Epi222: Measurement, an Abbreviated Lecture

Measurement Issues in Health Disparities and Social Epidemiology

- How do you measure the magnitude of disparities?
  1. Defining comparison groups
  2. Choosing the measure for a dimension of disparity
  3. Choosing the way to compare outcomes
  4. How the outcome influences your choices
- Challenges in measuring social constructs used in disparities research: What is the role of measurement in explaining disparities and how do insufficient measures compromise this?
- Some basic ideas for measuring latent variables

Measurement Issues in Health Disparities and Social Epidemiology

- How do you measure disparities?
  1. Defining comparison groups
  2. Choosing the measure for a dimension of disparity
  3. Choosing the way to compare outcomes
  4. How the outcome influences your choices
- 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

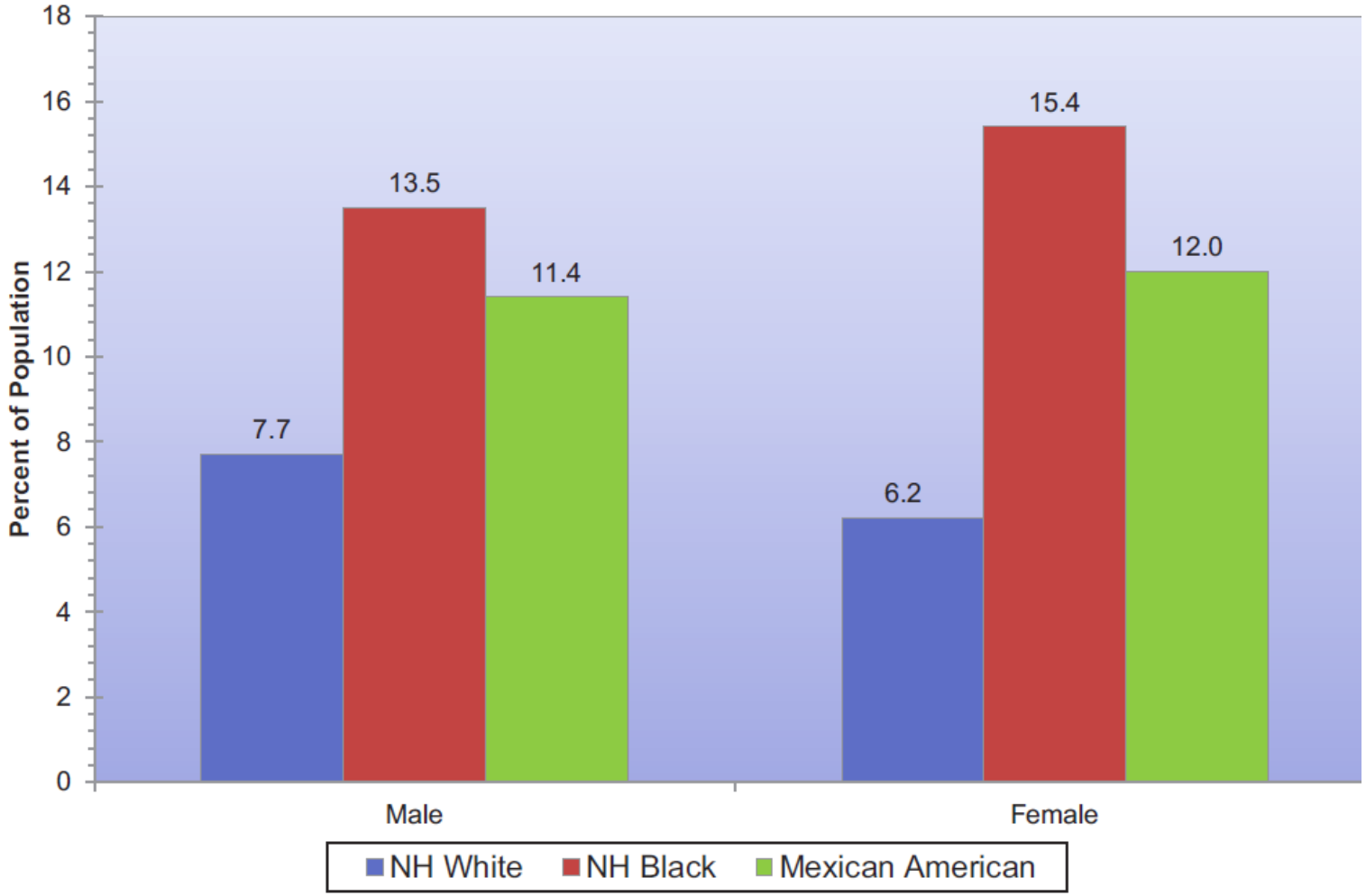
Measuring Disparities

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

Measuring disparities entails:

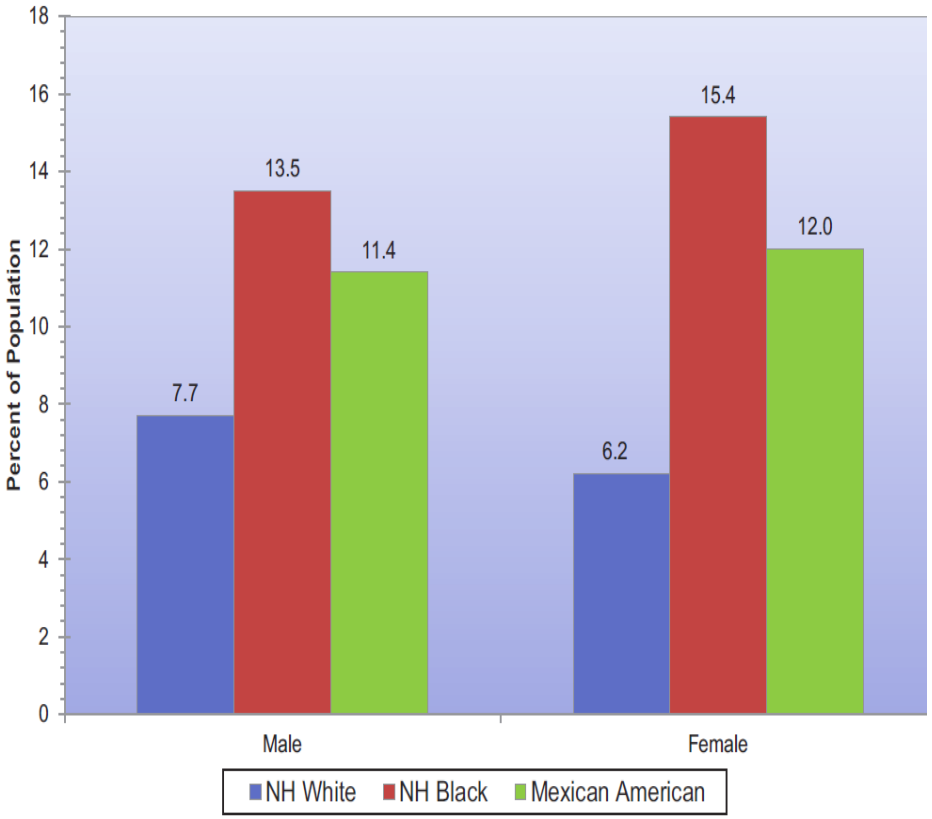
1. Choosing groups to compare
2. Showing a between-group health *difference*
3. Providing evidence that the difference is unjust (e.g., it is preventable, it arises from long-standing social disadvantages or historical power structures).

Measuring Disparities: Choosing a Comparison Group



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

Measuring Disparities : Choosing a Comparison Group



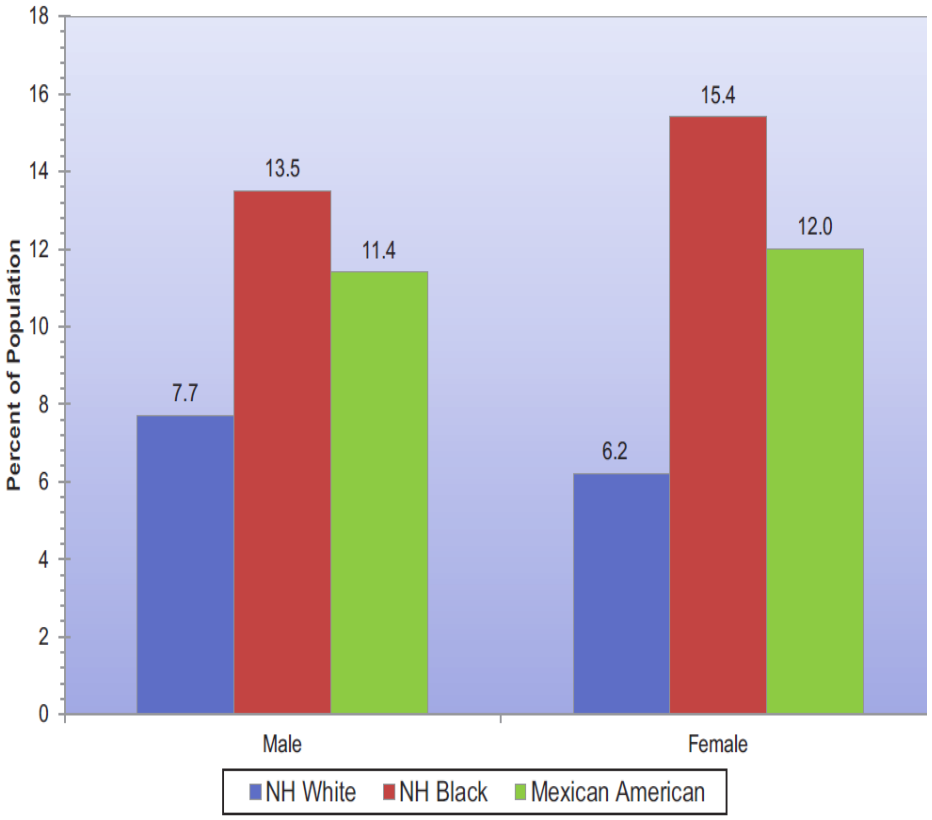
Men:
NH White Male: 7.7%
NH Black Male: 13.5%
Mexican American: 11.4%

