Conceptualizing and testing mediated effects

Steve Gregorich

Epi 222: Health Disparities Research Methods
May 8, 2014

Health disparities

Healthy People 2010: Goal #2 To *eliminate* health disparities among key demographic strata	*Healthy People 2020*: Social Determinants of Health
Gender	√
Race/ethnicity	√
Education or income	√ (now a *determinant*)
Disability	√
Geographic location	√
Sexual orientation	LGBT Health (also a new topic)

Healthy People 2000, 2010, and 2020

Healthy People 2000
Reduce health disparities.

Healthy People 2010
Eliminate health disparities.

Healthy People 2020
Achieve health equity,
Eliminate disparities, and
Improve the health of all groups.

Health disparities

- *Elimination* of health disparities requires

  Identification of health disparities, *and*

  Explanation of the mechanisms underlying those disparities

- Consider the demographic strata targeted by Healthy People 2010...
  (Gender - R/E - SES - Disability - Location - Sexual orientation)

  Fine for **identification** of health disparities

  What about **explanation**?

  What is missing?

Health Disparities

Explaining health disparities requires an understanding of causal pathways

Examples of generic factors that may lie along the causal pathway to health

Competing factors	
Genetic	Micro
Congenital	
Developmental	
Psychological	Mezzo
Behavioral/lifestyle	
Social	
Environmental	
Societal/structural	Macro

Health Disparities

Explaining health disparities requires...

...answering questions about causal mechanisms. For example,

What is the full causal chain of events?

Where do demographic strata fall along the causal chain of events?

What causes, is caused by, or is just correlated with demographic strata?

Testing competing hypotheses about underlying causal mechanisms

Using regression models to test causal hypotheses

Public Health Research Reality	Public Health Research Practice
Regression models usually reflect causal hypotheses	The causal nature of the hypotheses is not always explicitly stated
Regression framework is flexible and can address competing causal hypotheses	Flexibility is not always exploited
Causal inferences should be made cautiously, especially from observational data.	Reasonable alternative causal hypotheses are often left unaddressed

- If the goal is explanation, then causal hypotheses are being specified
- Attention to causal hypotheses helps to advance knowledge

Topics

- Different causal models and topics we will consider...

  Total effects model—bivariate regression

  Conditional effects model—standard multivariate regression

  Spurious correlation
    Shared causes between two explanatory variables

  Mediated effects model—direct and indirect causal effects

  Suppressor effects—AKA negative confounding

- Introductory material is supplemented with worked examples based upon the Duke, NC EPESE study data

- Finally, I will present a more advanced example from the literature

EPESE data

Established Populations for Epidemiological Studies of the Elderly (EPESE)

- Duke site
- Probability sample
- 65 years and older
- 54% African American
- $N \approx 2700$ (with complete data on key variables)
- Baseline data, circa 1981

EPESE data

Explanatory variables

- Race (0=White, 1=Black)
- Income (< 5K, 5-7K, 7-10K, 10-15K, 15K+)

Outcomes

- CES-D somatic symptoms scale (> is worse)
- Activities of daily living (> is worse)

Analyses

- Example analyses are for demonstration only
- I present standardized regression parameters throughout

Topics

- Different causal models and topics we will consider...

  Total effects model—bivariate regression

  Conditional effects model—standard multivariate regression

  Spurious correlation

  Shared causes between two explanatory variables

  Mediated effects model—direct and indirect causal effects

  Suppressor effects—AKA negative confounding

- A more advanced example

Total Effects: *Bivariate* regression model

$$\text{Somatic} = \text{intercept} + 0.117 \times \text{Race} + \varepsilon$$

- Race effect expressed as a causal diagram

$$\boxed{\textit{Race}} \xrightarrow{0.117^{**}} \boxed{\textit{Somatic}}$$

- Causal assumptions of the model:
  Race directly causes somatic symptoms
  Relationship is linear

- <u>Literal causal interpretation:</u>
  Black race causes significantly higher levels of reported symptoms

- The effects in bivariate models are often called **total effects**

Topics

- Different causal models and topics we will consider...

  Total effects model—bivariate regression

  Conditional effects model—standard multivariate regression

  An example of spurious correlation

  Shared causes between two explanatory variables

  Mediated effects model—direct and indirect causal effects

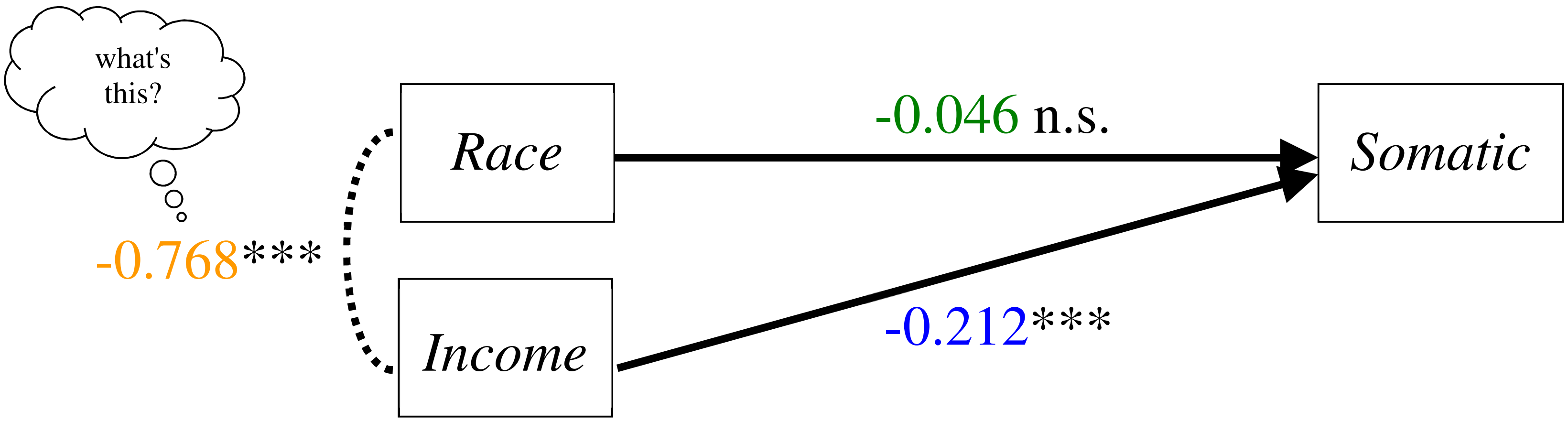
  Suppressor effects—AKA negative confounding

