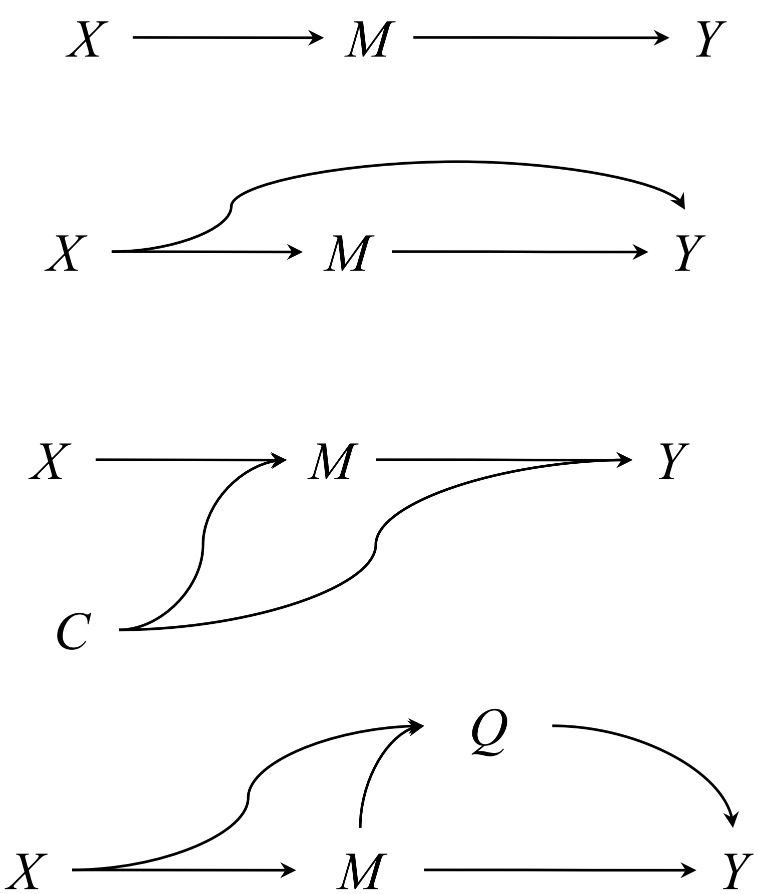
Mediation: Direct vs Indirect Effects Decomposition



- Would mother's education affect diabetes if you made everyone quit school at 5th grade?

- Would the direct effect of mother's education *differ* if you made everyone quit school at 5th grade vs if you made everyone finish college?

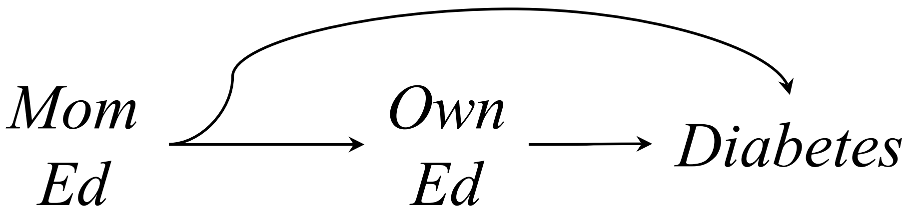
Mediation: lots of possible structures



Mediation

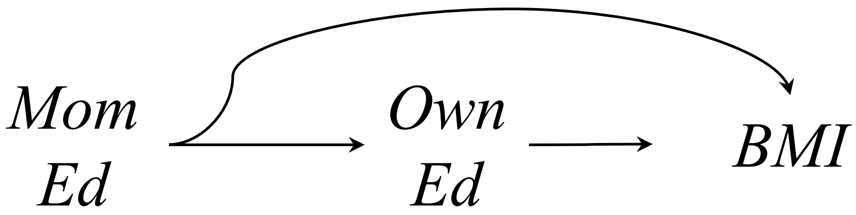
- Core idea and motivation
- Conventional approach
- Link with lifecourse models
- When conventional approach fails
- Counterfactual framework: CDE, NDE, and NIE

Mediation: Classic Decomposition Approach

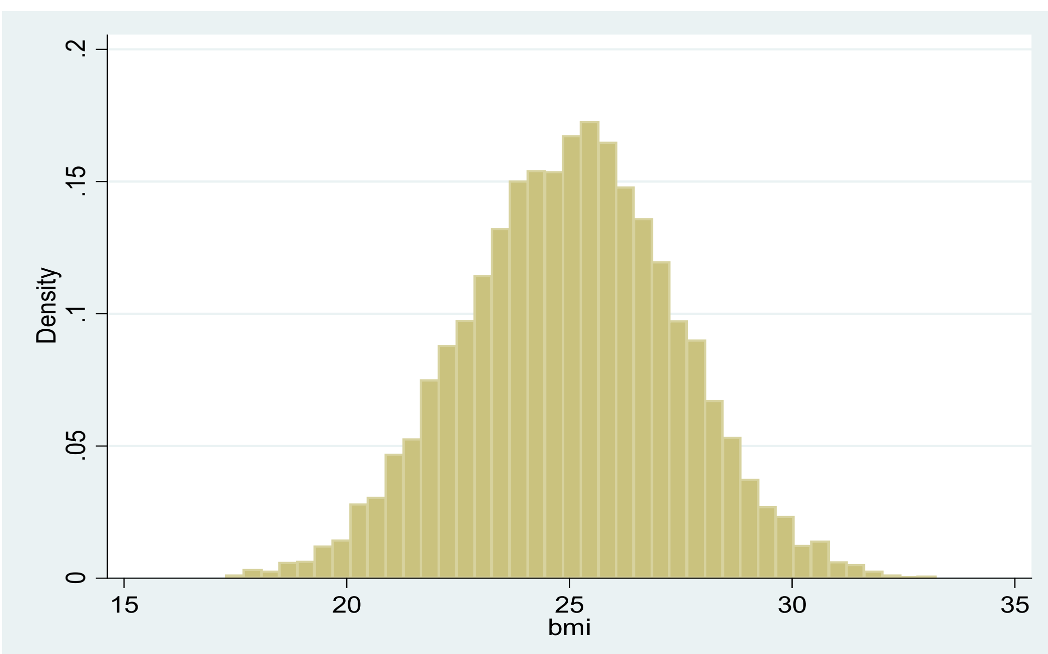


- Total effect= direct effect + indirect effect
- Indirect effect= total effect-direct effect
- Total effect= E(Db | MomEd=1)-E(Db|MomEd=0)
- Direct effect=E(Db|MomEd=1, OwnEd)-E(Db|MomEd=0, OwnEd)
- Sometimes called “Barron-Kenny” decomposition

Mediation: Classic Decomposition Example



E(Own Ed)=12+0.5*(Mom Ed-12)
E(BMI)= 25-0.5*(Own Ed-12) – 0.25*(Mom Ed-12)



Mediation: Classic Decomposition Example

$$\begin{tikzcd} \textit{Mom Ed} \arrow[r] \arrow[rr, bend left] & \textit{Own Ed} \arrow[r] & \textit{BMI} \end{tikzcd}$$

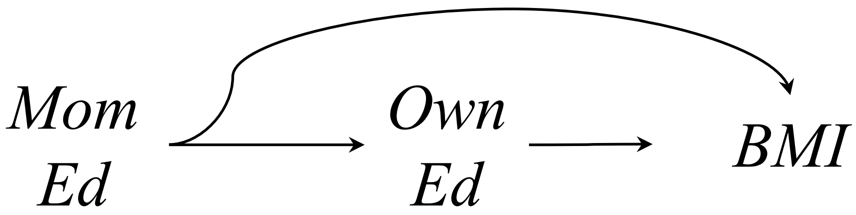
E(Own Ed)=12+0.5*(Mom Ed-12)
E(BMI)= 25-0.5*(Own Ed-12) – 0.25*(Mom Ed-12)

What is the direct effect of Mom Ed on BMI?
-0.25

What is the indirect effect of Mom Ed on BMI?
0.5*-0.5=-0.25

What is the total effect of Mom Ed on BMI?
-0.25 + 0.5*(-.5)=-0.5

Mediation: Classic Decomposition Example



E(Own Ed)=12+0.5*(Mom Ed-12)
E(BMI)= 25-0.5*(Own Ed-12) – 0.25*(Mom Ed-12)

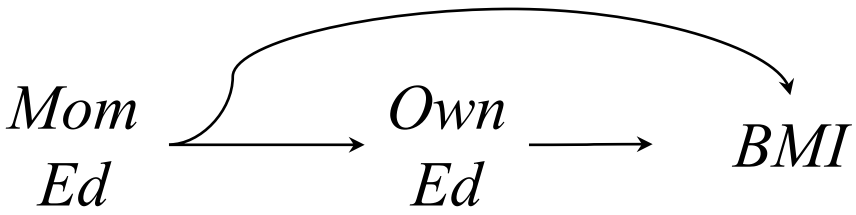
What is the total effect of Mom Ed on BMI?
-0.25 + 0.5*(-.5)=-0.5

```
. reg bmi momed

      Source |       SS       df       MS              Number of obs =   10000
-------------+------------------------------           F(  1,  9998) = 2102.71
       Model |  9760.68979     1  9760.68979           Prob > F      =  0.0000
    Residual |  46410.3396  9998  4.64196235           R-squared     =  0.1738
-------------+------------------------------           Adj R-squared =  0.1737
       Total |  56171.0294  9999  5.61766471           Root MSE      =  2.1545

------------------------------------------------------------------------------
         bmi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       momed |  -.4936806   .0107661   -45.86   0.000    -.5147842    -.472577
       _cons |   30.95059   .1309641   236.33   0.000     30.69387    31.20731
------------------------------------------------------------------------------
```

Mediation: Classic Decomposition Example



E(Own Ed)=12+0.5*(Mom Ed-12)
E(BMI)= 25-0.5*(Own Ed-12) – 0.25*(Mom Ed-12)

