Conceptualizing and testing mediated effects

Steve Gregorich

Epi 222: Health Disparities Research Methods
May 7, 2009

Health disparities

Healthy People 2010: Goal #2

- To *eliminate* health disparities among key demographic strata

    Gender

    Race/ethnicity

    Education or income

    Disability

    Geographic location

    Sexual orientation

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

<u>Competing factors</u>
- Genetic — Micro
- Congenital
- Developmental
- Psychological — Mezzo
- Behavioral/lifestyle
- Social
- 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

Reality	Practice
Regression models almost always reflect causal hypotheses	The causal nature of the hypotheses is not always explicitly stated
Regression framework is very 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 <u>explanation</u>, then causal hypotheses are being specified
- Attention to causal hypotheses helps to advance knowledge

Topics

- <u>Different causal models and topics we will consider...</u>

  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)
- Cognitive impairment (SPMSQ: > 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>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

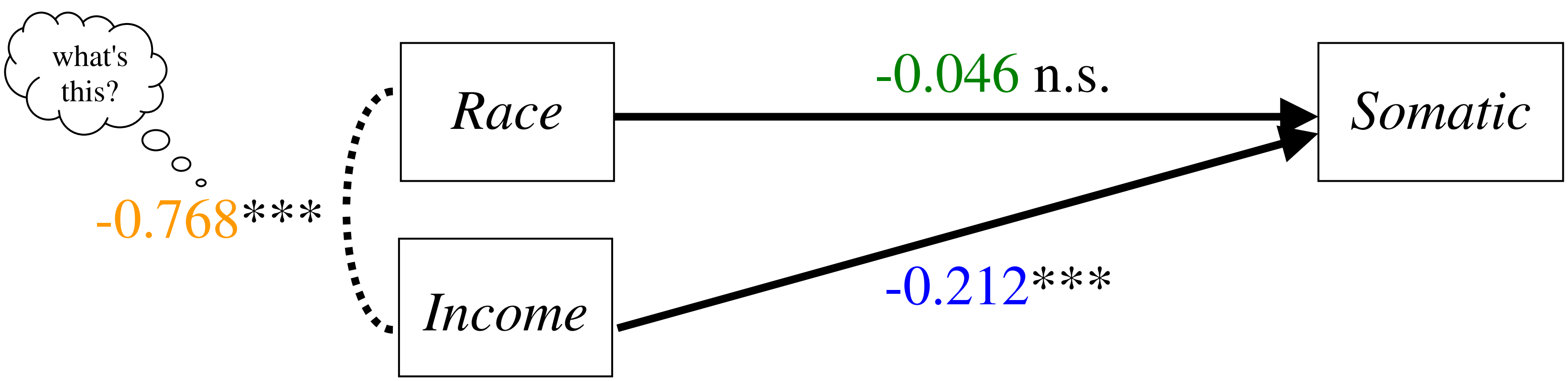
  Suppressor effects—AKA negative confounding

- A more advanced example

Conditional effects—standard multivariate regression Interpretation

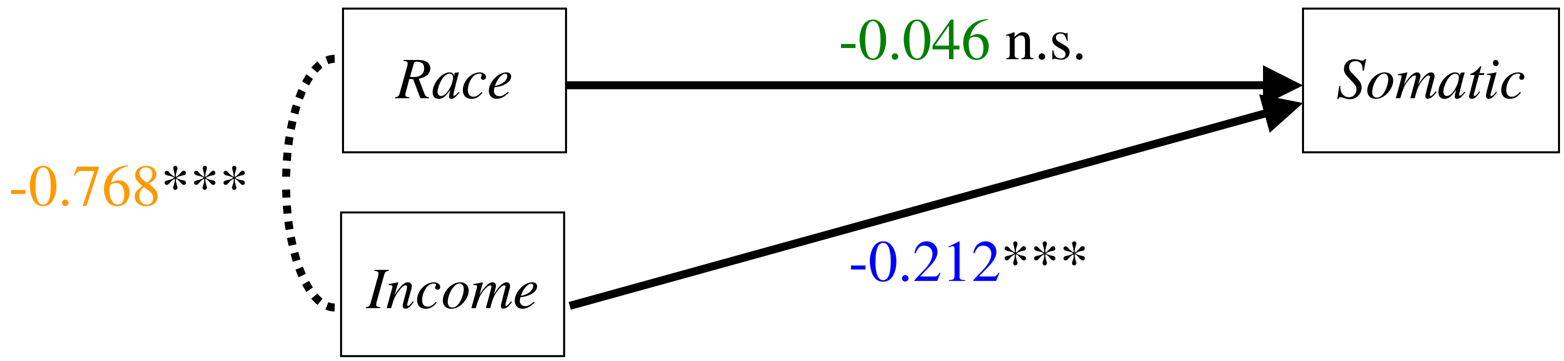
- Add *Income* to the model as an explanatory variable

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



- Note the *conditional* effect of Race is non-significant.
- Causal interpretation: Race is not an important cause of Somatic

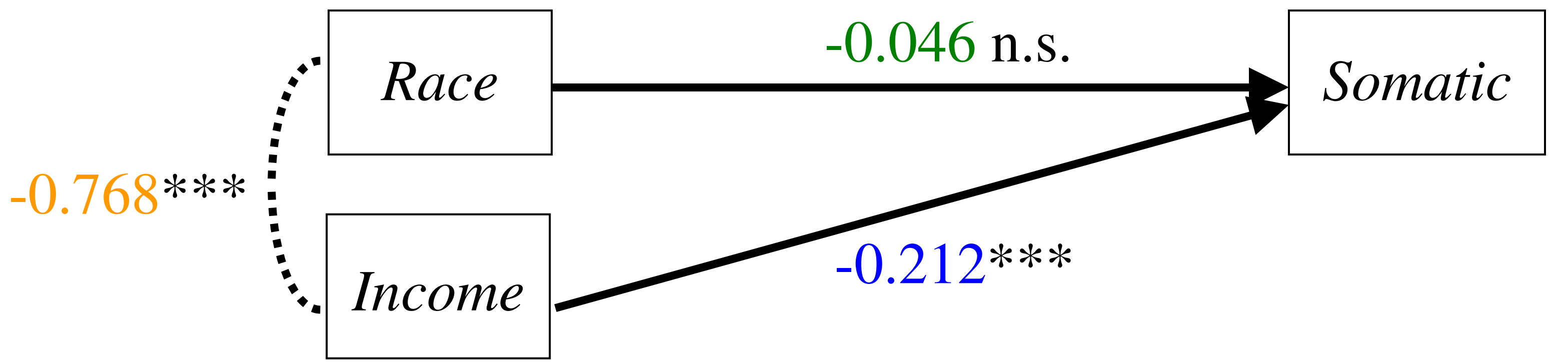
Conditional effects—standard multivariate regression Assumptions