- A more advanced example

Conditional effects—standard multivariate regression Interpretation

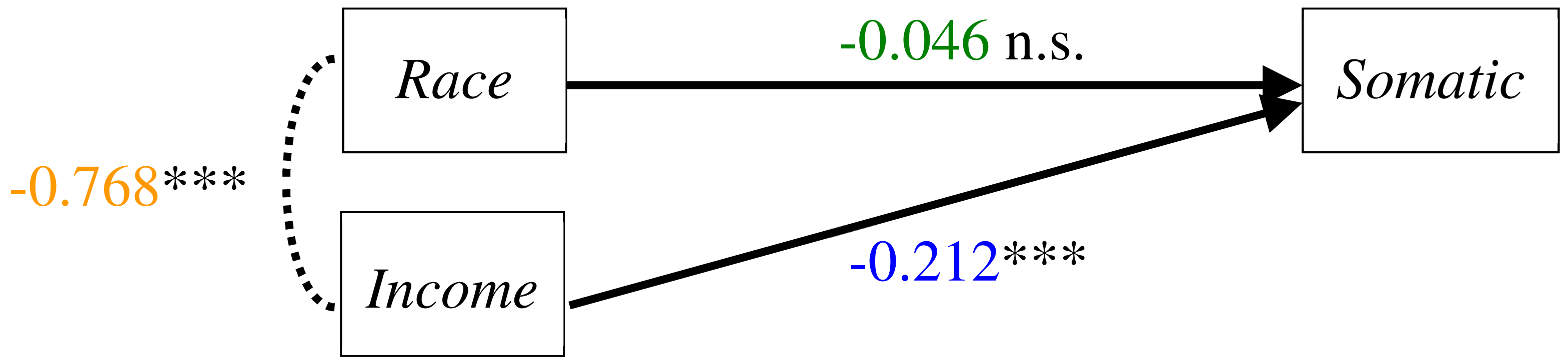
- Add *Income* to the model as an explanatory variable

$$\text{Somatic} = \text{intercept} - 0.046 \times \text{Race} - 0.212 \times \text{Income} + \varepsilon$$



- Note the *conditional* effect of Race is non-significant.
- <u>Literal causal interpretation:</u> Race is not an important cause of Somatic

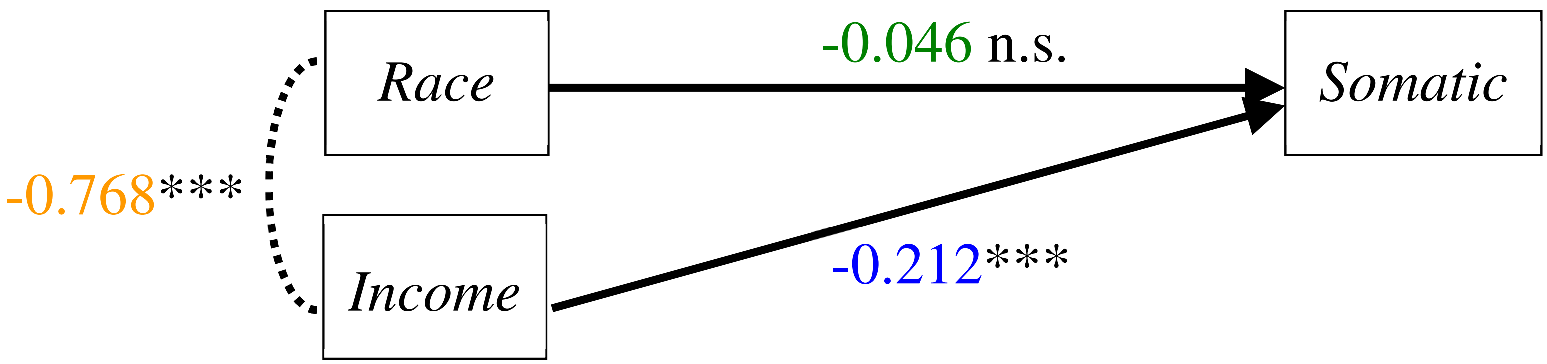
Conditional effects—standard multivariate regression Assumptions



- **Some causal assumptions of the model:**

1. Race and Income both directly, linearly cause reported somatic symptoms
2. Somatic symptoms do not cause Race or Income (*no* endogeneity)
3. Main effects only, no interaction between Race and Income
4. Race and Income are correlated, but not directly causally related

Conditional effects—standard multivariate regression
Assumptions



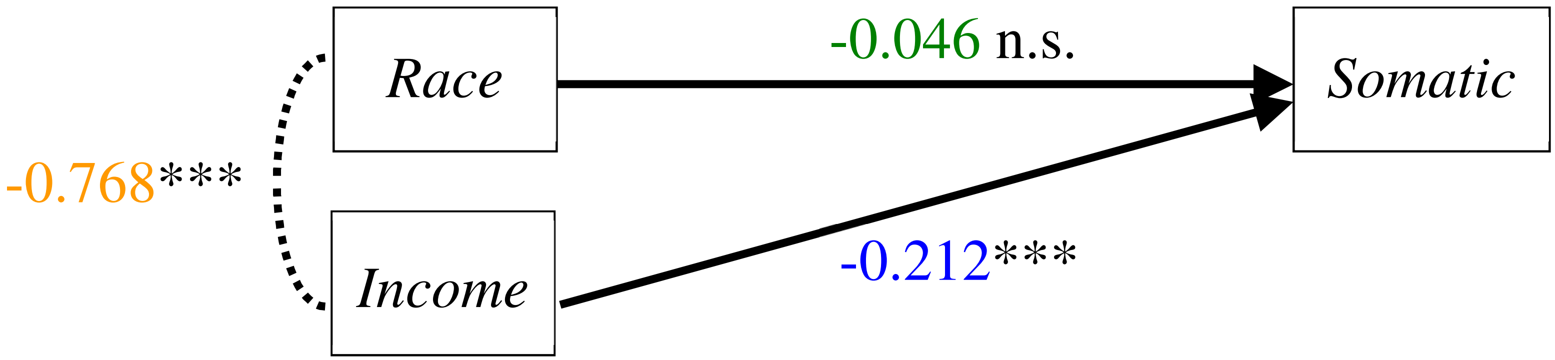
- **Causal assumption #2**

  Somatic symptoms do not cause Race or Income (*no* endogeneity)

- **Question**

  Is assumption #2 reasonable?

Conditional effects—standard multivariate regression
Assumptions



- **Causal assumption #4**

Race and Income are correlated, but not directly causally related

- **Question**

Is assumption #4 reasonable?

If we take assumption #4 as true, then **spurious** correlations exist between Race and Income as well as Race and Somatic

- **Next topic**

What circumstance would have to hold in order to support assumption 4?

Topics

- Different causal models we will consider...

  Total effects model—bivariate regression

  Conditional effects model—standard multivariate regression

  Spurious correlation

  Shared causes between two explanatory variables

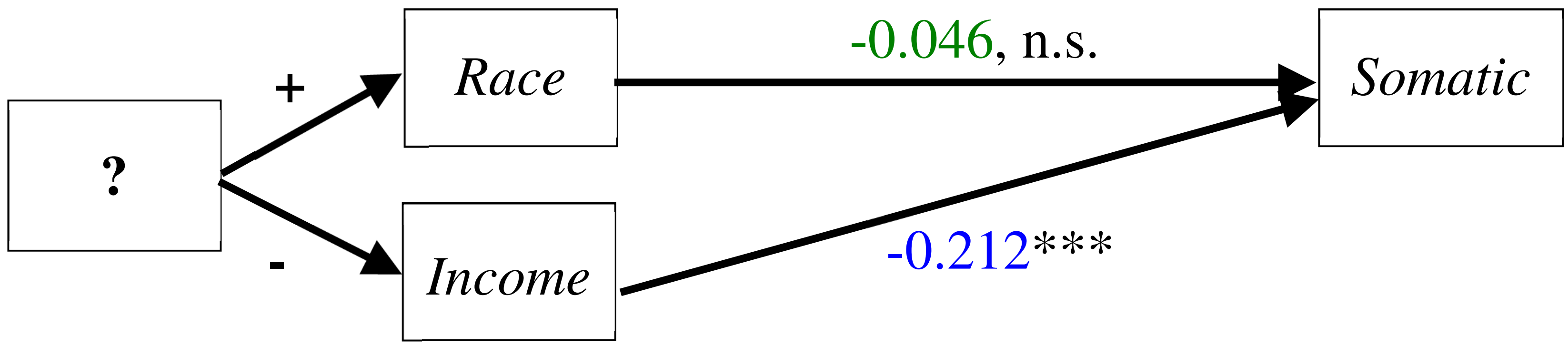
  Mediated effects model—direct and indirect causal effects

  Suppressor effects—AKA negative confounding

- A more advanced example

Shared causes: Correlated but not causally linked

- If 2 variables are correlated but not believed to be directly casually linked, they are often thought to share a common cause.



