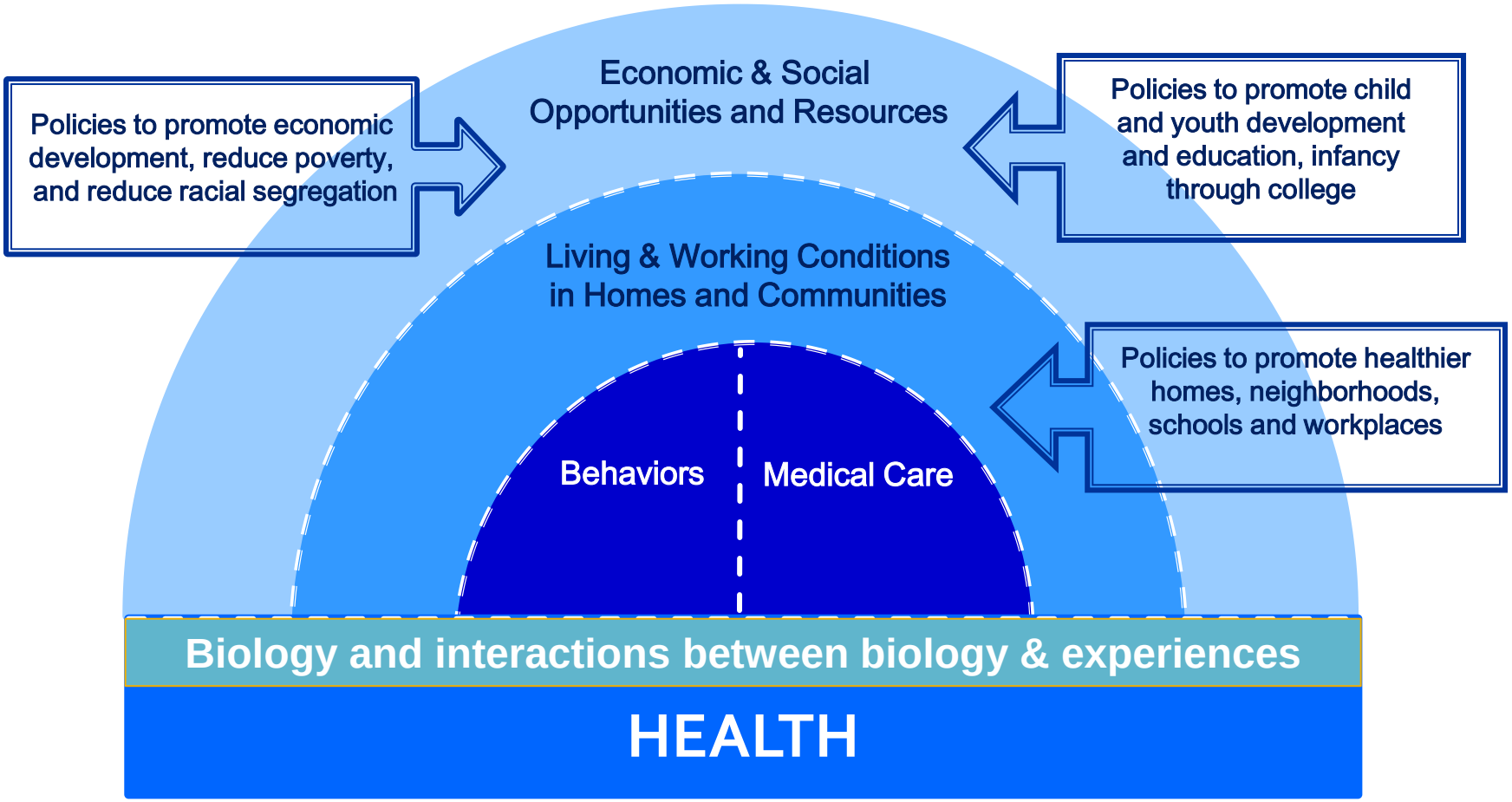
Epi222: Measurement of/in health disparities, an Abbreviated Lecture

Measurement Issues in Health Disparities and Social Epidemiology

- How do you measure disparities?
  1. Defining comparison groups
  2. Choosing the measure
  3. Outcomes
- What is the role of measurement in explaining disparities and how do insufficient measures compromise this?
- Challenges in measuring social constructs
- Some basic ideas for measuring latent variables

Where are we in the class?



Where are we in the class?

Nuts and bolts of conducting research on disparities and social inequalities.

- Recruitment and community engagement
- Measurement/quantitative research questions
- Asking important questions

HEALTH

Measurement Issues in Health Disparities and Social Epidemiology

- How do you measure disparities?
  1. Defining comparison groups
  2. Choosing the measure
  3. Outcomes
- What is the role of measurement in explaining disparities and how do insufficient measures compromise this?
- Challenges in measuring social constructs
- Some basic ideas for measuring latent variables

Measuring Disparities

A common (not the only) definition of a health disparity is that it is a difference in health outcomes between groups that is *unjust*.

Measuring disparities entails:

1. Choosing groups to compare
2. Showing a between-group health *difference*
3. Providing evidence that the difference is unjust (e.g., it is preventable, it arises from long-standing social disadvantages or historical power structures).

Measurement Issues in Health Disparities and Social Epidemiology

- How do you measure disparities?
  1. Defining comparison groups
  2. Choosing the measure
  3. Outcomes
- What is the role of measurement in explaining disparities and how do insufficient measures compromise this?
- Challenges in measuring social constructs
- Some basic ideas for measuring latent variables

Measuring Disparities : Defining the Dimension of Inequality

- Race, ethnicity, Socioeconomic position
- SEP (assuming you want SEP not social *class)*:
  - Select domain (education, occupation, financial)
  - define time in life (childhood, current, cumulative)
  - define level (own SEP, household, neighborhood)
  - Choose absolute vs relative (I feel better in 1st class because I am super-comfy *vs* I feel better in 1st class because I am comfier than those schmoes back in economy)
- Distinguish between the construct and how you will measure that construct

Distinguish between the construct and how you will measure that construct

- Origin of the measurement
  - Self-report, proxy report, administrative record, direct observation, instrument.
- Level of the measurement
  - Individual, group (neighborhood, school, state, country), time (repeated measures on the individual).
- For group measures: aggregate (average poverty rate) or structural (nuclear power plant, weather)
- Subjective vs objective (not so crisp)
- Important to be clear about what you aim to measure, even if you recognize you are not getting it perfectly
  - May use neighborhood SES as a proxy for individual SES
  - But sometimes problematic if neighborhood SES has a direct effect

SEP vs social class

Social class: “A social category referring to social groups forged by interdependent economic and legal relationships, premised upon people’s structural location within the economy—as employers, employees, self-employed, and unemployed, and as owners, or not, of capital, land, or other forms of economic investments; possession of educational credentials and skill assets also contribute to social class position”

SEP: “An aggregate concept that includes both resource-based and prestige-based measures, as linked to both childhood and adult social class position. Resource-based measures refer to material and social resources and assets, including income, wealth, educational credentials; terms used to describe inadequate resources include “poverty” and “deprivation”. Prestige-based measures refer to individual’s rank or status in a social hierarchy, typically evaluated with reference to people’s access to and consumption of goods, services, and knowledge, as linked to their occupational prestige, income, and education level” – Krieger, Williams and Moss, Annual Rev PH, 1997

Different dimensions of disadvantage are strongly correlated

Table 2. Mean Family Income by Educational Level and Racial/Ethnic Group Among Adults Aged 18-64 Years—Nationa Survey, 1989-1994 (n = 380 552)

Educational Level, y	Mean Income (95% Confidence Interval), $
	Overall
<9	20 551 (19 868-21 234)
9-11	25 945 (25 486-26 403)
12	36 022 (35 613-36 432)
13-15	41 857 (41 228-42 487)
≥16	54 606 (54 064-55 149)
Overall	39 716 (39 274-40 159)

- Highest education group 265% more income than lowest education group
- Graded continuously: any threshold is arbitrary

Braveman JAMA 2005

Different dimensions of inequality are not perfectly correlated

Table 2. Mean Family Income by Educational Level and Racial/Ethnic Group Among Adults Aged 18-64 Years—Na Survey, 1989-1994 (n = 380 552)

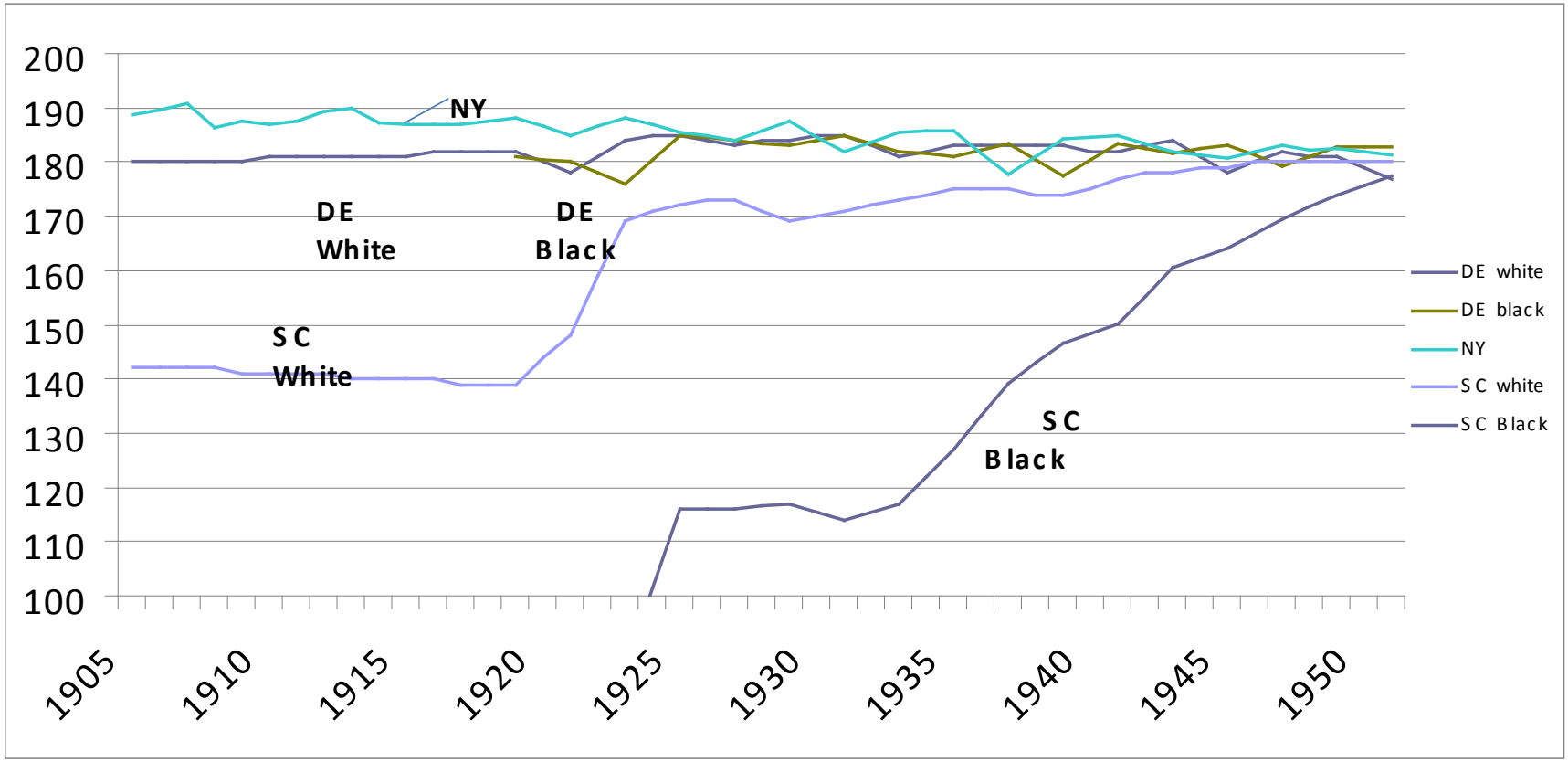
Educational Level, y	Mean Income (95% Confidence Interval), $		
	Black	Mexican American	White
<9	15 503 (14 851-16 154)	19 104 (17 549-20 659)	22 707 (21 935-23 479)
9-11	17 743 (17 051-18 434)	22 377 (20 807-23 948)	28 573 (28 062-29 083)
12	25 337 (24 529-26 144)	30 945 (29 791-32 098)	37 853 (37 400-38 305)
13-15	33 026 (31 898-34 154)	37 642 (36 089-39 195)	43 197 (42 510-43 884)
≥16	46 815 (45 535-48 095)	48 055 (46 137-49 973)	55 277 (54 725-55 829)
Overall	27 937 (27 123-28 752)	27 927 (26 498-29 356)	42 113 (41 626-42 600)

- Among people with 12 years of education, whites had 149% more income than blacks.
- Black-white and Mexican American-white disparity in income at every educational level