- **Causal assumptions of the model:**

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

Conditional effects—standard multivariate regression
Assumptions



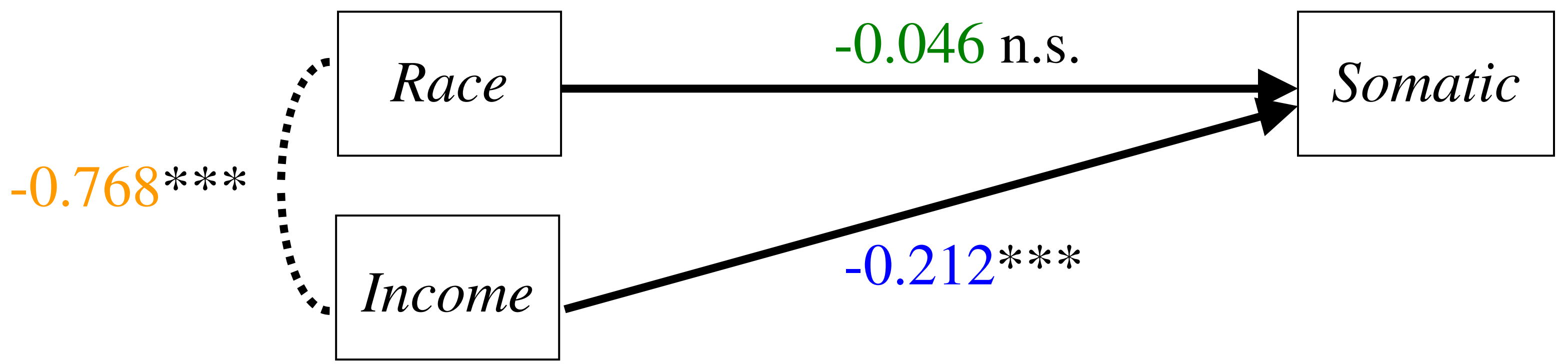
- **Causal assumption #3**

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

- **Question**

  Is assumption #3 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** relationships 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

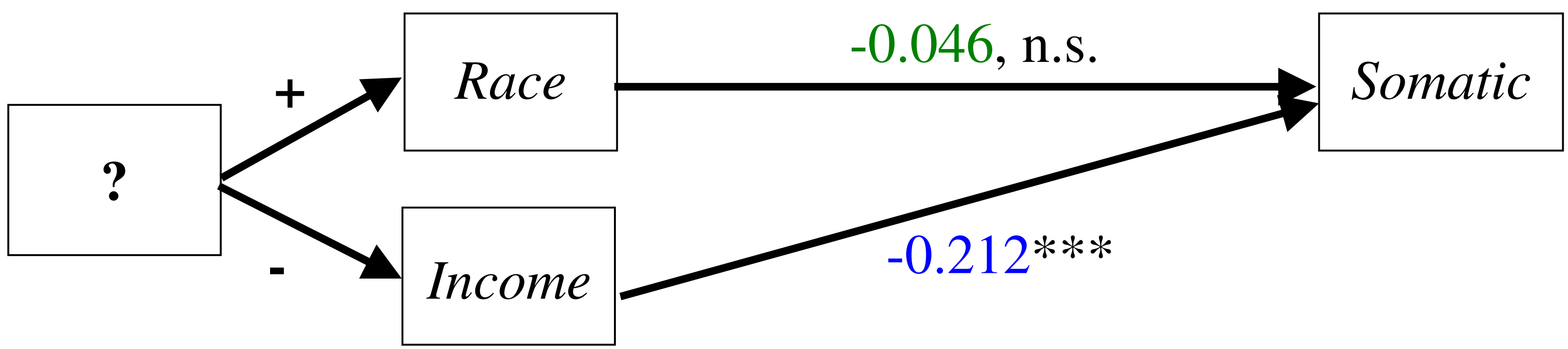
  Mediated effects model—direct and indirect causal effects

  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.

Topics

- Different causal models we will consider...

  Total effects model—bivariate regression

  Conditional effects model—standard multivariate regression

  Spurious correlation

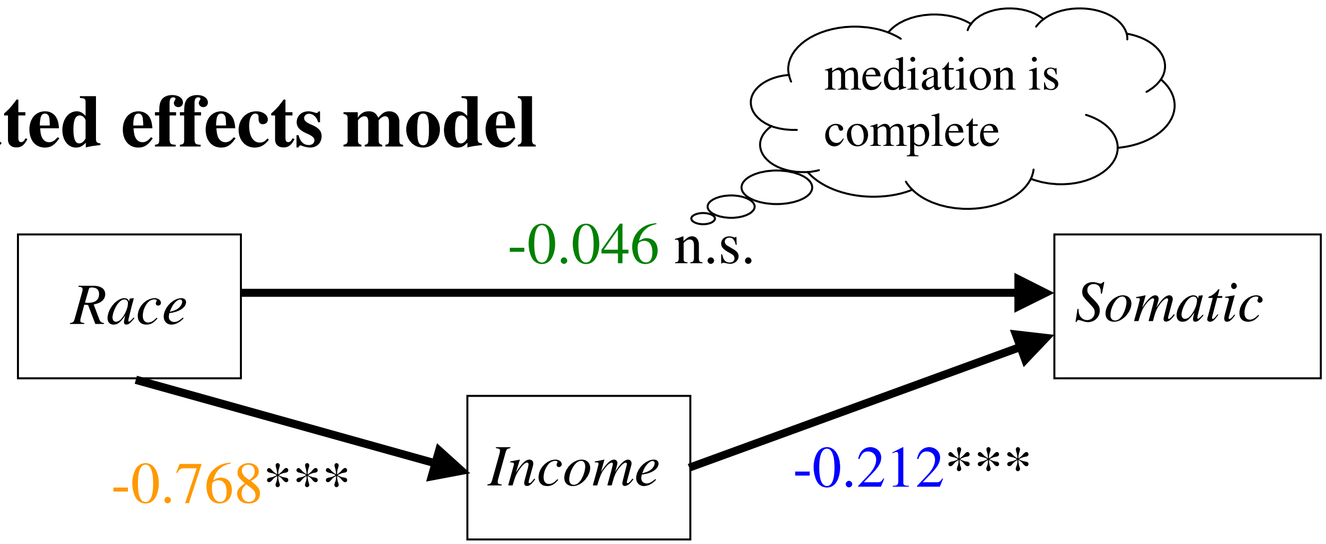
  Shared causes between two explanatory variables