- So, defense of the interpretation
  '*Race is not causally related to reports of somatic symptoms*'

  requires

  (i) identification of potential common causes of Race and Income

  (ii) demonstration of no causal link between Race and Income

This does not seem likely.

Shared causes: Correlated but not causally linked

"In the late 1940s, before there was a polio vaccine,
public health experts in America noted that polio cases
increased in step with the consumption of ice cream and soft drinks…

Eliminating such treats was even recommended as part of an 'anti-polio diet.'

It turned out that polio outbreaks were most common
in the hot months of summer, when people naturally ate more ice cream."

from: David Alan Grier, a historian and statistician, George Washington University. *For Today's Graduate, Just One Word: Statistics*: NYTimes, August 5, 2009

Topics

- Different causal models we will consider...

  Total effects model—bivariate regression

  Conditional effects model—standard multivariate regression

  Spurious correlation

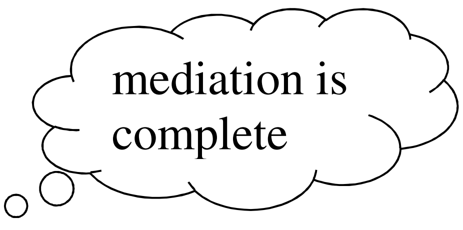
  Shared causes between two explanatory variables

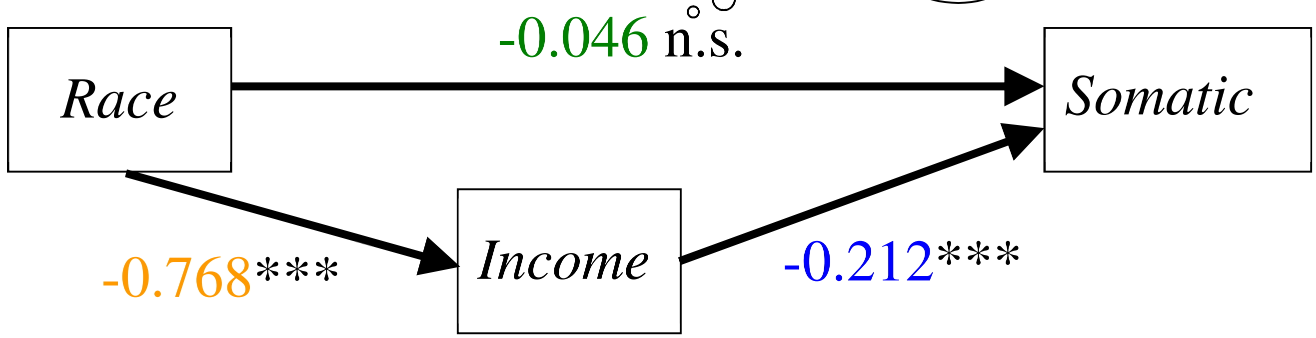
  Mediated effects model—direct and indirect causal effects

  Suppressor effects—AKA negative confounding

- A more advanced example

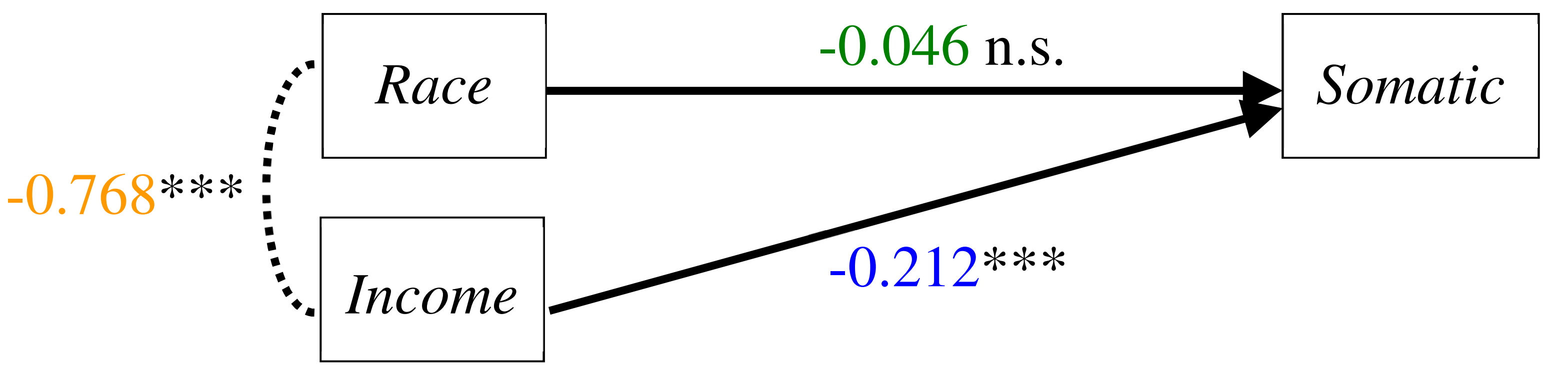
Mediated effects model





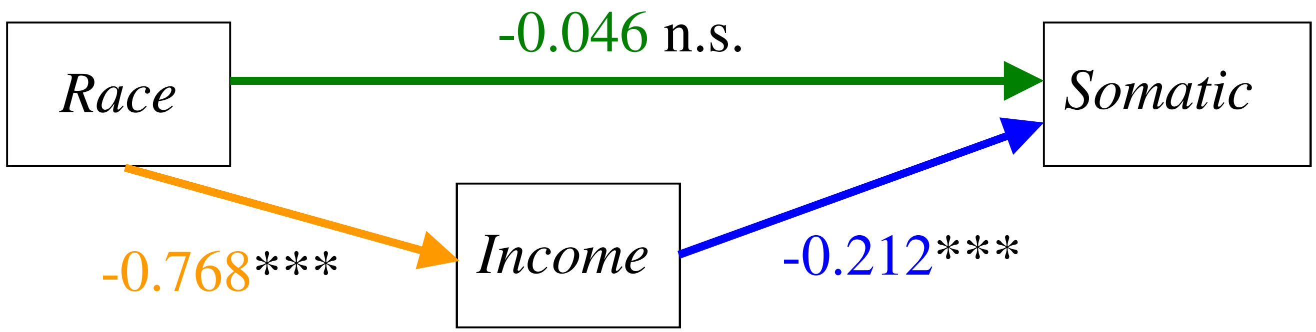
- What are the causal implications of the mediation model?

