



University of California
San Francisco

Basic measurement issues in health disparities research

Patience A. Afulani, MBChB, MPH, PhD

Assistant Professor, UCSF Department of Epidemiology & Biostatistics and
Department of Obstetrics, Gynecology & Reproductive Sciences

Institute for Global Health Sciences and Bixby Center for Global Reproductive Health

Introductions

Content credits

- Thanks to Maria Glymour

Why are we interested in measurement?

- Examine incidence and prevalence
 - Compare across settings and time
 - Examine determinants and consequences
 - Identify targets for improvement
 - Evaluate interventions
- **Identify and address disparities**

Learning objectives

1. Understand general issues in measuring the magnitude of disparities:
 - Defining the dimension of disparity
 - Choosing a measure for the dimension of disparity
 - Defining comparison groups
 - Choosing the outcome
 - Choosing the way to compare outcomes

2. Understand basic concepts in measuring social constructs in disparities research:
 - Distinguish between a construct and its measurement
 - Describe key steps in identifying/developing/adapting a measurement tool
 - Discuss how to assess the adequacy of a measurement tool

Measuring disparities

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)

Learning objectives

1. Understand general issues in measuring the magnitude of disparities:
 - **Defining the dimension of disparity**
 - Choosing the measure for a dimension of disparity
 - Defining comparison groups
 - Choosing the outcome
 - Choosing the way to compare outcomes
2. Understand basic concepts in measuring constructs in disparities research:
 - Distinguish between a social construct and its measurement
 - Describe key steps in identifying/developing/adapting a measurement tool
 - Discuss how to assess the adequacy of a measurement tool

Measuring disparities: Defining the dimension of inequality

- Commonly
 - Race / ethnicity
 - Socioeconomic position / social class

Race & ethnicity

- **Race:** a person's self-identification with one or more social groups. An individual can report as White, Black or African American, Asian, American Indian and Alaska Native, Native Hawaiian and Other Pacific Islander, or some other race. Survey respondents may report multiple races. (US Census Bureau)
- **Ethnicity:** determines whether a person is of Hispanic origin or not. For this reason, ethnicity is broken out in two categories, Hispanic or Latino and Not Hispanic or Latino. Hispanics may report as any race. (US Census Bureau)
- Race/Ethnicity in other contexts

Socioeconomic position 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”

Socioeconomic position

“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”

Defining the dimension of inequality

Dimensions of social class

- Race/ethnicity
- Socioeconomic status
- Immigration status
- Gender
- Sexual orientation
- Marital status
- Religion
- Caste

Dimensions of SES

- Education
- Financial (income vs wealth)
- Occupation

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—National Health Interview 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

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—National 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

Other considerations in defining the dimension of inequality for SES

- Define time in life (childhood, current, cumulative)
- Define level (own SES, household, neighborhood)
- Choose absolute vs relative position (working class/middle class, perceived social status)
- Distinguish between the construct (e.g., financial security) and how you will measure that construct (e.g., household income ratio to poverty threshold)

Learning objectives

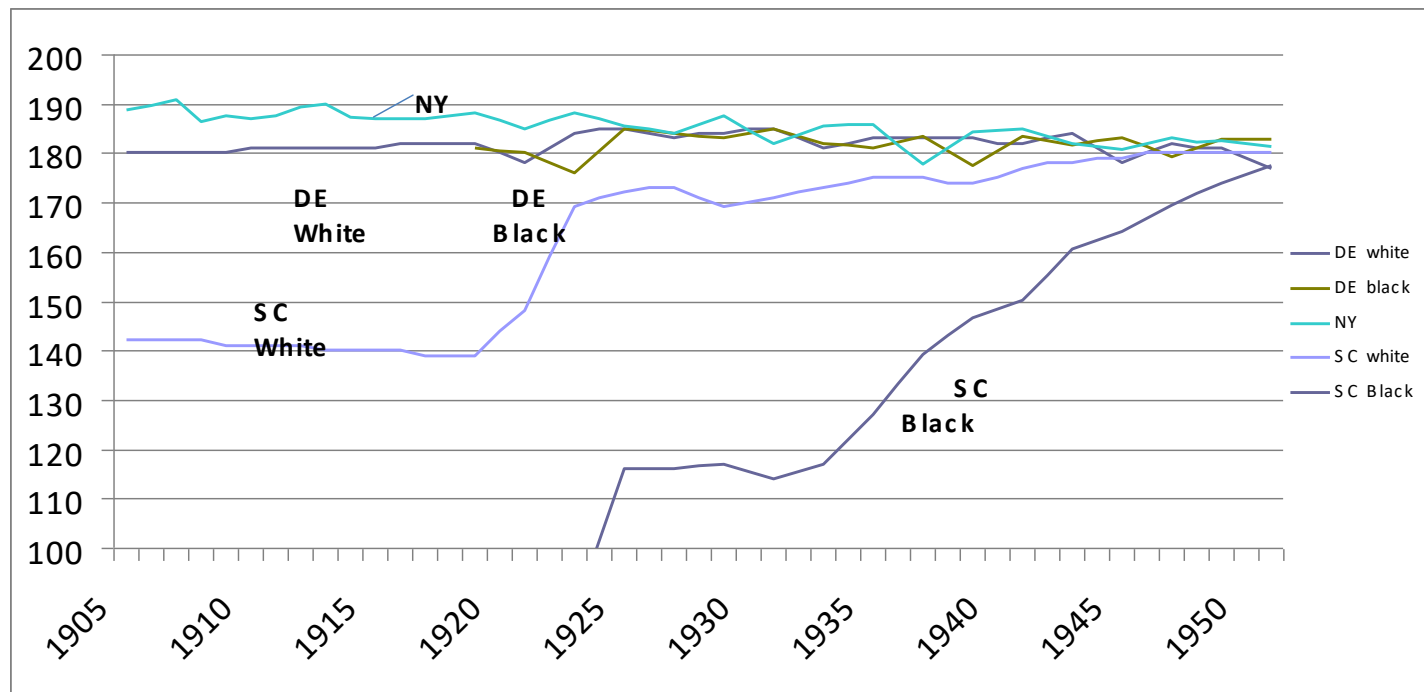
1. Understand general issues in measuring the magnitude of disparities
 - Defining the dimension of disparity:
 - **Choosing the measure for a dimension of disparity**
 - Defining comparison groups
 - Choosing the outcome
 - Choosing the way to compare outcomes
2. Understand basic concepts in measuring constructs in disparities research:
 - Distinguish between a social construct and its measurement
 - Describe key steps in developing/adapting a measurement tool
 - Discuss how to assess the adequacy of a measurement tool

Measures of SES

- Education
 - Years of formal education completed; certificate earned
- Income
 - Annual household and factor number of dependents, income to poverty ratio
- Wealth
 - Assets, Debt, Ownership (home, car)
- Occupation
 - Laborer, technical, professional, business, manual/non-manual

Even seemingly simple variables are hard to measure well: Education

Changes in school term length (in days)



Years of education $\neq$ 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

How do you choose a measure?

It depends on your research question

- What type of study you are conducting: etiologic/intervention research or surveillance/descriptive?
- If etiologic:
 - Theoretically, what is most likely to be relevant to your outcome?
 - Articulate mechanisms by which SES 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

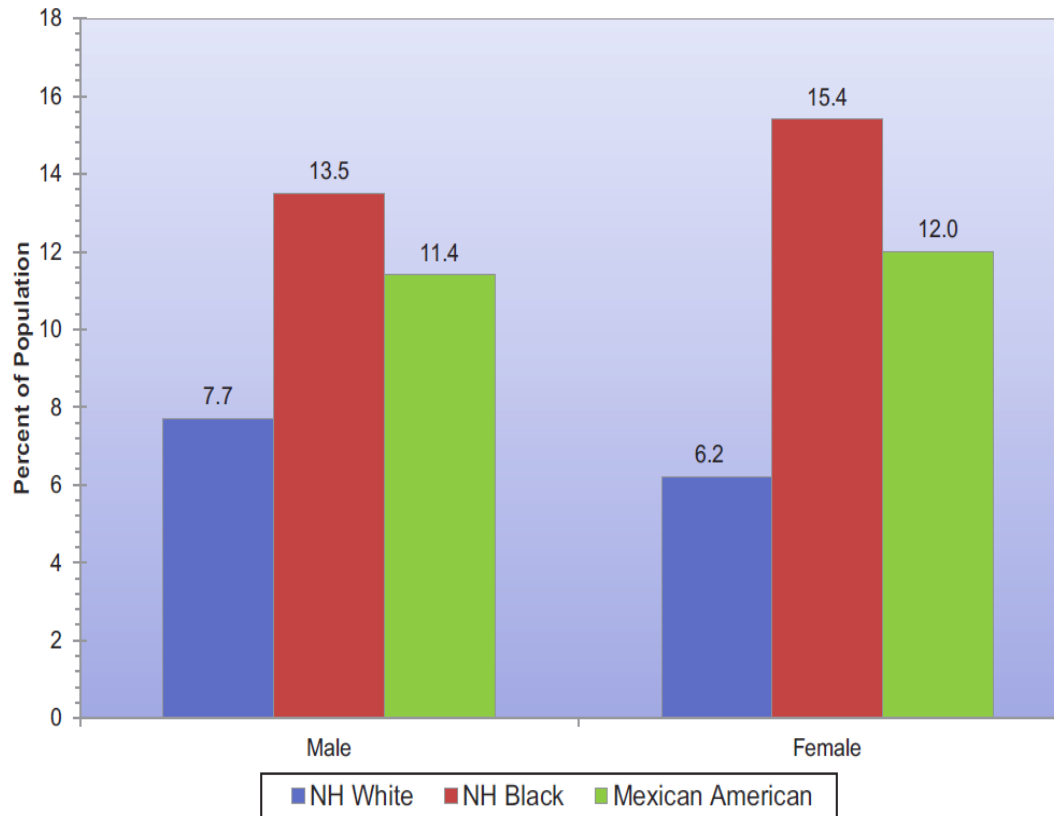
Learning objectives

1. Understand general issues in measuring the magnitude of disparities:
 - Defining the dimension of disparity
 - Choosing the measure for a dimension of disparity
 - **Defining comparison groups**
 - Choosing the outcome
 - Choosing the way to compare outcomes

2. Understand basic concepts in measuring constructs in disparities research:
 - Distinguish between a social construct and its measurement
 - Describe key steps in identifying/developing/adapting a measurement tool
 - Discuss how to assess the adequacy of a measurement tool

How do you decide on your comparison group?

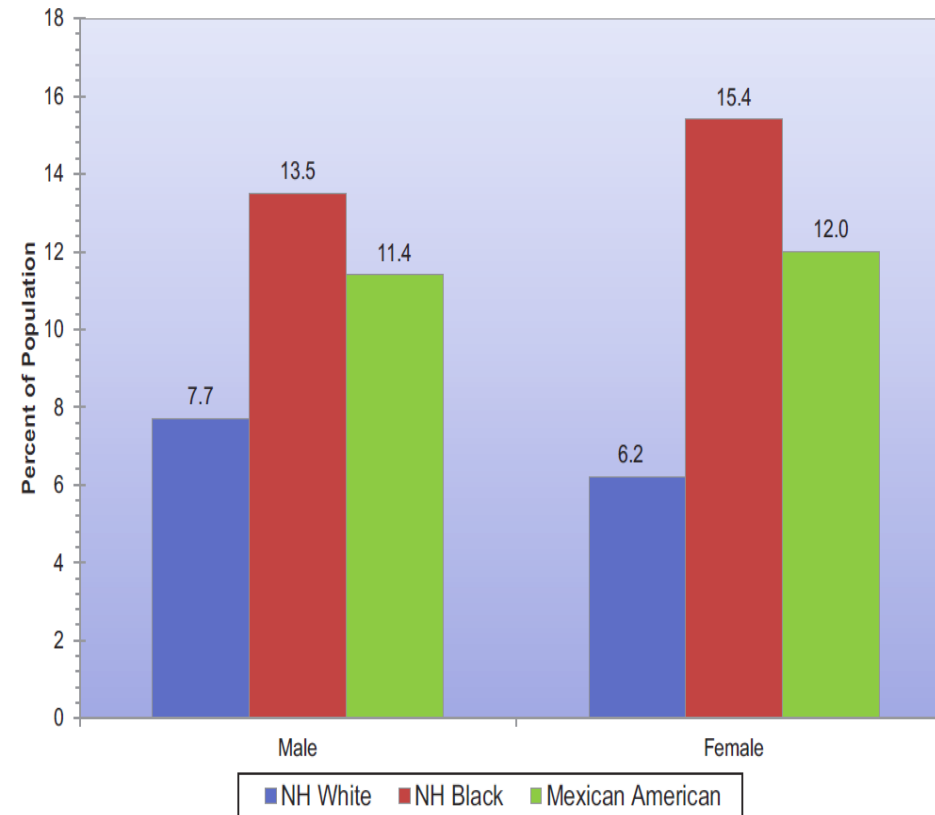
Measuring disparities: Choosing a comparison group



Which will you choose as your comparison group and why?