Braveman JAMA 2005

Even seemingly simple variables are hard to measure well: education

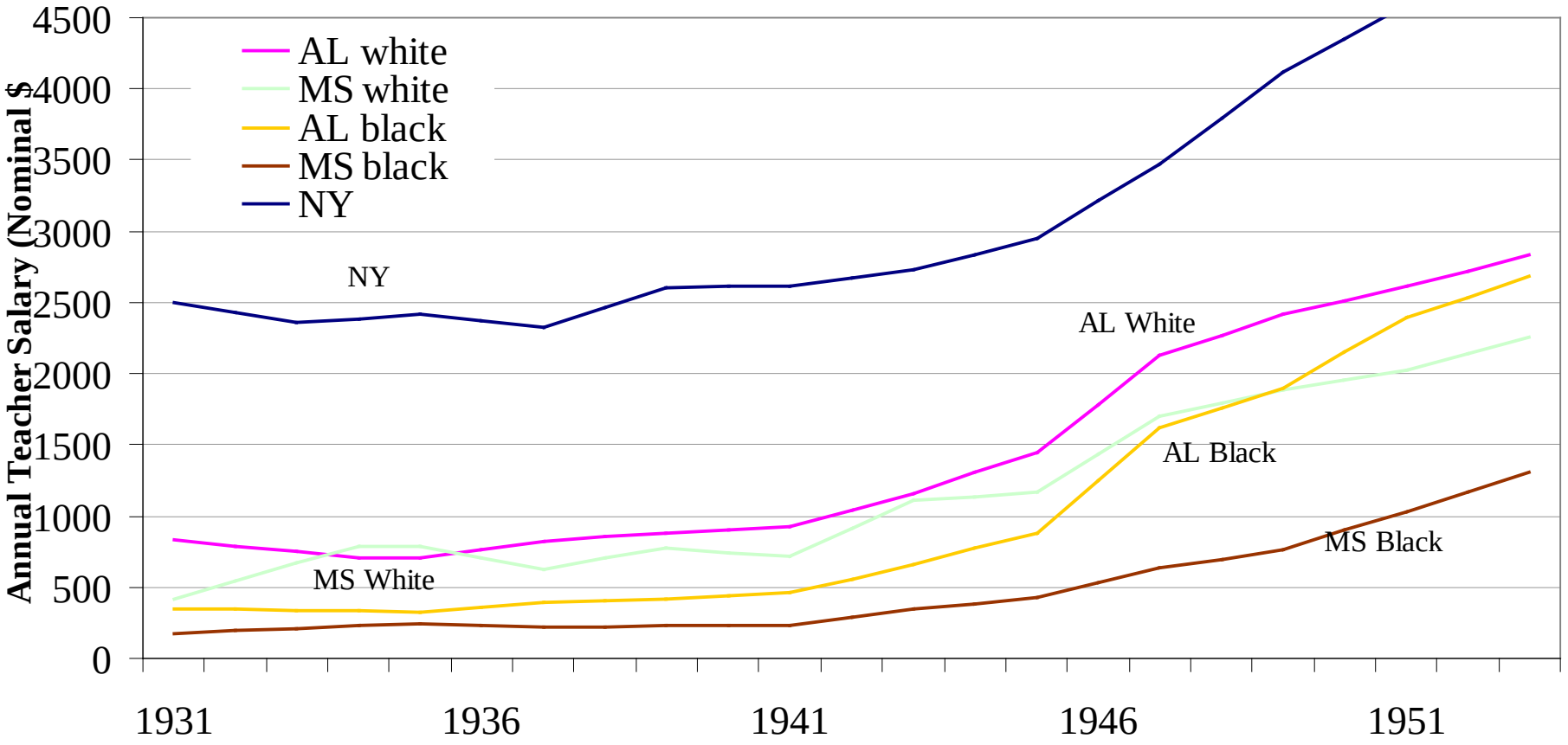
Changes in school term length (in days)



Years of Education ≠ Amount of Schooling

- Term length differed and not all days of schooling were attended.
  - A kid born in NY in 1925 who attended 10 years of school:
    - school was open a total of 1,830 days
    - on average he attended 1,598 days
    - =8.9 years (assuming 180 days / school year)
  - For a comparable white kid born in SC in 1925
    - School was open 1,735 days
    - On average he attended 1,418 days
    - = 7.9 years.
  - A comparable black kid born in SC in 1925:
    - School was open 1,288 days
    - On average he attended 958 days
    - =5.3 years (40% less than the NY kid).
- Other dimensions of quality differed as well, so this understates the magnitude of the problem.

Teacher Salaries



How do you choose a construct?

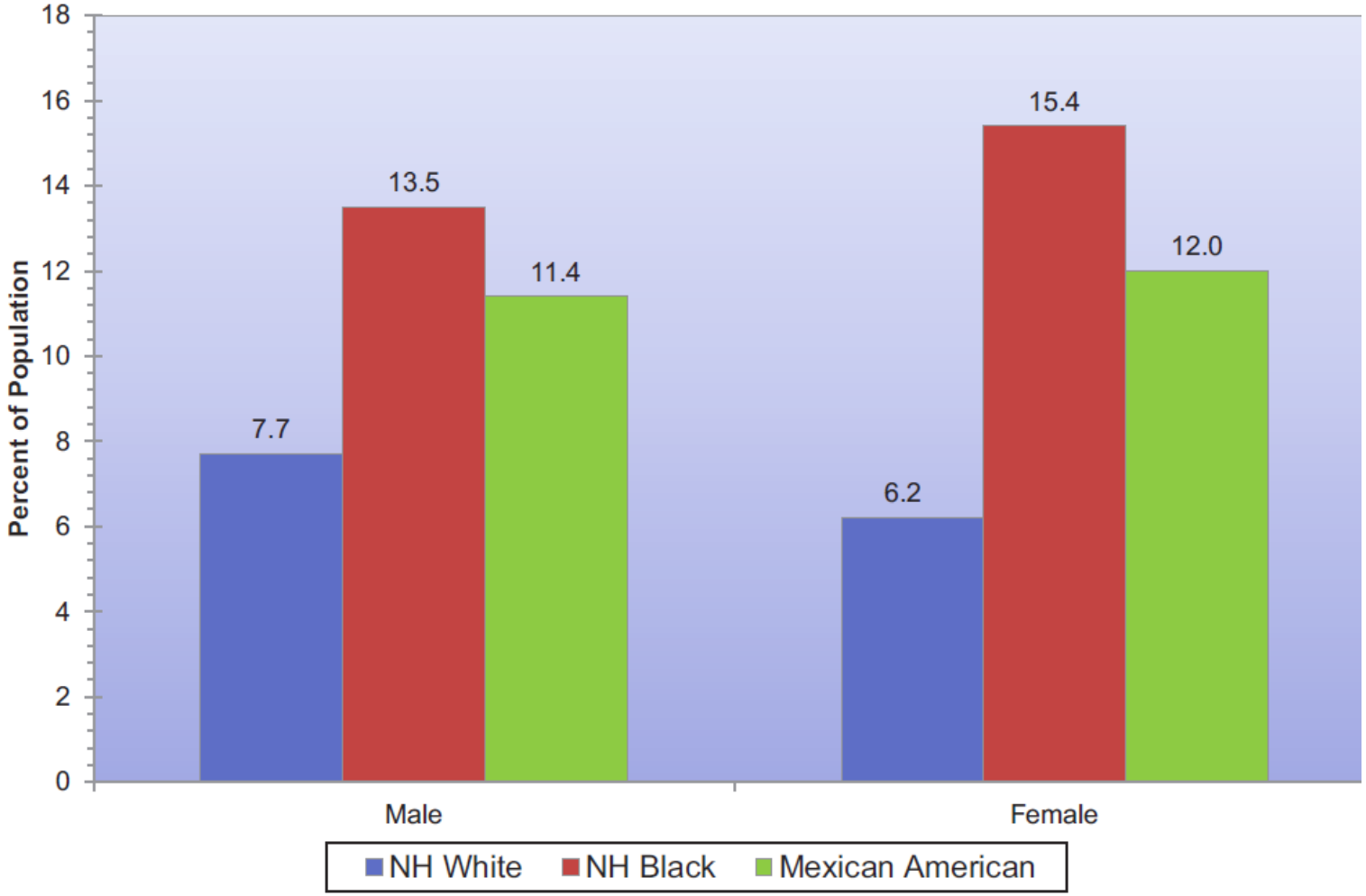
- Sometimes the choice of dimension will be obvious to you, but often not, and you should stop to justify the scientifically relevant dimension (SES, ethnicity)

- Are you conducting etiologic/intervention research or surveillance/descriptive?
  - If etiologic: theoretically, what is most likely to be relevant?
  - Articulate mechanisms by which SEP may influence your health outcome.

How do you choose a measure?

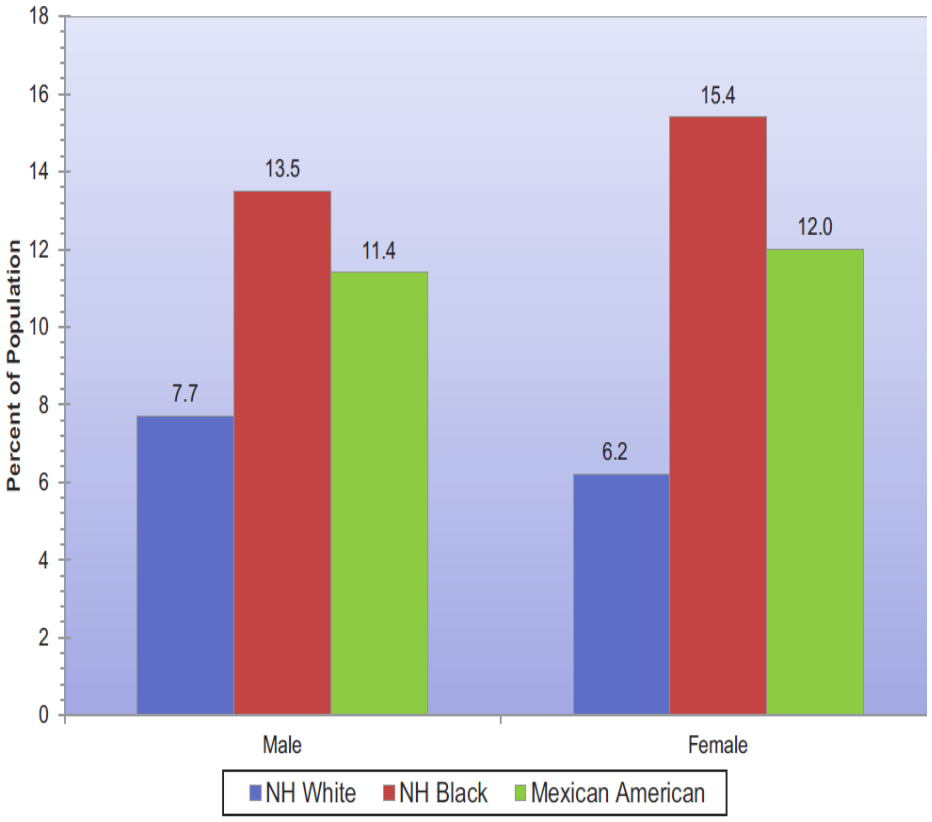
- After defining what you want to measure, consider the best you can do and the consequences of compromising.
  - Describe why your measure is inadequate/incomplete
  - Understand the bias likely to arise from inadequate measurement (next section)
- Is your goal to "explain away" another source of inequality?
  - Use a more comprehensive measurement variable(s)
- Is your goal to guide interventions?
  - Narrower, more clearly defined measurements are desirable

Measuring Disparities: Choosing a Comparison/Reference Group



Age adjusted prevalence of physician diagnosed diabetes mellitus, NHANES: 2007-2010

Measuring Disparities : Choosing a Comparison Group



Men:
NH White Male: 7.7%
NH Black Male: 13.5%
Mexican American: 11.4%

Mexican Americans vs whites?
Mexican Americans vs blacks?
Mexican Americans vs all others?

Three arguments:

1) Use whites as the comparator because they are the “default” or the most advantaged
2) Use the largest group as the comparator because they are most precisely estimated (often whites)
3) Use the healthiest group as the comparator

Differences between IM rates based on alternative references, US 2000

Race and Hispanic origin of mother	Number of infant deaths	Number of live births	Infant mortality rate[1]
American Indian or Alaska Native	346	41,668	8.3
Asian or Pacific Islander	977	200,544	4.9
Hispanic	4,564	815,883	5.6
Non-Hispanic black	8,212	604,367	13.6
Non-Hispanic white	13,461	2,362,982	5.7

[1] Infant deaths per 1,000 live births.

SOURCE: National Vital Statistics System, linked birth/infant death data file.

- Keppel et al 2005

Differences between IM rates based on alternative references, US 2000

Race and Hispanic origin of mother	Number of infant deaths	Number of live births	Infant mortality rate[1]	Referer: Another group rate (Non-Hispanic white) 5.7	Referer: Best group rate (Asian or Pacific Islander) 4.9
					Diffe
American Indian or Alaska Native	346	41,668	8.3	2.6	3.4
Asian or Pacific Islander	977	200,544	4.9	−0.8	0.0
Hispanic	4,564	815,883	5.6	−0.1	0.7
Non-Hispanic black	8,212	604,367	13.6	7.9	8.7
Non-Hispanic white	13,461	2,362,982	5.7	0.0	0.8

[1] Infant deaths per 1,000 live births.

SOURCE: National Vital Statistics System, linked birth/infant death data file.