- Note. standardized parameter estimates are identical to the previous model



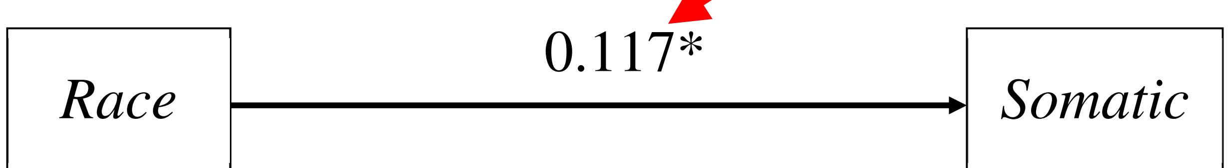
- What race-related inference do these two estimated models share?

Mediation: Decomposition of the total effect into direct and indirect effects



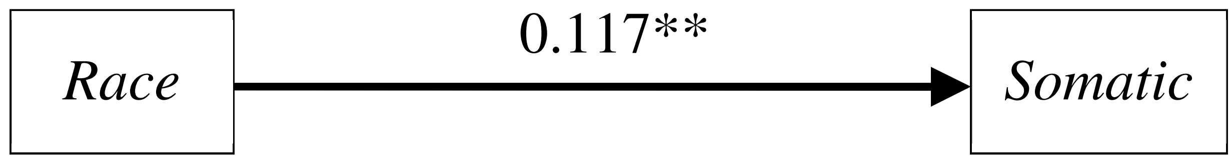
Direct effect: *Race* → *Somatic*	-0.046	
Indirect Effect: *Race* → *Income* → *Somatic*	0.163	(-0.768 × -0.212)
Total Effect	0.117	

- Remember the very first model?...

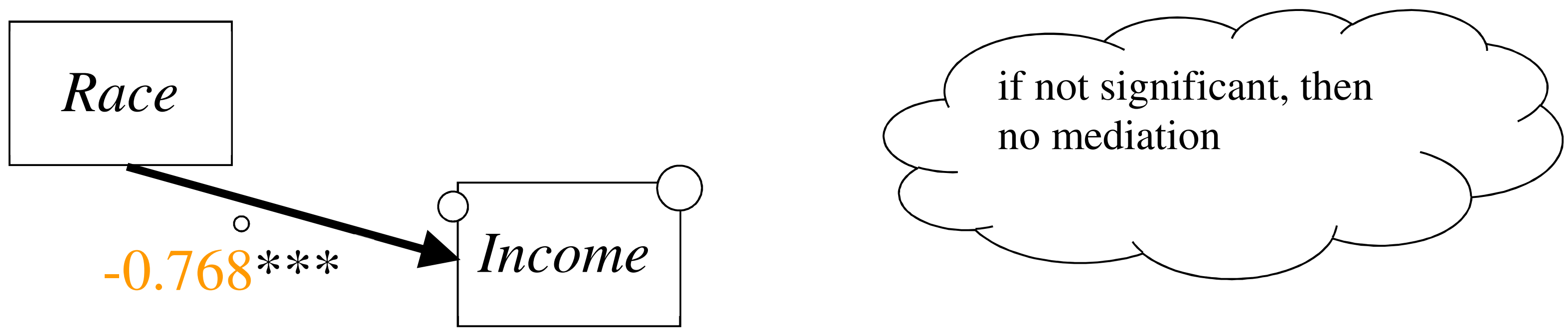


Mediation: Modeling steps

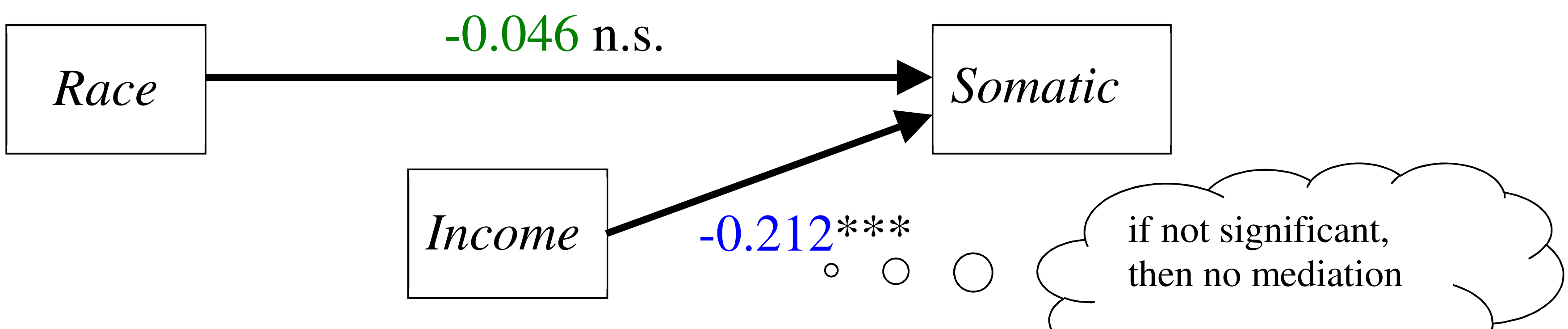
Model 1: Estimate **total effect** of *Race* on *Somatic*



Model 2: Estimate effect of *Race* on the candidate mediator, *Income*



Model 3: Estimate the **direct effect** of *Race* on *Somatic* (cond. on *Income*)

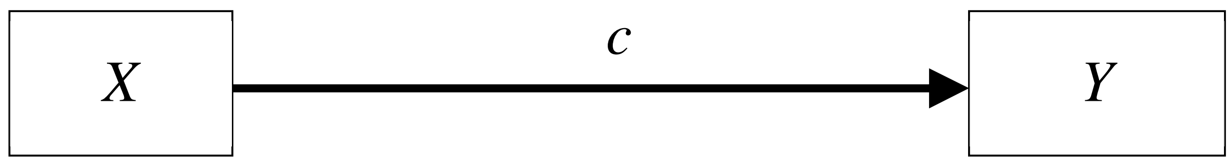


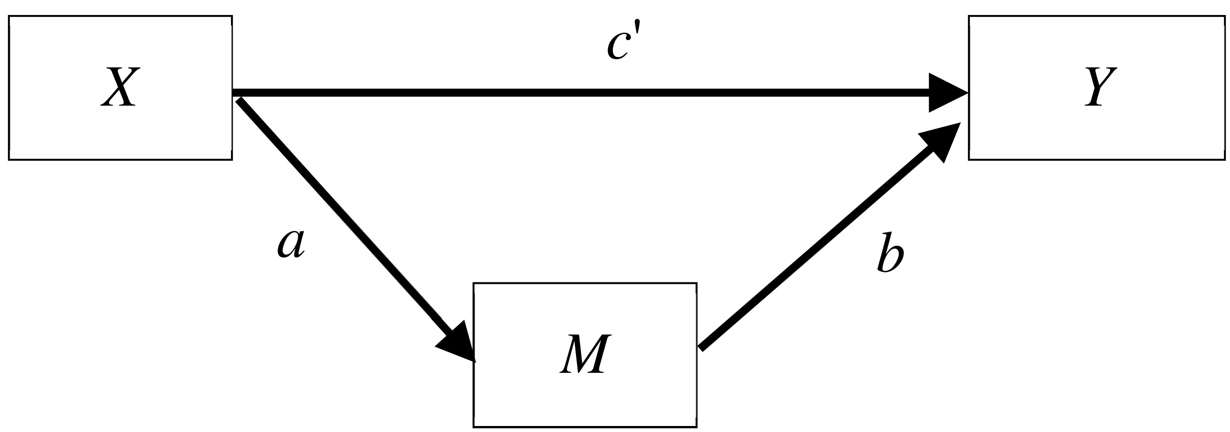
Mediation: Interpretation

Step 4: Assess the degree of mediation

Complete mediation

If the total effect (c) is significant and the direct effect (c') is not, then mediation is 'complete'