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).

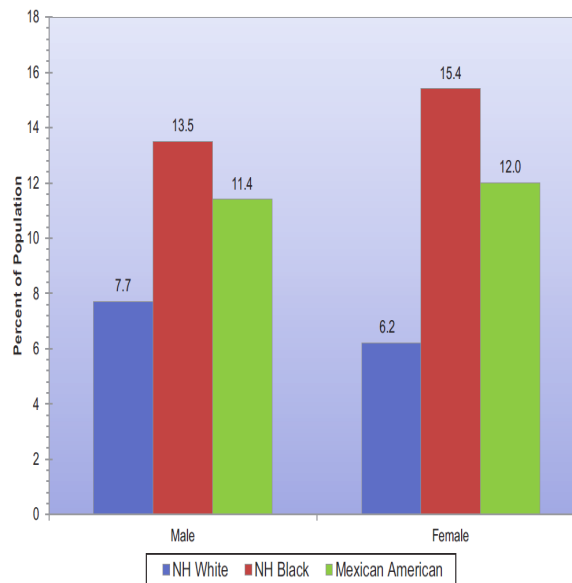
- Use the best off group as the comparator

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.

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

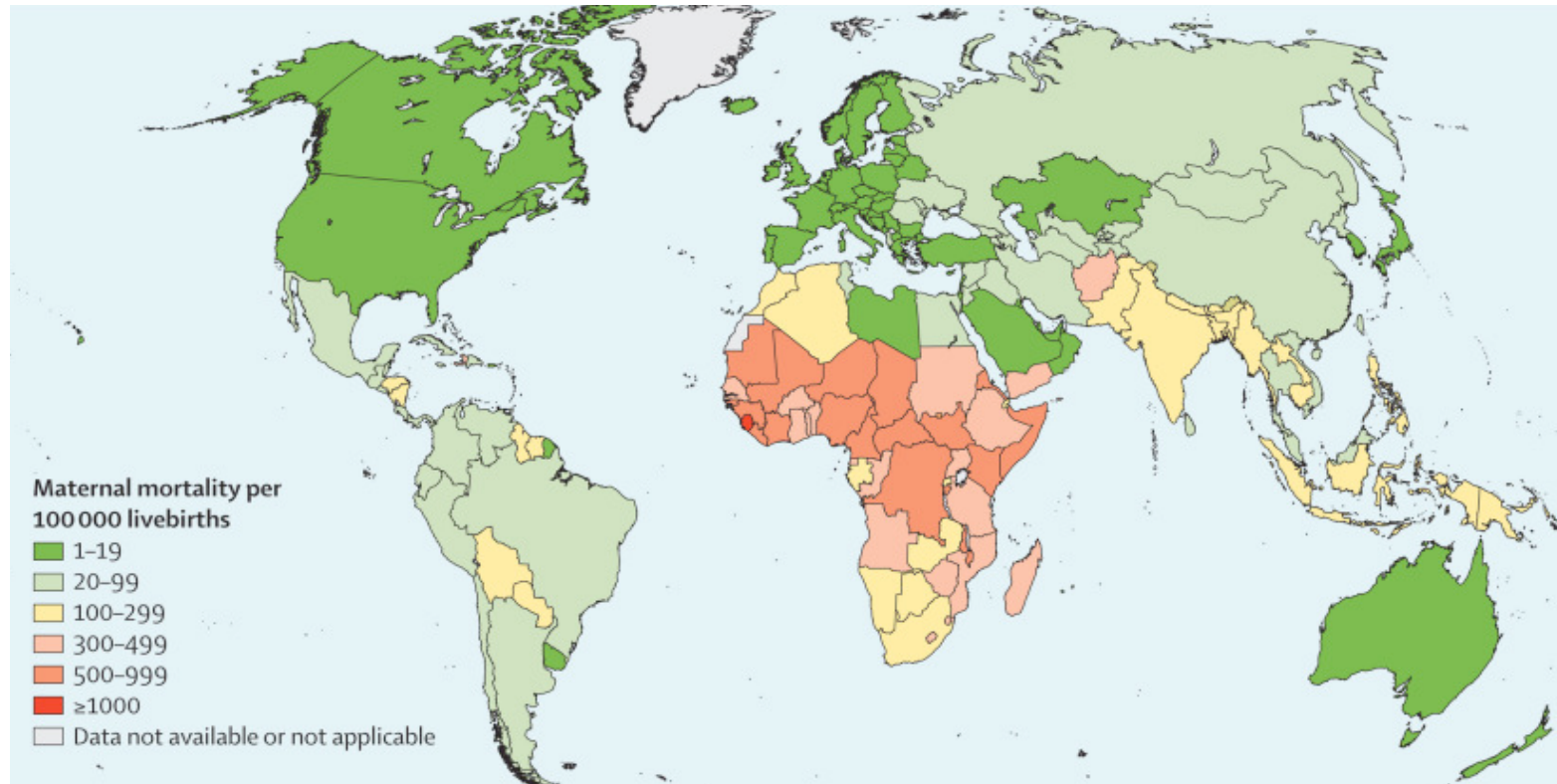
Facilitates comparison across studies. Easier to argue this is an unjust difference if it is a disadvantage compared to the most socially privileged group.

- What of comparing to the total?