  Mediated effects model—direct and indirect causal effects

  Suppressor effects—AKA negative confounding

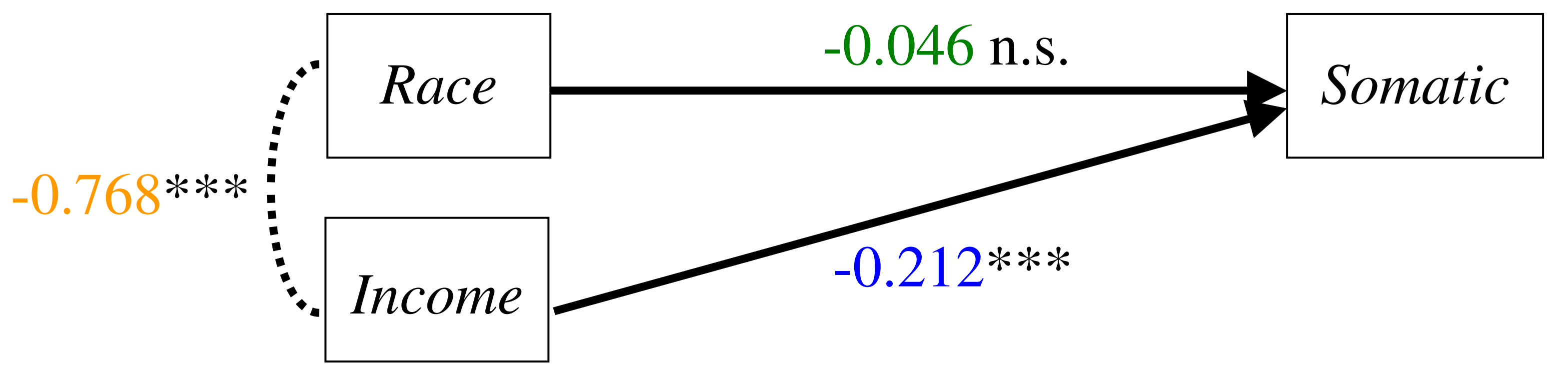
- A more advanced example

Mediated effects model



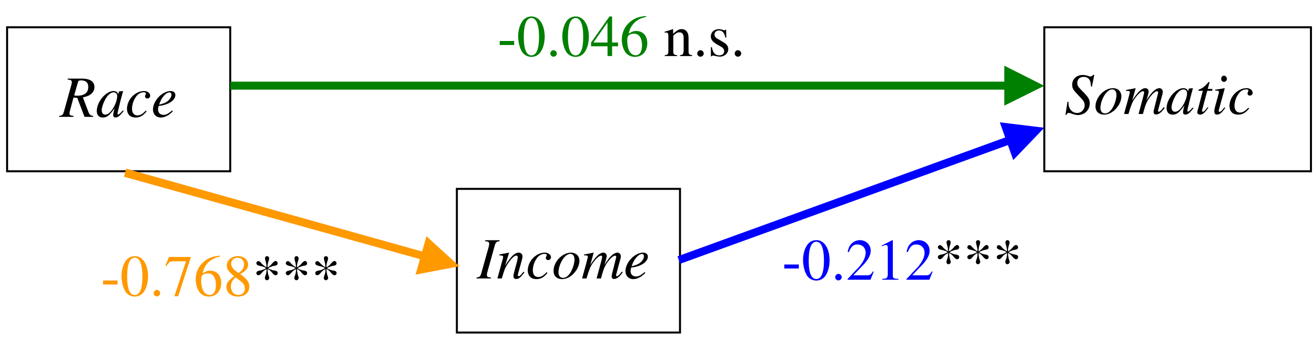
- What are the causal implications of the mediation model?

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



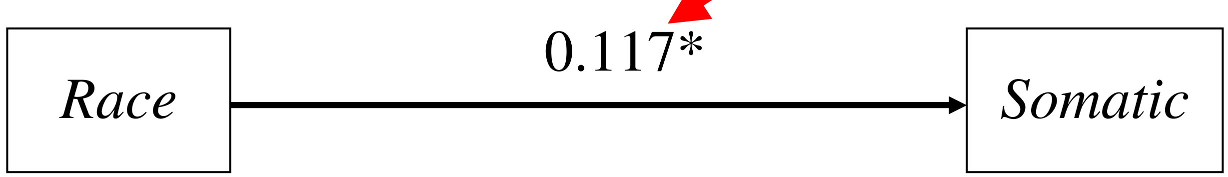
- What race-related inference do these two estimated models share?