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 (socially) privileged
2) Use the largest group as the comparator because they are most precisely estimated (often whites)
3) Use the healthiest group as the comparator

Measuring Disparities : Choosing a Comparison Group



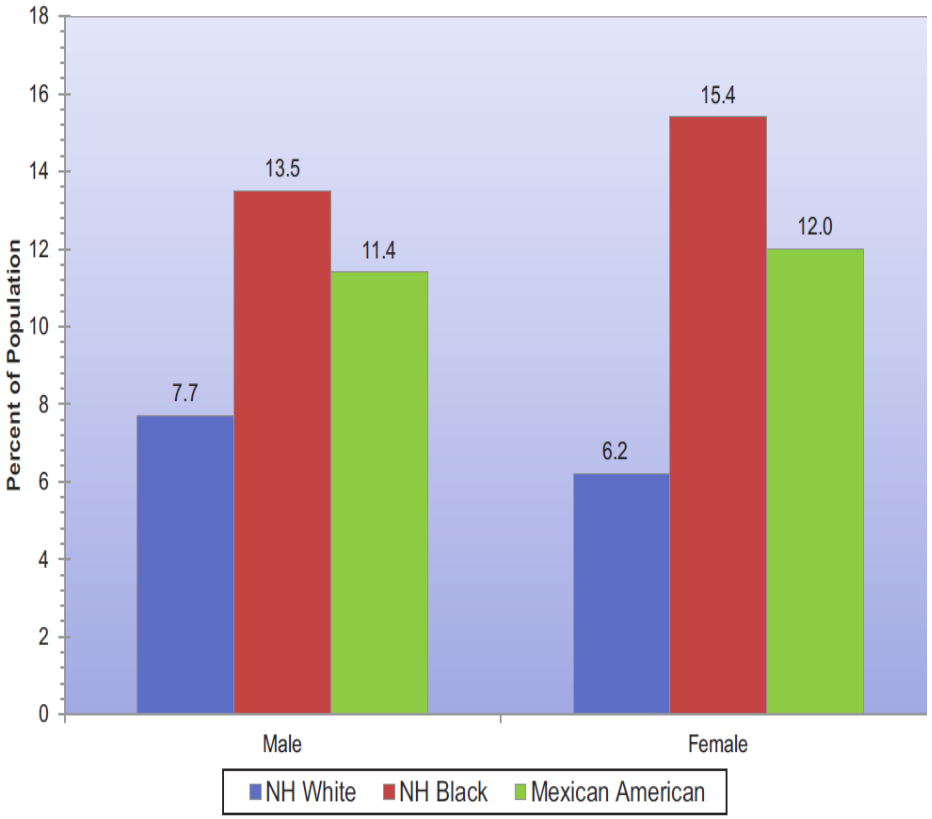
- Use the largest group as the comparator because they are most precisely estimated (usually whites)

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

Facilitates comparison across studies. Easier to argue this is an unjust difference if it is a disadvantage compared to the most socially privileged group.

Measurement Issues in Health Disparities and Social Epidemiology

- How do you measure disparities?
  1. Defining comparison groups
  2. Choosing the measure for a dimension of disparity
  3. Choosing the way to compare outcomes
  4. How the outcome influences your choices
- 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, SES
- SEP (assuming you want SEP not social *class)*:
  - define dimension (education, occupation, financial)
  - define time in life (childhood, current, cumulative)
  - define level (own SEP, household, neighborhood)
  - Choose absolute vs relative position (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 (e.g., financial security) and how you will measure that construct (e.g., household income ratio to poverty threshold)

How do you choose a measure?

- First ask: Are you conducting etiologic/intervention research or surveillance/descriptive?
- If etiologic:
  - theoretically, what is most likely to be relevant to your outcome?
  - Articulate mechanisms by which SEP may influence your health outcome.
- 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? (e.g., explain racial disparities with SES)
  - More comprehensive measurement is valuable
- Is your goal to guide interventions?
  - Narrower, more clearly defined and specific measurements are desirable

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)

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	Mean Income (95% Confidence Interval), $
	Overall
	20 551 (19 868-21 234)
	25 945 (25 486-26 403)
	36 022 (35 613-36 432)
	41 857 (41 228-42 487)
	54 606 (54 064-55 149)
	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)

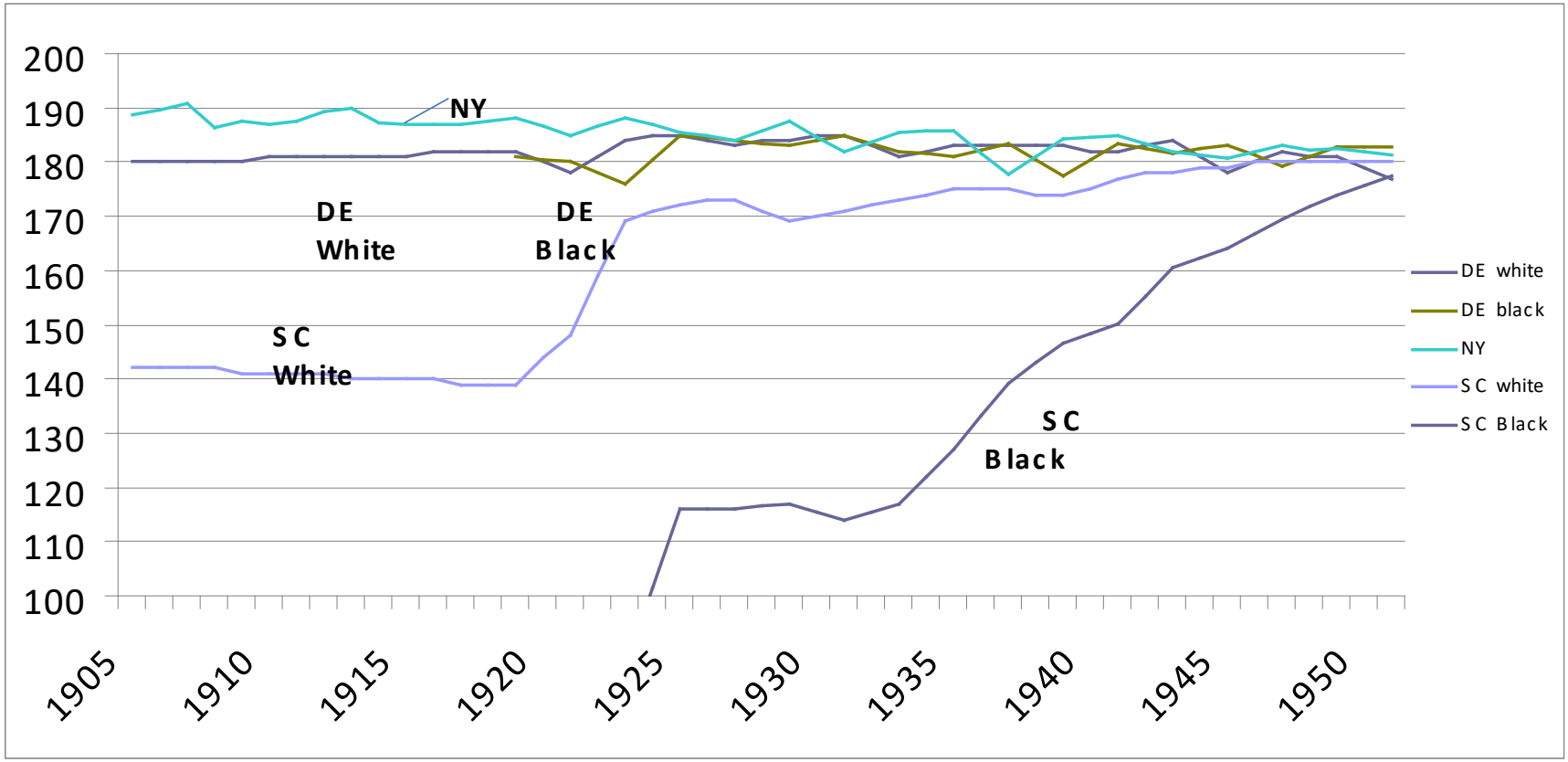
Educational Level, y	Mean Income (95% Confidence Interval), $		
	Black	Mexican American	White
<9	15 503 (14 851-16 154)	19 104 (17 549-20 659)	22 707 (21 935-23 479)
9-11	17 743 (17 051-18 434)	22 377 (20 807-23 948)	28 573 (28 062-29 083)
12	25 337 (24 529-26 144)	30 945 (29 791-32 098)	37 853 (37 400-38 305)
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