Other disparity examples and choosing reference groups

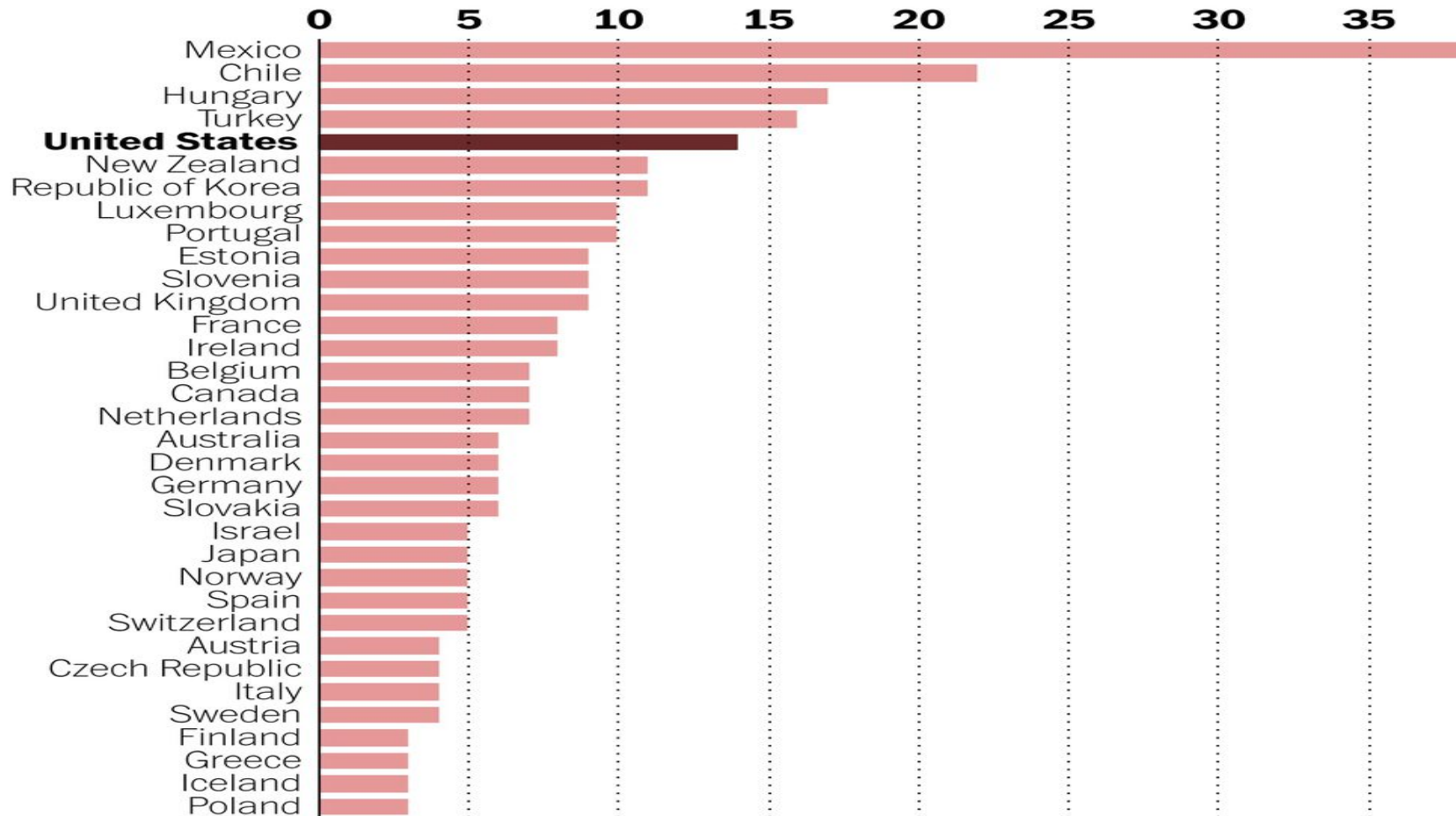
Maternal mortality



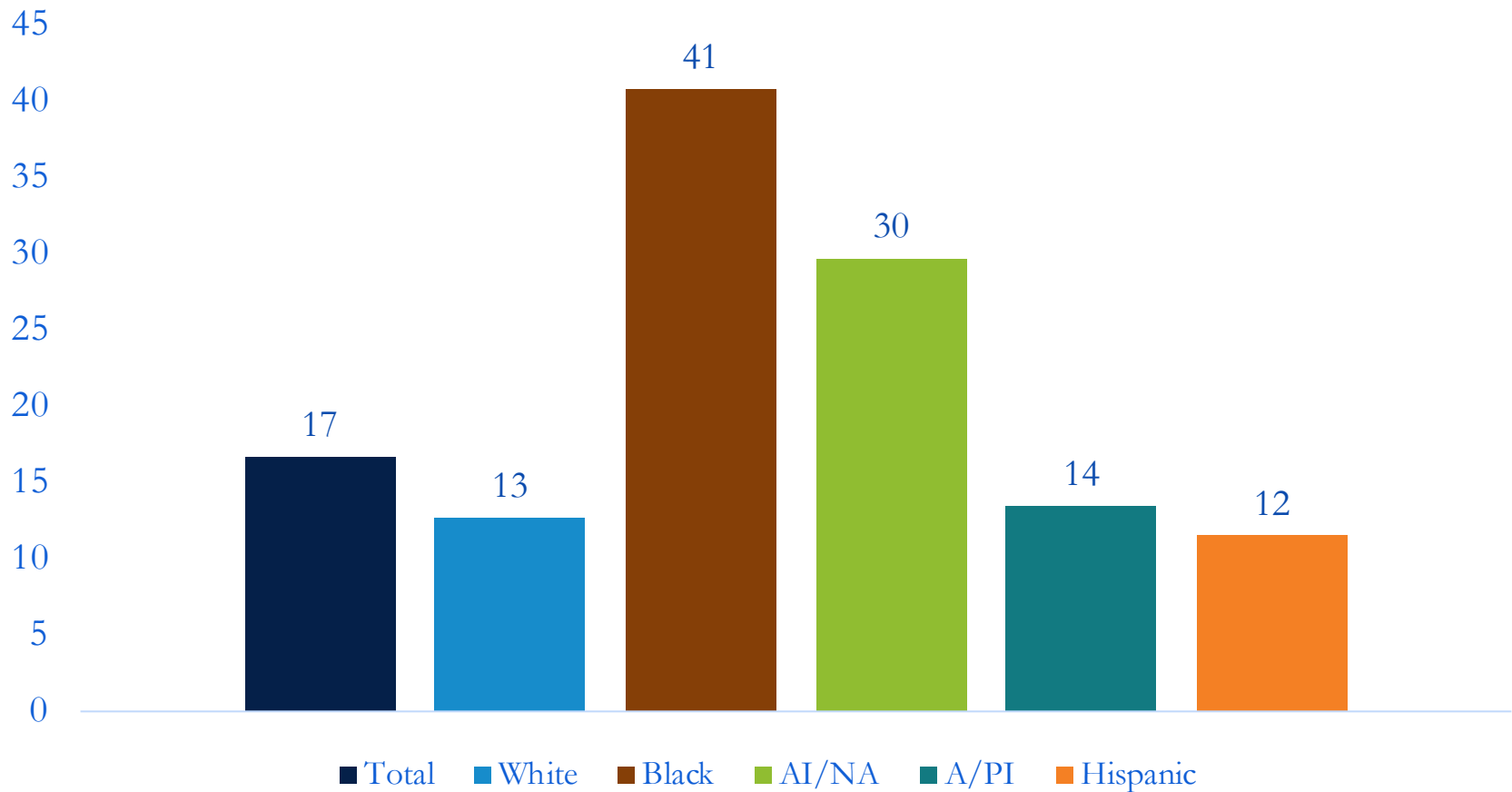
What does this hide?

U.S. lags behind other rich nations on maternal mortality

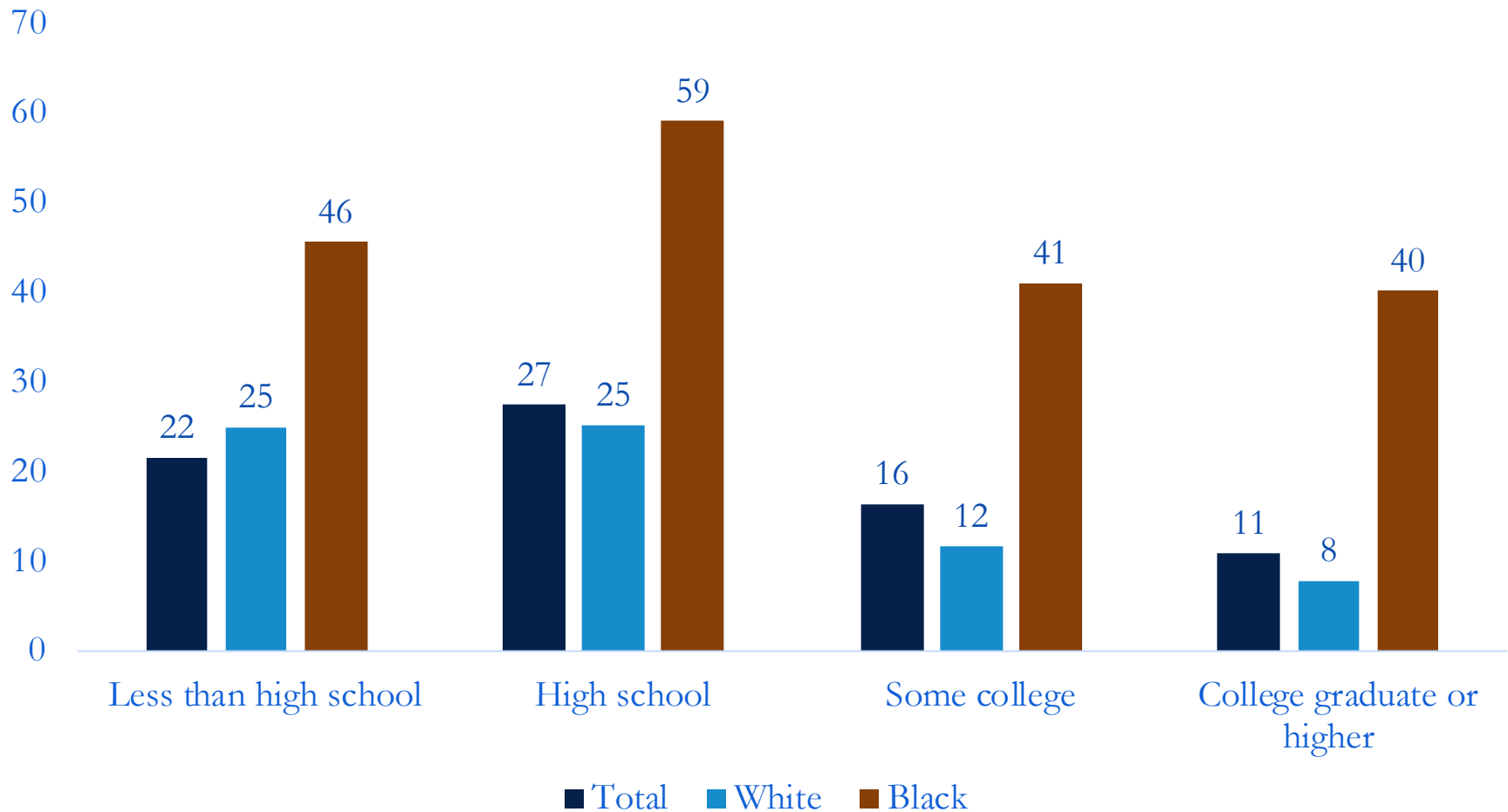
Maternal deaths due to pregnancy or labor complications per 100,000 live births, OECD countries



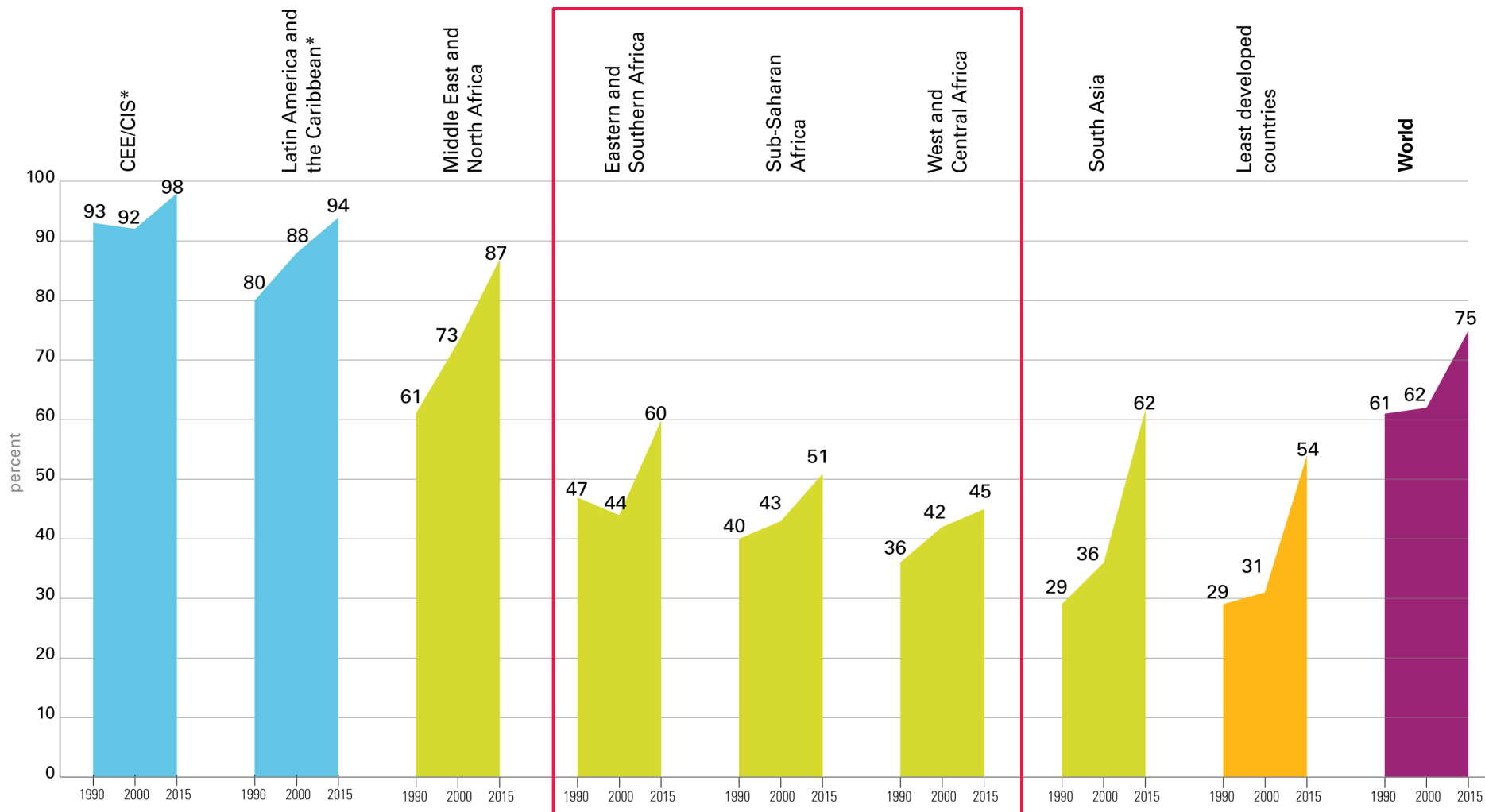
Racial/ethnic disparities in US maternal mortality