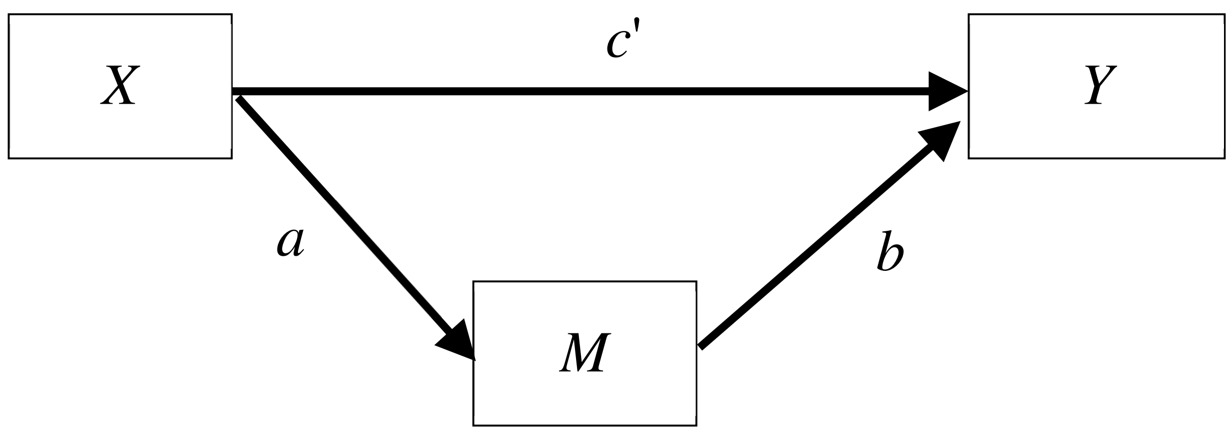
Mediation: Interpretation

Step 4: Assess the degree of mediation

Partial mediation
If the **direct effect** (c') is significant, then partial mediation may exist

Test the indirect effect

If both the direct (c') and indirect ($a \times b$) effects are significant, then partial mediation exists

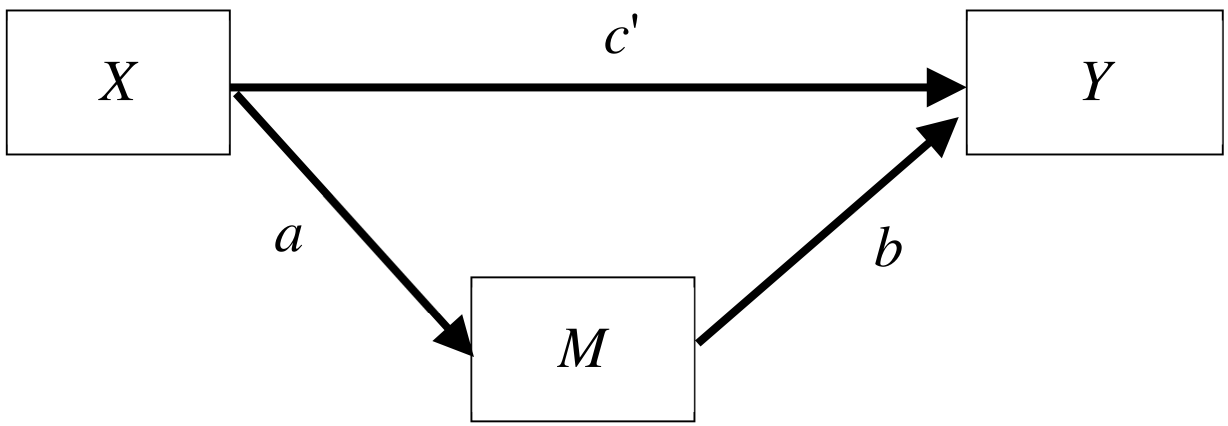


Mediation: Interpretation

Step 4: Assess the degree of mediation

No mediation

When the indirect effect ($a \times b$) is non-significant



Mediation: Estimating and testing the indirect effect

Obtaining a point estimate of an indirect effect

. Continuous mediator with

a continuous or binary X and any outcome distribution

Point estimate: $a \times b$

Binary mediator (more complex…)

Mediation: Estimating and testing the indirect effect

<u>Testing indirect effects</u>

1. Sobel and Aroian tests (test statistic and *p*-value)
   . Problem: low power
   . Problem: indirect effects can be skewed; *p*-values and CIs questionable

2. The joint-test of significance ('pass/fail' test of significance)
   . If both *a* and *b* are individually significant (p<.05)
   . Good power
   . Problems: does not provide a *p*-value or a confidence interval

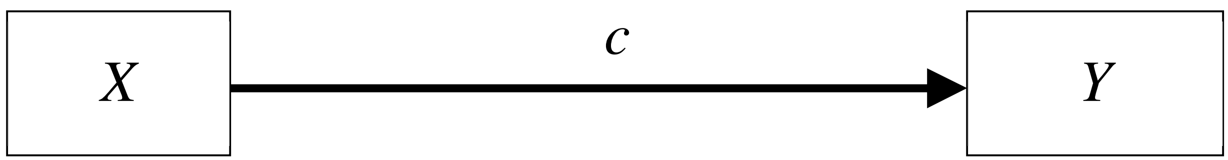
3. CIs based upon the distribution of the product of two normal variables (CI)
   . Better than Sobel- and Aroian-based CIs, but not the best option
   . Quick computation (http://www.amp.gatech.edu/RMediation)

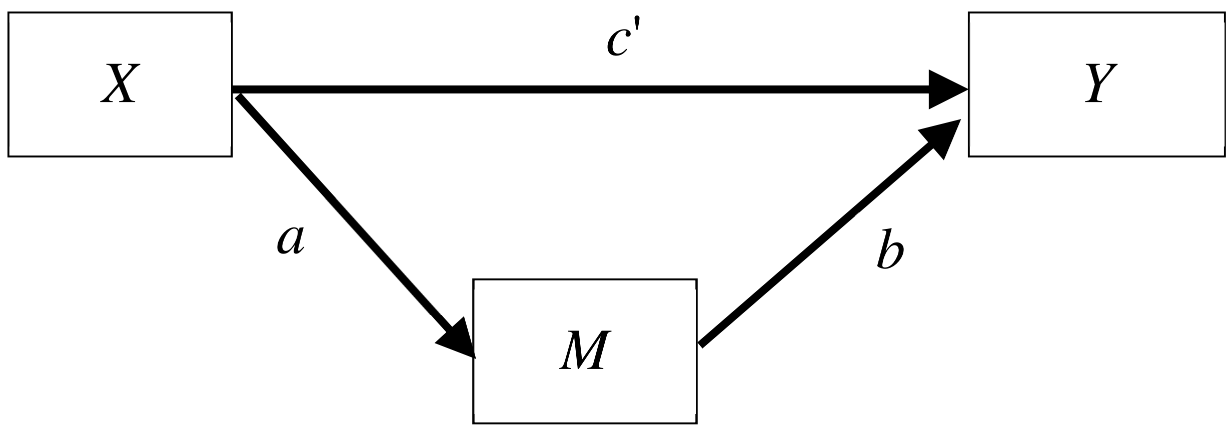
4. Bootstrap confidence intervals (CI only)
   . The best solution
   . Problem: computationally expensive