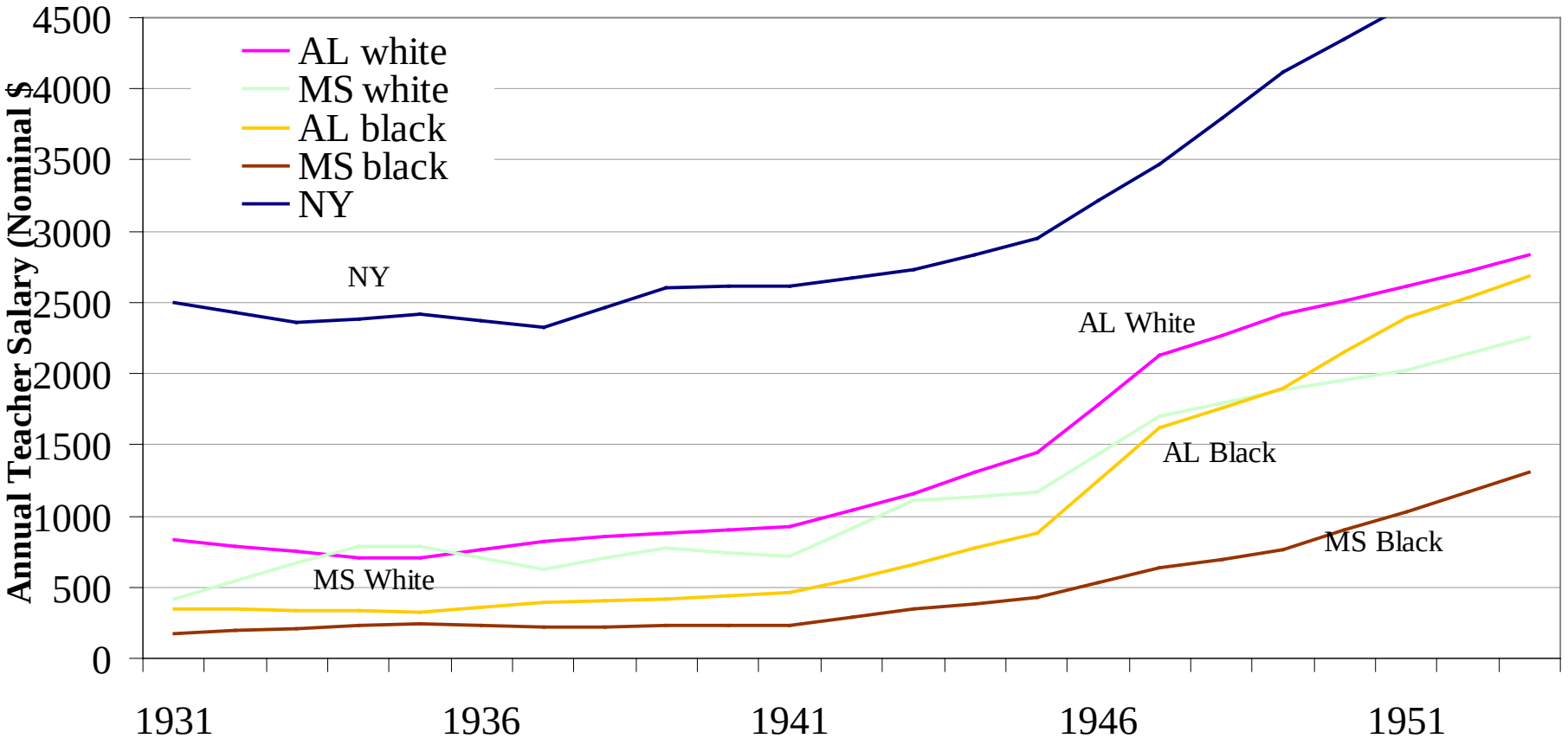
Changes in school term length (in days)



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



Measurement Issues in Health Disparities and Social Epidemiology

- How do you measure disparities?
  1. Defining comparison groups
  2. Choosing the measure for a dimension of disparity
  3. Choosing the way to compare outcomes
  4. How the outcome influences your choices
- 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 in health.

Note distinction from debate about absolute vs relative measures of social resources (the predictor variable)

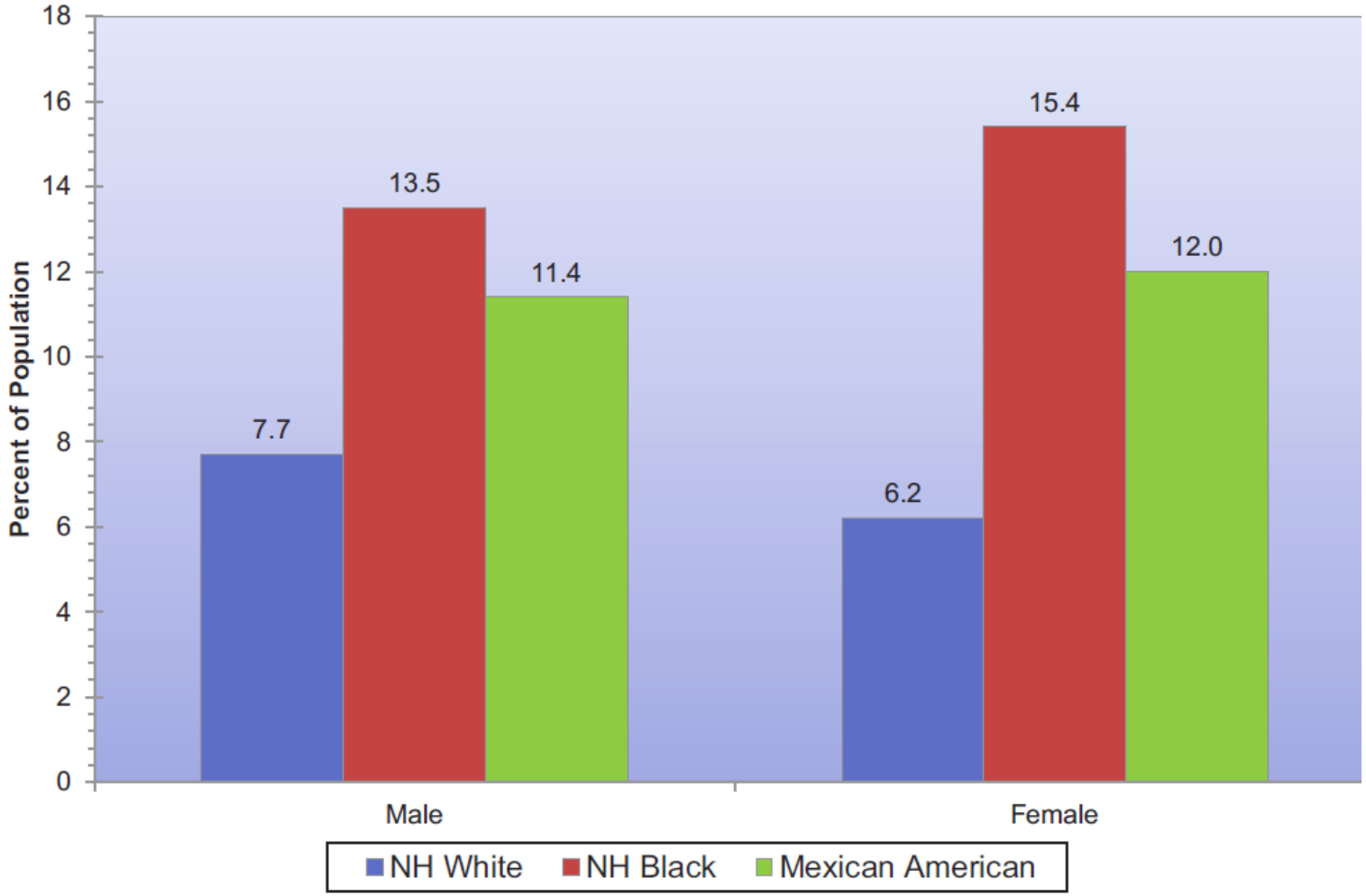
- Absolute: how much of the resource do you have?
- Relative: how much more or less than other people do you have?

Here we focus on absolute vs relative measures of health inequality (the outcome)

- Absolute: how many fewer cases per person
- Relative: ratio of rates or prevalence per person

This is controversial, and I have a strong personal opinion that is not universally accepted.

Measuring Disparities: Relative vs absolute comparisons of the groups?



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

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, or 75% higher prevalence.