Racial/ethnic disparities in US maternal mortality by education

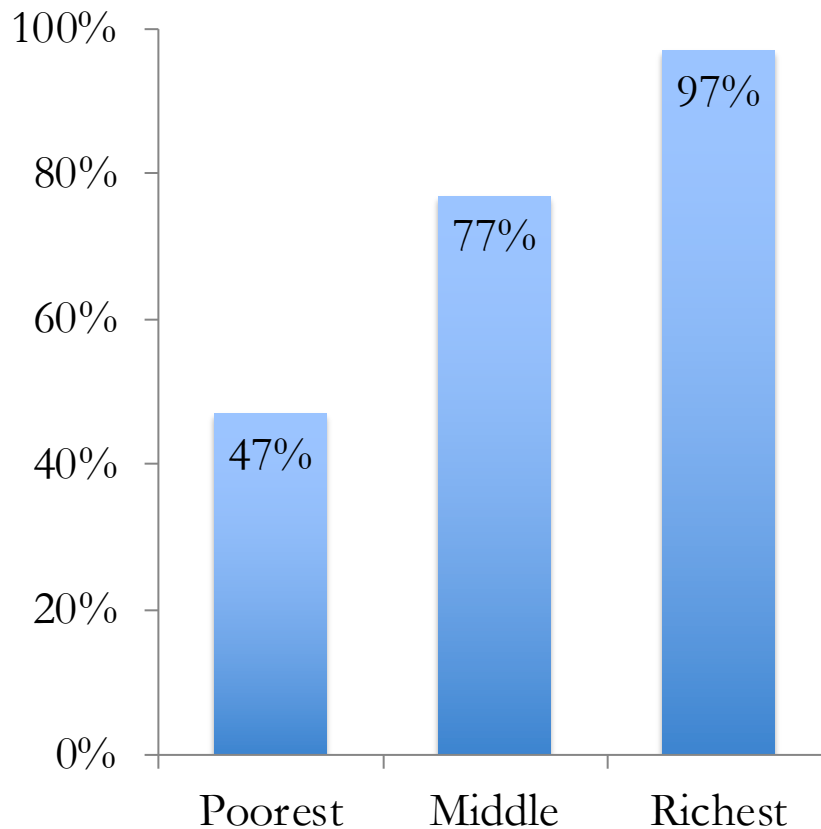


Births by skilled attendants globally

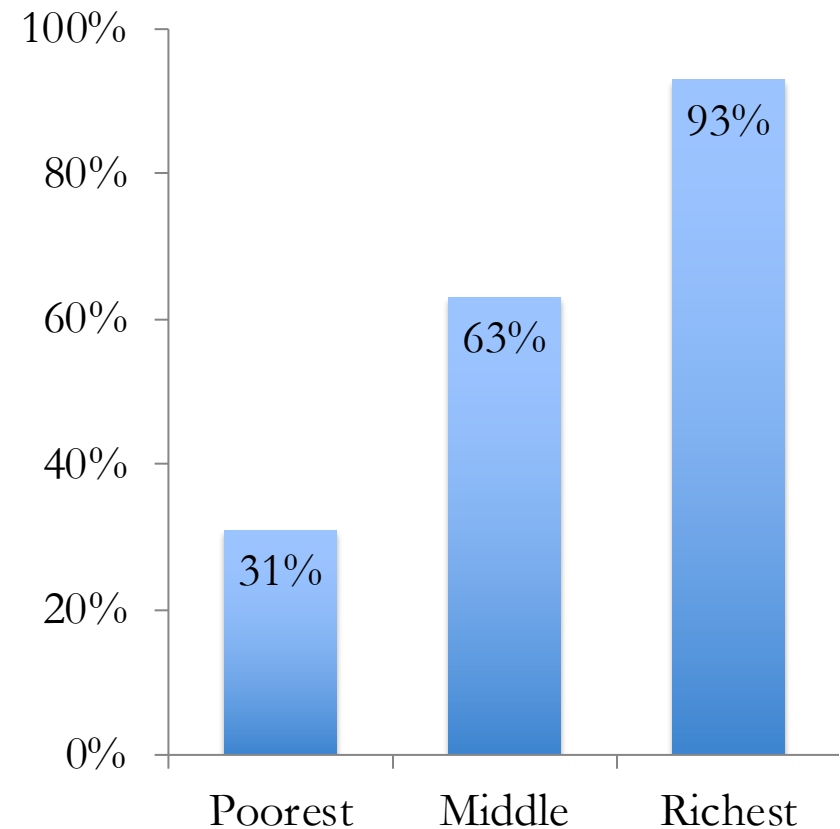


Births by skilled attendants by wealth quintile

Ghana



Kenya



Learning objectives

1. Understand general issues in measuring the magnitude of disparities:

- Defining the dimension of disparity
- Choosing the measure for a dimension of disparity
- Defining comparison groups
- **Choosing the outcome**
- Choosing the way to compare outcomes

2. Understand basic concepts in measuring social constructs in disparities research:

- Distinguish between a social construct and its measurement
- Describe key steps in identifying/developing/adapting a measurement tool
- Discuss how to assess the adequacy of a measurement tool

Choosing the outcome

- What outcome are you interested in?
- Distinguish between the construct (e.g., health) and how you will measure that construct (e.g., self-rated health, stroke, etc.)
- Note how the outcome influences your findings
- More on this later

Learning objectives

1. Understand general issues in measuring the magnitude of disparities:

- Defining the dimension of disparity
- Choosing the measure for a dimension of disparity
- Defining comparison groups
- Choosing the outcome
- **Choosing the way to compare outcomes**

2. Understand basic concepts in measuring social constructs in disparities research:

- Distinguish between a social construct and its measurement
- Describe key steps in identifying/developing/adapting a measurement tool
- Discuss how to assess the adequacy of a measurement tool

Measuring disparities: Showing a difference between groups

- *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

Measuring disparities: Showing a difference between groups

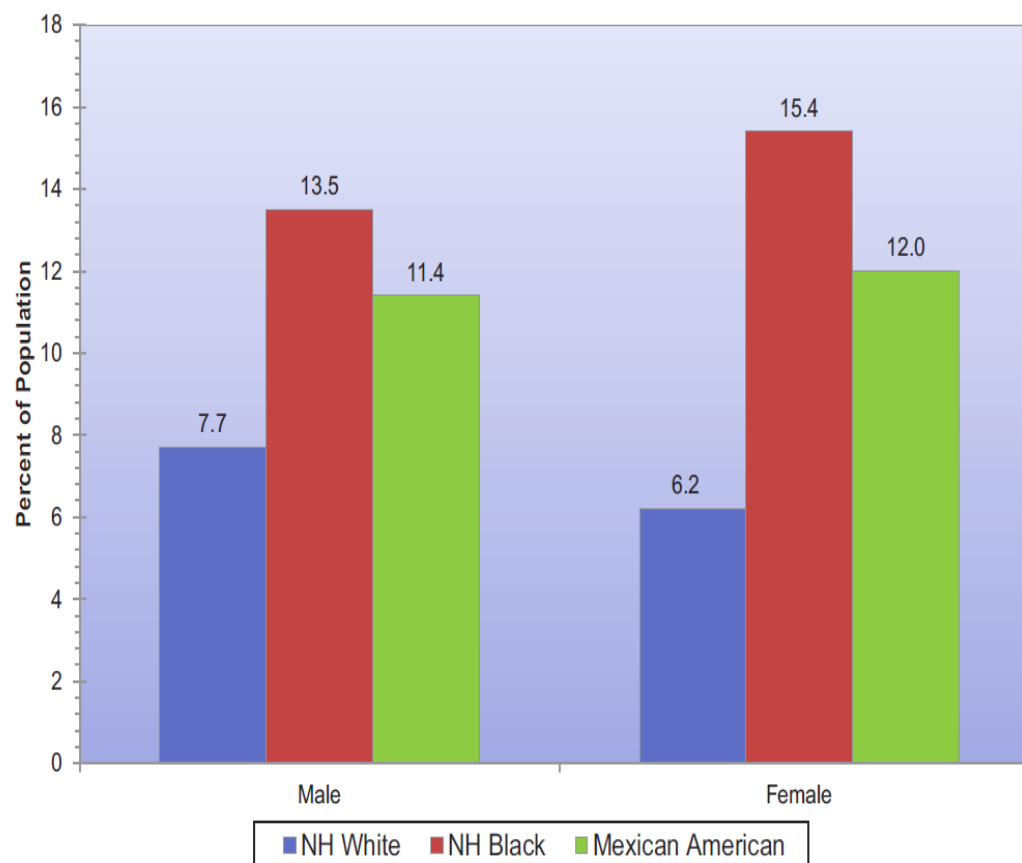
Relative measures

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

Absolute measures

- Risk difference
- Rate difference
- Years of life lost

Measuring disparities: Relative vs absolute comparisons of the groups?



Relative effects approach: black men have $13.5/7.7=1.75$ **times** the prevalence of diabetes as white men, or 75% higher prevalence. Women: 2.48 times

Absolute effects approach: black men have $13.5-7.7=5.8$ **percentage point** higher prevalence of diabetes than white men. Women: 9.2 percentage points

Which is better?

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

Measuring disparities: Relative vs absolute comparisons of the groups?

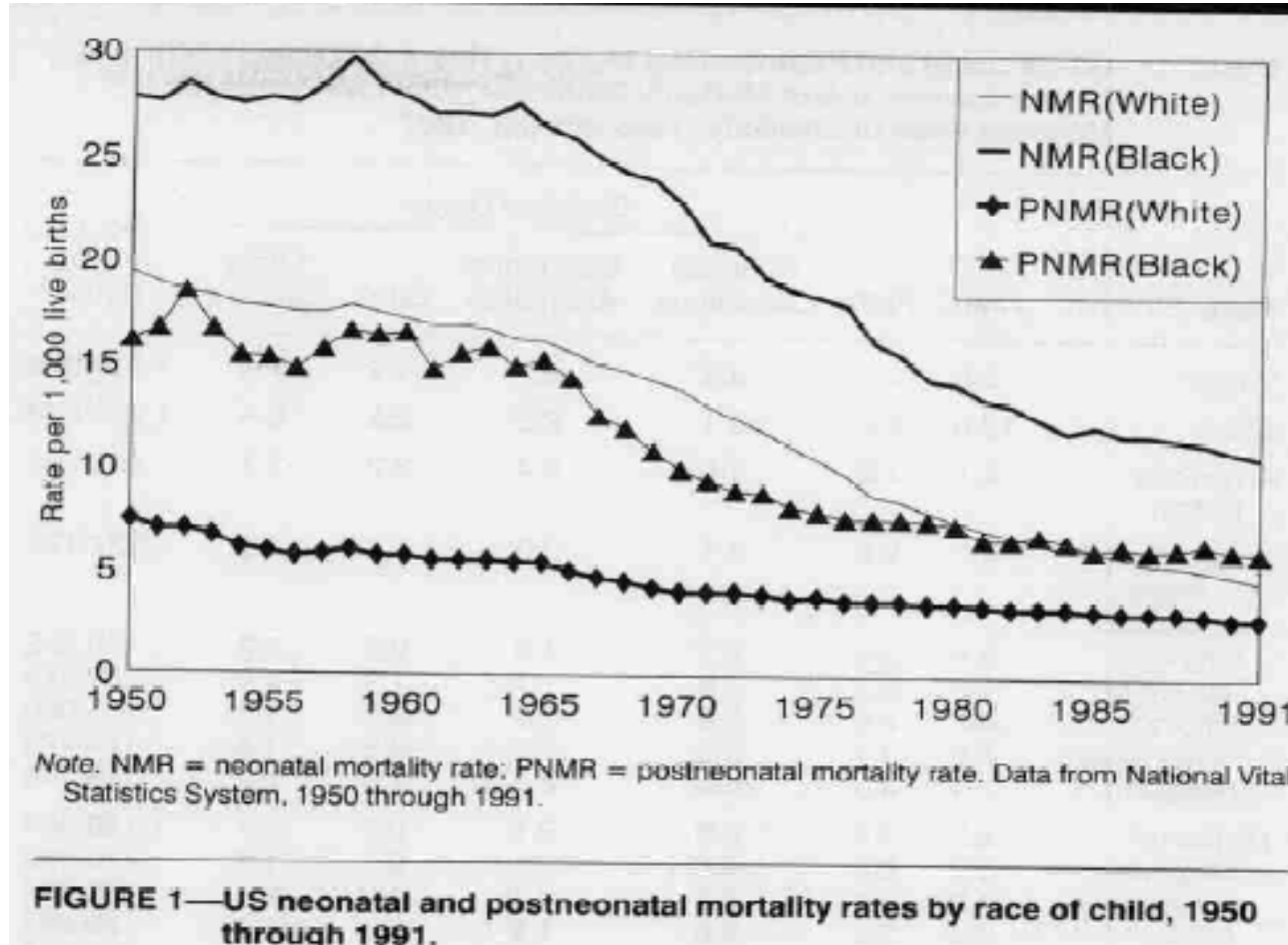
- Relative measures are extremely common, so when you see one, stop and think “how big is this effect in absolute terms?”
- Presentation of both the relative and absolute differences helpful in fully understanding differences

Measuring trends in disparities: Relative vs absolute measures

Claims that depend on comparing risk factor coefficients:

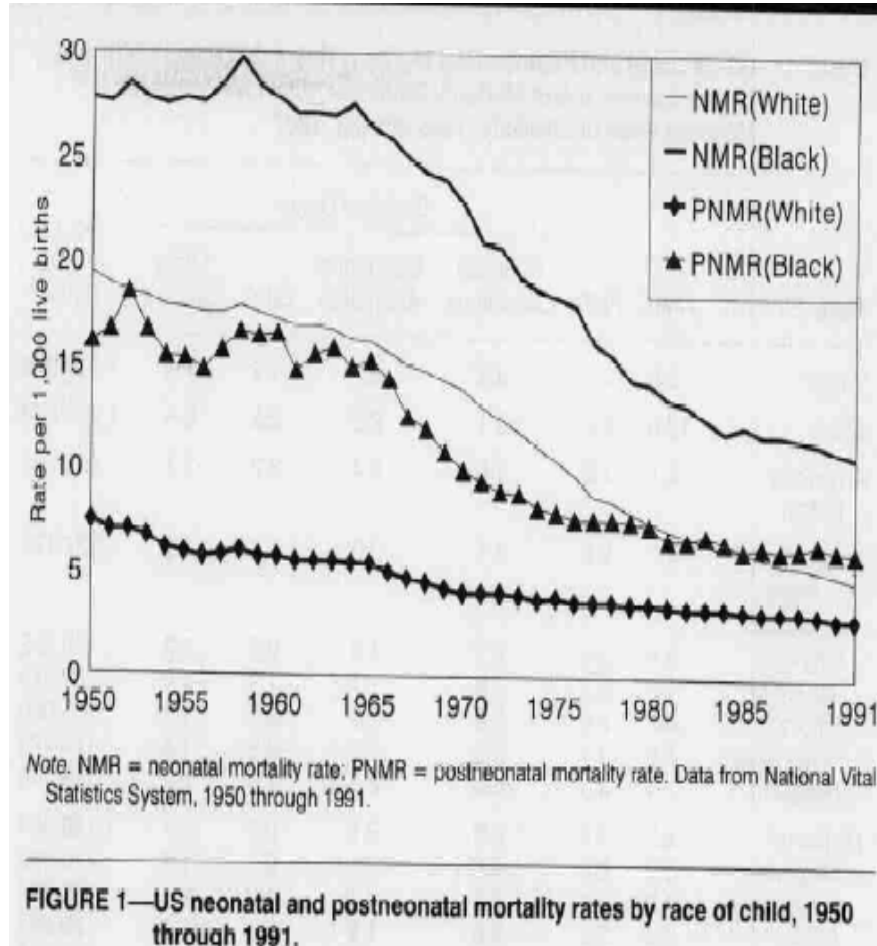
- **The black-white disparity in neonatal mortality got worse during the 20th century, despite the advances in civil rights.**
- Although the medical system may not work ideally for younger African-Americans, these problems are ameliorated among the elderly.
- After age 70, the role of socioeconomic 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



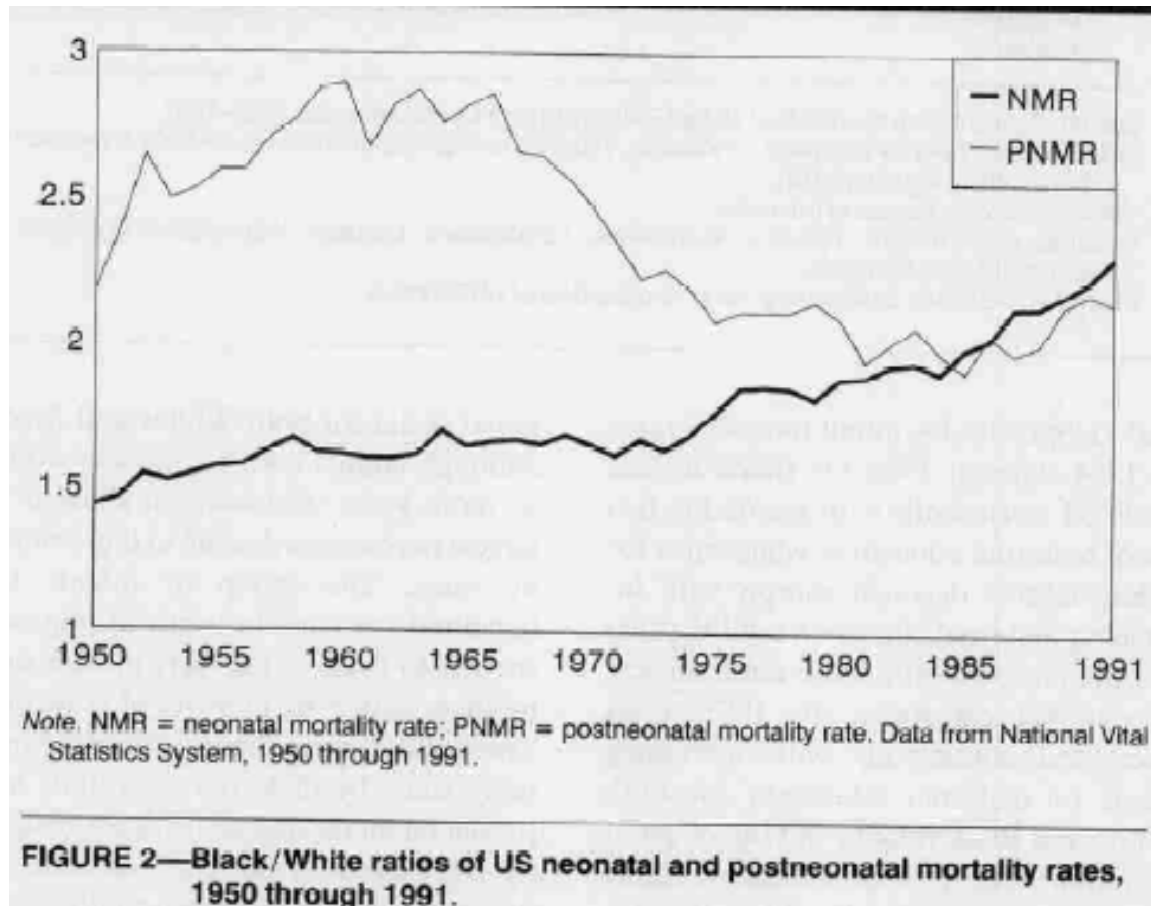
Did the black white disparity in infant mortality increase or decrease?

Trends in infant mortality



- In 1950, the rate was 43.9/1000 for black babies and 26.8/1000 for whites.
- relative=1.64; absolute=17.1
- By 1991 the rate was 16.5 /1000 for black babies and 7.5/1000 for whites.
- relative=2.2; absolute=9.0
- On a relative scale disparity got worse, but on an absolute scale it got better

Trends in infant mortality



- NMR disparity based on relative scale has been increasing
- PNMR disparity based on relative scale decreased to the 60s and appears to be increasing

Singh and Yu, 1995

Effect measure modification: Scale dependence

- Occurs when the magnitude of the effect of the primary exposure on an outcome differs depending on the level of a third variable
- 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?
- Depends on which measure is used: additive scale (absolute) or on a multiplicative scale (relative)
- The scale dependence of interaction is essential to recognize when interpreting disparities, and especially trends in disparities.

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.

Questions?

Learning objectives

1. Understand general issues in measuring the magnitude of disparities:

- Defining the dimension of disparity
- Choosing the measure for a dimension of disparity
- Defining comparison groups
- Choosing the outcome
- Choosing the way to compare outcomes

2. Understand basic concepts in measuring constructs in disparities research:

- Distinguish between a social construct and its measurement
- Describe key steps in identifying/developing/adapting a measurement tool
- Discuss how to assess the adequacy of a measurement tool

The challenge of measurement of health disparities

- Most of the analysis we do depend on having quantitative measures of health and risk factor attributes
- But often what we want to measure is not straightforward
- We are interested in questions between abstract concepts that cannot be assessed directly
 - Applies to both exposures and outcomes

What is a construct?

- An abstract idea one wishes to measure
- Cannot be observed directly
- Can be simple or complex
 - Simple constructs can be measured using only one or few questions
 - Complex constructs contain multiple dimensions that are bound together by some commonality that make up the construct
 - Mostly related to attitudes, perceptions, experiences, and psychosocial health outcomes
 - Require a whole battery of questions to fully operationalize

Examples of constructs/latent variables

- Racism
- Self-efficacy
- Stress
- Depression
- Food security
- Body image
- Food environment
- Neighborhood
- Health
- Social support
- Social capital
- Health literacy
- Pain
- Disability
- Cognitive function
- Quality of care
 - Person-centered care

Concepts/constructs/latent variables

- Concepts are the building blocks—the primary elements—of a theory
- Constructs are concepts developed or adopted for use in a particular theory. The key concepts of a given theory are its constructs.
- Variables are the operational forms of constructs. They define the way a construct is to be measured in a specific situation. Match variables to constructs when identifying what needs to be measured.

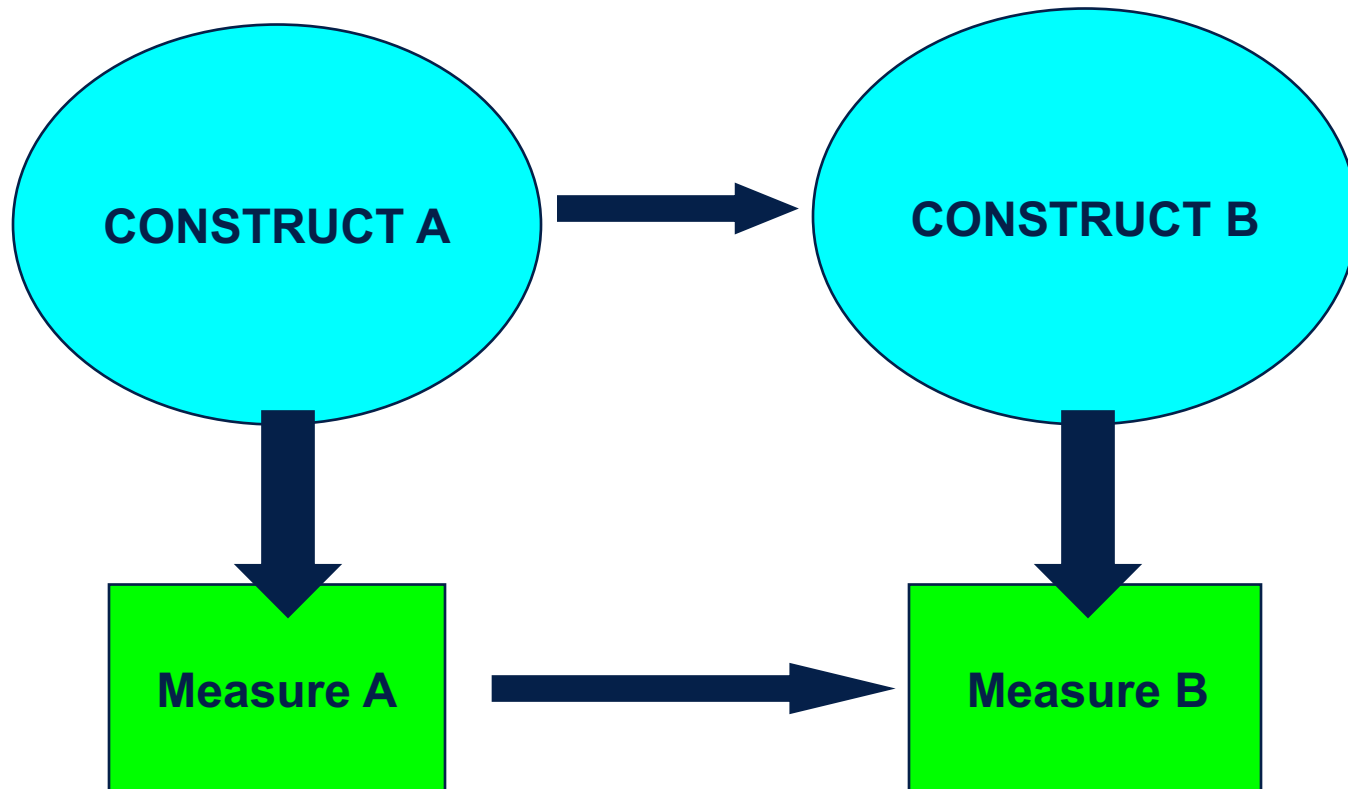
(USDHHS Theory at a glance)