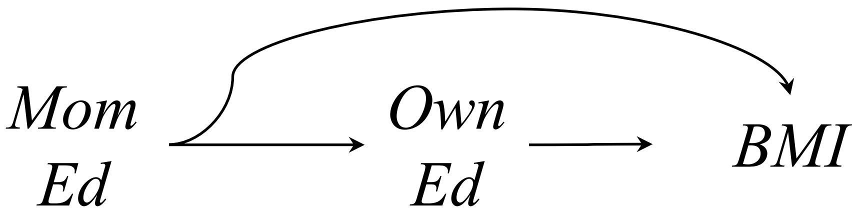
What is the direct effect of Mom Ed on BMI?
-0.25

```
. reg bmi momed own_ed

      Source |       SS       df       MS              Number of obs =   10000
-------------+------------------------------           F(  2,  9997) = 1871.21
       Model |  15300.174     2  7650.08698           Prob > F      =  0.0000
    Residual | 40870.8555  9997  4.08831204           R-squared     =  0.2724
-------------+------------------------------           Adj R-squared =  0.2722
       Total | 56171.0294  9999  5.61766471           Root MSE      =   2.022

------------------------------------------------------------------------------
         bmi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       momed |   -.245425   .0121478   -20.20   0.000    -.2692371   -.2216128
      own_ed |  -.4988862   .0135531   -36.81   0.000     -.525453   -.4723194
       _cons |   33.95703   .1475695   230.11   0.000     33.66777     34.2463
------------------------------------------------------------------------------
```

Mediation: Classic Decomposition Example



E(Own Ed)=12+0.5*(Mom Ed-12)
E(BMI)= 25-0.5*(Own Ed-12) – 0.25*(Mom Ed-12)

What is the indirect effect of Mom Ed on BMI?
total-direct= -0.5-(-0.25)= -0.25

```
. reg bmi momed own_ed

      Source |       SS       df       MS              Number of obs =   10000
-------------+------------------------------           F(  2,  9997) = 1871.21
       Model |  15300.174     2  7650.08698           Prob > F      =  0.0000
    Residual | 40870.8555  9997  4.08831204           R-squared     =  0.2724
-------------+------------------------------           Adj R-squared =  0.2722
       Total | 56171.0294  9999  5.61766471           Root MSE      =   2.022

------------------------------------------------------------------------------
         bmi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       momed |   -.245425   .0121478   -20.20   0.000    -.2692371   -.2216128
      own_ed |  -.4988862   .0135531   -36.81   0.000     -.525453   -.4723194
       _cons |   33.95703   .1475695   230.11   0.000     33.66777     34.2463
------------------------------------------------------------------------------
```

Mediation

- Core idea and motivation
- Conventional approach
- Link with lifecourse models
- When conventional approach fails
- Counterfactual framework: CDE, NDE, and NIE

Lifecourse timing

- Lifecourse models a good jumping off point for mediation, because alternative theories entail claims about mediation
- Alternative models can be represented as DAGs and tested against one another with divergent statistical claims implied by alternative the DAGs
- We will use the example of lifecourse SEP (childhood, early adulthood, midlife) and late life health, but you could apply this to any exposure sequence and outcome.

Etiologic Model: Immediate Risk

- After the exposure is removed, the risk starts to decline.

Childhood SEP (SEP_1)

Early Adult SEP (SEP_2)

Midlife SEP (SEP_3) $\longrightarrow$ Health

Etiologic Models: Immediate Risk + Social Trajectory

- After the exposure is removed, the risk starts to decline.
- Social conditions at each moment set the stage for the next period: the trajectory is sticky. The initial insult does not directly harm later health – but the first exposure leads to subsequent exposures that then directly affect health.

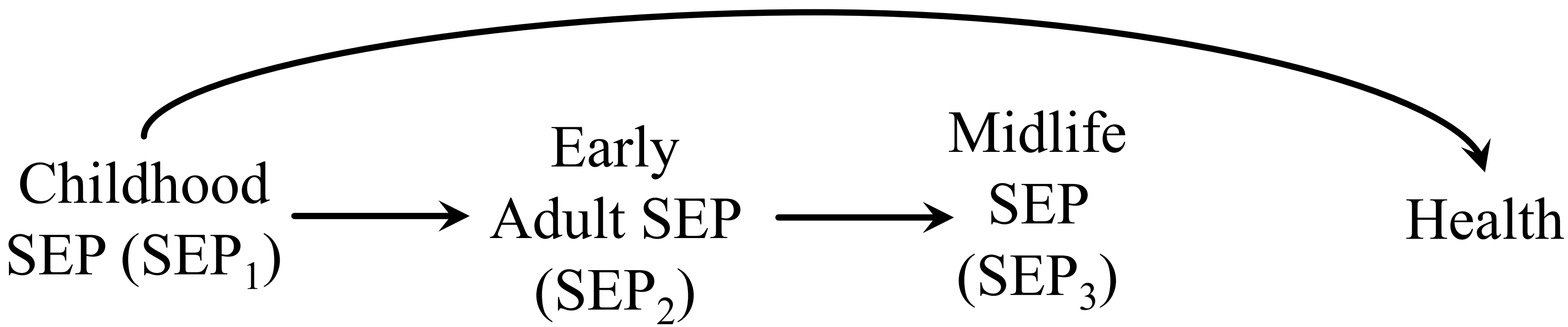
Childhood SEP (SEP_1) $\longrightarrow$ Early Adult SEP (SEP_2) $\longrightarrow$ Midlife SEP (SEP_3) $\longrightarrow$ Health

- Consistent with chain of risk model
- Adult SEP mediates effects of childhood SEP on health

21

Critical Period/Latency (+ Social Trajectory)

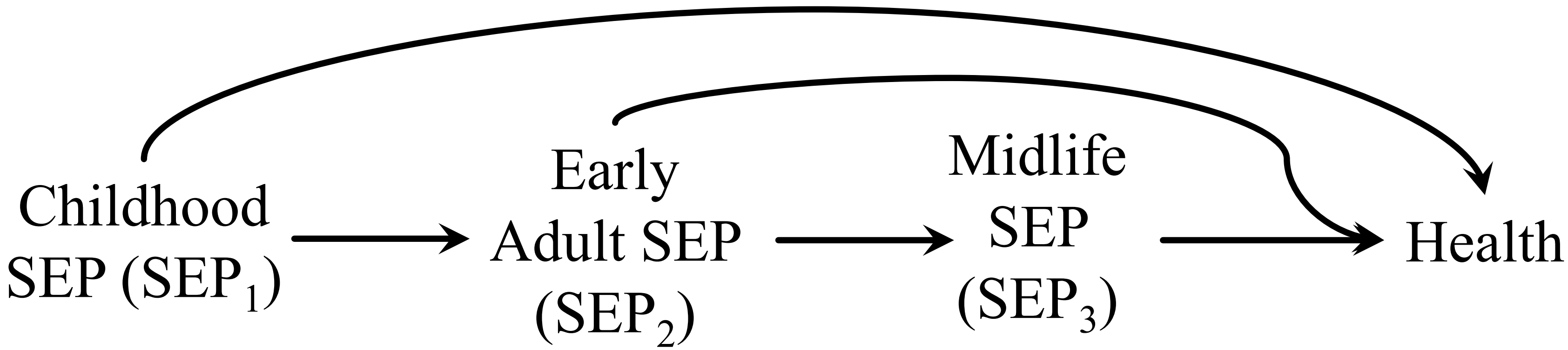
- Exposure in one period makes all the difference: subsequent exposures do not influence health.



- Also called programming models
- Example: fetal origins of disease
- Critical periods may occur in adulthood/old age, may be social, rather than biological
- Adult SEP does not mediate effect of child SEP on health

Etiologic Models: Cumulative Risk

- Each exposure period has an independent effect on risk.