- Keppel et al 2005

Differences between IM rates based on alternative references, US 2000

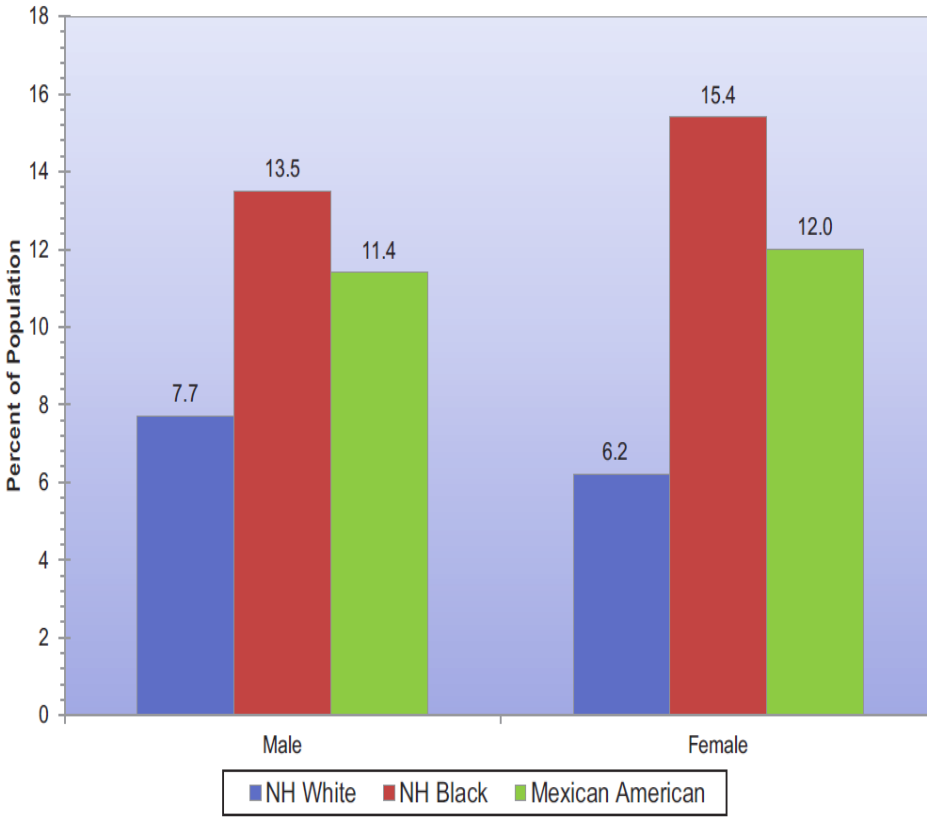
	Reference point[1]				
	Another group rate (Non-Hispanic white) 5.7	Best group rate (Asian or Pacific Islander) 4.9	Mean of group rates 7.6	Total population rate 6.9	Healthy People 2010 target 4.5
Race and Hispanic origin of mother	Difference[1]				
American Indian or Alaska Native	2.6	3.4	0.7	1.4	3.8
Asian or Pacific Islander	−0.8	0.0	−2.7	−2.0	0.4
Hispanic	−0.1	0.7	−2.0	−1.3	1.1
Non-Hispanic black	7.9	8.7	6.0	6.7	9.1
Non-Hispanic white	0.0	0.8	−1.9	−1.2	1.2

[1] Infant deaths per 1,000 live births.

SOURCE: National Vital Statistics Systen

- Keppel et al 2005

Measuring Disparities : Choosing a Comparison Group



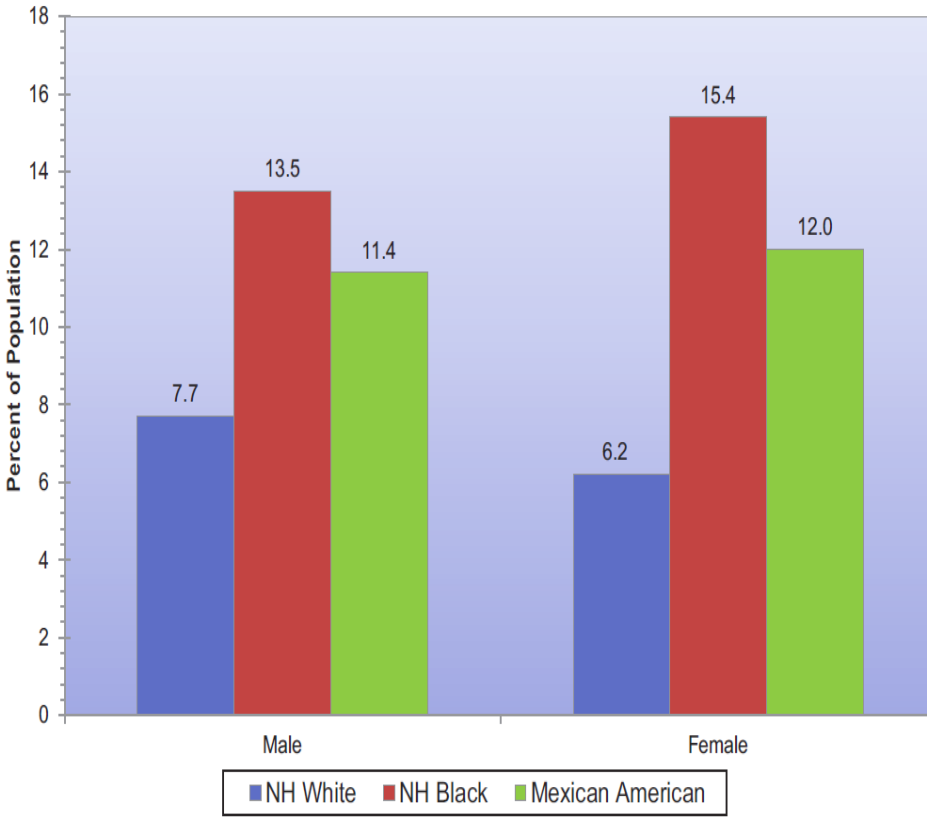
- Use the largest group as the comparator because they are most precisely estimated (usually whites)

This can be justified statistically and note that people can use the converse to hide disparities (if you want to hide a difference, split the groups, or test the comparison with a tiny group).

- Use the best off group as the comparator

In some sense, disparities are compelling because they tell us how much better we could be doing. So the contrast with the healthiest group is most important.

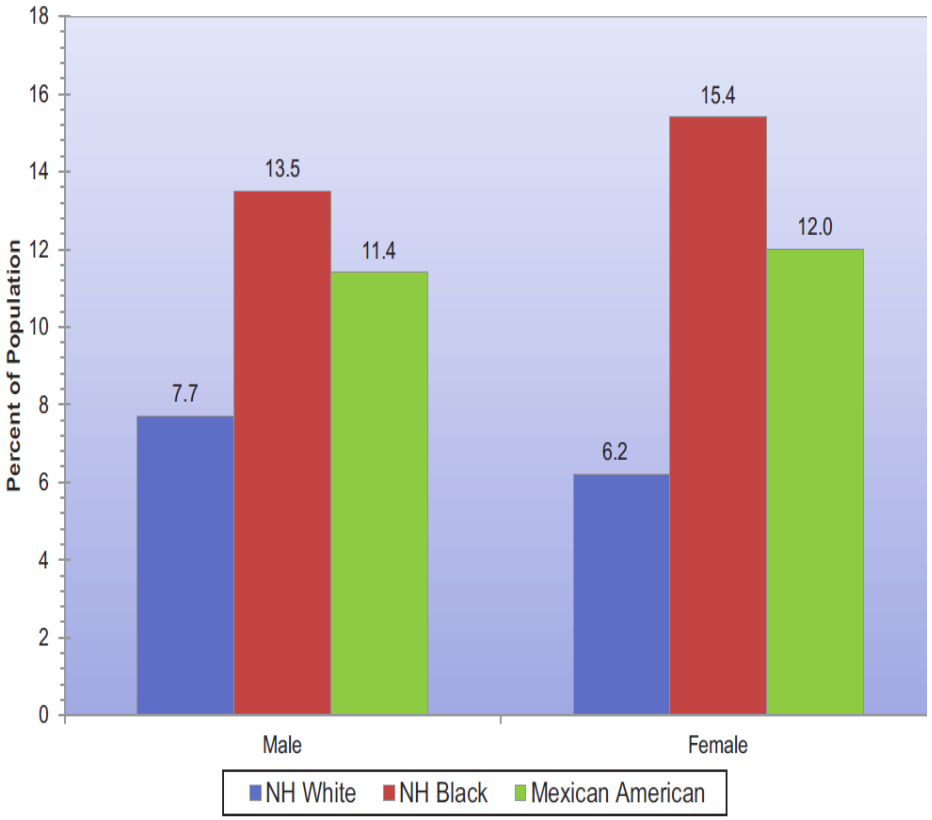
Measuring Disparities : Choosing a Comparison Group



- Use whites as the comparator because they are the “default” or the most advantaged

Most familiar contrasts
Implicitly, deviations from the most advantaged may correspond with “unjust” deviations. If a disadvantaged group is doing better, this is presumably not due to the injustices associated with disadvantage

Measuring Disparities : Choosing a Comparison Group



Three arguments:

1) Use whites as the comparator because they are the “default” or the most advantaged
2) Use the largest group as the comparator because they are most precisely estimated (often whites)
3) Use the healthiest group as the comparator

Consensus: be explicit about your choice and justification.

Measurement Issues in Health Disparities and Social Epidemiology

- How do you measure disparities?
  1. Defining comparison groups
  2. Choosing the measure
  3. Outcomes
- What is the role of measurement in explaining disparities and how do insufficient measures compromise this?
- Challenges in measuring social constructs
- Some basic ideas for measuring latent variables

Measuring Disparities: Showing a Difference Between Groups

A major distinction is between *absolute* differences and *relative* differences. This is controversial, and I have a strong personal opinion that is not universally accepted.

Relative effects approach: black men have 13.5/7.7=**1.75 times** the prevalence of diabetes as white men.

Absolute effects approach: black men have 13.5-7.7=**5.8 percentage point** higher prevalence of diabetes than white men.

Measuring Disparities: Showing a Difference Between Groups

RELATIVE MEASURES

- Odds ratio
- Relative risk
- Hazard ratio
- Incidence rate ratio
- Standardized mortality ratio

ABSOLUTE MEASURES

- Risk difference
- Rate difference
- Years of life lost

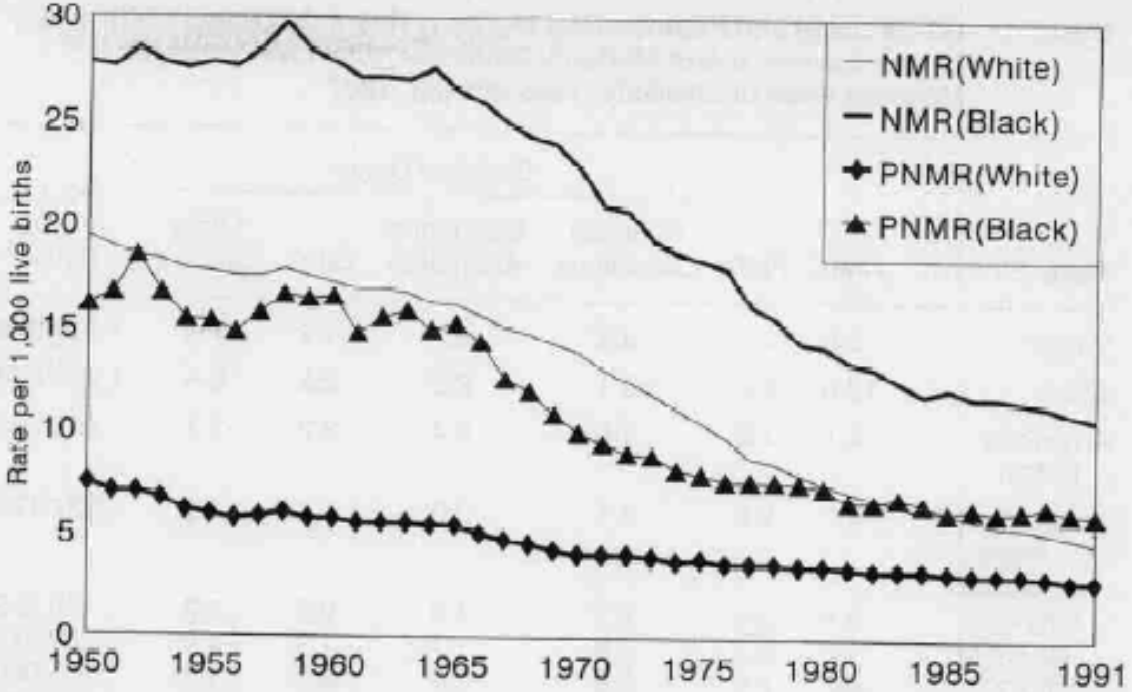
Relative measures are extremely common, so when you see one, stop and think “how big is this effect in absolute terms?”

Measuring Trends in Disparities: Relative vs Absolute Measures

Some claims that depend on comparing risk factor coefficients:

- The black-white disparity in neonatal mortality got worse during the 20th century, despite the advances in civil rights.
- Although the medical system may not work ideally for younger African-Americans, these problems are ameliorated among the elderly.
- After age 70, the role of socioeconomic factors in determining morbidity and mortality declines, and genetic risks predominate.
- From 1968-1996, Alabama made greater strides in reducing stroke mortality than Oregon.

Trends in Infant Mortality: Have disparities gotten larger or smaller?



Note. NMR = neonatal mortality rate; PNMR = postneonatal mortality rate. Data from National Vital Statistics System, 1950 through 1991.

FIGURE 1—US neonatal and postneonatal mortality rates by race of child, 1950 through 1991.