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



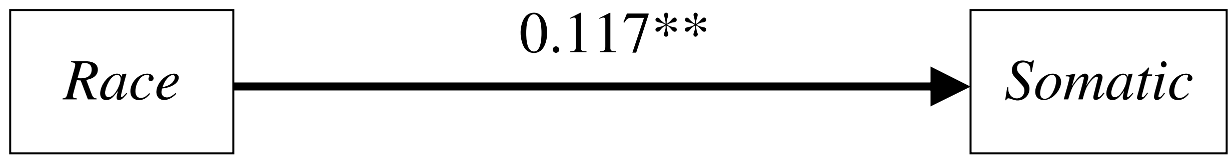
Direct effect: *Race* → *Somatic*	-0.046	
Indirect Effect: *Race* → *Income* → *Somatic*	0.163	(-0.768 × -0.212)
Total Effect	0.117	

- Remember the very first model?...

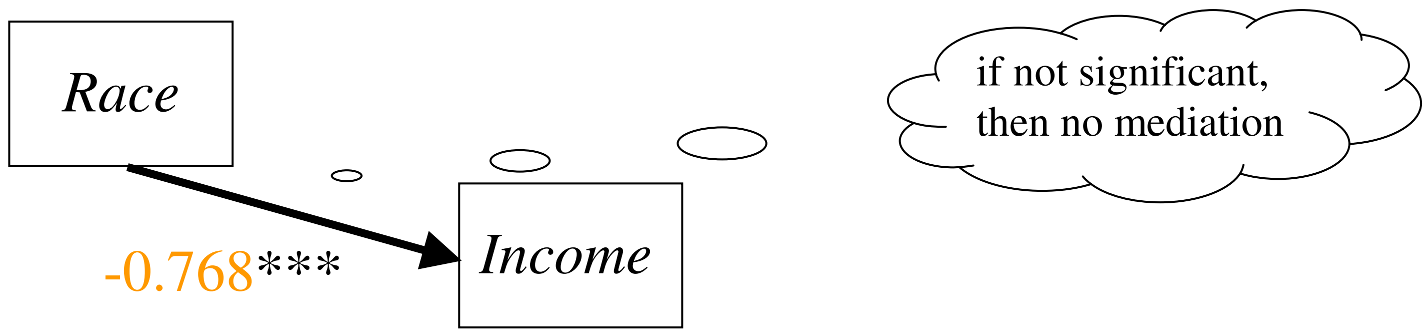


Mediation: Review of modeling steps

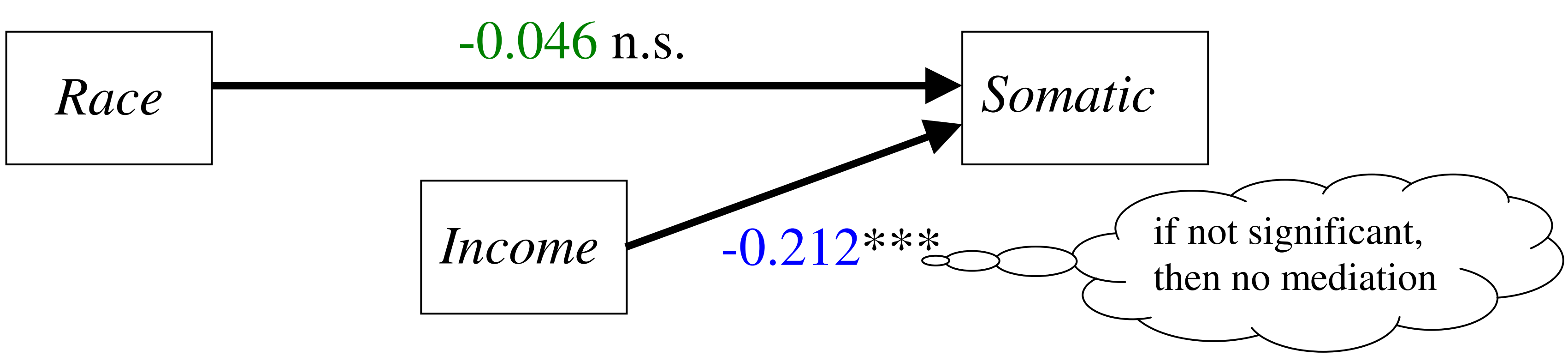
Model 1: Estimate **total effect** of *Race* on *Somatic*