- Concepts and constructs referred to as latent variables: the idea you want to measure
- Concepts and constructs are defined and operationalized by measures (observed indicators), “proxies” for what we cannot directly observe

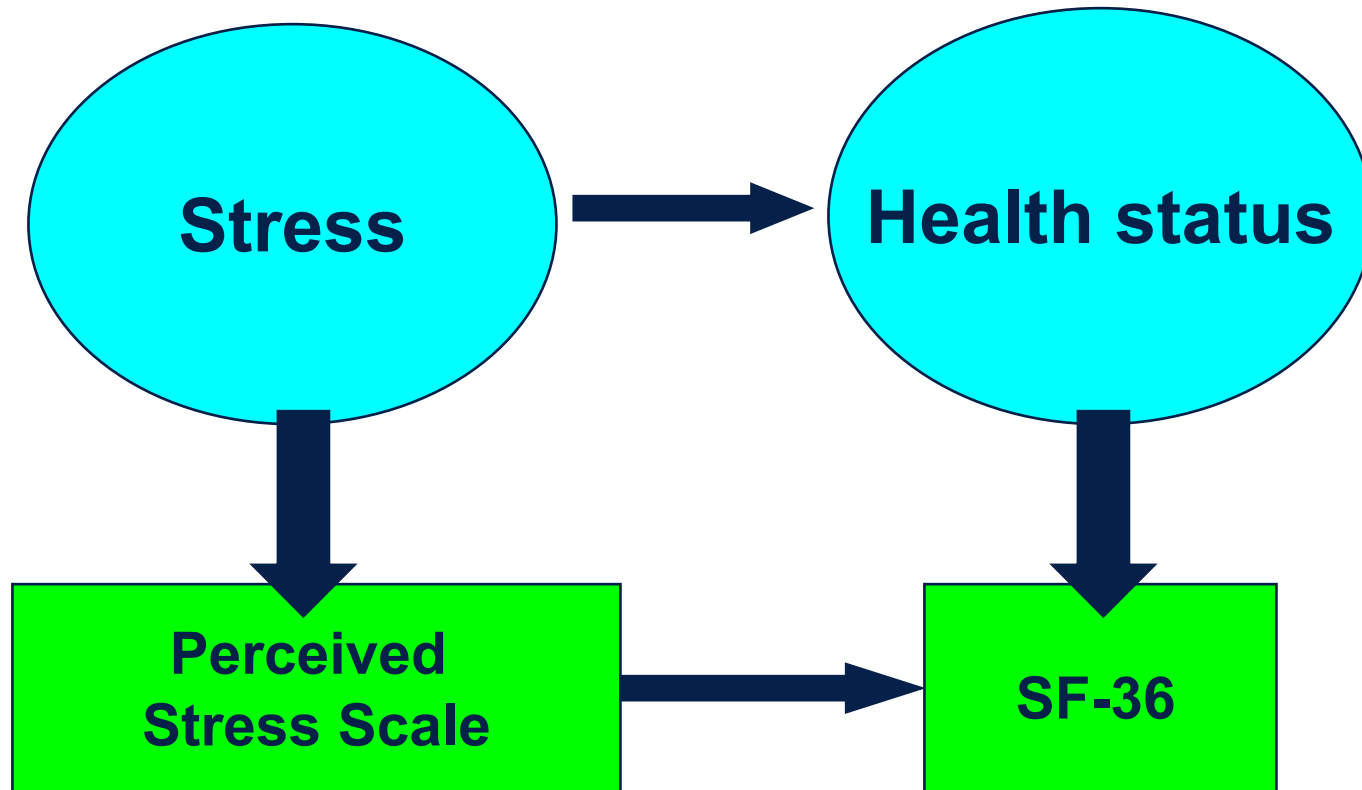
Distinguish between constructs & measures

- What is the the idea you would like to measure (construct/latent variables)?
- How you will measure it (the measure or instrument you will use)?
 - Operationalization: the measurable expression of the construct in the form of a question(s), checklists, etc.

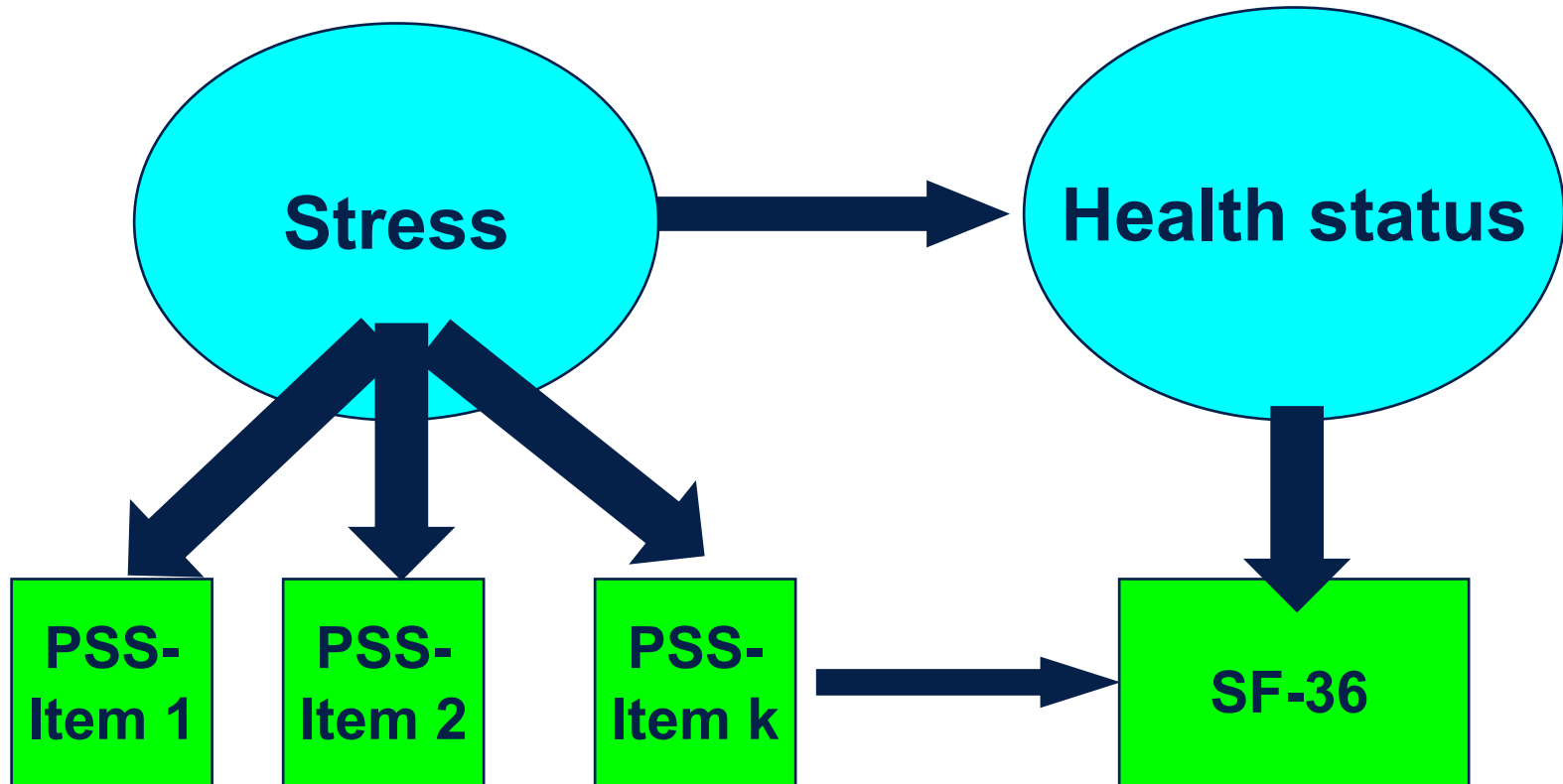
Depicting latent variables and measures



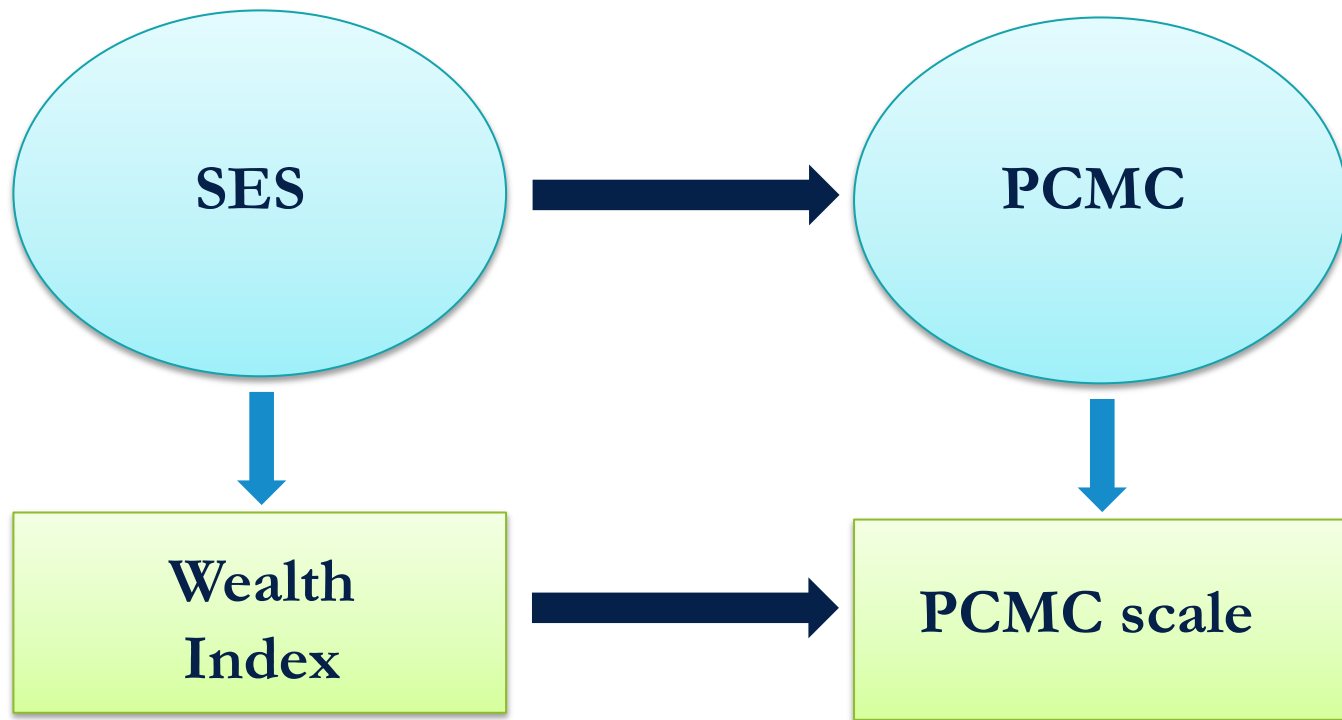
Depicting latent variables and measures



Depicting latent variables and measures



Depicting Latent Variables and Measures



Learning objectives

1. Understand general issues in measuring the magnitude of disparities:
 - Defining the dimension of disparity
 - Choosing the measure for a dimension of disparity
 - Defining comparison groups
 - Choosing the outcome
 - Choosing the way to compare outcomes

2. Understand basic concepts in measuring constructs in disparities research:
 - Distinguish between a social construct and its measurement
 - **Describe key steps in identifying/developing/adapting a measurement tool**
 - **Discuss how to assess the adequacy of a measurement tool**

Process of selecting good measures for your studies

- Define your construct
- Identify potential measures
- Review measures for conceptual adequacy
- Assess applicability and other potential issues in your population (cognitive interviews, pretesting)
- Review measures for psychometric adequacy (validity and reliability)
- Select final measure

Define your construct

- What exactly do you want to measure?
- How will you define it?
- Does it have sub-domains or dimensions?
- 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 constructs of scientific interest.
 - Racism, Gentrification, Segregation, Income inequality, Food environment

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
- What are dimensions of the construct?
- Distinguish between closely related constructs

Identify potential measures

- Extensive resources on measuring difficult concepts (should address conceptual adequacy)
- 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.