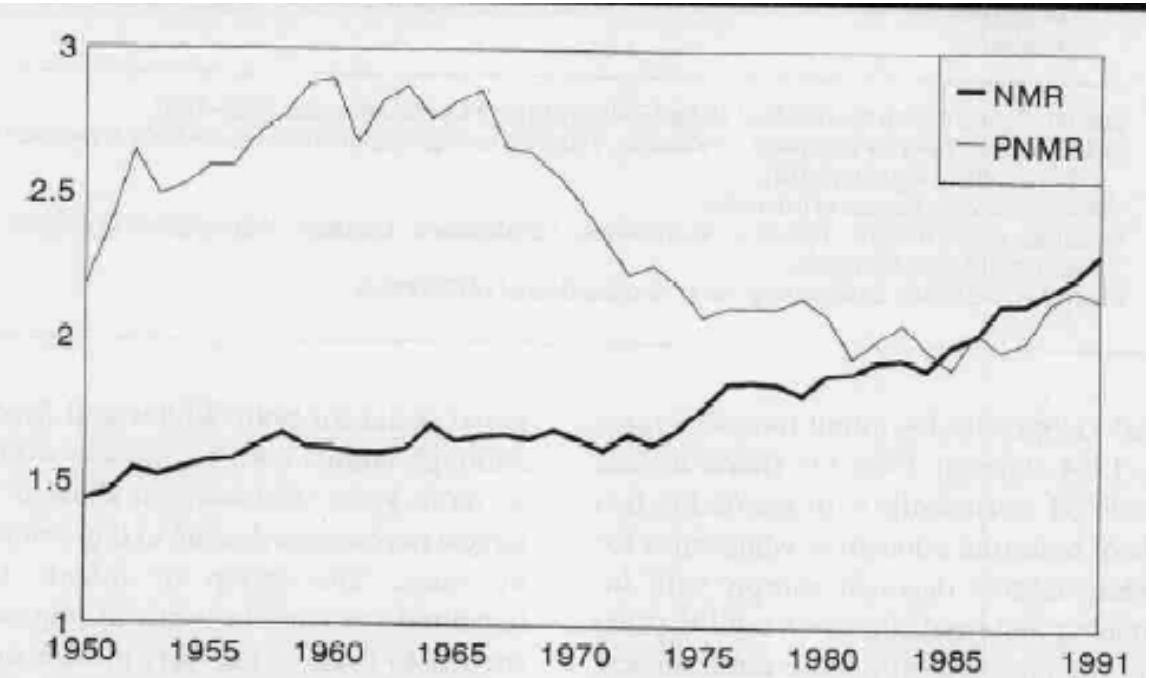
The IM for whites declined by 3.2% per year, but for blacks only by 2.9%.

"As a result, the racial disparity in the IM rate increased between 1950 and 1991." In 1950, the rate was 43.9/1000 for black babies and 26.8/1000 for whites. "By 1991, the relative infant mortality situation for Black infants had deteriorated, with their rate of 16.5 being 2.2 times greater than the (7.5) for white infants).

1950 RD=43.9-26.8=17.1.
1991 RD=16.5-7.5 = 9.0.

Singh and Yu, 1995

Trends in Infant Mortality: Interpretation depends on absolute vs relative measure



Note. NMR = neonatal mortality rate; PNMR = postneonatal mortality rate. Data from National Vital Statistics System, 1950 through 1991.

FIGURE 2—Black/White ratios of US neonatal and postneonatal mortality rates, 1950 through 1991.

Singh and Yu, 1995

Absolute vs Relative Changes

- Disparities should be measured in *both* absolute and relative terms in order to understand their magnitude, especially when making comparisons over time or across geographic populations or indicators.

- Absolute and relative estimates will always have the same sign (positive or negative) but estimates of *rate of change or trends* (eg comparisons of disparities across ages or calendar year) may not have the same sign.

Effect Measure Modification: Scale dependence

- Effect measure modification: is the effect measure the same in different strata?
  - Is the relative risk of stroke associated with obesity the same in men and women? Is it the same in blacks and whites?
  - Is the risk difference for stroke associated with obesity the same in men and women? Is it the same in blacks and whites?
- If there is a main effect (e.g. obesity predicts stroke and race also predicts stroke), usually no more than one effect measure will show *no* modification.
  - If the RRs are the same for blacks and whites, but blacks have a higher overall rate of stroke than whites, the RDs *must be different.*

Effect Modification: Scale dependence

- The scale dependence of interaction is essential to recognize when interpreting disparities, and especially trends in disparities.
- Relative effect measures will tend to be closer to the null in groups with higher baseline risk.
  - RRs in young routinely more extreme than RRs in old.
  - RRs tend to become more extreme over time if there is progress in reducing the population risk of a disease
- Which is more important: relative or absolute changes?

Measurement Issues in Health Disparities and Social Epidemiology

- How do you measure disparities?
  1. Defining comparison groups
  2. Choosing the measure
  3. Outcomes
- What is the role of measurement in explaining disparities and how do insufficient measures compromise this?
- Challenges in measuring social constructs
- Some basic ideas for measuring latent variables

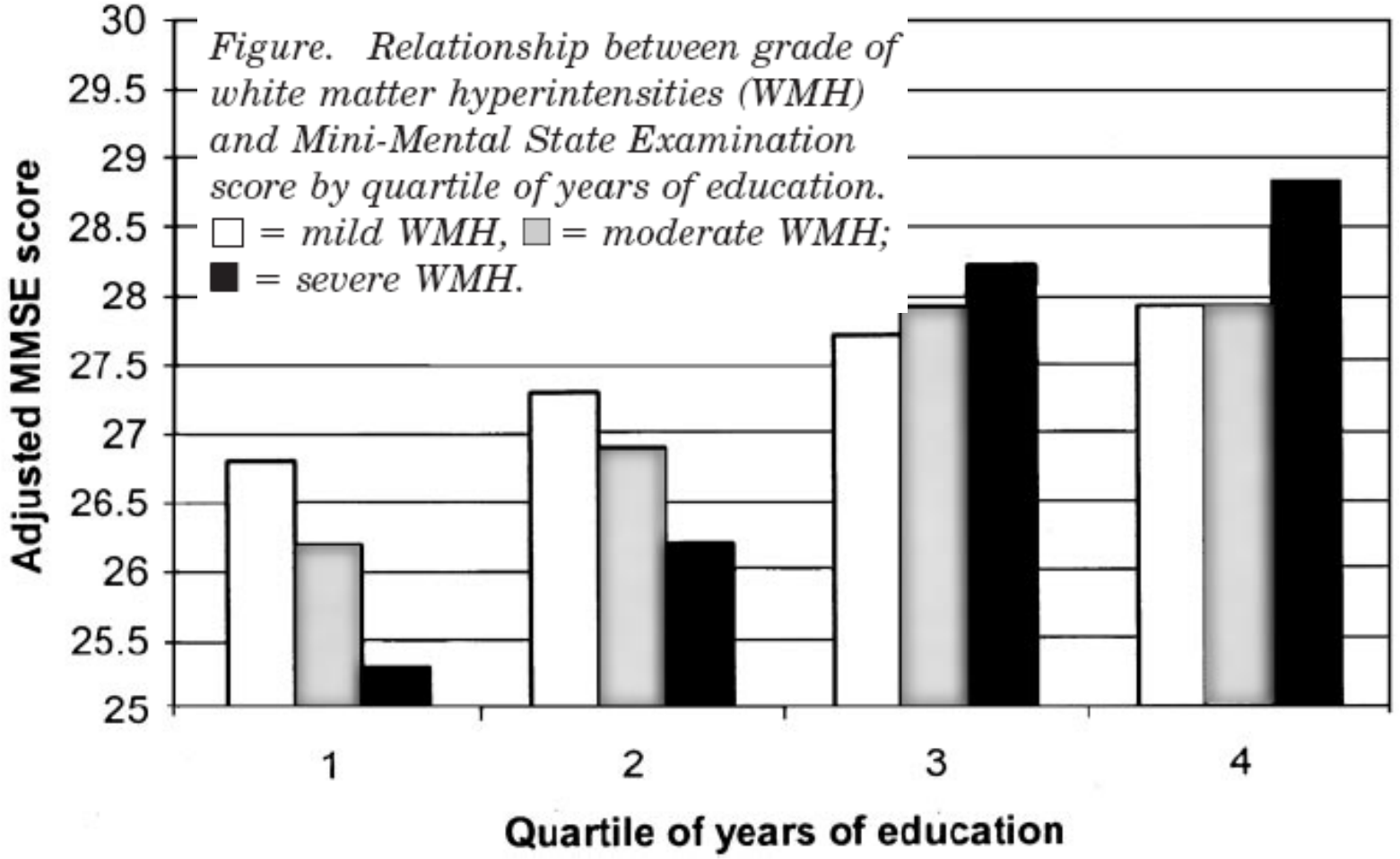
What should we measure?

- Determinants of Disease (eg behaviors)?
- Disease?
- Perceived health/well being?
- Health care utilization (of treatments)
- Health care utilization (of preventive care)
- Impairments (eg muscle weakness)?
- Disability/functioning (eg ADL)?
- Social participation/handicap/quality of life?
- Note diversity of vocabularies related to disability
  - Disability is created by a gap between the physical capacity of the person and the environmental demands
  - Outcomes should include participation in society, not just physical measures

"Objective" vs "subjective"

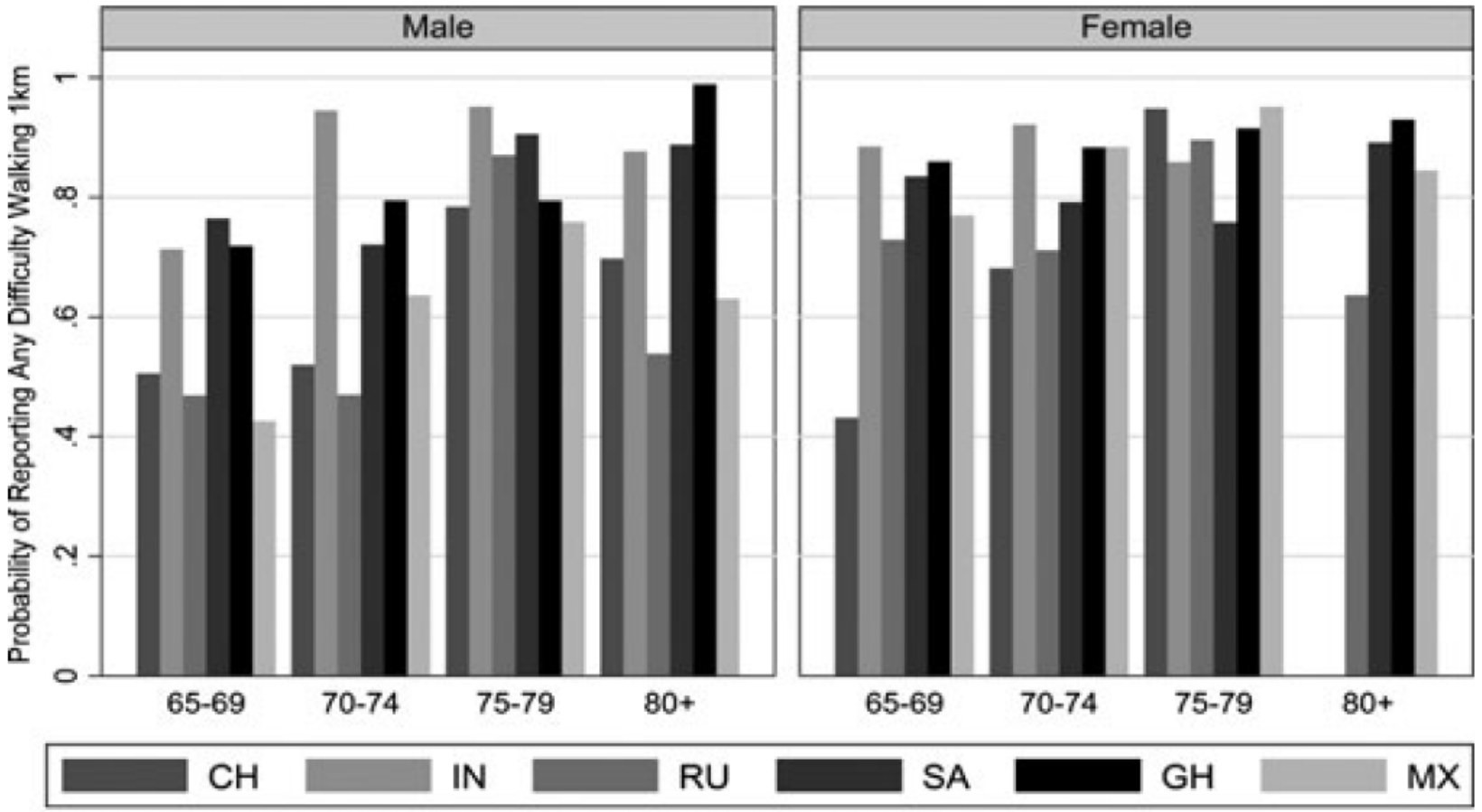
- MRI: white matter hyperintensities
- Cognitive test score
  - Which is most relevant clinically for assessing care needs?
  - Which is most relevant for identifying course of disease?
  - Which is most relevant for etiologic research?
- Which can be measured?
- What is the opportunity cost of measuring either?
  - Sacrifice study size, diversity (race, SES, geography), representativeness in exchange for more technical measures
  - Loss of measures of other variables due to time
  - Very tempting to always push for more technical measures, "objective" and splashier – not always better

Objective vs Subjective Measures



Dufouil, 2003

Probabilities of self-reporting difficulty walking 1 km in those with slow walking speed Study on Global Ageing and Adult Health, 2007–2010 Capistrant, JAGS 2014



CH=China IN=India RU=Russian Federation SA=South Africa GH=Ghana MX=Mexico
All women aged 80+ in China and India with slow walking speed reported difficulty; probabilities were not estimable.
Cross-national differences were statistically significant ($p<0.05$) for men and women at all ages, except 75-79 ($p>0.1$).