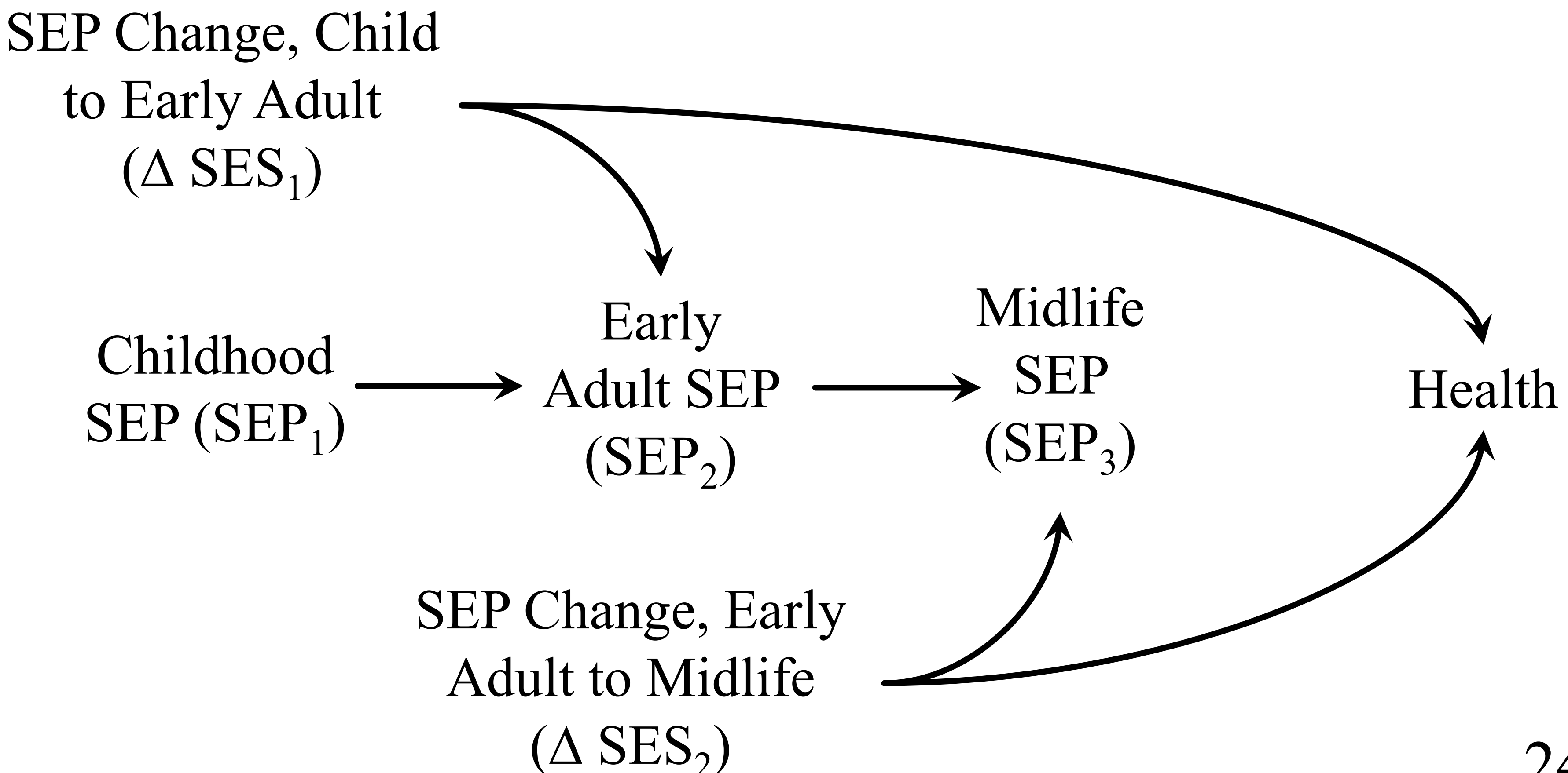


- Also consistent with chain of risk model
- Example: allostatic load
- Adult SEP partially mediates effects of child SEP on health

23

Physiological Effects of Trajectory/Change

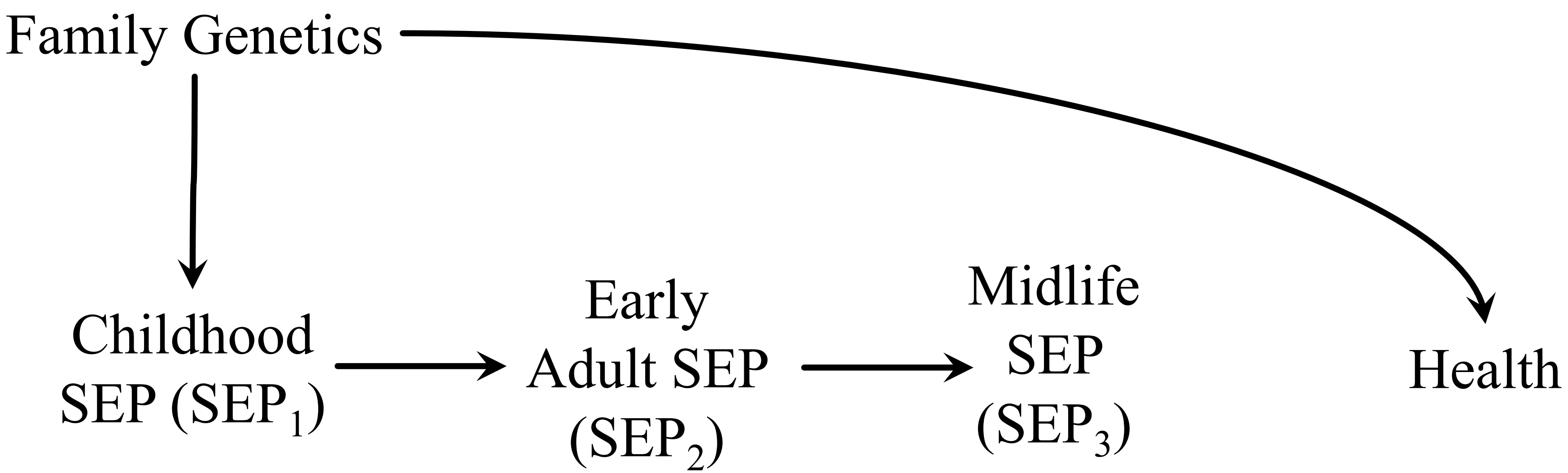
- Change itself, rather than level of SEP, influences health.



24

Etiologic Models: All Confounding

- Social conditions have no effect on adult health.
- The association between social conditions and adult health is entirely attributable to confounding by unmeasured characteristics such as genetic background.



25

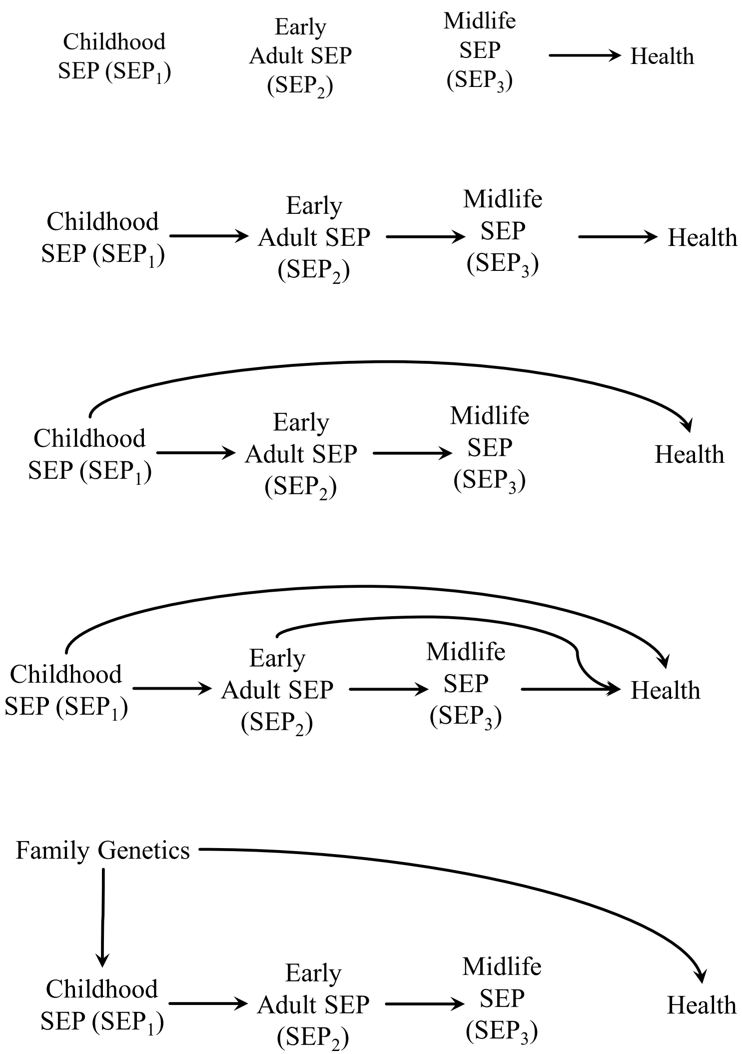
Implications of Alternatives

- All confounding: no effect of SES on adult health.
- Immediate risk: childhood SES does not affect adult health
- **Social trajectory: childhood SES has no direct effect on adult health other than that mediated by adult SES**
- Latency: adult SES has no effect on adult health, but the association between adult SES and adult health may be confounded by childhood SES
- **Cumulative: childhood SES indirectly affects adult health, via adult SES, and directly affects adult health, via pathways not mediated by adult SES.**

Lifecourse Models Have Inconsistent Empirical Predictions

- So you can provide evidence on which is most plausible for each outcome
- Mediation methods essential for evaluating

Conventional Approach: Total Effect of Child SES on Adult Health



- To distinguish *latency, cumulative, or trajectory* from *immediate* risk:
- E(health)=b0+b1*SEP1
- If you have some measured confounders (common causes or factors on a common cause pathway):
- E(health2)=b0+b1*SEP1+b2*Age +b3*Birth Regn
- Under immediate risk, b1=0

Conventional Approach: Total Effect of Child SES on Adult Health

- E(health)=b0+b1*SEP1
- E(health)=b0+b1*SEP1+b2*Age+b3*Birth Regns

- Important: to estimate the <u>total</u> effect of an exposure on an outcome, do not adjust for things that are affected by the exposure!
- Especially do not adjust for mediators on the pathway between the exposure and the outcome!

Conventional Approach: Direct Effect of Child SES on Adult Health

- To additionally rule out trajectory model
- $E(health)=a0+a1*SEP1+a2*SEP2$
- Under immediate risk or social trajectory model, $a1=0$ because effect of SEP1 is fully mediated by SEP2

- Challenge: should not adjust for mediators of childhood SES but most measured variables are not temporally prior to early life SES

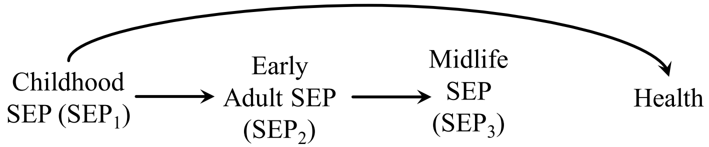
Conventional Approach: Indirect Effect of Child SES on Adult Health

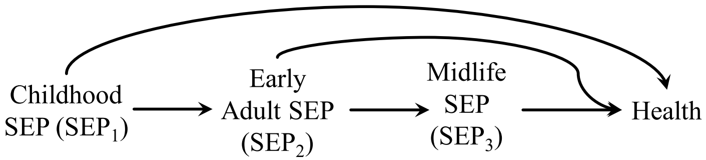
- $E(health) = b_0 + b_1 * SEP_1$
- $E(health) = a_0 + a_1 * SEP_1 + a_2 * SEP_2$

- Indirect Effect=Total Effect-Direct Effect
  - Indirect Effect$= b_1 - a_1$

- This decomposition assumes:
  - Indirect Effect % + Direct Effect % = 100%

Conventional Approach: Cumulative vs Latency





- Does adult SES predict health conditional on child SES:
  - E(health)=b0+b1*SES1+b2*Age
  - E(health)=a0+a1*SES1+a2*Age+a3*SES2
- Cumulative model implies b1>a1>0
- Latency model implies b1~=a1 because no mediation of SEP1 effect by SEP2
- Latency model implies a3=0 because SEP2 irrelevant once accounting for SEP1