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 rate
- Incidence rate ratio
- Standardized mortality ratio

ABSOLUTE MEASURES

- Risk difference
- Rate difference
- Years of life lost

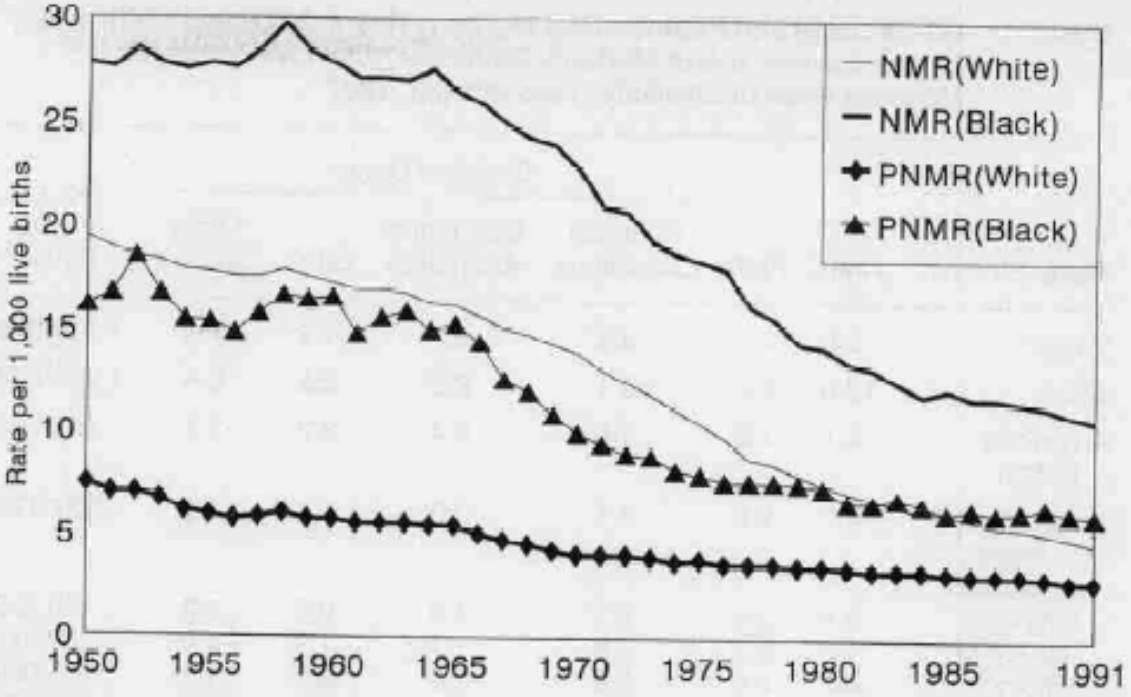
Relative measures are extremely common, so when you see one, stop and think “how big is this effect in absolute terms?”

Measuring Trends in Disparities: Relative vs Absolute Measures

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

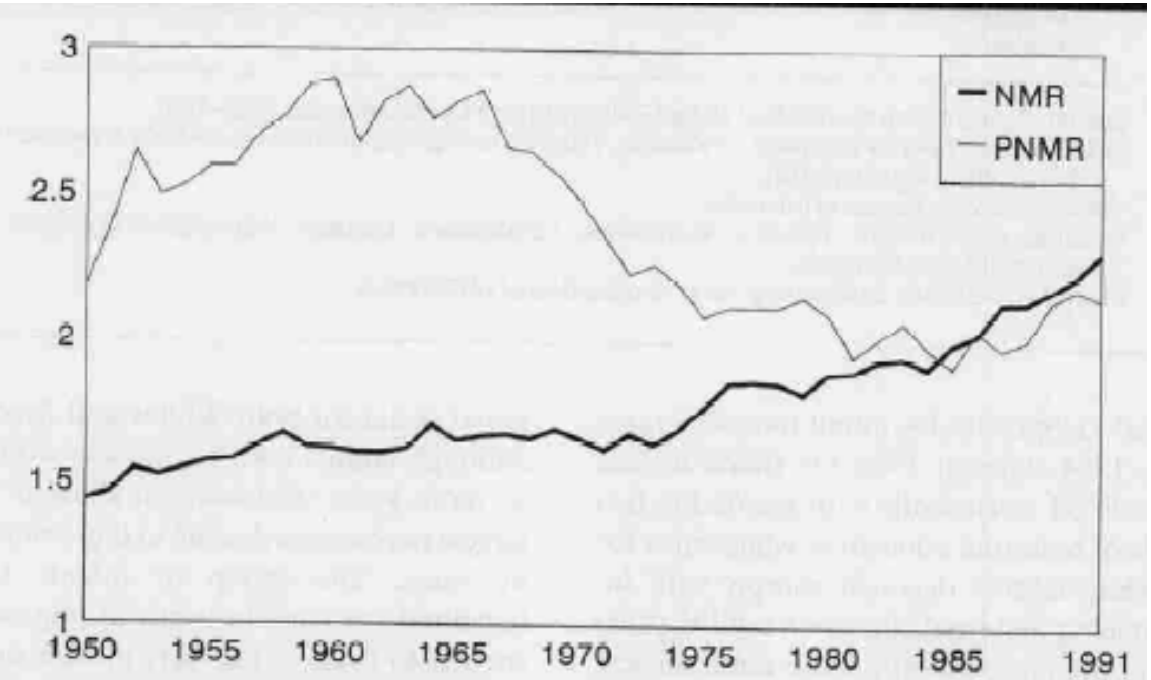


Note. NMR = neonatal mortality rate; PNMR = postneonatal mortality rate. Data from National Vital Statistics System, 1950 through 1991.

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

Singh and Yu, 1995

Trends in Infant Mortality



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

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 vs Interaction

- 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?

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 comparisons will always have the same sign (positive or negative) but estimates of **rate of change** in absolute and relative disparities may not have the same sign.

Slope index of inequality: Two ways to reduce inequality

- When there are multiple ordered levels, e.g., with education, how do you compare disparities between populations?
- Need to integrate information on
  - Prevalence of low vs medium vs high levels of education
  - Health difference between low vs medium and medium vs high levels of education
  - Is a society with only a small fraction of the population in the low education group more or less unequal than a society with a the same health "gap" between low and high education but a larger fraction of the population in the low education group?
- Slope index of inequality is the **absolute difference** in predicted values of a health indicator between those with the highest level of education and those with the lowest level of education or wealth
  - Takes into consideration the entire distribution of education using a regression model

Female mortality rate in Belgium is 5% lower than in the US (1989-2003). Why?

Education Level[a]	United States	Belgium
Percentage		
Low	20	67
Mid	61	19
High	19	14
Mortality rate per 100 000 person-years[b] (95% CI)		
Low	1023 (988, 1059)	801 (796, 806)
Mid	766 (745, 786)	628 (615, 641)
High	529 (495, 564)	582 (567, 597)
Total	806 (789, 823)	766 (761, 770)
Rate ratio (95% CI)		
Low	1.93 (1.80, 2.08)	1.38 (1.34, 1.41)
Mid	1.45 (1.35, 1.55)	1.08 (1.04, 1.11)
High (Ref)	1.00	1.00

Note. CI = confidence interval

Van Hedel, AJPH 2015

- More women in Belgium have low education than the US (67% vs 20%)
- But the mortality rate ratio associated wth low education in Belgium is smaller than in the US (1.38 vs 1.93)
- Population health differences be due to differences in exposure to adversity or differences in the effects of adversity

To calculate the slope index of inequality:

1) A weighted sample of the whole population is ranked from the most disadvantaged subgroup (at rank 0) to the most advantaged (at rank 1) according to, for example, education or wealth
2) The population of each education category is considered in terms of its range in the cumulative population distribution, and the midpoint of this range
3) The health indicator of interest is regressed against this midpoint value for education subgroups using an appropriate model (e.g. linear regression)
4) The predicted values of the health indicator are calculated for the two extremes (rank 1 and rank 0)
5) The difference between the predicted values at rank 1 and rank 0 (covering the entire distribution) generates the slope index of inequality value