Measurement Issues in Health Disparities and Social Epidemiology

- How do you measure disparities?
  1. Defining comparison groups
  2. Choosing the measure
  3. Outcomes
- What is the role of measurement in explaining disparities and how do insufficient measures compromise this?
- Challenges in measuring social constructs
- Some basic ideas for measuring latent variables

Inadequate measures can result in:

- Conceptual inadequacy: you are asking a silly question or a vaguely defined question→ meaningless answer to your fuzzy research question
- Poor data quality (e.g. missing data): people won't answer or can't answer → no answer to your question
- Poor variability: everyone answers the same → no answer to your question
- Poor reliability and validity: you measure the wrong thing or measure it badly → the wrong answer to your question or no power to answer the question

What to worry about depends on your question

- What is the effect of education on dementia risk?
  - Bad measure of education will tend to attenuate association
  - Dementia may be differentially measured – bias could go either way
  - Bad measures of confounders (mother's education?) leaves confounding bias
- How much of the black-white disparity in stroke is attributable to SES?
  - Bad measures of childhood SES lead to under-estimation of role of SES
- What is the effect of receiving treatment A on dementia risk?
  - Bad measure of social resources influencing receipt of treatment A leaves confounding bias – making hard to access treatments look good.

Excess Stroke in Blacks vs Whites

- "Risk factors for increased stroke mortality in blacks heretofore examined, primarily a higher prevalence of hypertension and diabetes and a lower socioeconomic status, explain only a fraction of this excess….

…It is possible that elevated blood pressure and insulin resistance/diabetes impact blacks more strongly, which perhaps suggests a genetic influence or susceptibility. …

… A more complete understanding of the mechanisms underlying the race/ethnic differences in stroke subtype will likely depend on advances in genomics. " – Wolf and Kannell, Circulation 2007

But how well are social risk factors measured?

- Place of birth/residence?
- Childhood SES?
- Comprehensive adult SES?

Overlap of Race and Place for Older Americans

		African-Americans		Whites	
		Birthplace		Birthplace	
		Non-Southern	Southern	Non-Southern	Southern
Current Residence	Non-Southern	15%	29%	60%	5%
	Southern	2%	54%	10%	25%
		17%	83%	70%	30%

2000 Census 5% sample, age 65+, race black or white, southern= census region.

Research Questions

- Does childhood SES predict stroke incidence in black and white adults?
- Do childhood social conditions "explain" racial disparities in stroke incidence?
- Do adult SES measures "explain" racial disparities in stroke incidence?
- Details in: Glymour et al Annals of Epi, 2009.

Childhood Social Conditions

Blacks (n=2,683)	Separate Models HR (95% CI)	Summary Models HR (95% CI)
Father's Education		
< 8 Years	1.39 (1.03 , 1.88)	
Unkown	1.39 (0.99 , 1.96)	
Father's Occupation		
Manual/Unskilled	0.83 (0.52 , 1.32)	
Farming	0.79 (0.49 , 1.27)	
Missing	0.82 (0.51 , 1.33)	
Childhood Health		
Fair or poor	1.49 (1.03 , 2.17)	1.49 (1.02 , 2.17)
Unknown	2.31 (1.53 , 3.48)	2.25 (1.49 , 3.40)
Height		
Inches	0.96 (0.92 , 1.00)	0.96 (0.92 , 1.00)

Childhood Social Conditions

Whites (n=15,906)	Separate Models HR	(95% CI)	Summary Models HR	(95% CI)
Father's Education				
< 8 Years	1.24	(1.07 , 1.44)		
Unkown	1.45	(1.20 , 1.76)		
Father's Occupation				
Manual/Unskilled	1.16	(1.00 , 1.34)		
Farming	1.07	(0.91 , 1.26)		
Missing	1.70	(1.44 , 2.01)		
Childhood Health				
Fair or poor	1.44	(1.18 , 1.75)	1.42	(1.17 , 1.73)
Unknown	2.85	(2.34 , 3.48)	2.79	(2.29 , 3.41)
Height				
Inches	0.99	(0.97 , 1.01)	0.99	(0.97 , 1.01)

Accounting for Racial Differences

	Model 1	
	HR	(95% CI)
Black race	1.46	(1.29 , 1.65)
Southern birthstate		
Low Parental SES Index		
Fair/poor child health		
Adult SES		
Income*		
Wealth*		
Main Occupation		
Manual/Unskilled		
Cardiovascular risk factors		
Hypertension		
Current smoker		
Vigorous activity		
Obese		
Diabetes		
Heart disease		

Accounting for Racial Differences

	Model 1		Model 4	
	HR	(95% CI)	HR	(95% CI)
Black race	1.46	(1.29 , 1.65)	1.30	(1.15 , 1.48)
Southern birthstate			1.27	(1.16 , 1.41)
Low Parental SES Index			1.27	(1.10 , 1.47)
Fair/poor child health			1.43	(1.21 , 1.71)
Adult SES				
Income*				
Wealth*				
Main Occupation				
Manual/Unskilled				
Cardiovascular risk factors				
Hypertension				
Current smoker				
Vigorous activity				
Obese				
Diabetes				
Heart disease				

Accounting for Racial Differences

	Model 1		Model 5	
	HR	(95% CI)	HR	(95% CI)
Black race	1.46	(1.29 , 1.65)	1.10	(0.96 , 1.26)
Southern birthstate			1.24	(1.12 , 1.37)
Low Parental SES Index			1.20	(1.03 , 1.39)
Fair/poor child health			1.38	(1.16 , 1.64)
Adult SES				
Income*			0.94	(0.91 , 0.98)
Wealth*			0.96	(0.95 , 0.98)
Main Occupation Manual/Unskilled			1.17	(1.01 , 1.34)
Cardiovascular risk factors				
Hypertension				
Current smoker				
Vigorous activity				
Obese				
Diabetes				
Heart disease				

Accounting for Racial Differences

	Model 1		Model 6	
	HR	(95% CI)	HR	(95% CI)
Black race	1.46	(1.29 , 1.65)	1.01	(0.88 , 1.16)
Southern birthstate			1.22	(1.11 , 1.35)
Low Parental SES Index			1.16	(1.00 , 1.35)
Fair/poor child health			1.34	(1.12 , 1.59)
Adult SES				
Income*			0.95	(0.92 , 0.98)
Wealth*			0.98	(0.96 , 0.99)
Main Occupation Manual/Unskilled			1.15	(1.00 , 1.32)
Cardiovascular risk factors				
Hypertension			1.53	(1.39 , 1.69)
Current smoker			1.54	(1.36 , 1.74)
Vigorous activity			0.84	(0.75 , 0.94)
Obese			1.04	(0.91 , 1.19)
Diabetes			1.87	(1.65 , 2.11)
Heart disease			1.44	(1.29 , 1.60)

Interpretation

- Racial disparity almost completely disappears after adjustment for SES.
- Just southern birth and wealth are quite powerful.
- Adjusting for rough SES measures (e.g. 3 levels of education) does not render race coefficient non-significant.
- Conclusion that racial disparities "not accounted for by SES" obviously unjustified.
- BUT, these models entail a serious potential bias due to adjusting for mediators.

Larger Take Home

- Demonstrating disparities often entails showing mechanisms that are unjust: ie testing mediation
- Measurement error in the mediators very severely compromises conventional mediation tests, even if "non-differential"
- I perceive this as among the most pervasive methodological problems in disparities research

Measurement Issues in Health Disparities and Social Epidemiology

- How do you measure disparities?
  1. Defining comparison groups
  2. Choosing the measure
  3. Outcomes
- What is the role of measurement in explaining disparities and how do insufficient measures compromise this?
- Challenges in measuring social constructs
- Some basic ideas for measuring latent variables

Many variables of most interest are hard to define and hard to measure

- Distinguish between a "construct" (what you would like to measure) and the instrument you use to measure it.
- Latent variables are not directly observed.
  - Cognitive function
  - Discrimination
  - Depression
  - Disability
  - Pain
  - Blood pressure (?)
- Most of the measurement challenges we will discuss relate to latent variables, but many of the lessons are relevant to directly observed and measurable variables, e.g., body weight.
- Most of the analysis methods you learn depend on having quantitative measures of health and risk factor attributes (even if the only two values the quantity takes are 0 and 1)

What's a latent variable or latent "construct"?

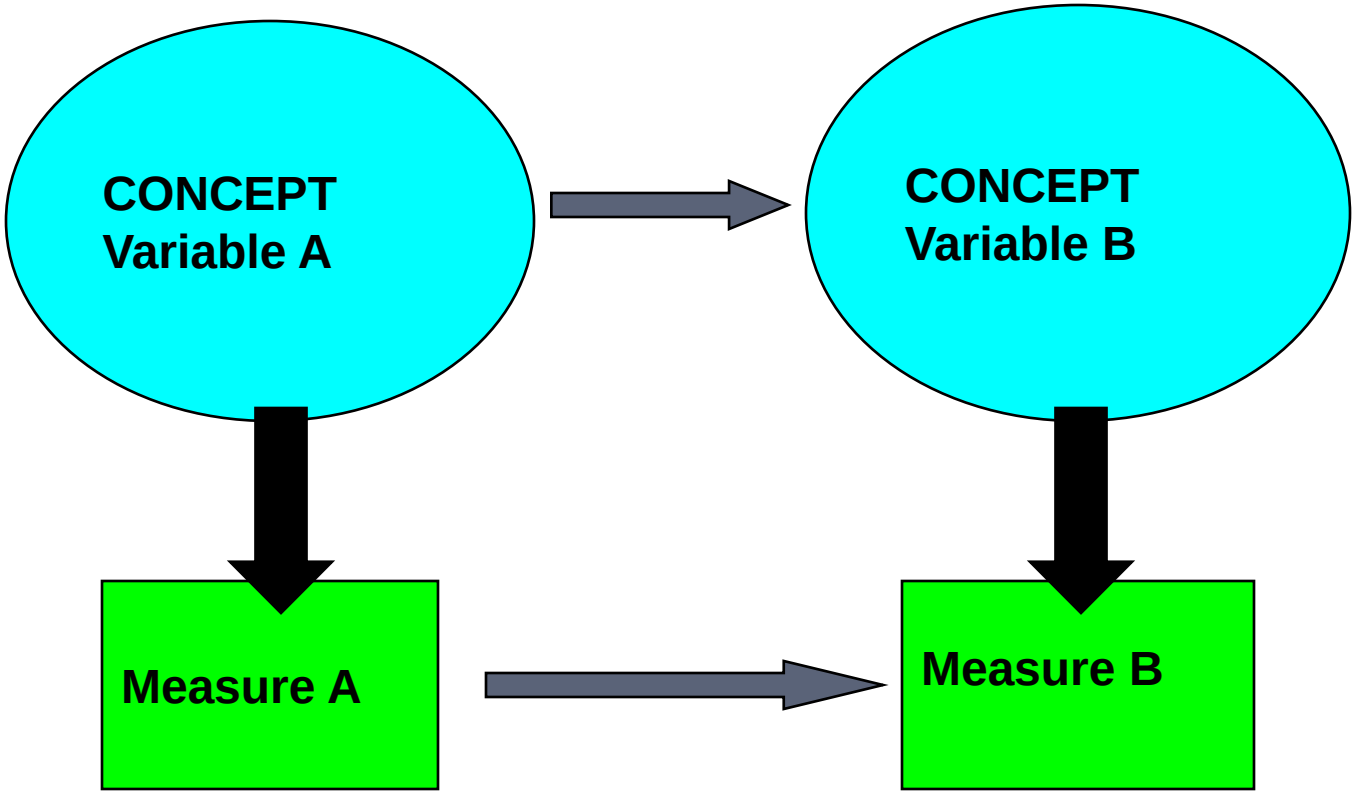
- A variable that is relatively abstract as opposed to concrete
- An abstraction based on observations of certain behaviors or characteristics
- Cannot be assessed directly
- Most of the analysis methods you learn depend on having quantitative measures of health and risk factor attributes (even if the only two values the quantity takes are 0 and 1)
- Often the greatest challenge is precisely defining the construct you wish to measure.
  - Gentrification, Segregation, Income inequality
  - Food environment
  - Racism
  - Memory, BMI, Depressive symptoms

What's a latent variable or latent "construct"?