Mediation

- Core idea and motivation
- Conventional approach
- Link with lifecourse models
- When conventional approach fails
- Counterfactual framework: CDE, NDE, and NIE

Assumptions for classic approach

1. No confounding of exposure-outcome
2. No confounding of mediator-outcome
3. No confounding of exposure-mediator
4. No interaction between mediator and the direct pathway
5. No measurement error of the mediator
6. Linear models

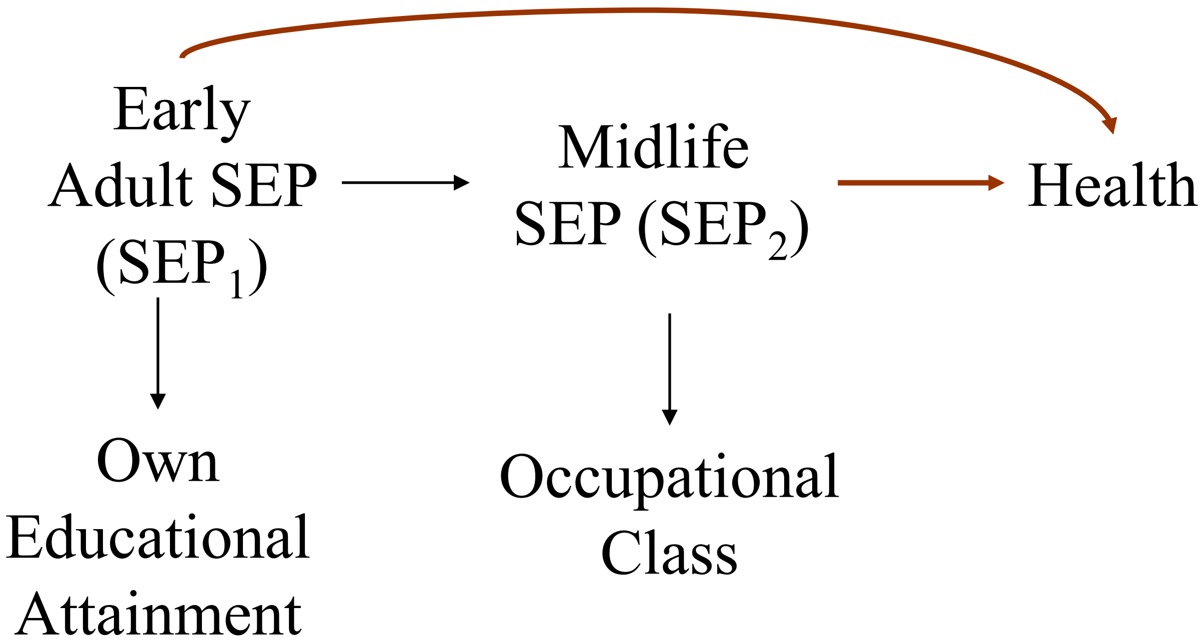
Commonly Ignored Problems with Classic Decomposition Approach

- Unmeasured confounding of the mediator and the outcome
- Effect modification of the direct pathway by the mediator
- Nonlinear models
- Inadequate measurement of the exposure or the mediator

Commonly Ignored Problems with Classic Decomposition Approach

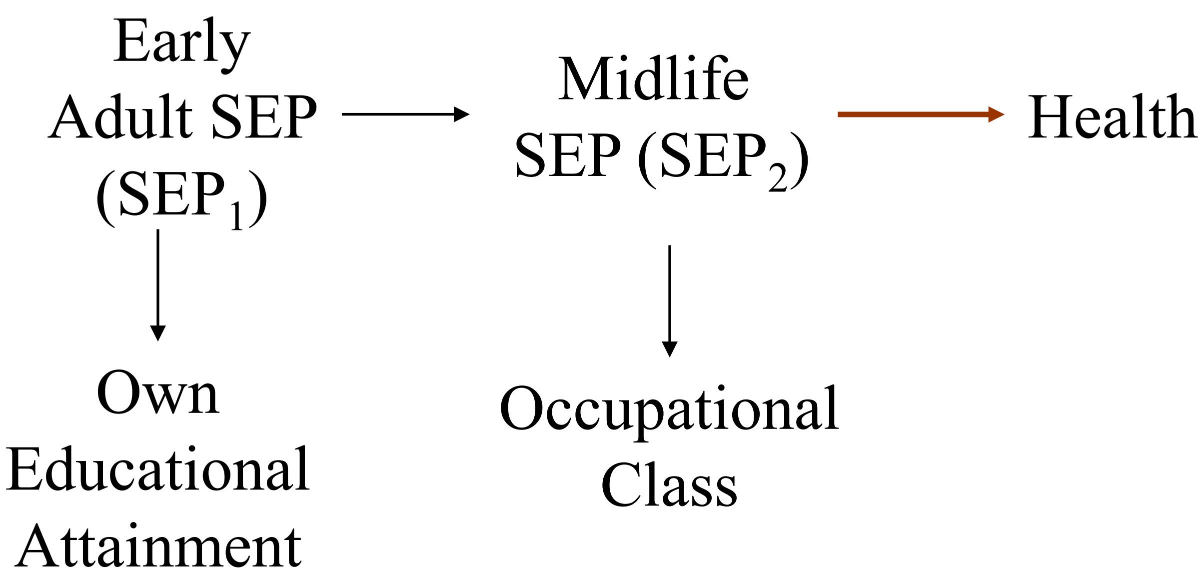
- Unmeasured confounding of the mediator and the outcome
- Effect modification of the direct pathway by the mediator
- Nonlinear models
- Inadequate measurement of the exposure or the **mediator**

Distinguishing Etiologic Models: Testing for Latency or Cumulative vs Immediate Risk



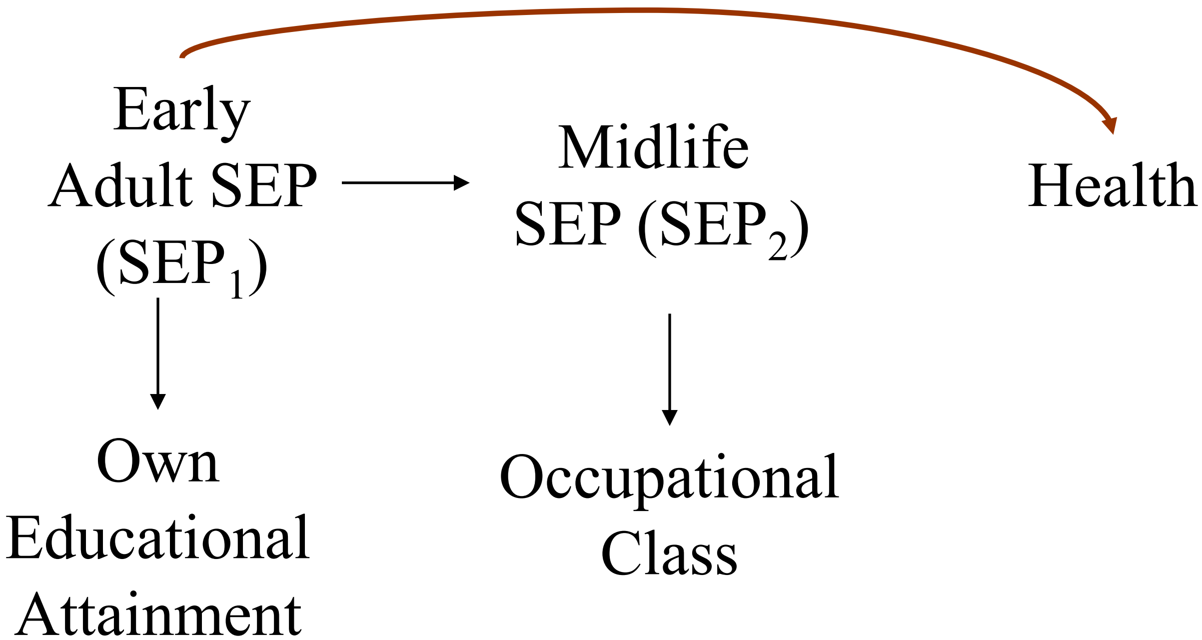
To test for a direct effect of SEP_1 on Health, we would compare the regression of Health on Education with and without adjustment for occupational class.

Distinguishing Etiologic Models: Testing for Latency or Cumulative vs Immediate Risk



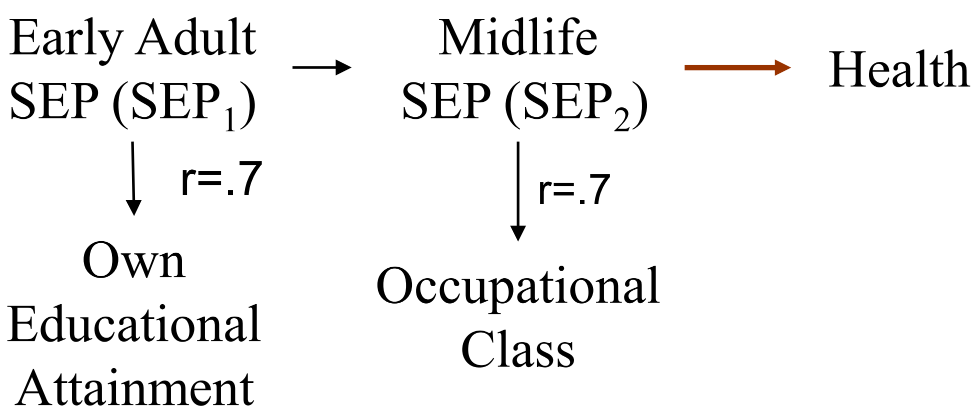
Even if there were no direct effect of SEP_1 on health, educational attainment would predict health independently of adult occupational class, due to measurement error.

Distinguishing Etiologic Models: Testing for Latency vs Cumulative Risk



Even if there were no effect of SEP_2 on health, occupational class would predict health independently of education, because occupational class is a correlate of the latent variable, SEP_1. In this DAG, SEP1 is a confounder of SEP2-health association.