Applied example: slope index of inequality

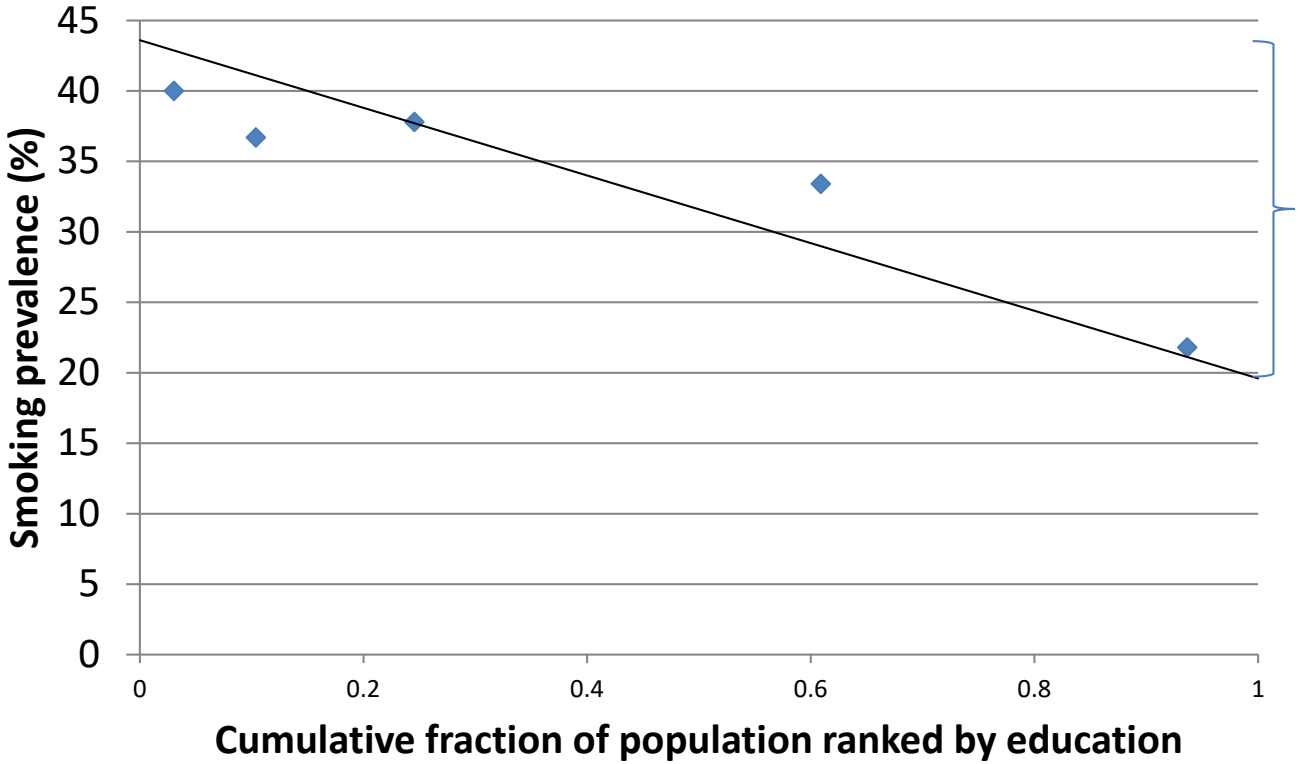
Step 1 for calculating SII based on education subgroups, for a population of men living in 27 middle-income countries and associated smoking prevalence, World Health Survey 2002–04

Education level	Proportional distribution of population	Cumulative range of population	Midpoint of cumulative range of population (x axis)	Smoking prevalence (%) (y axis)
No formal schooling	0.061	0–0.061	0.0305	40.0
Less than primary school	0.086	0.061–0.147	0.104	36.7
Primary school completed	0.198	0.147–0.345	0.246	37.8
Secondary/high school completed	0.529	0.345–0.873	0.609	33.4
College completed or more	0.127	0.873–1.000	0.937	21.8

Source: Data derived from Hosseinpoor AR et al. Socioeconomic inequalities in risk factors for noncommunicable diseases in low-income and middle-income countries: results from the World Health Survey. *BMC Public Health*, 2012, 12:912.

Applied example: slope index of inequality

Figure 1 Slope index of inequality: absolute inequality in smoking prevalence in a population of men living in 27 middle-income countries, World Health Survey 2002–04



RII = SII / population mean

Source: Data derived from Hosseinpoor AR et al. Socioeconomic inequalities in risk factors for noncommunicable diseases in low-income and middle-income countries: results from the World Health Survey. *BMC Public Health*, 2012, 12:912.

Measurement Issues in Health Disparities and Social Epidemiology

- How do you measure disparities?
  1. Defining comparison groups
  2. Choosing the measure for a dimension of disparity
  3. Choosing the way to compare outcomes
  4. How the outcome influences your choices
- 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

Considerations for outcome measures

- Example: Self-rated health
- Example: Stroke

Measurement Issues in Health Disparities and Social Epidemiology

- How do you measure disparities?
  1. Defining comparison groups
  2. Choosing the measure for a dimension of disparity
  3. Choosing the way to compare outcomes
  4. How the outcome influences your choices
- What is the role of measurement in explaining disparities and how do insufficient measures compromise this?
- 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 beneficial.

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?
- Some basic ideas for measuring latent variables
  - *With thanks throughout to Anita Stewart and Irene Yen*

What's a latent variable?

- 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
  - 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)
- Two basic goals: scaling and classification.

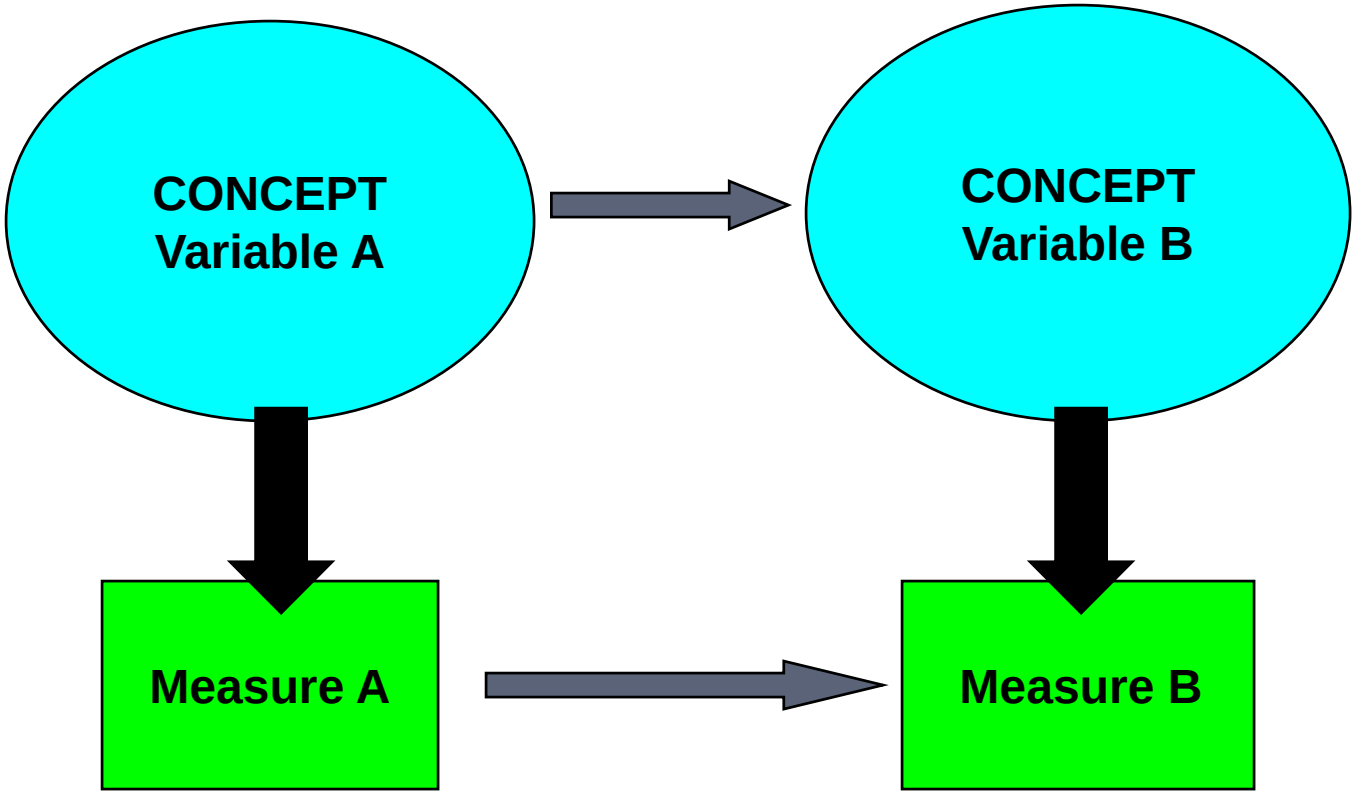
What's a latent variable or latent "construct"?

- 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)
- Two basic goals: scaling and classification.
- Often the greatest challenge is precisely defining the construct you wish to measure. Often it turns out that even well defined and popular measures do not clearly correspond with a well defined construct of scientific interest.
  - Gentrification, Segregation, Income inequality
  - Food environment
  - Racism
  - Memory, BMI, Depressive symptoms