Fitting the mediation model via piecewise regression

- Draw the path diagrams

- Identify the three equations
  Outcomes have arrows going into them
  Explanatory variables have arrows emanating from them
  One equation for each outcome





Fitting the mediation model via piecewise regression

1a. Estimate the total effect model using linear regression

$$Somatic = \text{intercept1} + 0.117 \times Race \qquad + \varepsilon 1$$

1b. Draw the corresponding causal diagram and include the parameter est.

$$\boxed{Race} \xrightarrow{\quad 0.117^{**} \quad} \boxed{Somatic}$$

Fitting the mediation model via piecewise regression

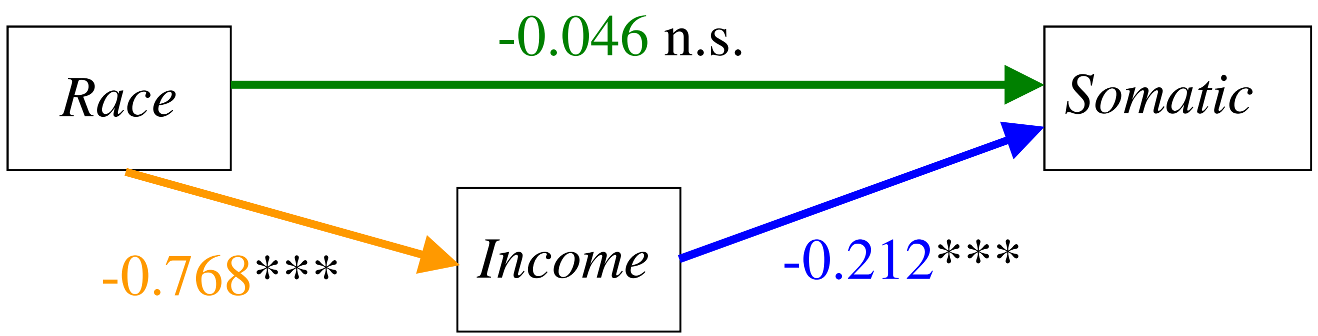
2a. Estimate the race-to-mediator linkage model

$$\text{Income} = \text{intercept2} - 0.768 \times \text{Race} + \varepsilon 2$$

2b. Estimate the direct effect model

$$\text{Somatic} = \text{intercept3} - 0.046 \times \text{Race} - 0.212 \times \text{Income} + \varepsilon 3$$

2c. Draw the corresponding causal diagram and include the parameter ests.



3. Test indirect effect (online calculator)

Topics

- Different causal models we will consider...

  Total effects model—bivariate regression

  Conditional effects model—standard multivariate regression

  Spurious correlation

  Shared causes between two explanatory variables

  Mediated effects model—direct and indirect causal effects

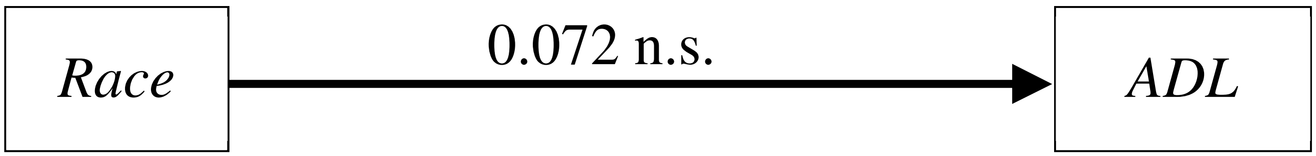
  Suppressor effects—AKA negative confounding

- A more advanced example

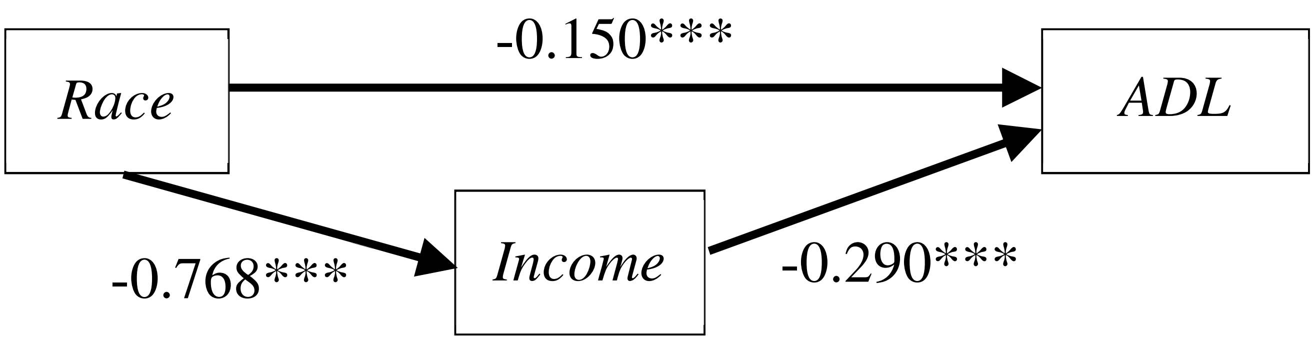
Suppressor effects—AKA negative confounding

- Sometimes a **total effect** can be non-significant
  even though the corresponding **direct effect** is significant

- Here *Race* has a *positive* but nonsignificant **total effect** on *ADL*



- but, conditional on *Income*, a **negative** significant **direct effect** on ADL



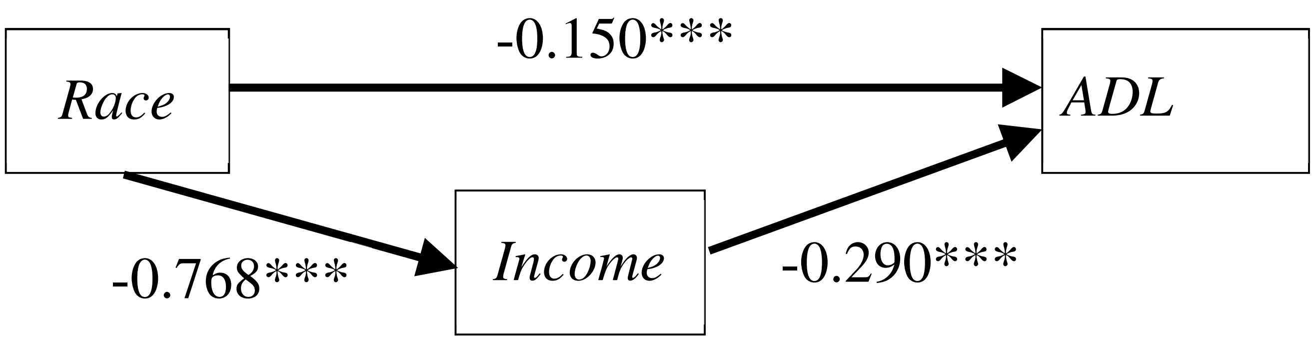
Suppressor effects—AKA negative confounding