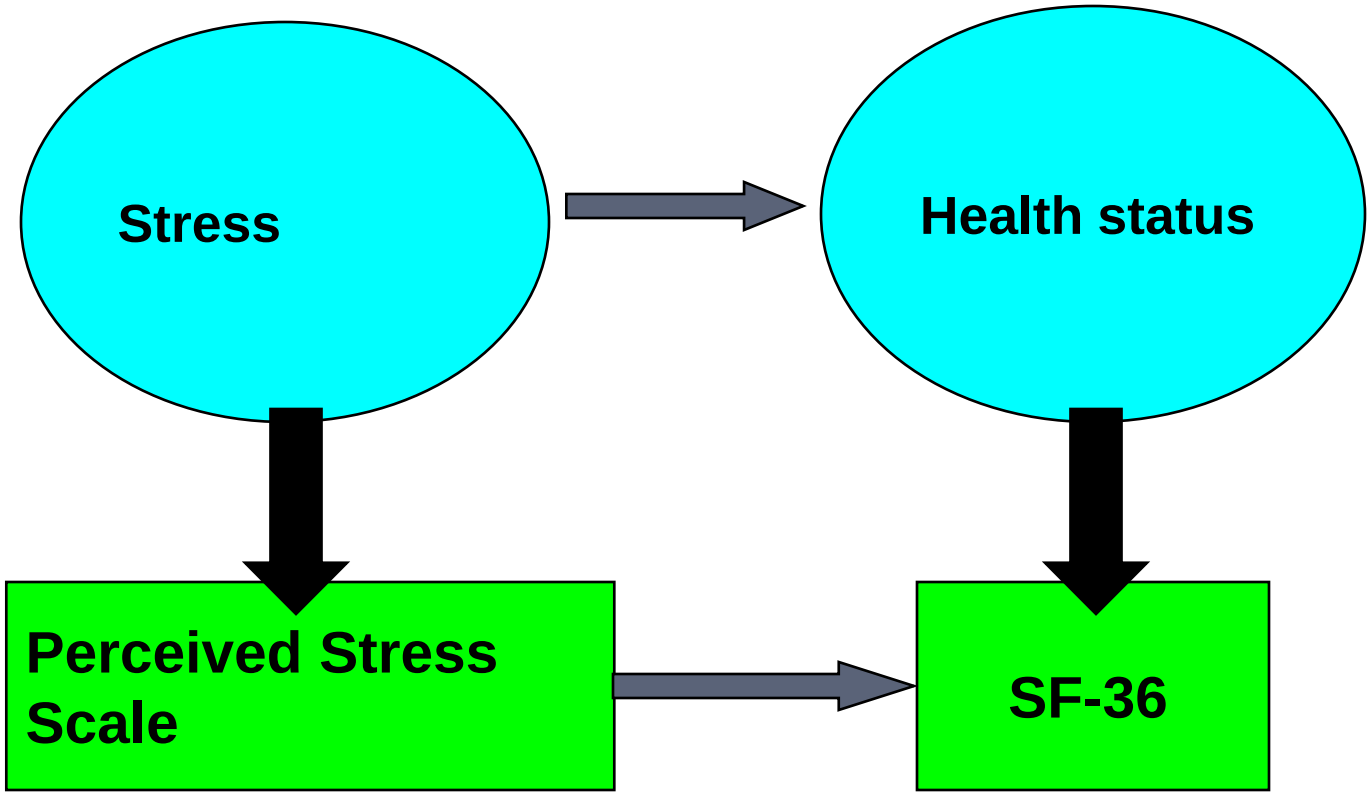
- Often the greatest challenge is precisely defining the construct you wish to measure. Gentrification, Segregation, Income inequality
  - Food environment
  - Racism
  - Memory, BMI, Depressive symptoms

Precise definitions and valid, reliable measurements form the basis for scientific inquiry. If you want to understand how something affects health or is affected by health, you must measure it well. Good measurement may be the most valuable contribution you can make.

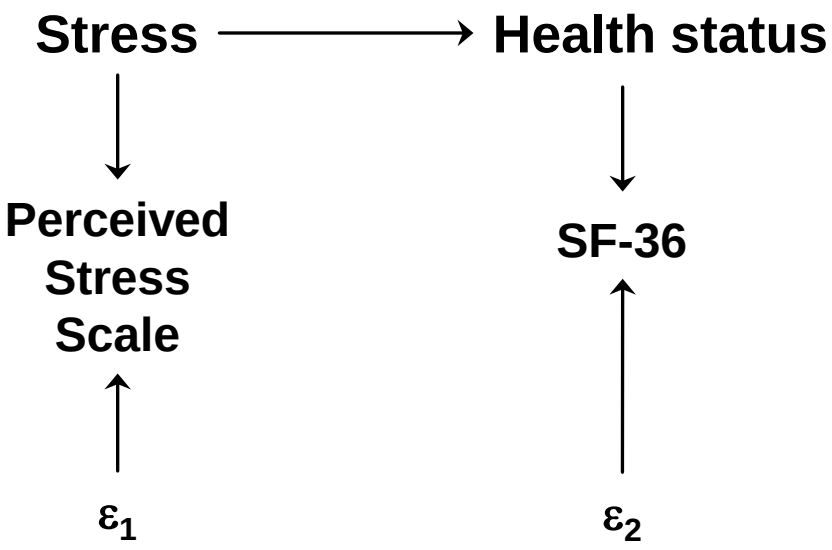
Depicting Latent Variables and Measures



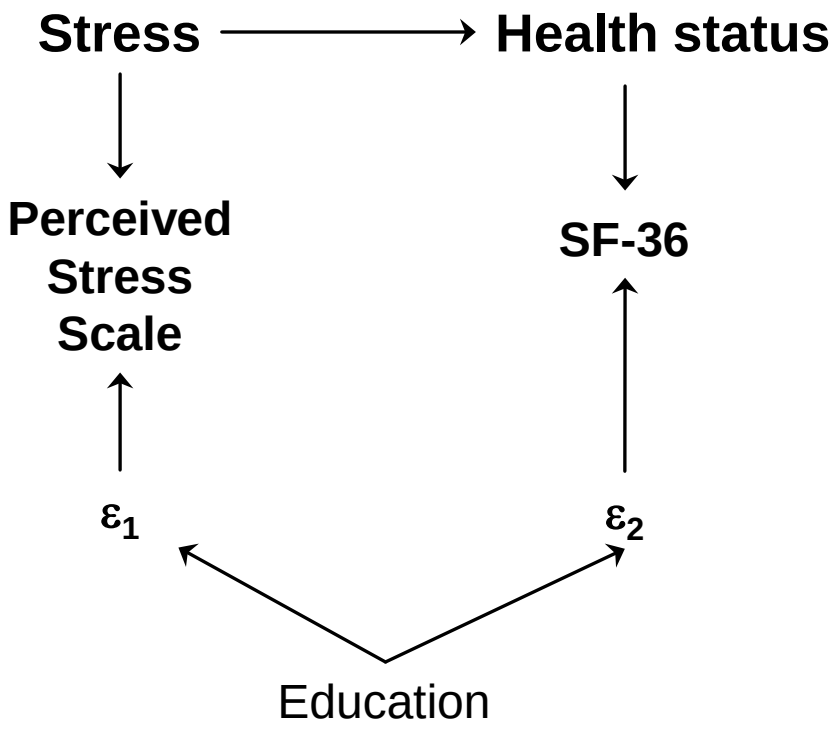
Depicting Latent Variables and Measures



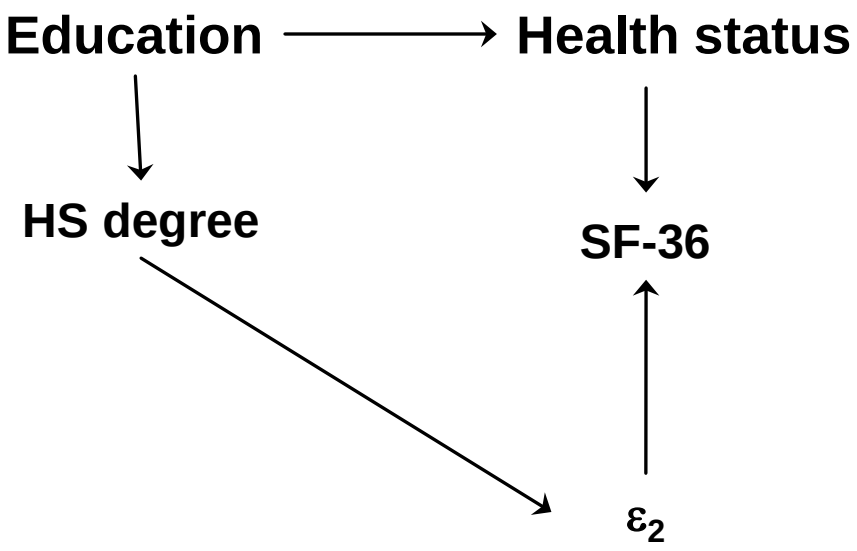
Most measurements are imperfect, ie measurement is affected by error



Lots of empirical work is devoted to evaluating sources of correlation between errors



Huge problem if primary exposure of interest influences errors in reporting



Well developed psychometric tools to evaluate problems such as correlated errors, differential item functioning.

Measurement Issues in Health Disparities and Social Epidemiology

- How do you measure disparities?
  1. Defining comparison groups
  2. Choosing the measure
  3. Outcomes
- What is the role of measurement in explaining disparities and how do insufficient measures compromise this?
- Challenges in measuring social constructs
- Some basic ideas for measuring latent variables (**Many thanks to Anita Stewart and Irene Yen!**)

Process of Selecting Good Measures for Your Studies

Define concept (variable)

Identify potential measures

Review measures' properties
--conceptual adequacy
--psychometric adequacy

Pretest best 1-2 measures

Select final measure

Defining hard to describe constructs

- State Role of Concept in Your Research
  - Evaluate intervention (outcome)
  - Determinant (e.g., social support) of an outcome
  - Predict future event (e.g. functioning to predict hospitalization)
  - Describe population
  - Covariate
- Define Concept for Your Research
  - Define it first from your point of view: How you would define it based on your experience and understanding of your research question
  - For dependent variables: Describe how intervention or independent variables might affect it - types of changes you expect
  - For independent (predictor) variables: Describe how it might be predictive
- Are all dimensions represented?
- Distinguish between closely related constructs

Qualitative methods often invaluable

- ethnography
- direct observation
- in-depth interviews
- focus groups
- Data consist of words (text), useful for
  - Concept development
  - Item construction

Mixed Methods: Iterative Sequence to Develop Structured Measures

- Focus groups
- In-depth interviews
- Input from quantitative studies
- Develop preliminary items
- Review by expert panel
- Cognitive interview pretest
- Pilot study
- Nationwide survey
- Psychometric testing

N Krause, *J Geront: Soc Sci*, 2002;57B:S263-274

Process of Selecting Good Measures for Your Studies

Define concept (variable)

Identify potential measures

Review measures' properties
--conceptual adequacy
--psychometric adequacy

Pretest best 1-2 measures

Select final measure

Locating Potential Measures

- Extensive resources on measuring difficult concepts.
- Look for measures with known psychometric properties.
- Check the psychometric properties in your population (if possible).
- Preferable to use established high-quality measures
  - To facilitate comparisons across time, populations, settings
  - To draw on previous evidence of validity and reliability
- Sometimes you may have to develop a new measure or modify an existing measure.
  - Modify with caution – the cost of the migrating measure is real.
  - Create overlaps whenever possible

Best Compendium

- McDowell I and Newell C, *Measuring Health: A Guide to Rating Scales and Questionnaires.* Third Edition, Oxford University Press, 2006.
  - Reviews measures of physical disability, social health, psychological well-being, anxiety, depression, mental status, pain, and general health status and quality of life
  - Includes actual instruments where possible and selected items in some

Reviews of Measures: Examples

- Structured reviews:
  - AC Mui et al. Cross-cultural assessment of geriatric depression: A review of the CES-D and the GDS. *J Ment Health Aging*, 2001;7:137-164.
  - SM Martinez et al. A review of physical activity measures used among U.S. Latinos: guidelines for developing culturally appropriate meausres. *Ann Beh Med*, 2008;36:195-207.
- Narrative reviews:
  - JH Skinner. Acculturation: measures of ethnic accommodation to the dominant American culture. *J Ment Health Aging*, 2001;7:41-52.

Measure, Instrument, Battery, Questionnaire, Survey, Tool, Inventory

- Measure - single- or multi-item scale or index (umbrella term)
- Instrument - a published, named measure or set of measures
- Battery - collection of measures from diverse sources
- Questionnaire - any of the above (instrument, profile, battery) formatted into a set of measures
- Survey - same as questionnaire
- Tool = Questionnaire = Survey?
- Inventory?

When evaluating, consider type of measurement scales

- Categorical (nominal)
  - Classification (race/ethnicity; health insurance source)
  - Number answer categories on the survey to improve data quality/management
  - But numbers are simply labels for categories
- Continuous
  - Ordinal (order matters)
  - Rank
  - Interval (each 1 unit interval equivalent meaning)
  - Ratio (measure has a meaningful zero point, so ratio of values is informative: MMSE vs fertility rate)

Typical Sequence of Developing New Self-Report Measures

Develop/define concept

Create item pool

Pretest/revise

Field survey

Psychometric analyses

Final measures

Scales vs Indexes (per the Devellis framework)

- A usual approach to measuring a latent variable is to develop several items. When considering such item pools, it is important to ask: is this a scale or an index?
- Items on a scale are thought to all have a common cause (the latent variable). They are assumed to correlate with one another because they share this common cause. This is a useful application of confounding!
- Items on an index are thought to all have a common *effect.*
- This distinction is critical because there is no reason to believe the items on an index should be correlated with one another (although they might be).
- Some validation approaches and approaches to missing data are relevant only for scales

Things to Pay Attention to in Items

- Item stem
  - Time frame
  - Complexity
- Response scales
  - Type
  - Number of choices
  - Specific choices, distance between choices
  - Format
- Match of response choices to item stem

Composition of an Item

During the past month, how much of the time have you felt tired?