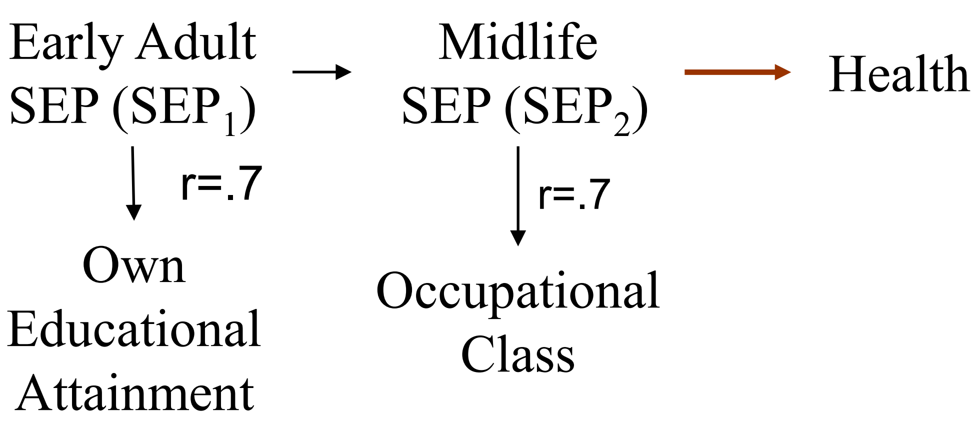
Empirical Example: spurious evidence for latent effects of SEP1




```
. corr
(obs=10000)

             |     sep1      sep2    health      educ     occpn
-------------+--------------------------------------------------
        sep1 |   1.0000
        sep2 |   0.7027    1.0000
      health |   0.2233    0.3191    1.0000
        educ |   0.7053    0.4978    0.1550    1.0000
       occpn |   0.4900    0.6998    0.2333    0.3377    1.0000
```

Empirical Example: spurious evidence for latent effects of SEP1




```
. reg health educ

      Source |       SS       df       MS
-------------+------------------------------
       Model |  2365.92653     1  2365.92653
    Residual |  96130.9624  9998  9.61501924
-------------+------------------------------
       Total |  98496.8889  9999  9.85067396

------------------------------------------------------
      health |      Coef.   Std. Err.      t    P>|t|
-------------+----------------------------------------
        educ |   .4873077   .0310655    15.69   0.000
       _cons |   .0052761   .0310092     0.17   0.865
```

Empirical Example: spurious evidence for latent effects of SEP1

Early Adult SEP (SEP_1) → Midlife SEP (SEP_2) → Health

Early Adult SEP (SEP_1) ↓ r=.7 Own Educational Attainment

Midlife SEP (SEP_2) ↓ r=.7 Occupational Class

. reg health educ

Source	SS	df	MS
Model	2365.92653	1	2365.92653
Residual	96130.9624	9998	9.61501924
Total	98496.8889	9999	9.85067396

health	Coef.	Std. Err.	t	P>|t|
educ	.4873077	.0310655	15.69	0.000
_cons	.0052761	.0310092	0.17	0.865

. reg health educ occpn

Source	SS	df	MS
Model	6006.87944	2	3003.43972
Residual	92490.0095	9997	9.25177648
Total	98496.8889	9999	9.85067396

health	Coef.	Std. Err.	t	P>|t|
educ	.2704272	.0323748	8.35	0.000
occpn	.648817	.032706	19.84	0.000
_cons	.0069007	.0304179	0.23	0.821

Empirical Example: spurious evidence for latent effects of SEP1

Early Adult SEP (SEP_1) → Midlife SEP (SEP_2) → Health

r=.7 r=.7

Own Educational Attainment

Occupational Class

. reg health educ

Source	SS	df	MS
Model	2365.92653	1	2365.92653
Residual	96130.9624	9998	9.61501924
Total	98496.8889	9999	9.85067396

health	Coef.	Std. Err.	t	P>|t|
educ	.4873077	.0310655	15.69	0.000
_cons	.0052761	.0310092	0.17	0.865

. reg health educ occpn

Source	SS	df	MS
Model	6006.87944	2	3003.43972
Residual	92490.0095	9997	9.25177648
Total	98496.8889	9999	9.85067396

health	Coef.	Std. Err.	t	P>|t|
educ	.2704272	.0323748	8.35	0.000
occpn	.648817	.032706	19.84	0.000
_cons	.0069007	.0304179	0.23	0.821

Conclusion:

Less than half the effect of childhood SEP is mediated by adult SEP.

Empirical Example: spurious evidence for latent effects of SEP1

Early Adult SEP (SEP_1) → Midlife SEP (SEP_2) → Health

Early Adult SEP (SEP_1) ↓ r=.7 Own Educational Attainment

Midlife SEP (SEP_2) ↓ r=.7 Occupational Class

```
. reg health educ
```

Source	SS	df	MS
Model	2365.92653	1	2365.92653
Residual	96130.9624	9998	9.61501924
Total	98496.8889	9999	9.85067396

health	Coef.	Std. Err.	t	P>|t|
educ	.4873077	.0310655	15.69	0.000
_cons	.0052761	.0310092	0.17	0.865

```
. reg health educ occpn
```

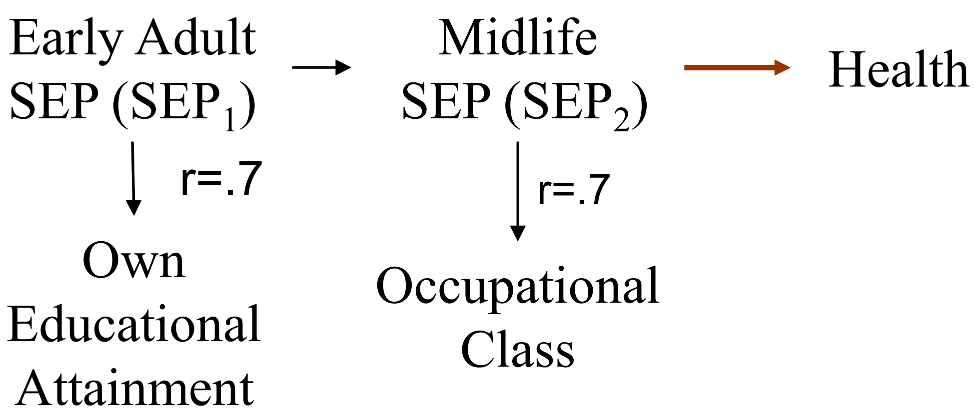
Source	SS	df	MS
Model	6006.87944	2	3003.43972
Residual	92490.0095	9997	9.25177648
Total	98496.8889	9999	9.85067396

health	Coef.	Std. Err.	t	P>|t|
educ	.2704272	.0323748	8.35	0.000
occpn	.648817	.032706	19.84	0.000
_cons	.0069007	.0304179	0.23	0.821

Completely Incorrect

Conclusion:
Less than half the effect of childhood SEP is mediated by adult SEP.

Empirical Example: spurious evidence for latent effects of SEP1

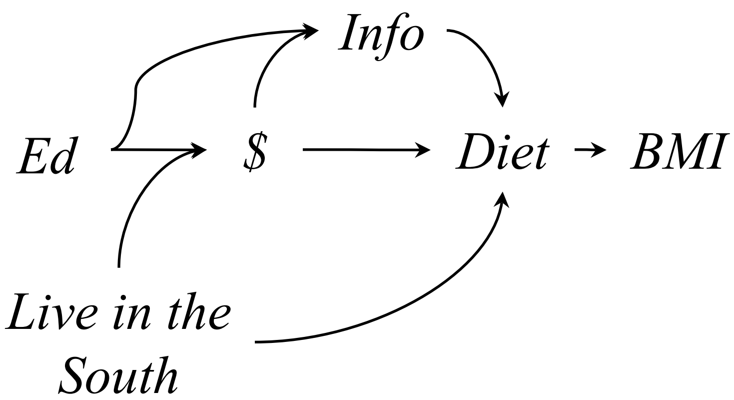


- Even fairly modest measurement error on the mediator compromises the conventional B-K approach to testing mediation.
- If you explicitly assess the measurement error, you can correct for this problem.
- If you have no idea how bad your measurement error is, recognize the limitations of your inferences.

Commonly Ignored Problems with Classic Decomposition Approach

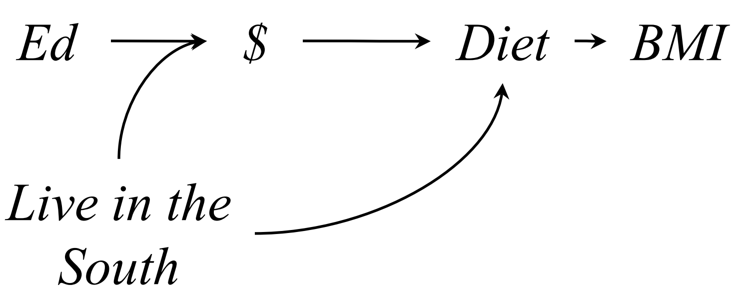
- Unmeasured confounding of the mediator and the outcome
- Effect modification of the direct pathway by the mediator
- Nonlinear models
- Inadequate measurement of the exposure or the mediator

Mediation



We wish to estimate the direct effect of education on BMI, not mediated by $. Earnings are also affected by place of residence, and place of residence influences diet.

Confounded Mediators

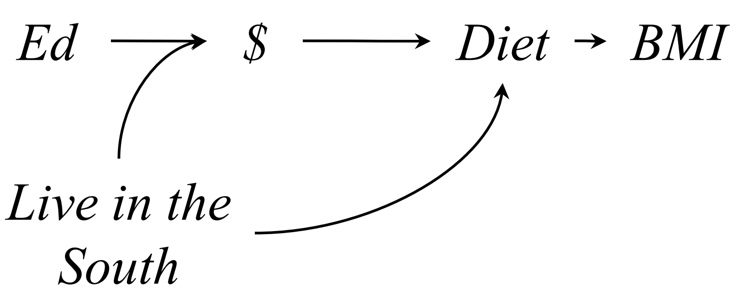


Earnings are also affected by place of residence, and place of residence influences diet.