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



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

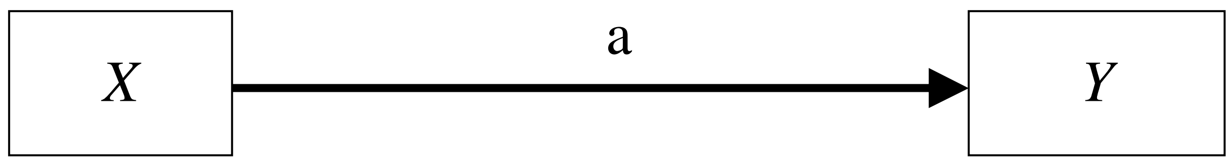


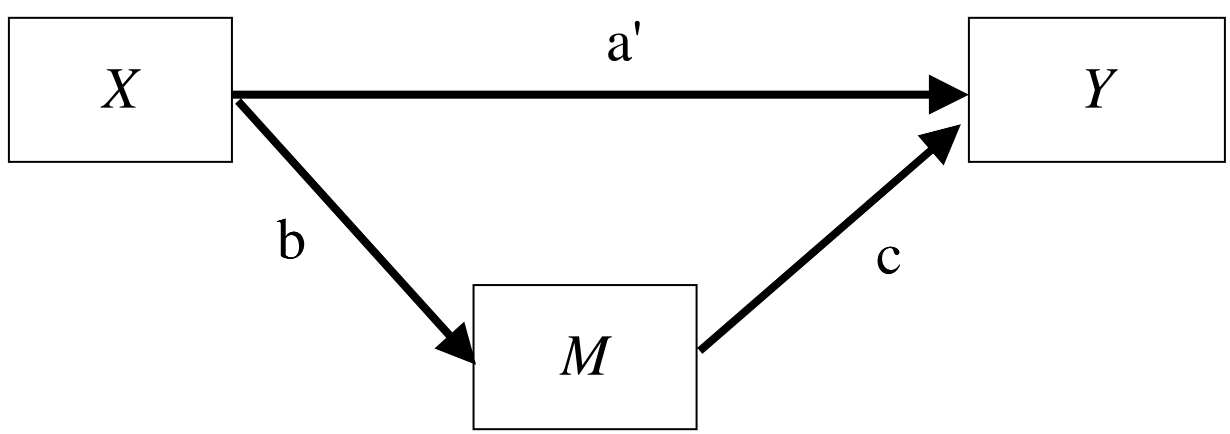
Mediation: Review of modeling steps

Step 4: Assess the degree of mediation

Complete mediation

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



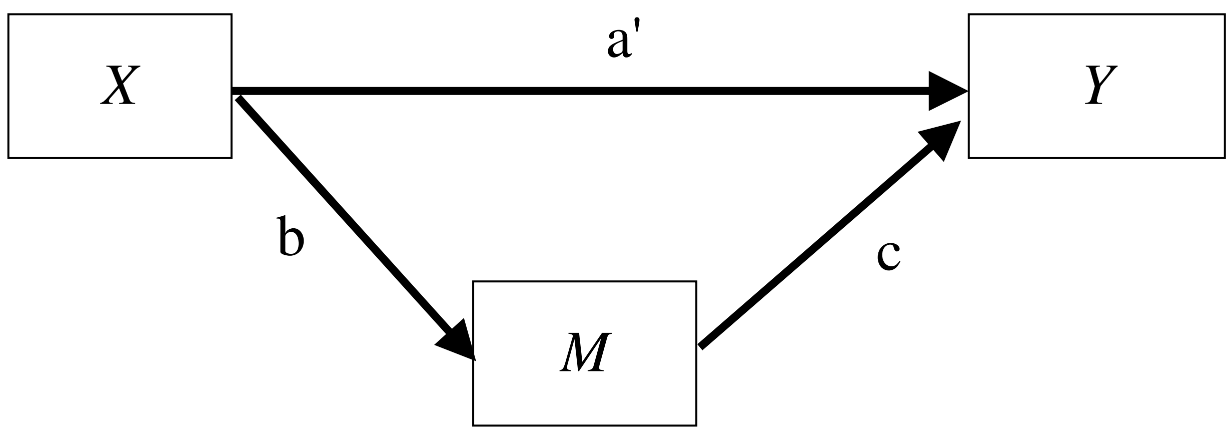


Mediation: Review of modeling steps

Step 4: Assess the degree of mediation

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

Test the indirect effect
If both the direct (a') and indirect (b×c) effects are significant, then partial mediation exists

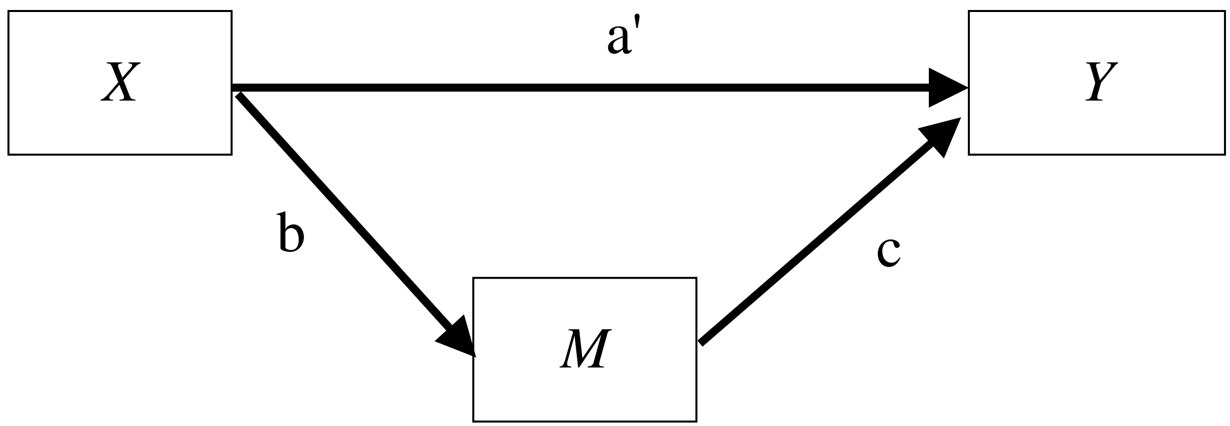


Mediation: Review of modeling steps

Step 4: Assess the degree of mediation

No mediation

The indirect effect (b×c) is non-significant



Mediation: Testing the indirect effect (b×c)

Estimating and testing b×c

- Continuous outcome and mediator: Sobel, Aroian, & Goodman tests
- Binary outcome and continuous mediator: Rescaling methods
- Binary mediator: more complex

Another method, assesses the significance of the product of two *t*-values

- Alternatively, you can 'eyeball' the p-values for paths *b* and *c*

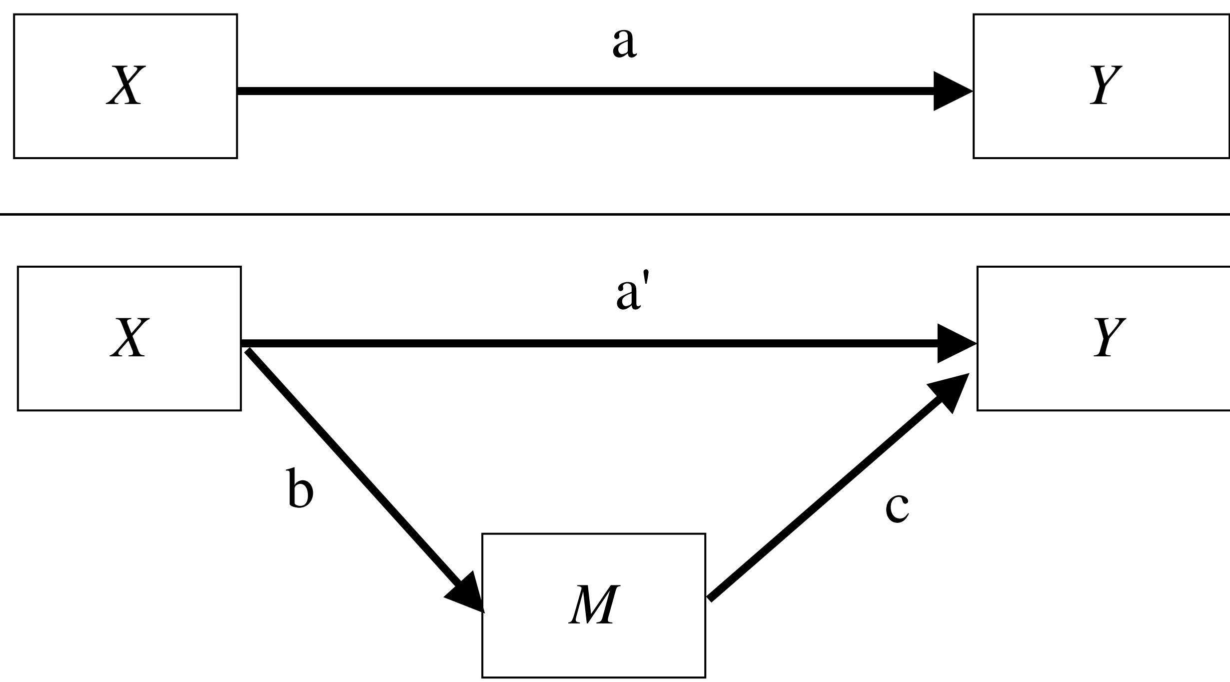
Minimum p-values for a significant indirect effect