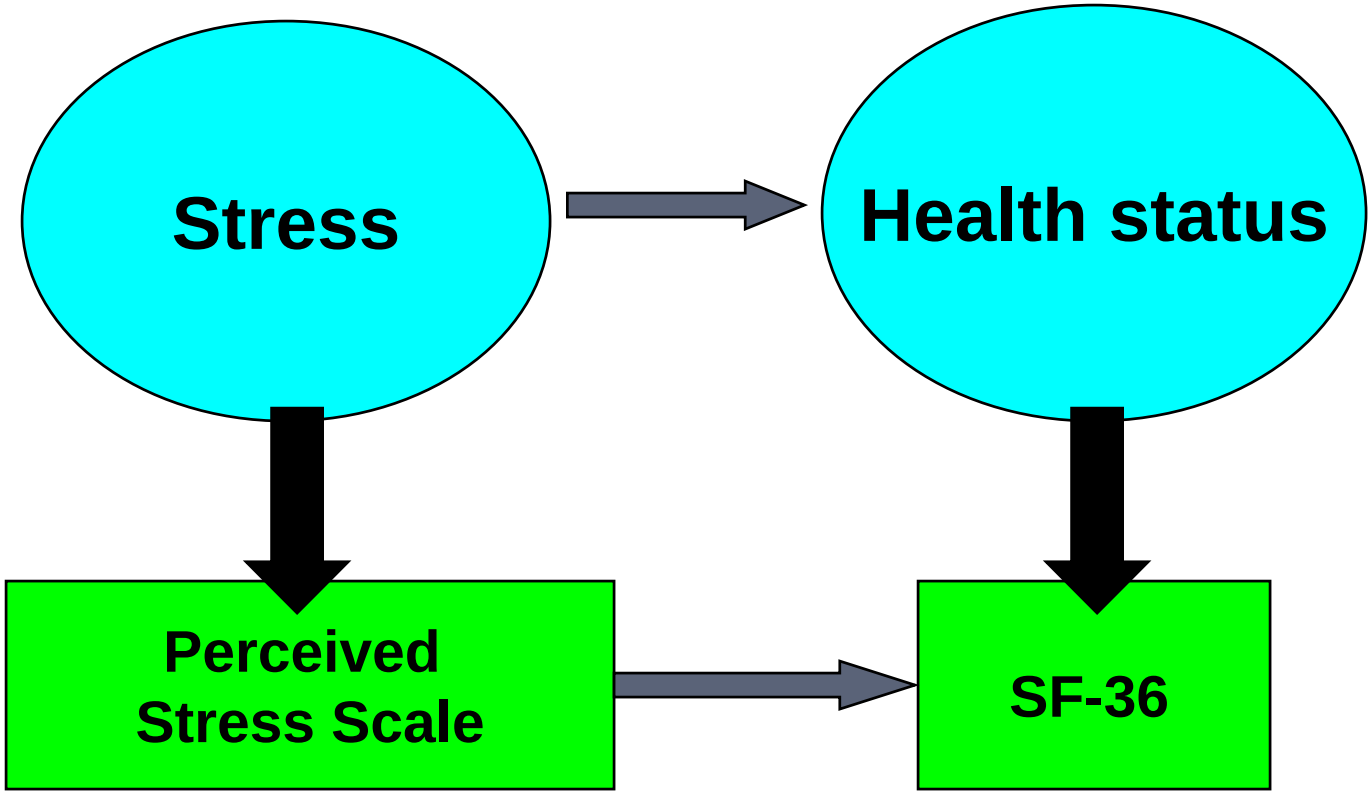
Measures of Concepts

- Concepts are defined and operationalized in terms of observed indicators
- Measures are "proxies" for the latent variables we cannot directly observe

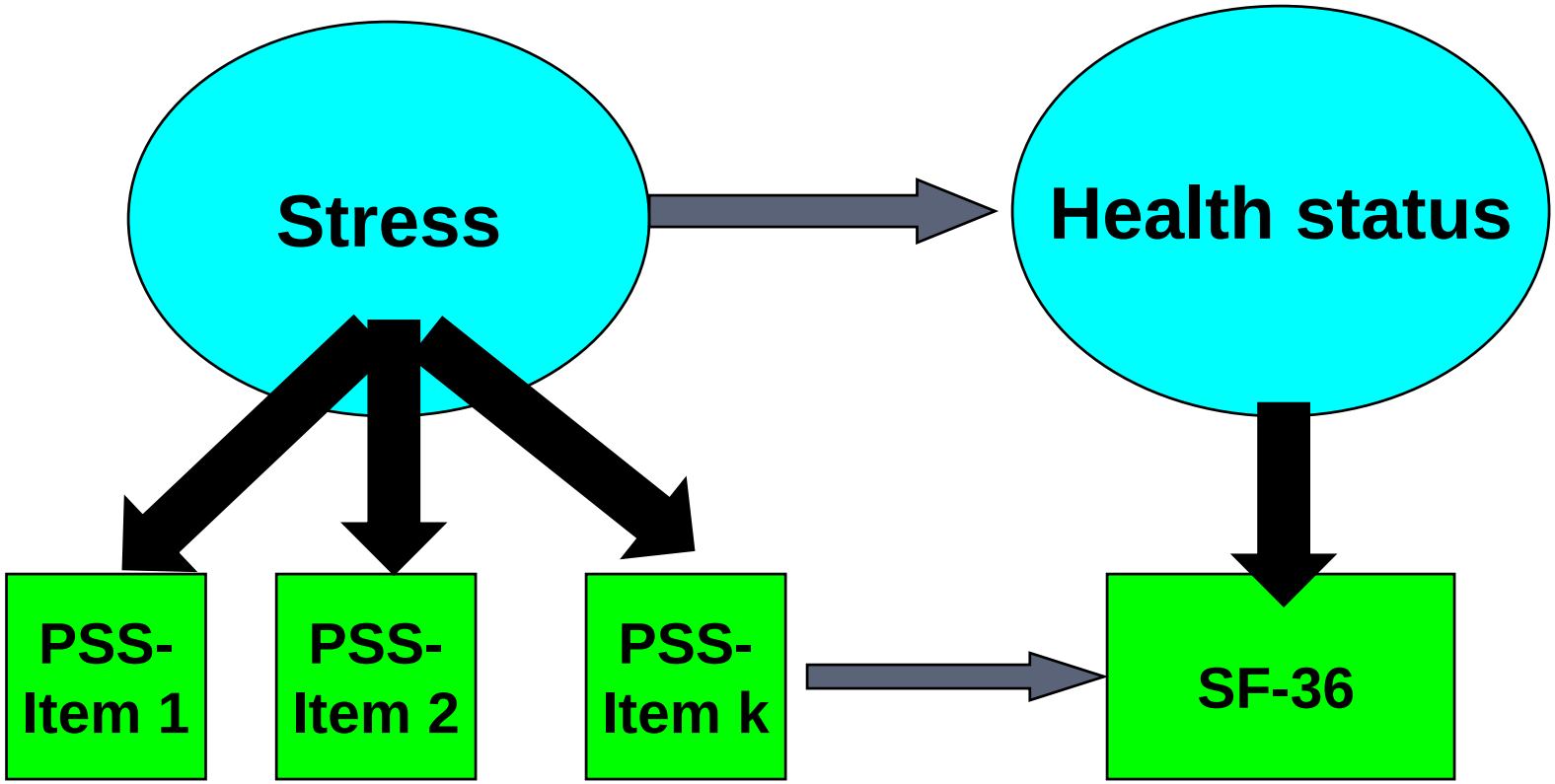
Depicting Latent Variables and Measures



Depicting Latent Variables and Measures



Depicting Latent Variables and Measures



Process of Selecting Good Measures for Your Studies

Define concept (variable)

Identify potential measures

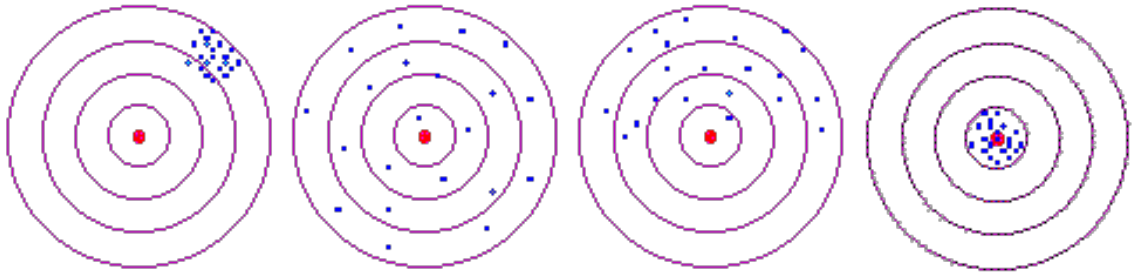
Review measures' properties
--conceptual adequacy
--psychometric adequacy

Pretest best 1-2 measures

Select final measure

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, precisely stated, construct
- Reliability: The extent to which a measure is precise; the proportion of variance in a measurement that is due to variation in the construct of interest.