❖Education increases earnings by $5,000.
❖Southerners earn $10,000 less than non-southerners.
❖Every $5K earnings improves diet enough to reduce BMI by 1.
❖Southerners' diets are so bad they have avg BMI 2 units higher than others.

E(Earnings)= 30,000-10,000*(Southern) +5,000*(Education)
E(BMI) = 31 -1*(Earnings/5,000) +2*(Southern)

Confounded Mediators



$$E(\text{Earnings}) = 25{,}000 - 5{,}000*(\text{Southern}) + 5{,}000*(\text{Education})$$

$$E(\text{BMI}) = 30 \quad -1*(\text{Earnings}/5{,}000) + 2*(\text{Southern})$$

Group	Education	Southern	Earnings	BMI
A	Low=0	No=0	$25,000	25
B	High=1	No=0	$30,000	24
C	Low=0	Yes=1	$20,000	28
D	High=1	Yes=1	$25,000	27

Confounded Mediators

$$\begin{tikzcd} Ed \arrow[r] & \$ \arrow[r] & Diet \arrow[r] & BMI \\ \text{Live in the South} \arrow[ur, bend left] \arrow[urr, bend right] & & & \end{tikzcd}$$

E(Earnings)= 25,000-10,000*(Southern) +5,000*(Education)
E(BMI) = 30 -1*(Earnings/5,000) +2*(Southern)

% of the population	Group	Education	Southern	Earnings	BMI
25%	A	Low=0	No=0	$25,000	25
25%	B	High=1	No=0	$30,000	24
25%	C	Low=0	Yes=1	$15,000	29
25%	D	High=1	Yes=1	$20,000	28

Confounded Mediators

$$Ed \longrightarrow \$ \rightarrow Diet \rightarrow BMI$$

Live in the South

E(Earnings)= 25,000-5,000*(South) +5,000*(Ed)
E(BMI) = 30 -1*(Earnings/5,000) +2*(South)

Does “southern” confound the relationship between education and BMI?

% of the population	Group	Education	Southern	Earnings	BMI
25%	A	Low=0	No=0	$25,000	25
25%	B	High=1	No=0	$30,000	24
25%	C	Low=0	Yes=1	$20,000	28
25%	D	High=1	Yes=1	$25,000	27

Confounded Mediators

$$
\begin{tikzcd}
\mathit{Ed} \arrow[r] & \$ \arrow[r] & \mathit{Diet} \arrow[r] & \mathit{BMI} \\
\textit{Live in the South} \arrow[ur, bend left] \arrow[urr, bend right] & & &
\end{tikzcd}
$$

E(Earnings)= 25,000-5,000*(South) +5,000*(Ed)
E(BMI) = 30 -1*(Earnings/5,000) +2*(South)

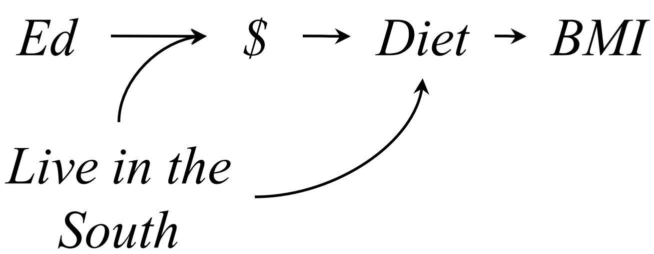
Does “southern” confound the relationship between education and BMI?

-Is “southern” a common cause of education and BMI or on the path of a common cause of education and BMI? (NO)

-Is “southern” associated with education? (NO, not unless condition on a collider)

% of the population	Group	Education	Southern	Earnings	BMI
25%	A	Low=0	No=0	$25,000	25
25%	B	High=1	No=0	$30,000	24
25%	C	Low=0	Yes=1	$20,000	28
25%	D	High=1	Yes=1	$25,000	27

Confounded Mediators



E(Earnings)= 25,000-5,000*(South) +5,000*(Ed)
E(BMI) = 30 -1*(Earnings/5,000) +2*(South)

Does "southern" confound the relationship between earnings and BMI?

% of the population	Group	Education	Southern	Earnings	BMI
25%	A	Low=0	No=0	$25,000	25
25%	B	High=1	No=0	$30,000	24
25%	C	Low=0	Yes=1	$20,000	28
25%	D	High=1	Yes=1	$25,000	27

Confounded Mediators: total effect estimates

Ed → *$* → *Diet* → *BMI*

Live in the South

E(Earnings)= 25,000-5,000*(South) +5,000*(Ed)
E(BMI) = 30 -1*(Earnings/5,000) +2*(South)

What you measure:

% of the population	Group	Education	Southern	Earnings	BMI
50%	A + C	Low=0	No=0	$22,500	26.5
50%	B + D	High=1	No=0	$27,500	25.5

Full data:

% of the population	Group	Education	Southern	Earnings	BMI
25%	A	Low=0	No=0	$25,000	25
25%	B	High=1	No=0	$30,000	24
25%	C	Low=0	Yes=1	$20,000	28
25%	D	High=1	Yes=1	$25,000	27

Confounded Mediators: total effect estimates

Ed → *$* → *Diet* → *BMI*

Live in the South

E(Earnings)= 25,000-5,000*(South) +5,000*(Ed)
E(BMI) = 30 -1*(Earnings/5,000) +2*(South)

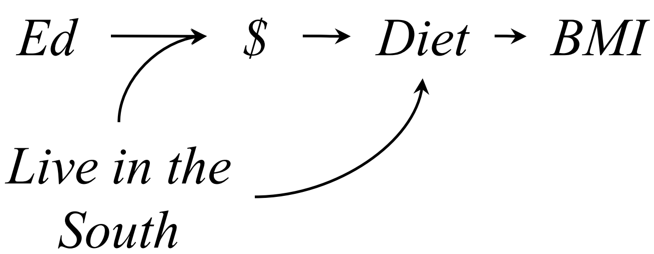
What you measure:

% of the population	Group	Education	Southern	Earnings	BMI
50%	A + C	Low=0	No=0	$22,500	26.5
50%	B + D	High=1	No=0	$27,500	25.5

Compare the High vs Low Education individuals to estimate the effect of education:

- ❖Estimated effect of education on earnings: $5,000
- ❖Estimated effect of education on BMI (fully mediated by earnings): -1

Confounded Mediators: Direct Effect Estimate



E(Earnings)= 25,000-5,000*(South) +5,000*(Ed)
E(BMI) = 30 -1*(Earnings/5,000) +2*(South)

What is the effect of education on BMI that is not mediated by earnings?
Conventional analyses compare the high and low education individuals *with the same earnings.*

% of the population	Group	Education	Southern	Earnings	BMI
25%	A	Low=0	No=0	$25,000	25
25%	B	High=1	No=0	$30,000	24
25%	C	Low=0	Yes=1	$20,000	28
25%	D	High=1	Yes=1	$25,000	27

Confounded Mediators: Direct Effect Estimate

Ed → *$* → *Diet* → *BMI*

Live in the South

E(Earnings)= 25,000-5,000*(South) +5,000*(Ed)
E(BMI) = 30 -1*(Earnings/5,000) +2*(South)

What is the effect of education on BMI that is not mediated by earnings?
Conventional analyses compare the high and low education individuals *with the same earnings.*

% of the population	Group	Education	Southern	Earnings	BMI
25%	A	Low=0	No=0	$25,000	25
25%	B	High=1	No=0	$30,000	24
25%	C	Low=0	Yes=1	$20,000	28
25%	D	High=1	Yes=1	$25,000	27

Estimated effect of education on BMI among people with the same earnings= +2

Confounded Mediators: Direct and Indirect Effect Decomposition

- Estimated total effect of education on BMI: -1
- Estimated direct effect of education on BMI: +2
- Direct + Indirect=Total
- Estimated indirect effect of education on BMI: -3

% of the population	Group	Education	Southern	Earnings	BMI
25%	A	Low=0	No=0	$25,000	25
25%	B	High=1	No=0	$30,000	24
25%	C	Low=0	Yes=1	$20,000	28
25%	D	High=1	Yes=1	$25,000	27

Estimated effect of education on BMI among people with the same earnings= +2

Confounded Mediators: Direct and Indirect Effect Decomposition

- ❖Estimated total effect of education on BMI: -1
- ❖Estimated direct effect of education on BMI: +2
- ❖Estimated indirect effect of education on BMI: -3

```
. reg bmi owned

      Source |       SS       df       MS              Number of obs =  100000
-------------+------------------------------           F(  1, 99998) = 2334.87
       Model |  26194.1795     1  26194.1795           Prob > F      =  0.0000
    Residual |  1121845.26 99998   11.218677           R-squared     =  0.0228
-------------+------------------------------           Adj R-squared =  0.0228
       Total |  1148039.44 99999  11.4805092           Root MSE      =  3.3494

------------------------------------------------------------------------------
         bmi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       owned |  -1.023621    .021184   -48.32   0.000    -1.065141   -.9821003
       _cons |   26.52335   .0150206  1765.80   0.000     26.49391    26.55279
------------------------------------------------------------------------------

. reg bmi owned earnings

      Source |       SS       df       MS              Number of obs =  100000
-------------+------------------------------           F(  2, 99997) =14169.08
       Model |  253502.867     2  126751.433           Prob > F      =  0.0000
    Residual |  894536.576 99997  8.94563413           R-squared     =  0.2208
-------------+------------------------------           Adj R-squared =  0.2208
       Total |  1148039.44 99999  11.4805092           Root MSE      =  2.9909

------------------------------------------------------------------------------
         bmi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       owned |   1.986342    .026728    74.32   0.000     1.933956    2.038729
    earnings |  -.0006031   3.78e-06  -159.41   0.000    -.0006105   -.0005957
       _cons |     40.091   .0861646   465.28   0.000     39.92212    40.25988
------------------------------------------------------------------------------
```

Confounded Mediators: Direct and Indirect Effect Decomposition

```
. reg bmi owned earnings

      Source |       SS       df       MS              Number of obs =  100000
-------------+------------------------------           F(  2, 99997) =14169.08
       Model |  253502.867     2  126751.433           Prob > F      =  0.0000
    Residual |  894536.576 99997  8.94563413           R-squared     =  0.2208
-------------+------------------------------           Adj R-squared =  0.2208
       Total |  1148039.44 99999  11.4805092           Root MSE      =  2.9909

------------------------------------------------------------------------------
         bmi |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       owned |   1.986342    .026728    74.32   0.000     1.933956    2.038729
    earnings |  -.0006031   3.78e-06  -159.41   0.000    -.0006105   -.0005957
       _cons |     40.091   .0861646   465.28   0.000     39.92212    40.25988
------------------------------------------------------------------------------
```

❖This only looks so extreme because I set the numbers to illustrate, with almost all of the variation in earnings due to either own ed or southern residence.