if one p-value equals…	*then, the other p-value must be…*
.0055	<.0055
.010	<.0025
.025	<.0001
.030	<.000005
.040	<.0000000005
.045	$\rightarrow 0$

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

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

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

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

Fitting the mediation model via piecewise regression

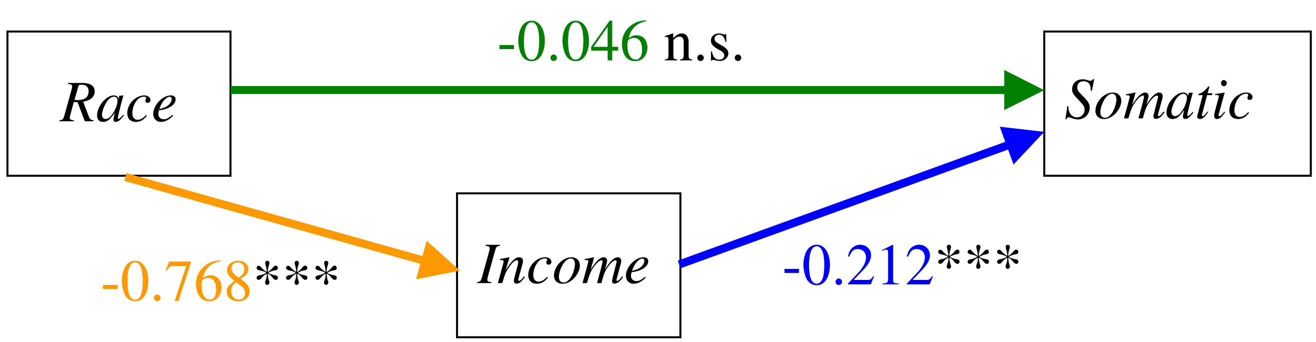
2a. Estimate the race-to-mediator linkage model

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

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.



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

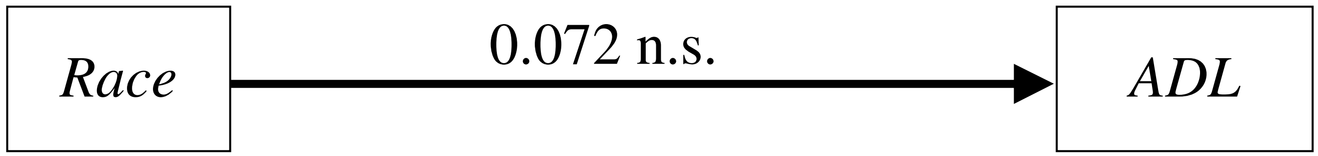
  Suppressor effects—AKA negative confounding

- A more advanced example

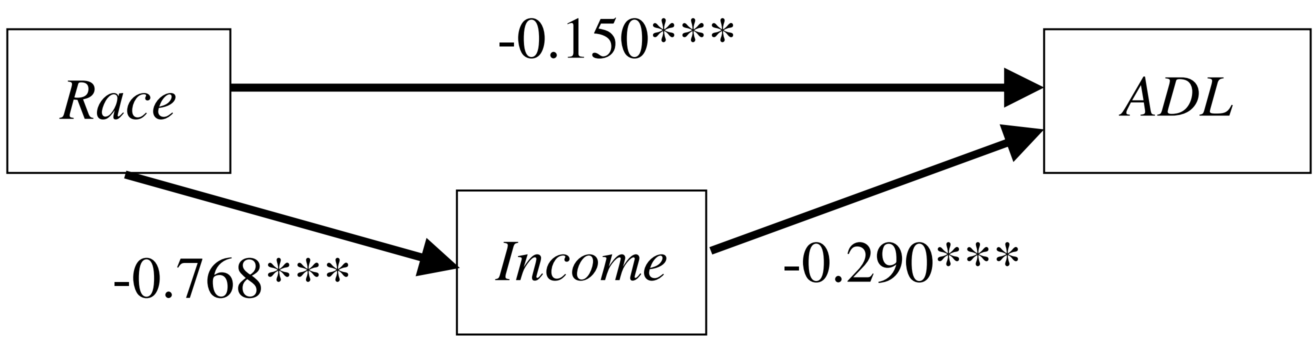
Suppressor effects—AKA negative confounding