Item stem

1 Never
2 A little of the time
3 Some of the time
4 Most of the time
5 All of the time

Response scale

Optimal Number of Response Choices

- Are 100 levels better than 5?
- 5-points is optimal for items
  - Especially if items combined into multi-item scales
- For single-item measures
  - 7 points preferred
  - Allows distribution over 5 points since people seldom use endpoints

Response Scale Choices: "Vague, Imprecise Quantifiers"

- How often?
  - Very often, pretty often, not too often
  - Sometimes, often, never
- How much?
  - Too little, about right, too much
  - Below average, average, above average

Bradburn NM. *Public Opinion Quart* 1979, 92-101.

Typical Sequence of Developing New Self-Report Measures

Develop/define concept

Create item pool

Pretest/revise

Field survey

Psychometric analyses

Final measures

Pretest in Target Population

- Pretesting essential for measures being applied to any new population group
  - Especially priority measures (e.g., outcomes)
- Pretest is to identify:
  - problems with procedures
    - method of administration, respondent burden
  - problems with questions
    - Item stems, response choices, and instructions

Typical Sequence of Developing New Self-Report Measures

Develop/define concept

Create item pool

Pretest/revise

Field survey

Psychometric analyses

Final measures

Distance Between Levels: "In general, how would you rate your health?"

Self-rated health item:	Mean 7-item current health scale: Screening N=~11,000		Mean 7-item current health scale: Baseline N=3,054
1 – poor	10.8	-- 20 --	10.8
2 – fair	30.0	-- 26 --	30.6
3 – good	57.6	-- 18 --	55.9
4 – very good	75.5	-- 11 --	75.4
5 – excellent	87.9		86.9

largest diff

Distance Between Levels: "In general, how would you rate your health?"

Self-rated health item:	Mean: current health scale		
	Screening N=~11,000		Baseline N=3,054
1 – poor	10.8	-- 20 --	10.8
2 – fair	30.0	-- 26 --	30.6
3 – good	57.6	-- 18 --	55.9
4 – very good	75.5	-- 11 --	75.4
5 – excellent	87.9		86.9

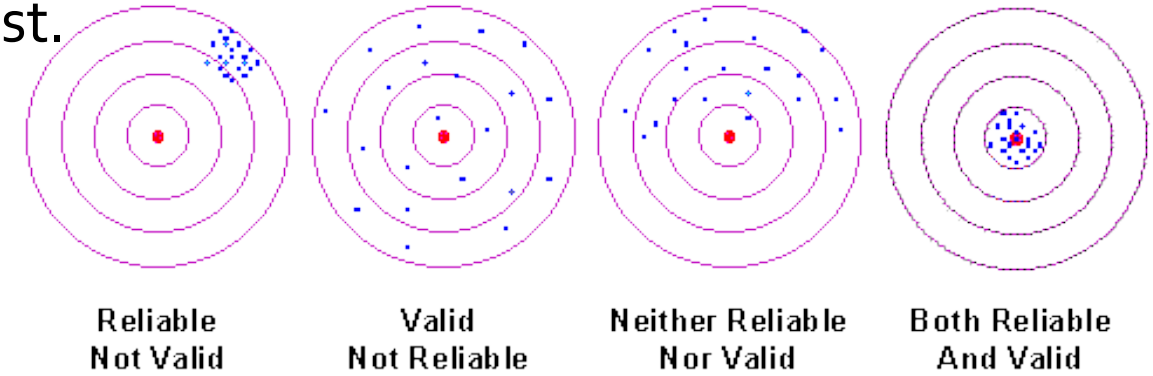
smallest diff

Distance Between Levels: Relevant for many continuous variables

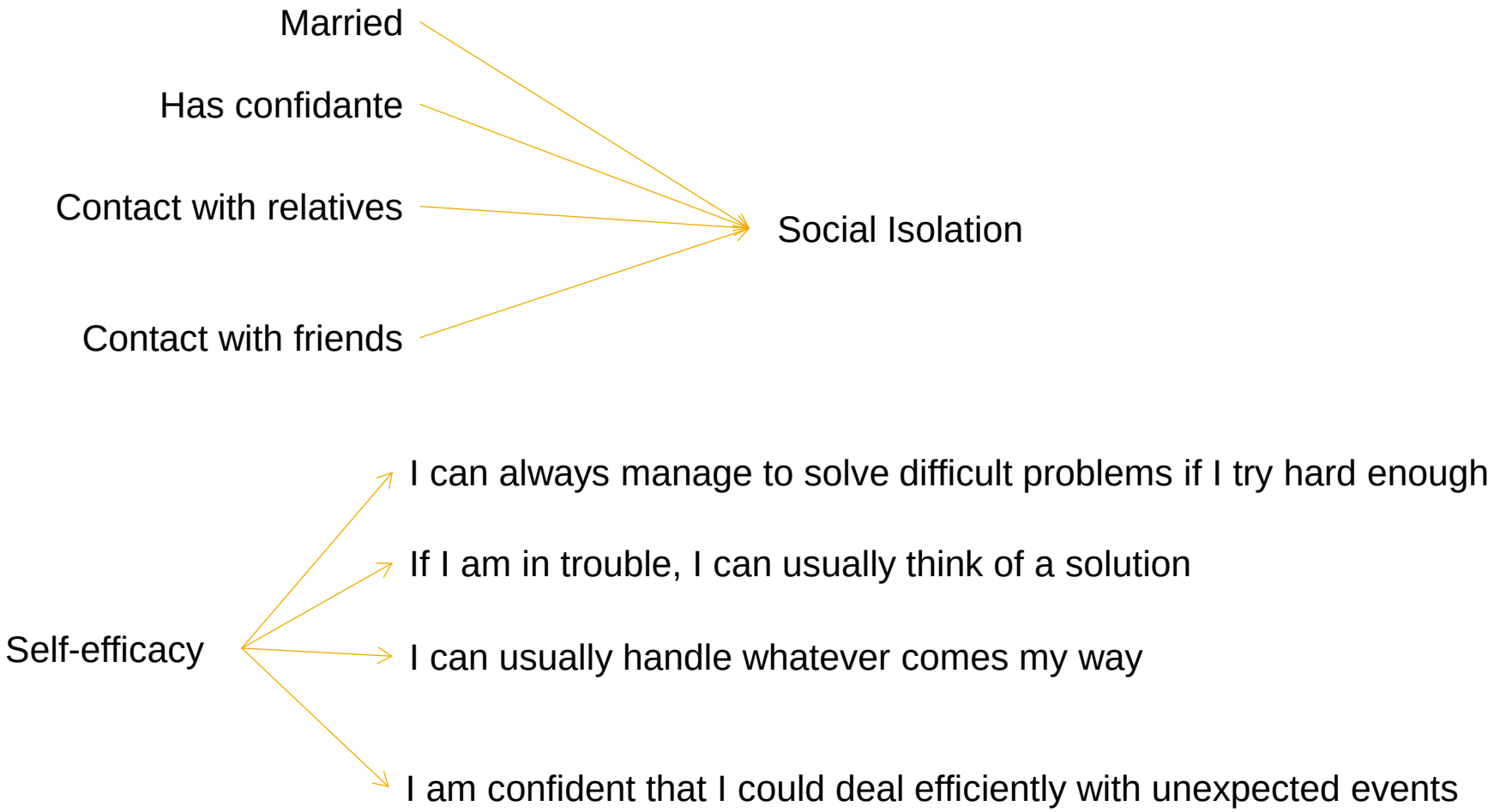
- MMSE: ceiling=30, impaired=24
  - Highly correlated with education, race, literacy, age…
- Non-normally distributed
  - Log transform?
- Is a decline from 30 to 29 the same as a decline from 24 to 23?
- Item response theory/rasch models intended to equalize intervals.

How good is my measure?

- Distinguish between validity and reliability
- Validity: The extent to which an assessment measures what it is intended to measure. (Systematic error)
  - Everything starts with a clear, precise, concept
- Reliability: The extent to which a measure is precise; the proportion of variance in a measurement that is due to the construct of interest.

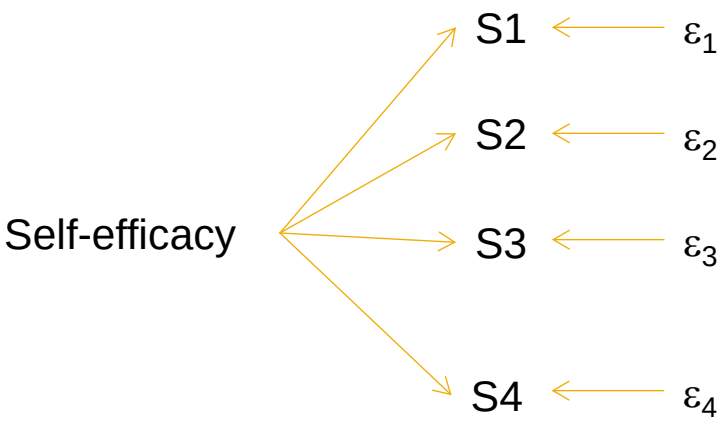


Scales vs Indexes (per the Devellis framework)



Why repeat basically the same question in different ways?

- Classical measurement theory:
  - The person's response to each item is the latent construct (say, efficacy) plus some item specific noise.
  - S1=b*self-efficacy+ε_1
  - S2=b*self-efficacy+ε_2
  - S3=b*self-efficacy+ε_3
  - S4=b*self-efficacy+ε_4



Classic Measurement Model Assumes:

- Errors are random
- Errors are not correlated with each other
- Errors are not correlated with the true score
- Latent variable affects all items equally
- Variance of each error is the same

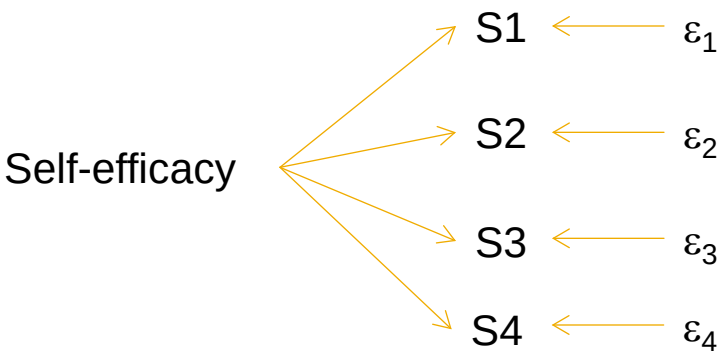
Many modern tools to loosen these assumptions, but classic measurement still conceptually helpful

Why repeat basically the same question in different ways?

- Classical measurement theory:
  - S1=b*self-efficacy+ε_1
  - S2=b*self-efficacy+ε_2
  - S3=b*self-efficacy+ε_3
  - S4=b*self-efficacy+ε_4

Under these assumptions, you can estimate the latent variable very well with a few items.

  - self-efficacy~$N(0,\sigma^2)$
  - ε~$N(0,\sigma^2)$
  - Assume b=1
  - What % of the variance in S1 is due to self-efficacy?

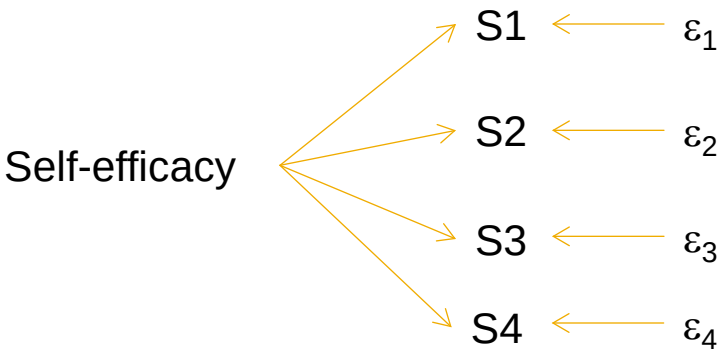


Adding items reduces measurement error and improves reliability

- Classical measurement theory:
  - S1=b*self-efficacy+ε_1
  - S2=b*self-efficacy+ε_2
  - S3=b*self-efficacy+ε_3
  - S4=b*self-efficacy+ε_4

Under these assumptions, you can estimate the latent variable very well with a few items.

  - self-efficacy~$N(0,\sigma^2)$
  - ε~$N(0,\sigma^2)$
  - Variance of S1=$\sigma^2+\sigma^2$ (because ε_1 is independent of self-efficacy)
  - What % of the variance in S1 is due to self-efficacy? 50%



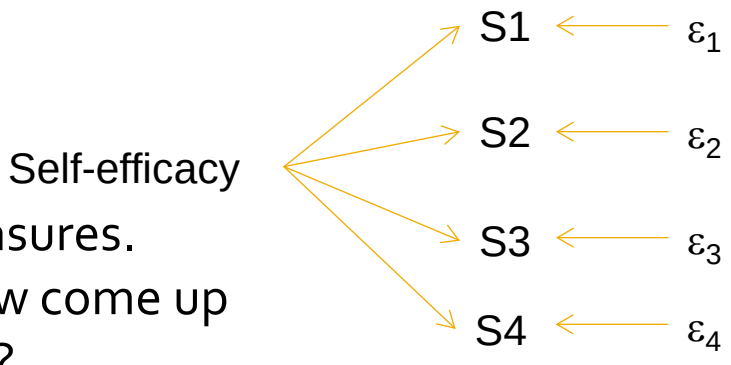
- Now calculate (S1+S2)/2
- What % of the variance in (S1+S2)/2 is due to self-efficacy?