- This causal system is said to be *inconsistent*

  The **direct effect** is negative, but the **indirect effect** is positive

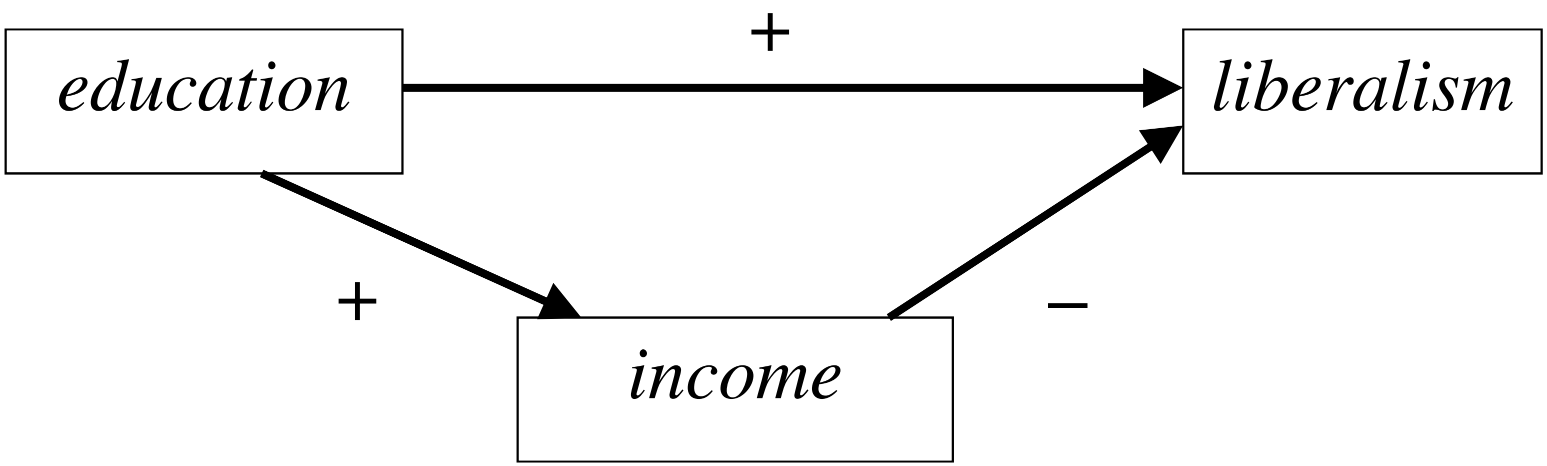
  These two effects tend to cancel each other out

  Relative to the **direct effect**, **t**he **total effect** is *suppressed* toward zero



- Generally, we expect **direct** and **indirect effects** to have the same sign but this is not always the case...

Suppressor effect: Classic example



direct effect of education on liberalism is positive

indirect effect of education on liberalism is negative

total effect of education on liberalism is 'suppressed'
positive direct effect and negative indirect effect act to cancel each other

Topics

- Different causal models we will consider...

  Total effects model—bivariate regression

  Conditional effects model—standard multivariate regression

  Spurious correlation

  Shared causes between two explanatory variables

  Mediated effects model—direct and indirect causal effects

  Suppressor effects—AKA negative confounding

- A more advanced example

Advanced Example: Kuppermann et al

Kuppermann, M. et al. (2006)
Beyond Race or Ethnicity and Socioeconomic Status: Predictors of Prenatal Testing for Down Syndrome. *Obstetrics and Gynecology*, 107, 1087-1097.

Objective
- Demographic, knowledge, and attitudinal predictors of prenatal test choice

Design
- Recruited women presenting for prenatal care prior to 20 weeks gestation
- 23 SF Bay Area obstetric clinics and practices
  (UCSF, SFGH, Kaiser, and community practices)
- Asian, African American, Latina, and White women
- Test use assessed after 30 weeks
- 344 women > 35 years of age

Advanced Example: Kuppermann et al

Binary outcome

- *Initial* choice of prenatal testing: invasive versus no prenatal testing

Explanatory variables

race/ethnicity	maternal age	language	site of care
income	education	occupation status	parity

Continuous candidate mediators

- *Knowledge* about prenatal testing and Down syndrome
- *Perceived risks* of Down syndrome and procedure-related miscarriage
- *Perceived understanding* of prenatal testing, plus decisional uncertainty
- *Attitudes*:
    . Value of prenatal testing information
    . Faith-based/fatalistic perspective on birth outcomes
    . General distrust of the health care system
    . "Rather have child w/ Down syndrome than no child"
    . "Modern medicine interferes too much with my pregnancy"
    . Pregnancy termination attitude

Advanced Example: Kuppermann et al

Results

Effects of demographic indicators on prenatal test choice
(i.e., before conditioning on candidate mediators)

• Conditional on all other demographic indicators, only race/ethnicity and income had significant effects on invasive testing

• Here, I focus on the effect of race/ethnicity, specifically Blacks versus all other groups combined

• African American women had significantly lower rates of initially choosing invasive testing compared to the other racial/ethnic groups.

$$\boxed{\text{Race}} \xrightarrow{\text{OR = .25*}} \boxed{\text{Invasive v None}}$$

Advanced Example: Kuppermann et al

Results

Multivariate model of prenatal test choice including all demographic indicators and candidate mediators

- The following candidate mediators had significant effects on invasive testing

  Attitudes

  . Higher value of information provided by prenatal testing

  . General distrust of the health care system

  . Lower levels of faith/fatalism

- Race/ethnicity and income no longer had significant direct effects

  . The effects of race/ethnicity and income were completely mediated

- Which of the 3 candidate mediators were mediating these effects?

  What can be determined at this point? Not much.

Advanced Example: Kuppermann et al

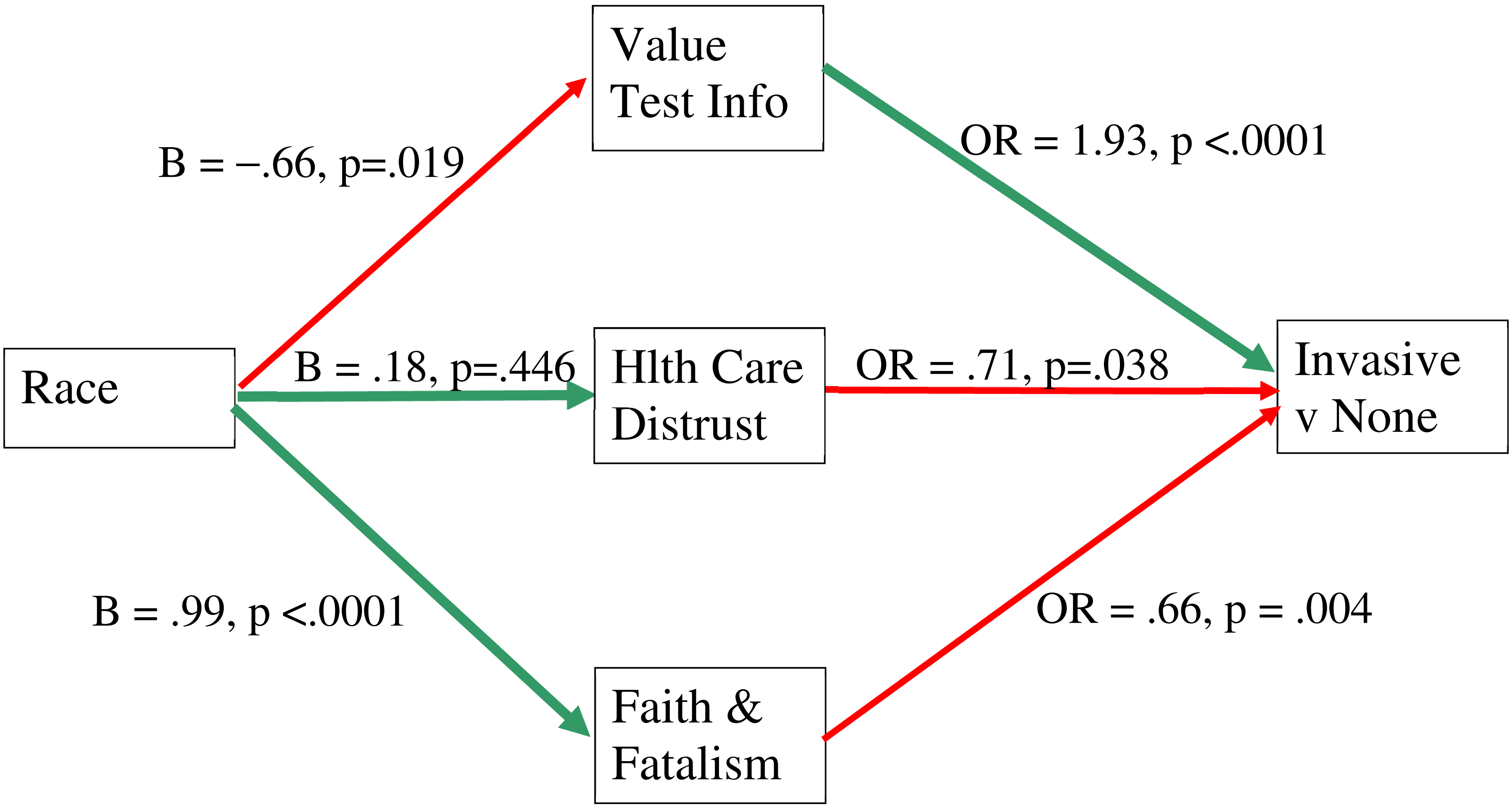
- We already have a model of the binary invasive testing outcome
  The explanatory variables include all demographics and
  the 3 candidate mediators; save results.