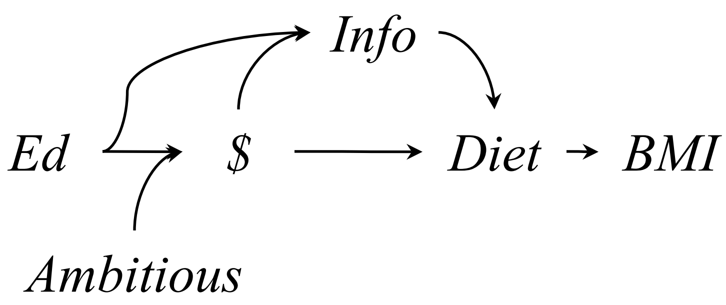
❖In real situations (i.e. with modest correlations among the variables), this phenomenon is usually far less marked.

❖The severity depends on the strength and mechanisms of the relationships, so if you don't know these, you don't know how bad the bias is.

Commonly Ignored Problems with Classic Decomposition Approach

- Unmeasured confounding of the mediator and the outcome
- Effect modification of the direct pathway by the mediator
- Nonlinear models
- Inadequate measurement of the exposure or the mediator

Mediation if the Mediator Modifies the Direct Pathway



Higher earnings also enable people to access the most current nutritional research from the best physicians. This is especially valuable for people with high education.

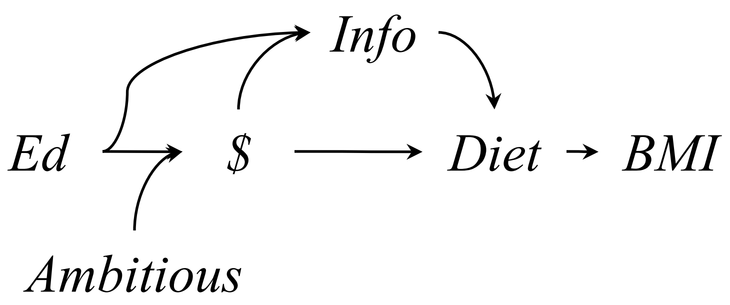
Assume everything is Z-scored (it's really hard to get plausible #s)

E(Earnings) = 2*Education + 3*Ambitious

E(Information)= 4*Education + 2*Earnings + 2*Earnings*Education

E(BMI) =-.2*(Earnings) - 0.3*Information

Mediation if the Mediator Modifies the Direct Pathway

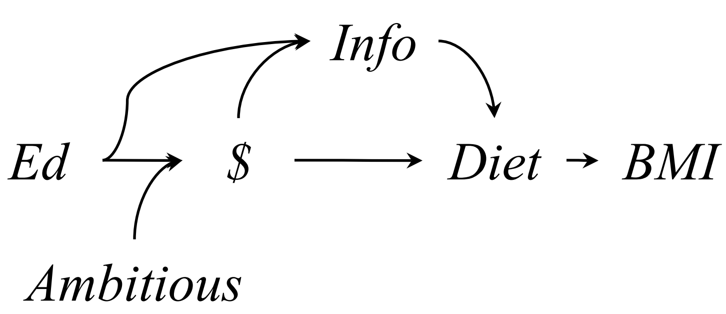


Higher earnings also enable people to access the most current nutritional research from the best physicians. This is especially valuable for people with high education.

Assume everything is Z-scored (it's really hard to get plausible #s)
E(Earnings) = 2*Education + 3*Ambitious
E(Information)= 4*Education + 2*Earnings + 2*Earnings*Education
E(BMI) =-.2*(Earnings) - 0.3*Information

E(BMI|Ed=0, Earnings=0)=-.2*0-0.3*0=0
E(BMI|Ed=1, Earnings=0)=-.2*0-.3*(4)=-1.2
Direct Effect=-1.2

Mediation if the Mediator Modifies the Direct Pathway

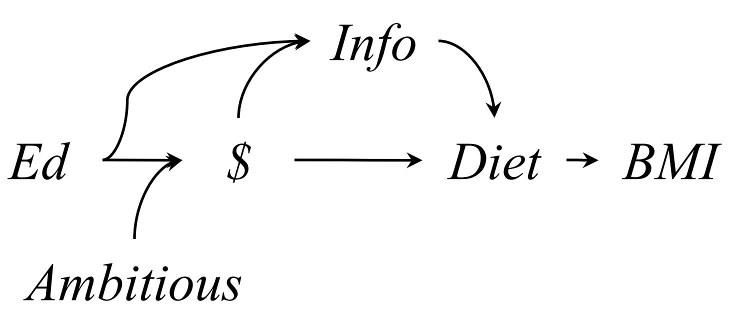


Higher earnings also enable people to access the most current nutritional research from the best physicians. This is especially valuable for people with high education.

Assume everything is Z-scored (it's really hard to get plausible #s)
E(Earnings) = 2*Education + 3*Ambitious
E(Information)= 4*Education + 2*Earnings + 2*Earnings*Education
E(BMI) =-.2*(Earnings) - 0.3*Information

E(BMI|Ed=0, Earnings=1)=-.2*1-0.3*(2)=-0.8
E(BMI|Ed=1, Earnings=1)=-.2*1-0.3*(4+2+2)=-2.6
Direct Effect=-1.8

Mediation if the Mediator Modifies the Direct Pathway



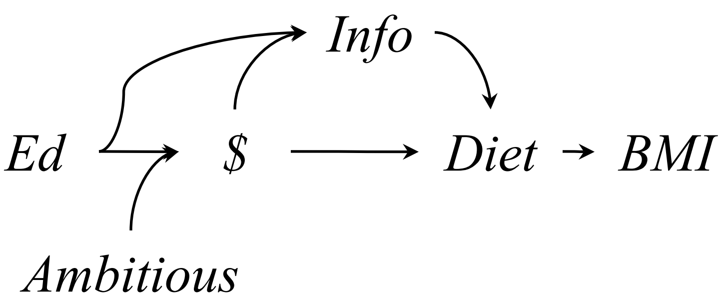
Higher earnings also enable people to access the most current nutritional research from the best physicians. This is especially valuable for people with high education.

Direct Effect if Earnings=0 is -1.2
Direct Effect if Earnings=1 is -1.8

The "controlled" direct effect is the effect of education on BMI if we forced everyone to have the same particular value of earnings.
This depends on which value of earnings we assign everyone in the population.

There may be a unique "controlled" direct effect for every possible value of earnings.

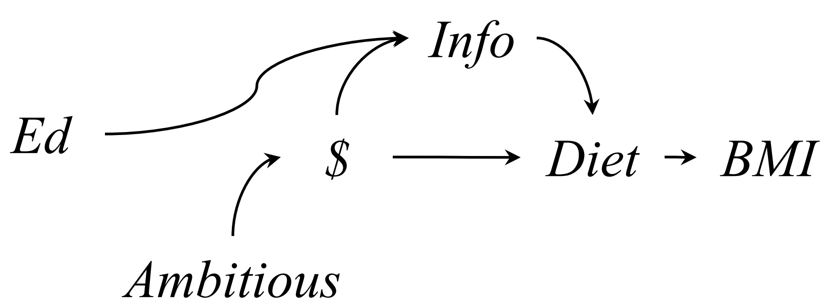
Mediation if the Mediator Modifies the Direct Pathway



But how can we make everyone have the same earnings?
We would have to eliminate variations in ambition!

Impossible!

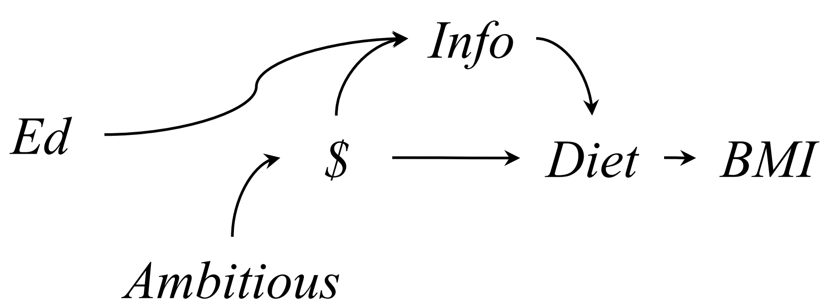
But, perhaps we can prevent employers from compensating extra for education:



Mediation if the Mediator Modifies the Direct Pathway

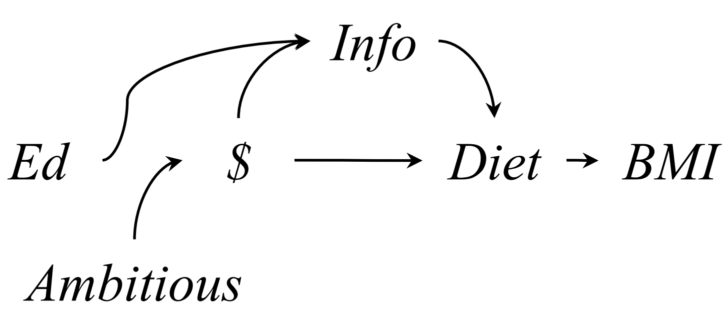
But how can we make everyone have the same earnings?
We would have to eliminate variations in ambition!
Impossible!