Defining hard to describe constructs

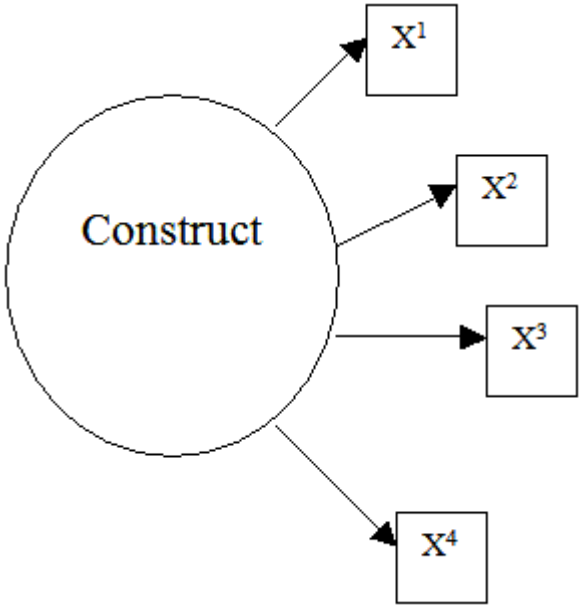
- State Role of Concept in Your Research
  - Evaluate intervention (outcome)
  - Determinant (e.g., social support) of an outcome
  - Predict future event (e.g. functioning to predict hospitalization)
  - Describe population
  - Covariate
- Define Concept for Your Research
  - Define it first from your 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

Scales vs Indexes (per the DeVellis framework)

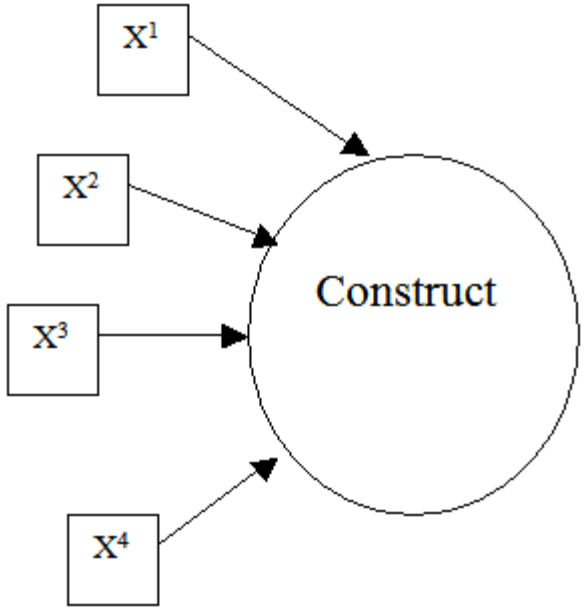
- A usual approach to measuring a latent variable is to develop several items. When considering such item pools, it is important to ask: is this a scale or an index?
- Items on a scale are thought to all have a common cause (the latent variable). They are assumed to correlate with one another because 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

Scales vs Indexes=Reflective vs Formative

- If you encounter the SEM or psych literature, they often use the terms formative and reflective, corresponding to scale vs index



Reflective: items reflect the underlying construct



Formative: items form the construct

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
- Sometimes you may have to develop a new measure or modify an existing measure.

Typical Sequence of Developing New Self-Report Measures

Develop/define concept

Create item pool

Pretest/revise

Field survey

Psychometric analyses

Final measures

Things to Pay Attention to in Items

- Item stem
  - Time frame
  - Complexity
  - Ambiguity or double barreled items (feel sad or lonely)
- Response scales
  - Type
  - Number of choices
  - Specific choices, distance between choices
  - Format
- Match of response choices to item stem

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?"

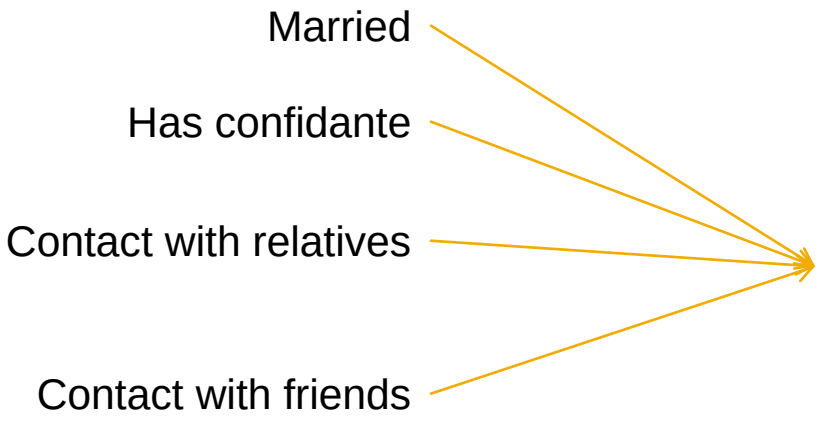
Self-rated health item:	Mean: current health scale		
	Screening N=~11,000		Baseline N=3,054
1 – poor	10.8	-- 20 --	10.8
2 – fair	30.0	-- 26 --	30.6
3 – good	57.6	-- 18 --	55.9
4 – very good	75.5	-- 11 --	75.4
5 – excellent	87.9		86.9

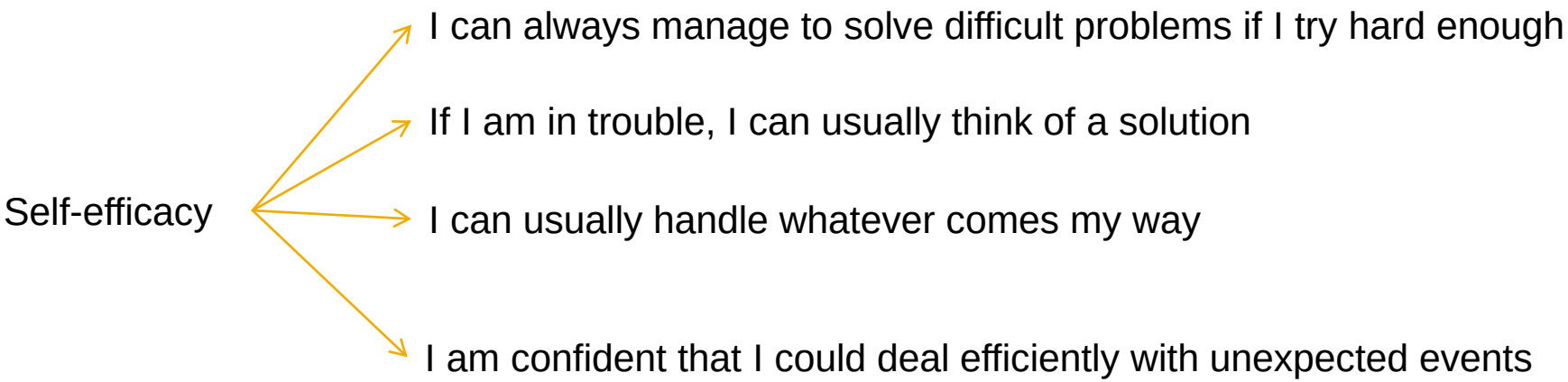
smallest diff

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

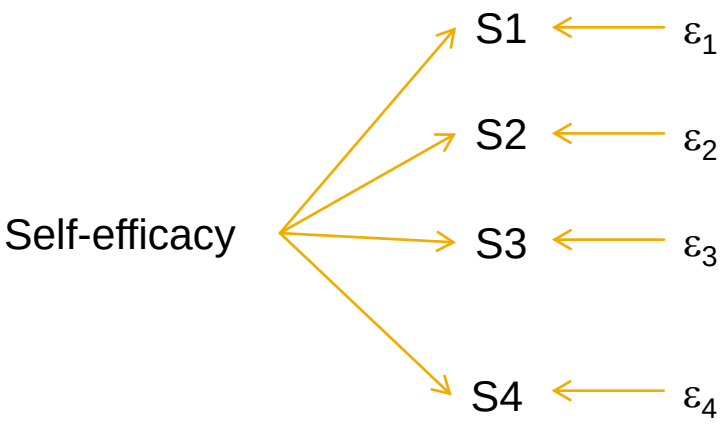
Scales vs Indexes (per the Devellis framework)





Why repeat basically the same question in different ways?

- Classical measurement theory:
  - The person's response to each item is the latent construct (efficacy) plus some item specific noise.
  - $S1 = b*\text{self-efficacy} + \varepsilon_1$
  - $S2 = b*\text{self-efficacy} + \varepsilon_2$
  - $S3 = b*\text{self-efficacy} + \varepsilon_3$
  - $S4 = b*\text{self-efficacy} + \varepsilon_4$



Classic Measurement Model Assumes:

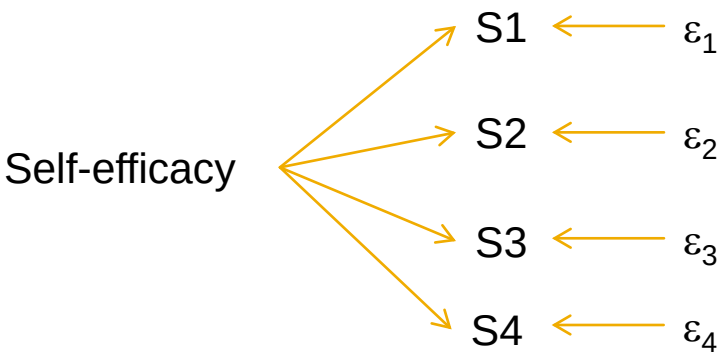
- Errors are random
- Errors are not correlated with each other
- Errors are not correlated with the true score
- Latent variable affects all items equally
- Variance of each error is the same

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

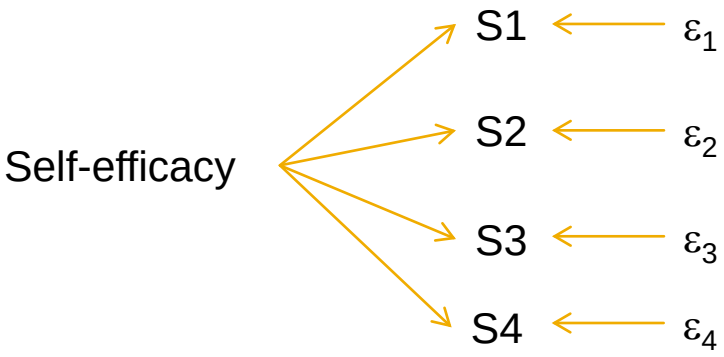


Adding items reduces measurement error and improves reliability

- 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~$N(0,\sigma^2)$
  - $\varepsilon \sim N(0,\delta^2)$
  - Variance of S1=$\sigma^2+\delta^2$ (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?
- S1+S2=(2*b*self-efficacy+$\varepsilon_1+\varepsilon_2$)/2
- Var (S1+S2)=Var(S1)+Var(S2)+2*Cov(S1,S2)
- The Cov(S1,S2) is entirely due to self-efficacy