Important considerations

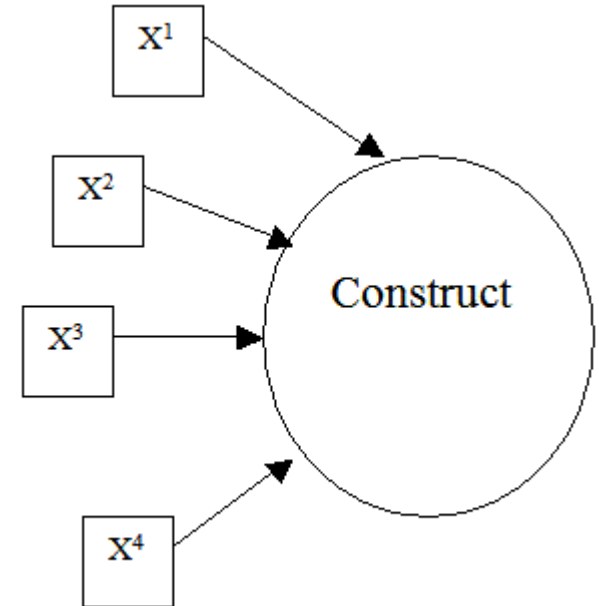
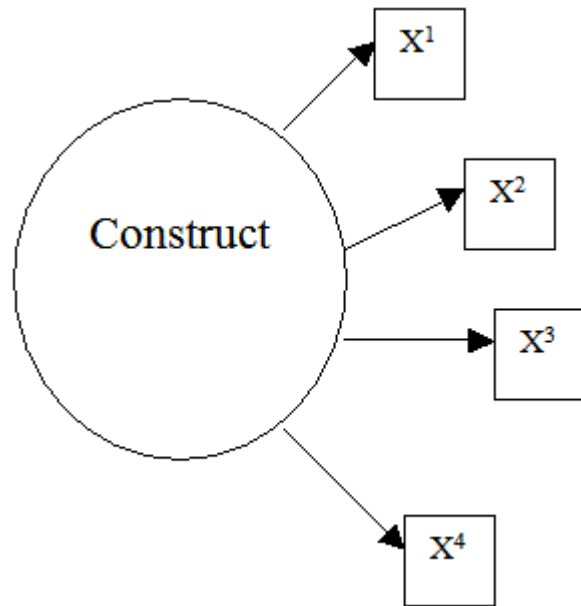
- How will you obtain your data?
 - Self-report, proxy report
 - Direct observations, administrative records, standard tests
 - Physical measurement apparatus,
 - Level of the measurement
 - Individual, group (neighborhood, school, state, country), time (repeated measures on the individual).
 - For group measures: aggregate (average poverty rate) or structural (nuclear power plant, weather)
 - Subjective vs objective measures
 - Pros and cons of each method
-

Scales vs indexes (DeVellis)

- 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.
- 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

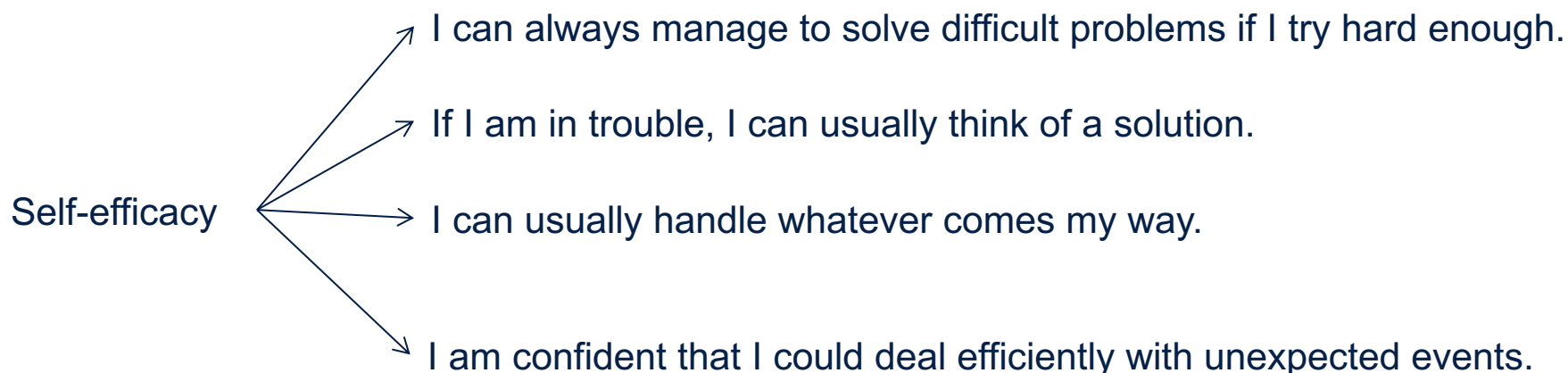
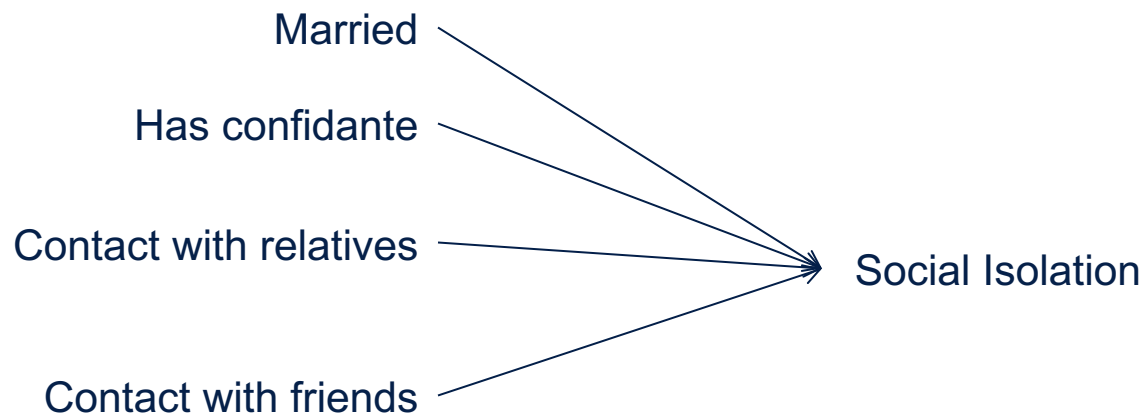
- If you encounter the SEM or psych literature, they often use the terms formative and reflective, corresponding to scale vs index



Scales- Reflective: items reflect the underlying construct

Index-Formative: items form the construct

Scales vs indexes (per the Devellis framework)



Identify potential measures

- Do you need to develop a new measure or modify an existing measure?
 - What type of tool do you want to develop?

General steps to developing measures

- 1 • Define construct and identify domains
- 2 • Develop list of items to operationalize your construct
- 3 • Assess content validity
- 4 • Ensure questions/items are clearly worded
- 5 • Iterative revision
- 6 • Administer items in target population
- 7 • Conducted psychometric analysis

Define your construct

- What exactly do you want to measure?
- How will you define it?
- Does it have sub-domains or dimensions?

Defining the construct (e.g.,)

- Person-centered maternity care (PCMC): Providing maternity care that is respectful and responsive to women's preferences, needs, and values

Initial PCMC domains	
1	Dignity/Respect
2	Autonomy
3	Privacy/Confidentiality
4	Communication
5	Social Support
6	Supportive Care
7	Predictability/Transparency of Payments
8	Trust
9	Stigma/Discrimination
10	Health Facility Environment

Developing the items

- You need items that adequately measure your construct based on your definition and the domains you have identified
- **How do you ensure items measure the construct?**

Developing the items

- How do you ensure items measure the construct?
 - Literature review
 - Expert reviews
 - Prior tools
 - Prior qualitative research

- Ensures content validity

Domain	Question
Dignity/respect	Did the doctors, nurses, or other staff at the facility treat you with respect?
Privacy/	During examinations in the labor room, were you covered up with a cloth or blanket or screened with a curtain so that you did not feel exposed?
Autonomy	Did you feel like the doctors, nurses or other staff at the facility involved you in decisions about your care?
Communication	Did the doctors and nurses explain to you why they were doing examinations or procedures on you?
Social support	Were you allowed to have someone you wanted (outside of staff at the facility, such as family or friends) to stay with you during labor?
Supportive care	Did the doctors and nurses at the facility talk to you about how you were feeling?
Trust	Did you feel the doctors, nurses or other staff at the facility took the best care of you?
Predictability & transparency of payments	During your time at the facility, did any staff at the facility ask you or your family for kitu kidogo? (colloquial translation for bribe)
Stigma & discrimination	During your time in the health facility, would you say you were treated differently because of any personal attribute... like your age, marital status, number of children, your education, wealth, your connections with the facility, or something like that?
Facility environment	Thinking about the wards, washrooms and the general environment of the health facility, will you say the facility was very clean, clean, dirty, or very dirty?

Developing the items

- How do you ensure respondents are understanding the questions the way you intend?
- Cognitive interviews and Pretesting essential for measures being applied to any new population group
 - Especially priority measures (e.g., outcomes)

Cognitive interviews and pretesting

■ Cognitive interviews

- To assess comprehension, clarity, sources of confusion in stems, items, and response options, and relevance to inform item revision decisions
- Can help identify misalignment between participant interpretation of items and the developer's intentions and to identify ways to modify those items based on participant's response

■ Pretest

- To identify problems with procedures, method of administration, respondent burden
- Pretest may also help with problems with instructions, item stems, questions, and response choices, **but may not**

Things to pay attention to

- Item stem
- Complexity
- Ambiguous or double-barreled items
- Response options
 - Type/format/order/number/directionality
- Match of response choices to item stem
- Time frame
- Place of interview

Iterative revisions

- Don't fall in love with your questions too early

Administer items in target population

- What is your population of interest?
- Sampling issues

How good is my measure?

- Psychometric analysis

- Goal: assess validity and reliability of the tool
- Does your tool measure “what it is supposed to measure” and does so consistently?