- 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



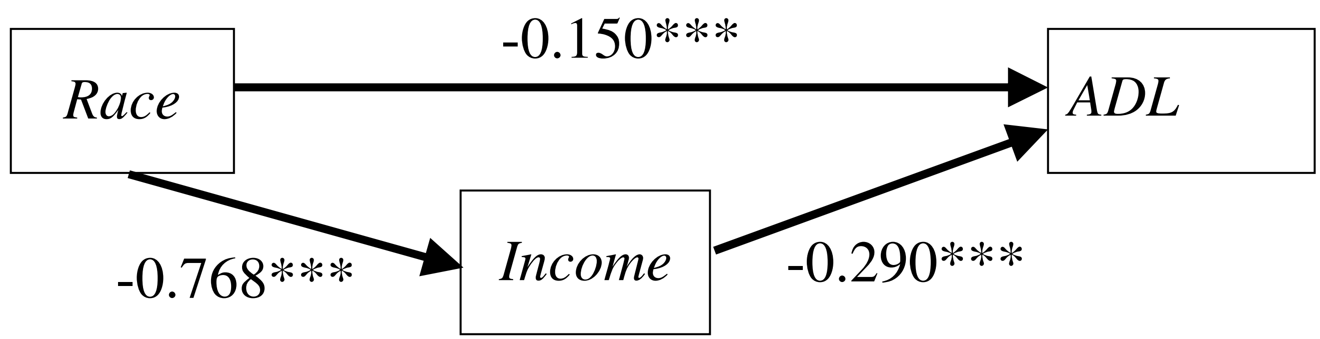
Suppressor effects—AKA negative confounding

- This causal system is said to be *inconsistent*

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

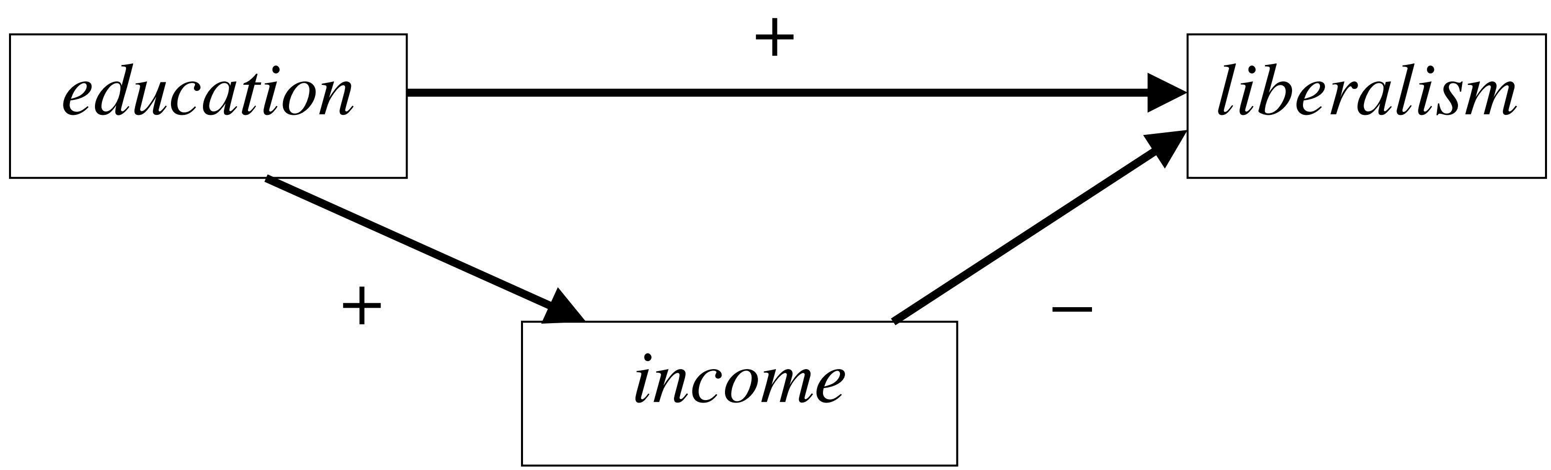
  These two effects tend to cancel each other out

  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, inconsistent system



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

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

<u>Objective</u>

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

<u>Design</u>

- 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

<u>Binary outcome</u>

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

<u>Explanatory variables</u>

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

<u>Continuous candidate mediators</u>

- *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

<u>Results</u>

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 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

<u>Results</u>

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

- Candidate mediators are those with significant effects on invasive testing
  - <u>Attitudes</u>
    - . Higher value of information provided by prenatal testing
    - . Lower levels of risk of procedure-related miscarriage
    - . Lower levels of faith/fatalism

- Race/ethnicity and income no longer had significant 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?