But, perhaps we can prevent people from compensating extra for education:



Now some people will have low earnings, and some will have high earnings, but this will be unrelated to education.
What is the effect of education on BMI now?

Mediation if the Mediator Modifies the Direct Pathway

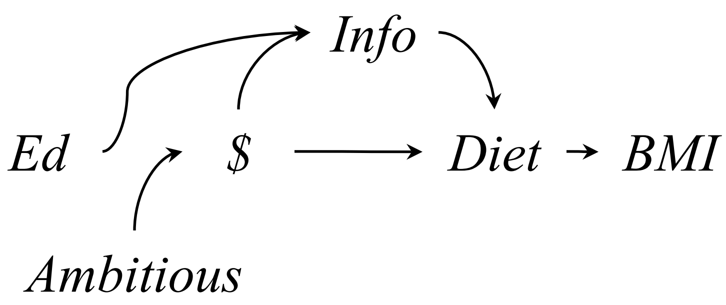


Higher earnings also enable people to access the most current nutritional research from the best physicians. This is especially valuable for people with high education.

Assume everything is Z-scored (it's really hard to get plausible #s)
E(Earnings)   = 3*Ambitious
E(Information)= 4*Education + 2*Earnings + 2*Earnings*Education
E(BMI)   =-.2*(Earnings) - 0.3*Information

E(BMI|Ed=0, Ambitious=0)=-.2*0-0.3*0=0
E(BMI|Ed=1, Ambitious=0)=-.2*0-.3*(4)=-1.2
Effect of education on BMI among the non-ambitious=-1.2

Mediation if the Mediator Modifies the Direct Pathway

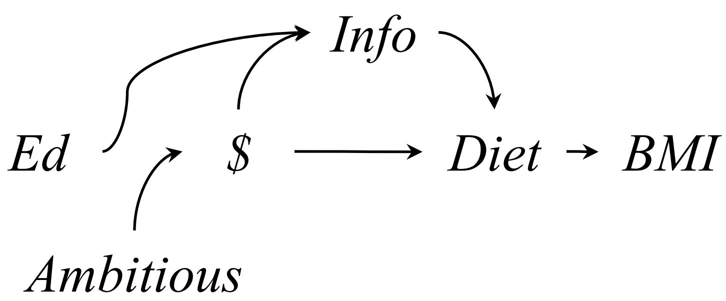


Higher earnings also enable people to access the most current nutritional research from the best physicians. This is especially valuable for people with high education.

Assume everything is Z-scored (it's really hard to get plausible #s)
E(Earnings)   = 3*Ambitious
E(Information)= 4*Education + 2*Earnings + 2*Earnings*Education
E(BMI)   =-.2*(Earnings) - 0.3*Information

E(BMI|Ed=0, Ambitious=1)=-.2*3-0.3*(0+6)=-2.4
E(BMI|Ed=1, Ambitious=1)=-.2*3-.3*(4+6+6)=-5.4
Effect of education on BMI among the ambitious=-3

Mediation if the Mediator Modifies the Direct Pathway

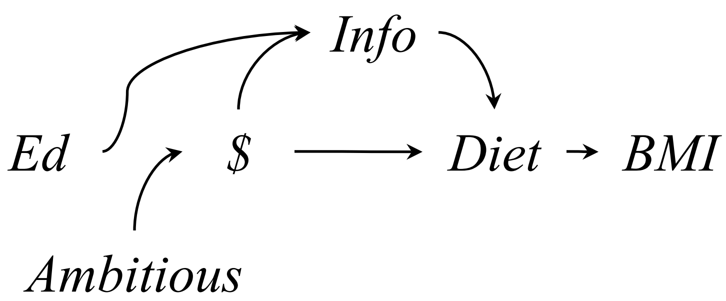


Effect of education on BMI among the non-ambitious=-1.2
Effect of education on BMI among the ambitious=-3

The population average effect of education on BMI *depends on the prevalence of ambition.*

The population average effect of education on BMI if you eliminate the influence of education on earnings is called the “natural” direct effect.

Mediation if the Mediator Modifies the Direct Pathway



❖If Earnings do not modify the direct pathway from education to BMI, all values of the “controlled” direct effect are the same.

❖Furthermore, the controlled direct effect is the same as the natural direct effect.

Mediation

- Core idea and motivation
- Conventional approach
- Link with lifecourse models
- When conventional approach fails
- Counterfactual framework: CDE, NDE, and NIE

Mediation: Direct vs Indirect Effects Decomposition

$$\begin{tikzcd} X \arrow[r] \arrow[rr, bend left] & M \arrow[r] & Y \end{tikzcd}$$

- Controlled direct effect: The difference in counterfactual Y if you set X to 1 vs X to 0 but in both circumstances also set M=0
  - $Y_{x=1,m=0}$ and $Y_{x=0,m=0}$
- Controlled direct effect: The difference in counterfactual Y if you set X to 1 vs X to 0 but in both circumstances also set M=1
  - $Y_{x=1,m=1}$ and $Y_{x=0,m=1}$
- These two don't have to be the same, so there are as many values of the "controlled direct effect" as there are possible values of M.
- Controlled "indirect" effect doesn't make sense because we are setting M to the same value for everyone in the controlled framework.

Natural Direct and Indirect Effects: Logistic Model

- Natural direct effect: The difference in counterfactual Y if you set X to 1 vs X to 0 but in both circumstances also set M to what it would be for a reference value of X, e.g., X=0
- M will differ between different people, but theoretically unrelated to X
- Sometimes criticized because it requires a cross-world contrast: the potential outcome Y if X=1 but M= what it would be if X=0
- If no X-Y, X-M, or M-Y confounding, no interaction between X and M, and rare outcome, then classic BK is approximately NDE/NIDE
  - $\text{Logit}(p(Y|X,M,C))=a0+a1*X+a2*M+a3*C$
  - $E(M|x,c)=b0+b1*X+b2*C$
  - $NDE=CDE=\exp(a1*(x-x'))$
  - $NIDE=\exp(b1*a2*(x-x'))$

Natural Direct and Indirect Effects: Logistic Model

- If interaction between X and M estimate:
- $\text{Logit}(p(Y|X,M,C))=a_0+a_1*X+a_2*M+a_3*X*M+a_4*C$
- $E(M|x,c)=b_0+b_1*X+b_2*C$

- $CDE(m)=\exp((a_1+a_3*m)*(x-x'))$
- $NIDE=\exp((a_2*b_1 + a_3*b_1*x)(x-x'))$

Controlled vs Natural Direct Effects

- CDE requires fewer assumptions
- NDE cannot be estimated with recanting witness that is influenced by X and a common cause of M and Y.
- Debate about which is more policy relevant
- Often not that different, so debate is perhaps overblown, but check for interactions

Estimation

- Now lots of packages, but if you have the formulas, you can always calculate what you need, so don't get too tied to a particular software package.
- Imai's generalized approach:
  - To estimate any direct or indirect effect, just estimate the potential outcomes.
  - Let $Y_{x=1,m=1}$ represent the counterfactual value of Y setting X to 1 and M to 1
  - Let $Y_{x=1,m(x=1)}$ represent the counterfactual value of Y setting X to 1 and M to whatever value M would take if X were set to 1.

Estimation

1. Fit a model for the mediator, as a function of exposure and confounders
   - For example reg M X C
   - Make two fake-world copies of each observation, but set X to 0 in one world and X to 1 in the other
   - Generate fake-world predictions using the regression equation from the real world
   - So you have predicted values of $M_{x=0}$ and $M_{x=1}$
2. Fit a model for the outcome, as a function of exposure, mediator, and confounders (if you need to control for variables you cannot include in the outcome model, use IPW). Remember to allow for exposure*mediator interactions if those are plausible.
   - For example reg Y M X M*X C
   - In your fake world where you used X=0 to predict M, use the outcome regression equation to predict the value of Y when X=0 and $M=M_{x=0}$
   - In your fake world where you used X=1 to predict M, switch everyone to X=0 and use the outcome regression equation to predict the value of Y when X=0 and $M=M_{x=1}$
   - So you have predicted values of $Y_{x=0}$ and $M_{x=1}$
3. Now you have $Y_{x=0,m=m(x=1)}$ and $Y_{x=0,m=m(x=0)}$ and the difference between these two is the indirect effect of X on Y mediated by M when X is 0, i.e., the Natural indirect effect.
4. Bootstrap the Cis.

Estimation

- Now you have $Y_{x=0,m=m(x=1)}$ and $Y_{x=0,m=m(x=0)}$ and the difference between these two is the indirect effect of X on Y mediated by M when X is 0, i.e., the Natural indirect effect.
- You can do this for any definition of treatment vs no treatment,
- Simulate the potential values of the mediator under treatment versus no treatment (whatever contrast you would like, e.g., "all treated" vs "nobody treated" or "everyone has education=12" vs "everyone has education=8 years" or "one third of the population has educ=16, one third educ=12, and one third=8" vs "one half has educ=12 and one half has educ=16").
- You can also do this to calculate $Y_{x=1,m=0}$ and $Y_{x=0,m=0}$ the controlled direct effect setting M to 0 or $Y_{x=1,m=1}$ and $Y_{x=0,m=1}$ the controlled direct effect setting M to 1

Other approaches

- Inverse Odds Ratio Weighting
  - Use the mediator to predict the exposure first, and use the odds ratio to characterize that association because the odds ratio is the same even if you switch exposure-outcome.
  - So M->X is the same as X->M
  - Now calculate the inverse of the odds that the mediator took the value it did, given the value of X for each individual.
  - In the weighted population, X has no effect on M, so calculate the effect of X on Y in this weighted population to estimate the direct effect of X on M
  - Uses the same IPW trick!
  - Very clever. Convenient for multiple mediators. Convenient for any functional form of the outcome model.
  - Weights lead to imprecise estimates.

end