Spearman-Brown Prophecy Formula:

- α=k*average inter-item correlation)/
- (1+(k-1)*average inter-item correlation
- =2*.5/(1+1*.5)=1/1.5=.66

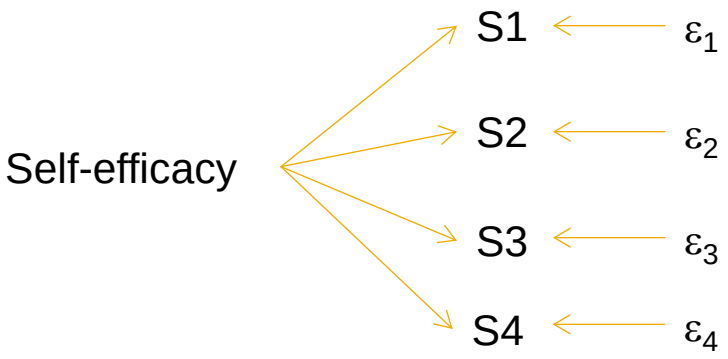
How can I assess reliability?

- What we did is inter-item consistency, a measure of reliability. Often called Cronbach's alpha.
  - In stata: **alpha S1-S4 , item generate(SE_hat)**
  - Pay attention to the direction of items,
  - evaluate whether you might want to dump some of your items
- Test-retest reliability
  - Give the test
  - Give it again
  - Examine the correlation between the two assessments – consider whether latent construct might have *actually changed* in the delay interval.
  - Seem familiar?
- Inter-rater reliability
  - Have Joe give the test
  - Have Elaine give the test
  - Examine the correlation between measures.
  - Note all the issues diagnostic tests now come up
  - Does Elaine always give higher scores?



How can I assess reliability?

- What we did is inter-item consistency, a measure of reliability. Often called Cronbach's alpha.
  - In stata: **alpha S1-S4 , item generate(SE_hat)**
  - Pay attention to the direction of items,
  - evaluate whether you might want to dump some of your items
- Test-retest reliability
- Inter-rater reliability
- Correlation, coefficient of variation (within person SD/mean), Bland Altman (deviations plotted against mean), Kappa coefficients (observed-expected)/(1-expected).
- Plot them.



How can I assess validity?

- Content validity: the extent to which a specific set of items reflects a content domain, e.g., symptoms of anxiety
  - A scale has content validity when its items are a randomly chosen subset of the universe of appropriate items.
- Predictive validity: does the measure predict some other, external measure of the same thing?
- Criterion validity: if you have a gold standard measure of the construct, does the new measure predict the gold standard?
  - Sensitivity and specificity if you have a diagnostic criterion.
- Construct validity: Does the measure correlate with other things that, theoretically, you think it ought to correlate with? Does it *not* correlate with things you think should be independent? It's hard to make a strong case for validity using only construct validity.

Unreliable measures in regressions

- Assume everything is mean centered, so:

- $Y=\beta X+\varepsilon$, $\varepsilon \sim N(0,\sigma^2)$
- $E(Y|X)=\beta X$
- $E(\varepsilon|X)=0$
- (side note: the correlation of Y and X is: $r=\sigma_{xy}/\sigma_x\sigma_y$

- Now suppose we haven't measured X, but just $X'=X+\varepsilon_2$ $\varepsilon_2 \sim N(0,\sigma^2_2)$
- $Y=\gamma X'+\varepsilon$
- $Y=\gamma(X+\varepsilon_2)+\varepsilon$

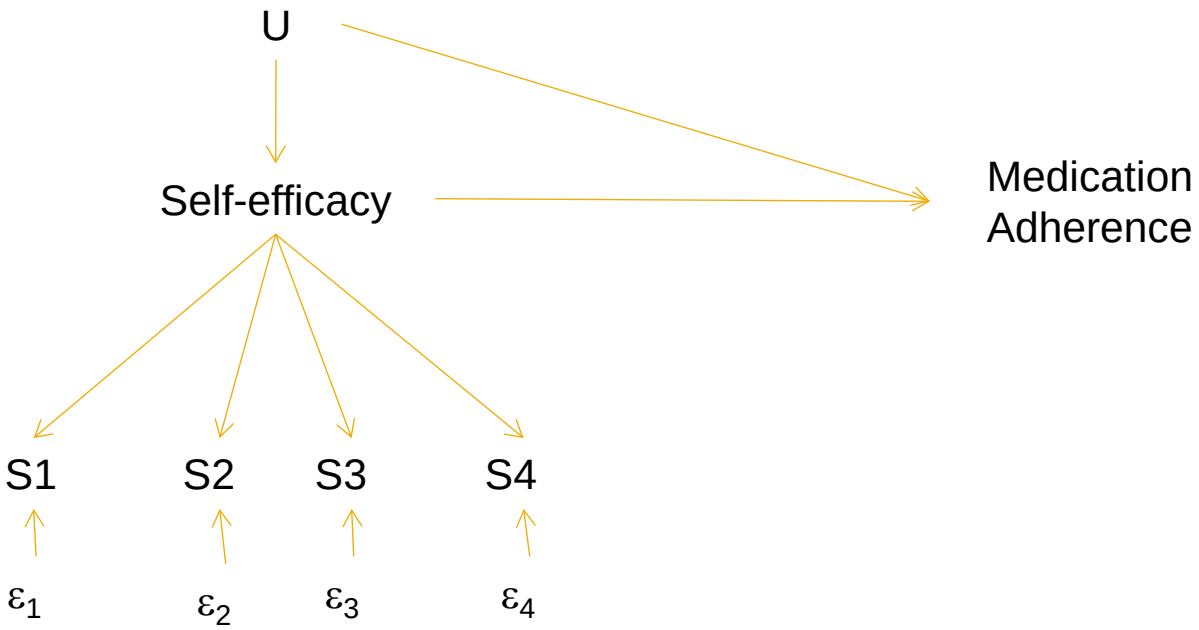
- How is γ (what you estimated) related to β (the truth)?

- $\gamma=\beta\sigma^2_x/(\sigma^2_{x+}\sigma^2_2)$

What you estimate is equal to the truth attenuated by the fraction of the variance of the measure that is due to the "true" underlying value (reliability).

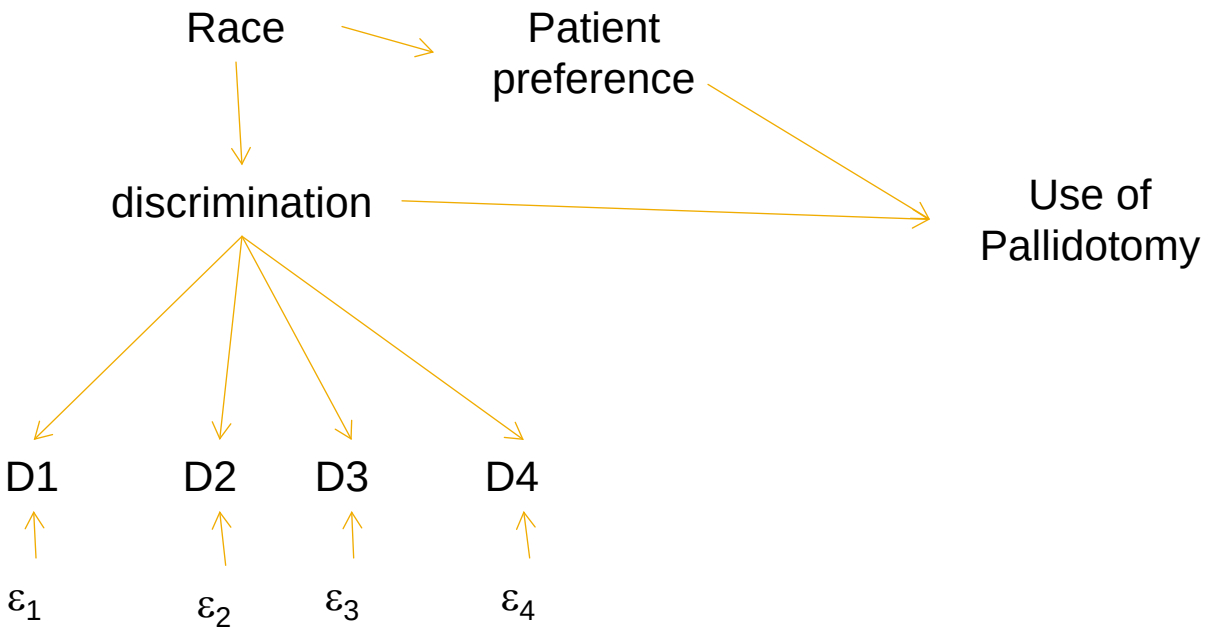
Unreliable measures in regressions

- What does this do to statistical power?
- What does measurement error in the dependent variable do?
- What does measurement error in confounders do?
- What does measurement error in mediators do?



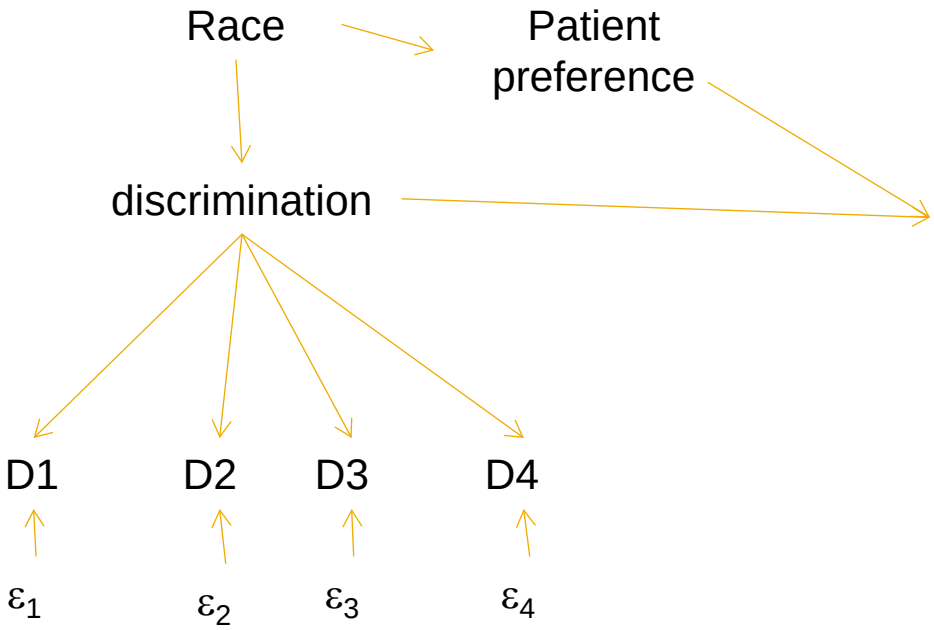
Unreliable measures in regressions

- What does measurement error in mediators do?
- Effects mediated by discrimination are clearly unjust, and constitute part of the disparity.



Unreliable measures in regressions

- What does measurement error in mediators do?
- Effects mediated by discrimination are clearly unjust, and constitute part of the disparity.



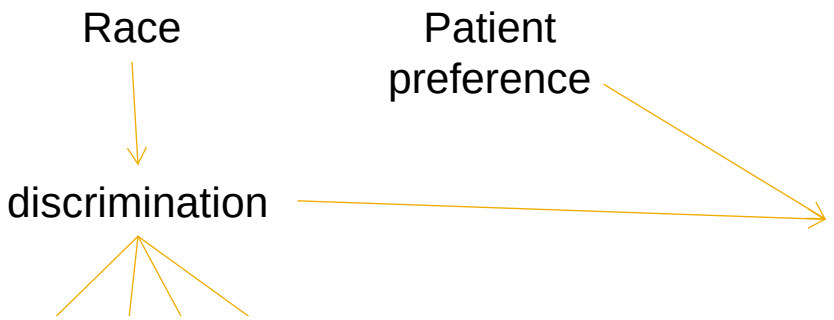
Commonly, no measure of patient preference, but an imperfect measure of discrimination, so test for "mediation" by discrimination

Compare association of Race and Pallidotomy with and without adjustment for discrimination measure (say, D-average)

Unreliable measures in regressions

- What does measurement error in mediators do?
- Effects mediated by discrimination are clearly unjust, and constitute part of the disparity.

Commonly, no measure of patient preference, but an imperfect measure of discrimination, so test for "mediation" by discrimination



Even if effect is FULLY MEDIATED by discrimination, measurement error in discrimination will leave an association between race and Pallidotomy.
Size of residual bias proportional to reliability of measure.

D

ε_1 ε_2 ε_3 ε_4

Discussion

1. Write a conceptual definition of perceived neighborhood gentrification
2. Write survey questions (interviewer- or self-administered) to measure the concept.
3. List possible approaches to validating the measure.
4. List possible approaches to assessing the reliability of the measure

Homework

Part 1:

1. Choose a paper describing the development or validation of a measure of relevance in health disparities research (please give the full citation).
2. What was the definition of the construct?
3. How did the authors provide evidence on the validity of the measure? Could you think of additional approaches to validating the measure?
4. How did the authors provide evidence on the reliability of the measure? Could you think of additional approaches to evaluating the reliability of the measure?
5. Describe the implications of a lack of measurement reliability for future research applications.

Part 2:

1. Find a paper describing a health disparity (please give the full citation)
2. Summarize the construct and measurement of the dimension of disparity (e.g., race, SES) and the outcome measured (e.g., self-rated health).
3. What is the evidence for the validity and reliability of the measures?
4. What is the reference category used for the disparity measure? Why does this reference category make sense (or not) for this research question?
5. How is the disparity quantified? Is this an absolute or relative measure or are both provided? Describe which type of measure you would prefer for this research area, or, if both, why.