Advanced Example: Kuppermann et al

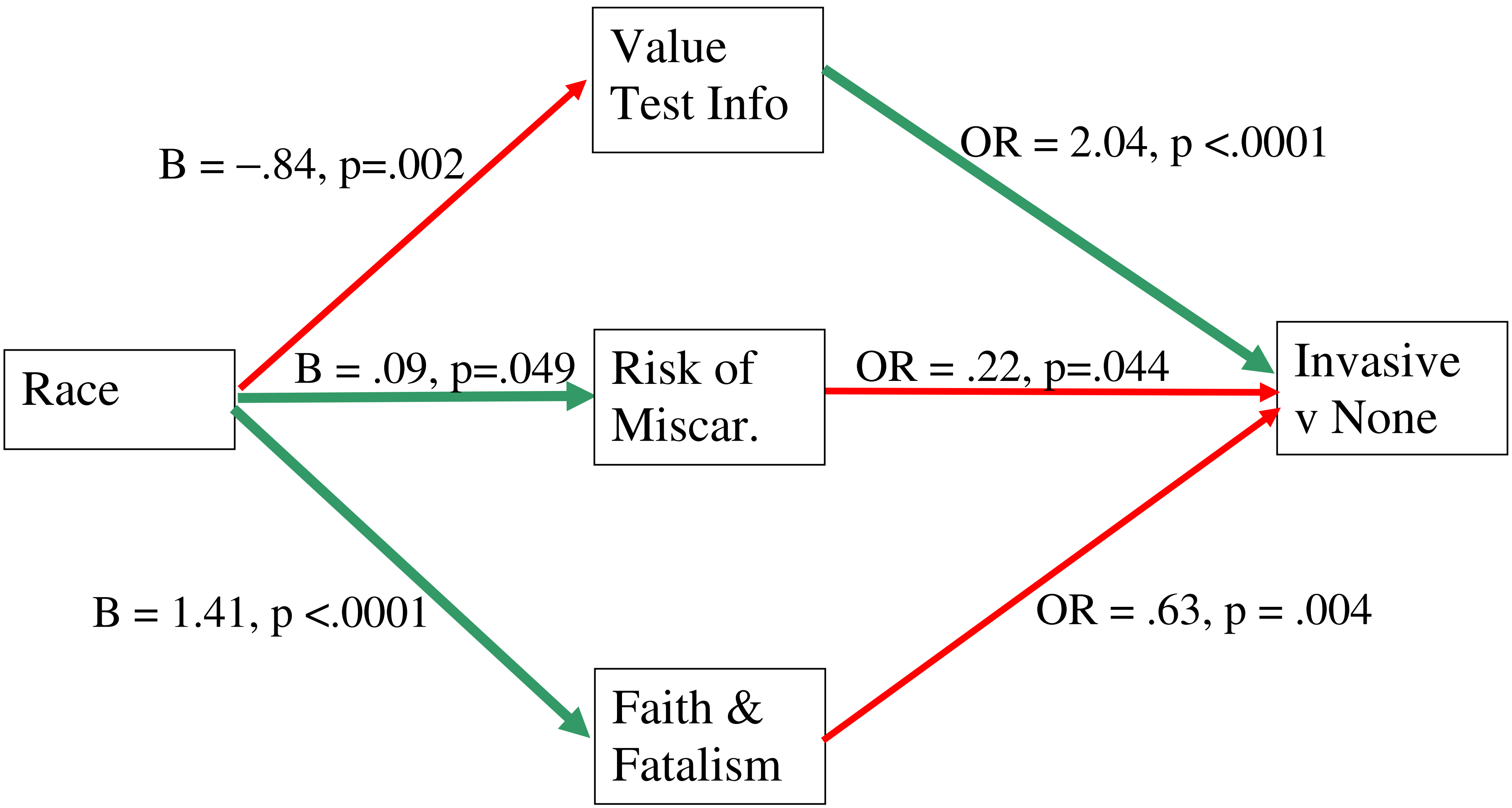
Additional steps in testing mediation

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

- Predict binary outcome
  Predict invasive testing outcome with a logistic regression model.
  The explanatory variables include all demographics and the 3 candidate mediators; save results.

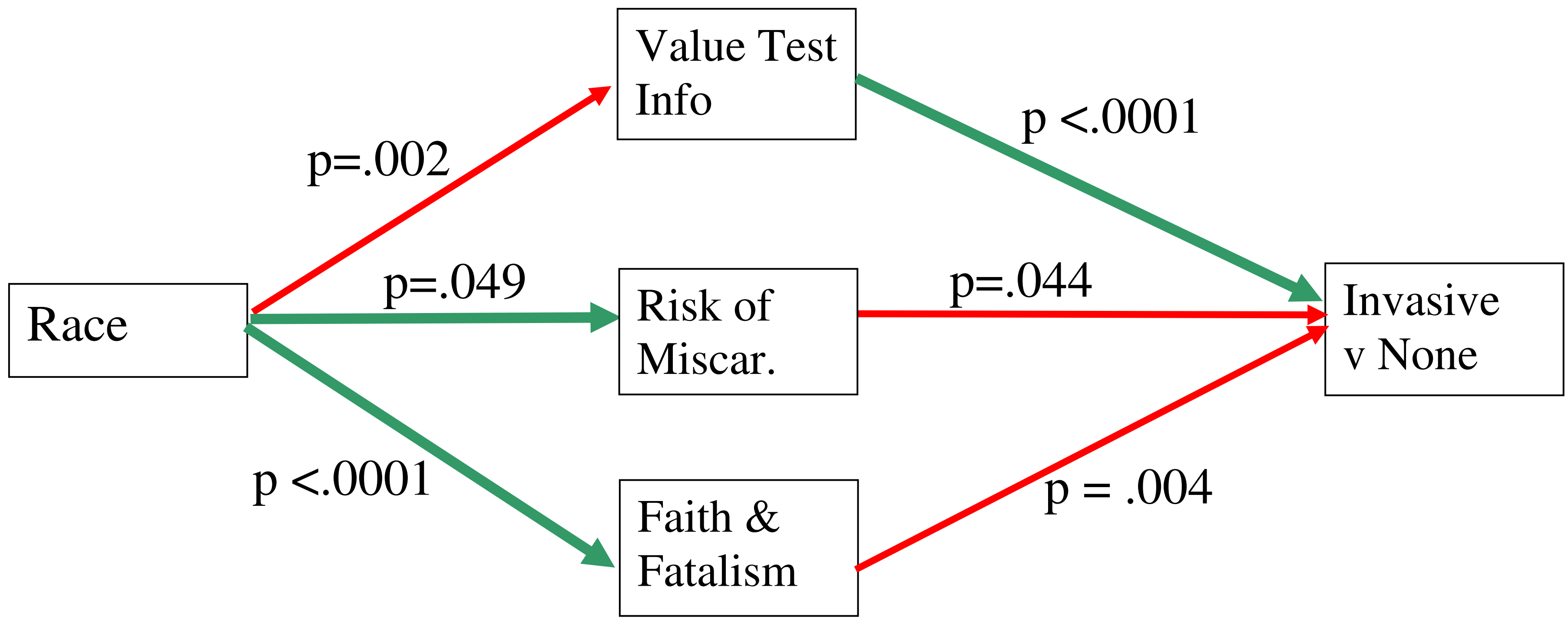
- Graphically integrate results

Advanced Example: Kuppermann et al



(negative effect; positive effect)

- Which candidate mediators explain the effect of race/ethnicity?



Minimum p-values for a significant indirect effect

if one p-value equals…	*then, the other p-value must be…*
.0055	<.0055
.010	<.0025
.025	<.0001
.030	<.000005
.040	<.0000000005
.045	→ 0

Advanced Example: Kuppermann et al

- This application required a mix of linear and logistic regression
  Cannot easily calculate indirect effects, but it can be done

Summary

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

- Secondary and administrative data are OK for identifying disparities
  but they are much more limited at offering explanations

- Race/ethnicity and other demographic strata are 'packed' constructs

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