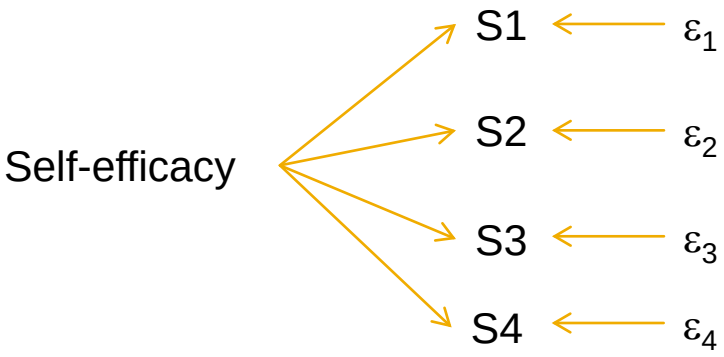
So self-efficacy contributes disproportionately.

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.

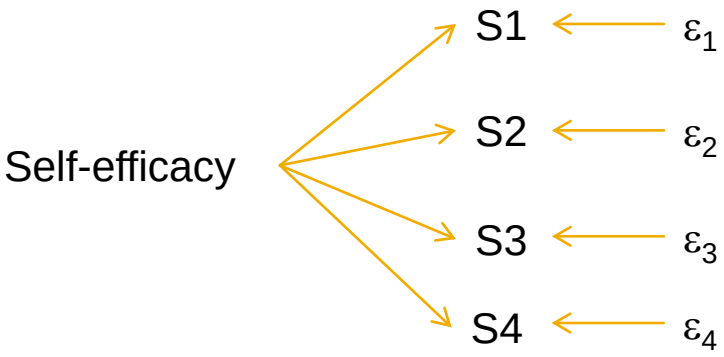
  - self-efficacy~$N(0,\sigma^2)$
  - $\varepsilon \sim N(0,\delta^2)$
  - Variance of S1=$\sigma^2+\delta^2$ (because ε_1 is independent of self-efficacy)
  - What % of the variance in S1 is due to self-efficacy? 50%



- Now calculate (S1+S2)/2
- What % of the variance in (S1+S2)/2 is due to self-efficacy?
- **Spearman-Brown:**
- **α=k*average inter-item correlation)/**
- **(1+(k-1)*average inter-item correlation**
- =2*.5/(1+1*.5)=1/1.5=.66

Adding items reduces measurement error and improves reliability

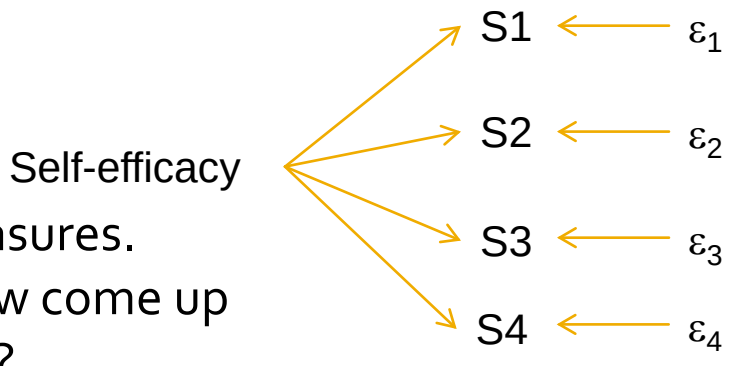
- Try it
  - set obs 1000
  - gen SE=invnorm(uniform())
  - gen S1=SE+invnorm(uniform())
  - gen S2=SE+invnorm(uniform())
  - gen S3=SE+invnorm(uniform())
  - gen S4=SE+invnorm(uniform())
  - gen S12_avg=(S1+S2)/2
  - sum
  - Corr



What if you add 3 items?
What if you make each item primarily due to noise?
eg
gen S1=SE+3*invnorm(uniform())

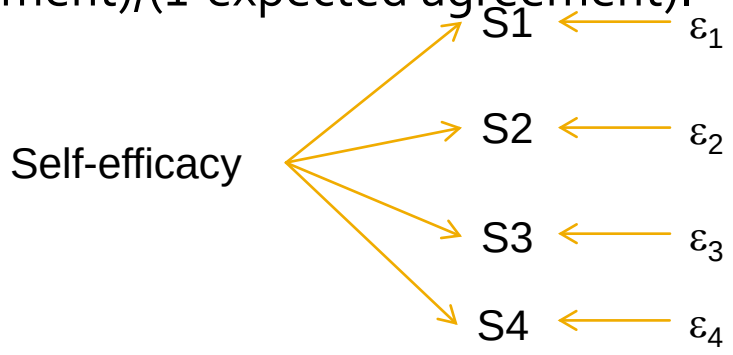
How can I assess reliability?

- 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 between two assessments (serious limitations)
- Coefficient of variation: within person SD/mean
- Bland Altman: deviations plotted against mean
- Kappa coefficients:
  - (observed agreement-expected agreement)/(1-expected agreement).
- 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 related to β?
- $\gamma=\beta\sigma^2_x/(\sigma^2_{x+}\sigma^2_2)$

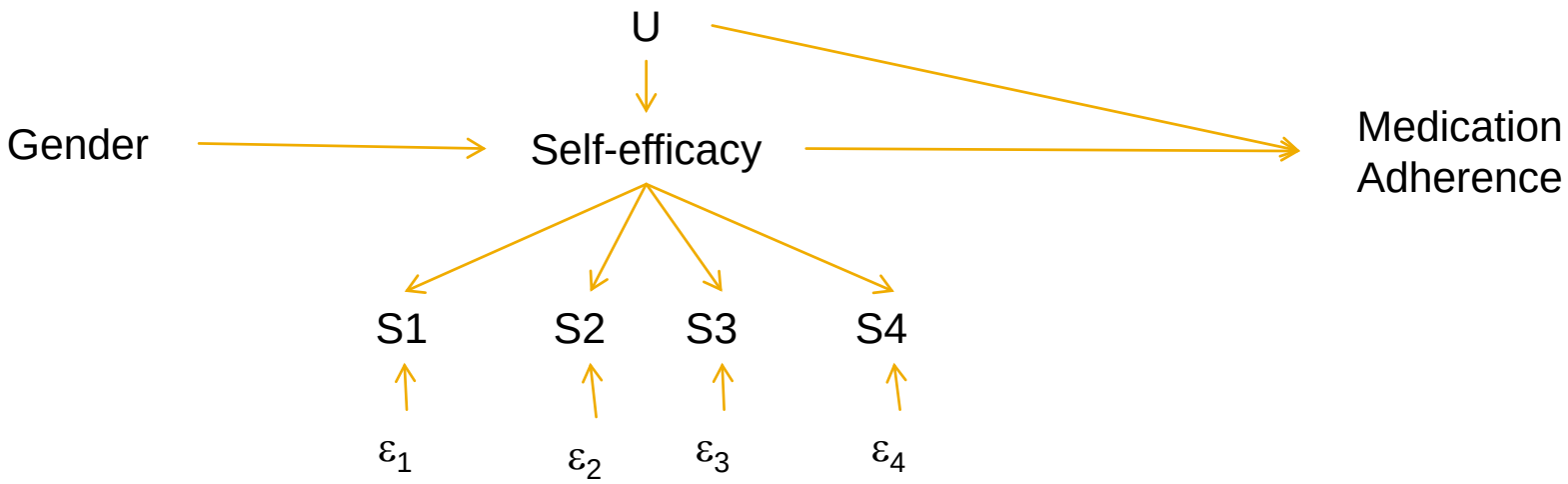
ie, attenuated in roportion to the % of variance in X' explained by X

Try it:

- gen y=2*SE+invnorm(uniform())
- reg y SE
- reg y S1
- reg y SE_hat

Unreliable measures in regressions

- What does this do to statistical power to evaluate effects of efficacy on adherence?
- What does measurement error in the dependent variable do, e.g., when evaluating effects of gender on self-efficacy??
- What does measurement error in confounders do?
- What does measurement error in a mediator do (eg when evaluating whether effect of gender on adherence is mediated by self-efficacy?)



Conclusion

- Measurement is a critical task in rigorous disparities research
- Good measurement starts with conceptual clarity
- Formal approaches to evaluate the validity and reliability of measurement instrument in your population
- Large evidence base to draw on, but there are always compromises

Group Activity

1. Break into at least 2 groups
2. Write a conceptual definition of perceived neighborhood gentrification
3. Write survey questions (interviewer- or self-administered) to measure the concept.
4. List possible approaches to validating the measure.
5. List possible approaches to assessing the reliability of the measure
6. Merge questions and try out with the class