Additional steps in testing mediation

- Predict continuous candidate mediators
  . Fit 3 linear regression models predicting candidate mediators from
  race/ethnicity, income, and all other demographic variables;
  save results

- Graphically integrate results

Advanced Example: Kuppermann et al



(negative effect; positive effect)

- Which candidate mediator(s) most explain(s) the effect of race/ethnicity?

Advanced Example: Kuppermann et al

- This application included 'continuous' mediators and a binary outcome.

In <u>this case</u>, you can still estimate the indirect effect point estimate as $a \times b$ and use the joint test (and/or online calculator, bootstrap)

E.g., the effect of race on invasive testing via Value of Testing Information

$a = -0.66$ and $b = \ln(1.93) = 0.658$,

$a \times b = -0.66 \times 0.658 = -0.434$, and

$OR_{indirect.TEST.INFO} = e^{-0.434} = 0.648$

why take the natural log?

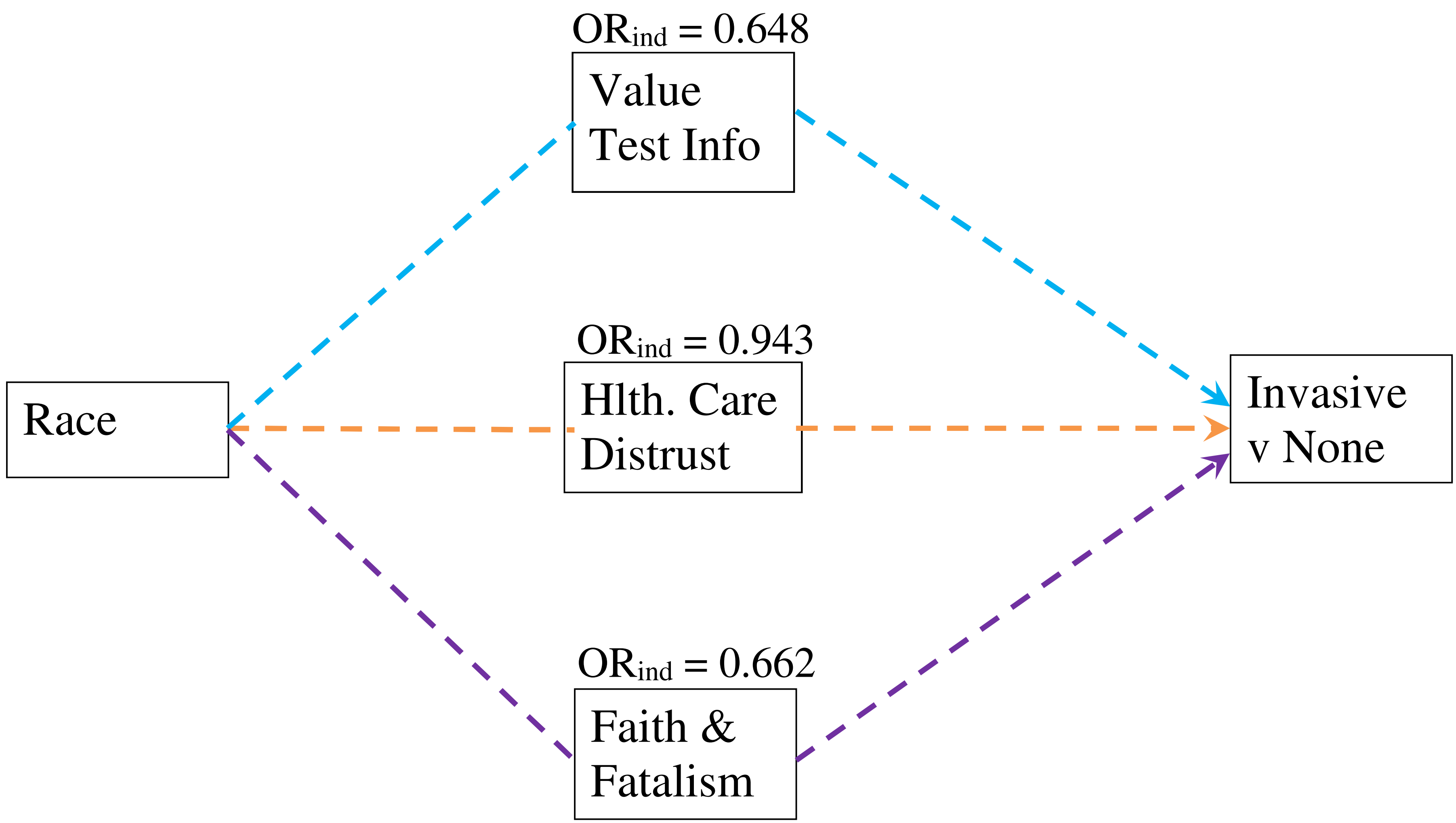
<u>Summarizing all indirect effect estimates</u>

$OR_{indirect.TEST.INFO} = e^{-0.434} = 0.648$

$OR_{indirect.HxCARE.DISTRUST} = e^{-0.059} = 0.943$

$OR_{indirect.FAITH/FATE} = e^{-0.412} = 0.662$

Advanced Example: Kuppermann et al



Summary

- In practice, there are many untested causal assumptions
  Causal direction cannot always be known
  Feedback loops are possible
  Longitudinal data helps

- Think about plausible causal relationships among your X variables,
  as well as between your X and Y variables.
  Is your model too simplistic?

Thinking causally about mechanisms that may lead to health disparities will help to clarify the effects of demographic strata and allow for explanation of health disparities

Thank you