Distinguish between validity and reliability

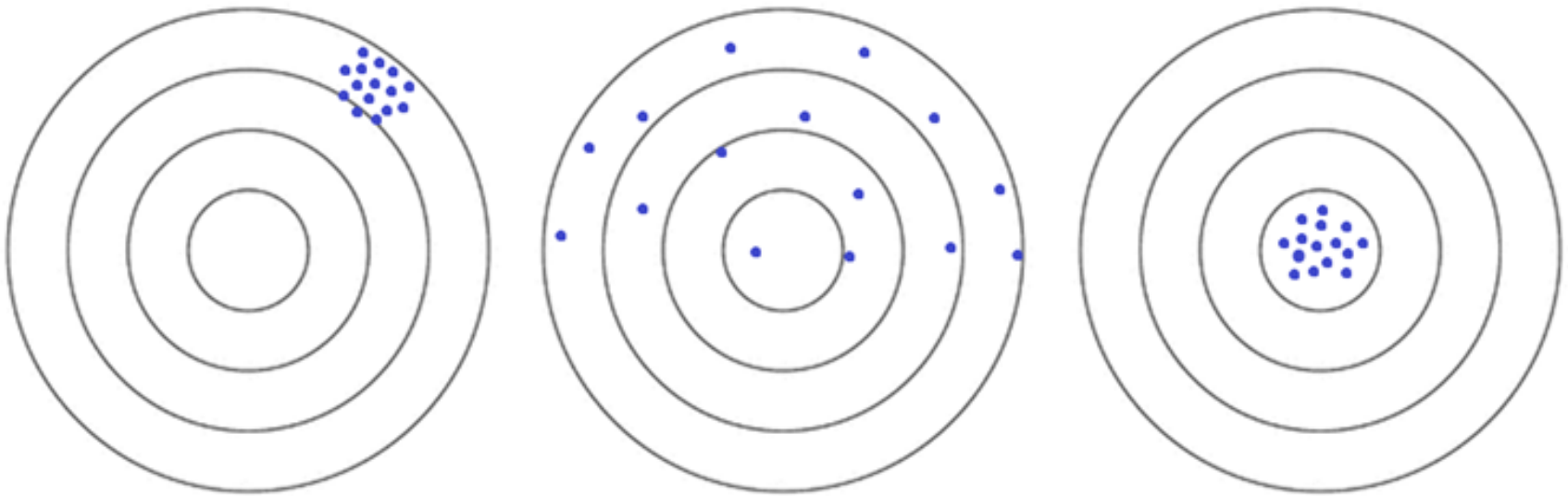
Validity:

- The extent to which a measure/tool measures what it is intended to measure
 - Starts with a clear, precisely stated, construct

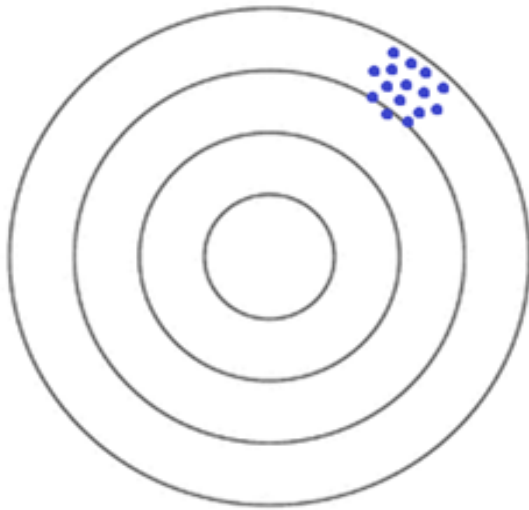
Reliability:

- The extent to which a measure produces the same results when used repeatedly to measure the same thing (under the circumstances it should not change)
 - 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.
-
- A valid and reliable measure measures “what it is supposed to measure” and does so consistently

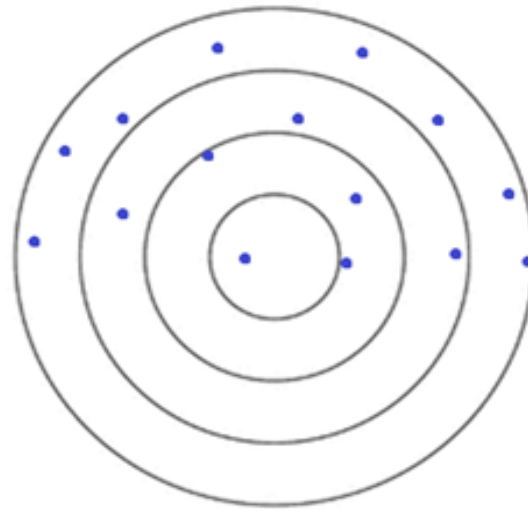
Valid or reliable?



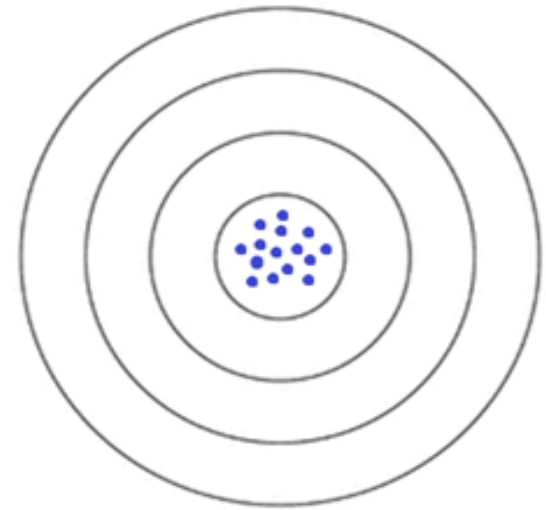
Valid or reliable?



Reliable but not valid



Neither reliable nor valid



Valid and reliable

- Validity relates to the accuracy of a method, while reliability relates to the consistency.
- **Is it possible for a tool with poor validity to be very reliable?**
- A valid and reliable measure measures “what it is supposed to measure” and does so consistently.

Assessing validity

- **Face validity:**

- **Content validity:**

- Assess if the items represent a universe of all possible indicators relevant to the construct.
- 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.

- Does not start after you have collected your data. Most important is content validity which is based on everything you do before you collect your data.

Assessing validity

■ Construct validity:

- How well items represent underlying latent (conceptual) structure
 - Factor analysis often used to assess construct validity
- Assesses the degree to which a measure relates to other measures in theoretically predictable ways
 - Does the measure correlate with other things that, theoretically, you think it ought to correlate with (*convergent validity*)?
 - Does it *not* correlate with things you think should be independent (discriminant validity) ?
- Hard to make a strong case for validity using only construct validity.

Assessing validity

- **Predictive validity:** does the measure predict some other, external measure of the same thing?
- **Criterion validity:** assesses the relationship of a measure to a gold standard or to an outcome which by the definition of the measure it should be related to.
 - Sensitivity and specificity if you have a diagnostic criterion.

Sensitivity and specificity

- **Sensitivity:** the extent to which the value on a measure change when there is a change or difference in the thing being measured.
- **Specificity:** the extent to which a measure is specific to the construct of interest.
- The sensitivity of a clinical test refers to the ability of the test to correctly identify those patients with the disease. (true positive rate)
- The specificity of a clinical test refers to the ability of the test to correctly identify those patients without the disease. (true negative rate)

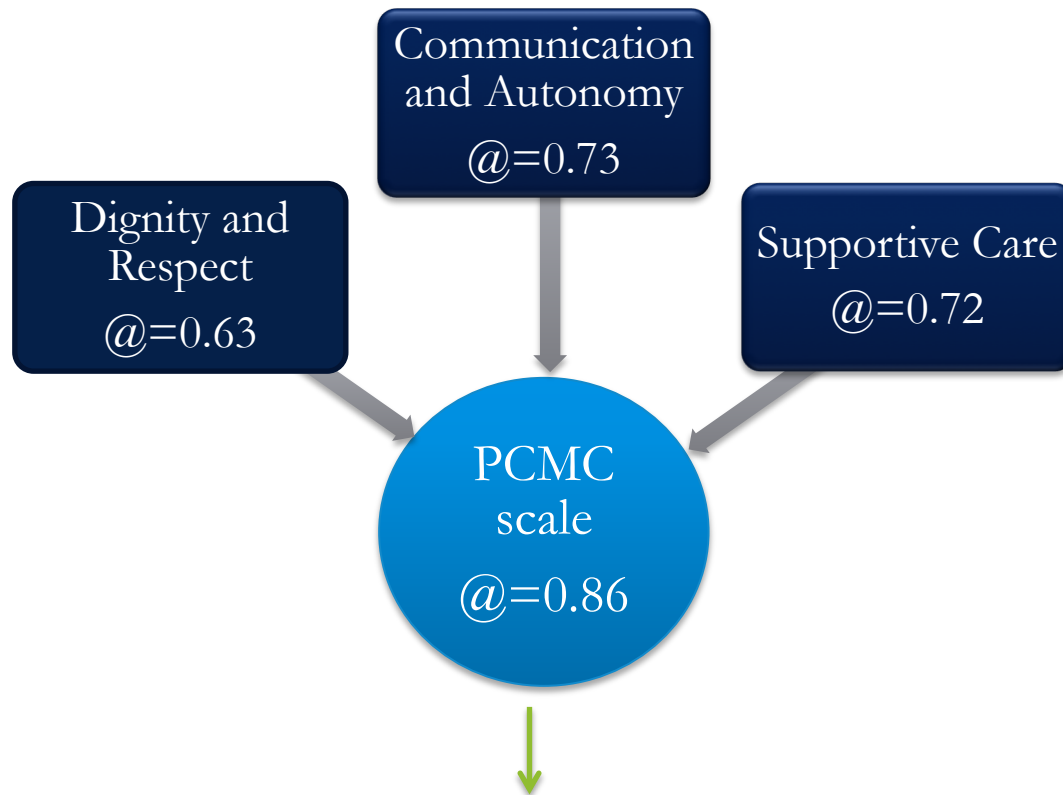
Assessing reliability

- **Internal consistency reliability:** The extent to which items within a tool measure the same construct
 - Measured using Cronbach's alpha, which is expressed as a number between 0 and 1, with values closer to 1 indicating that items are more related to each other--implies greater reliability
 - Pay attention to the direction of items
 - Evaluate whether you might want to dump some of your items
 - Increases with more correlated items, but does not necessarily make it a more valid measure; very high over 9 might just mean same items
 - In stata: alpha S1-S4, item generate (SE_hat)

Assessing reliability

- **Test-retest reliability:** extent to which a method produces consistent data in similar conditions across multiple time points.
 - 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.
- **Inter-rater reliability:** similarity between data from different observers of the same phenomenon
 - Have Joe give the test
 - Have Elaine give the test
 - Examine the correlation between measures. Does Elaine always give higher scores?

PCMC scale: Kenya, 30 items



PCMC subscales and items	
Sub-scale	Items
Dignity and respect (6)	
	Treated with respect
	Friendly
	Verbal abuse
	Physical abuse
	Visual privacy
	Record confidentiality
Communication and autonomy (9)	
	Introduce self
	Called by name
	Involvement in care
	Consent to procedures
	Delivery position choice
	Language
	Explain exams/ procedures
	Explain medicines
	Able to ask questions
Supportive Care (15)	
	Time to care
	Labor support
	Delivery support
	Talk about feeling
	Support anxiety
	Attention when need help
	Took best care
	Control pain
	Trust
	Enough staff
	Crowded
	Clean
	Water
	Electricity
	Safe

How can insufficient measures compromise identifying/explaining health disparities?

Inadequate measures can result in

- Poor validity: you measure the wrong thing or measure it badly
 - The wrong answer to your question
 - Conceptual inadequacy due to vaguely defined questions
 - Meaningless answer to your fuzzy research question
 - Poor data quality (e.g. missing data): people won't answer or can't answer, so no answer to your question
 - Poor variability: everyone answers the same
- Poor reliability weakens observed associations between exposure and outcome variables which can conceal true relationships

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.

Recap/summary

- Measurement is a critical task in rigorous disparities research
- Often the things we want to measure are abstract (constructs) and challenging to measure but need to be measured
- Good measurement starts with conceptual clarity: Be clear about what you want to measure and how you can measure it
- Use existing validated tools if possible. But always ask: Does the tool adequately measure what you want to measure (validity) consistently (reliability)?
- Use formal approaches to develop and/or evaluate the validity and reliability of measurement instrument in your population
- Large evidence base to draw on, but there are always compromises
- Your findings are as good as your measurement tools
- To have credible data, you need to: Ask the right questions; Ask the questions the right way; Ask the right people; Ask enough people

Questions?

Group Discussion on 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 and/or upload the paper if that's possible).
- 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 validity or reliability for future research applications.

Group Discussion on homework: Part 2

- 1. Find a paper describing a health disparity (please give the full citation or, even better, upload the paper so everyone else can look at it too)
- 2. Summarize the construct and measurement of the dimension of disparity (e.g., racial inequalities?, SES inequalities?) 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 (ie, who is the comparison group)? Why does this reference category make sense (or not) for this research question?
- 5. How is the disparity quantified or measured? 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.

—

Thank you!



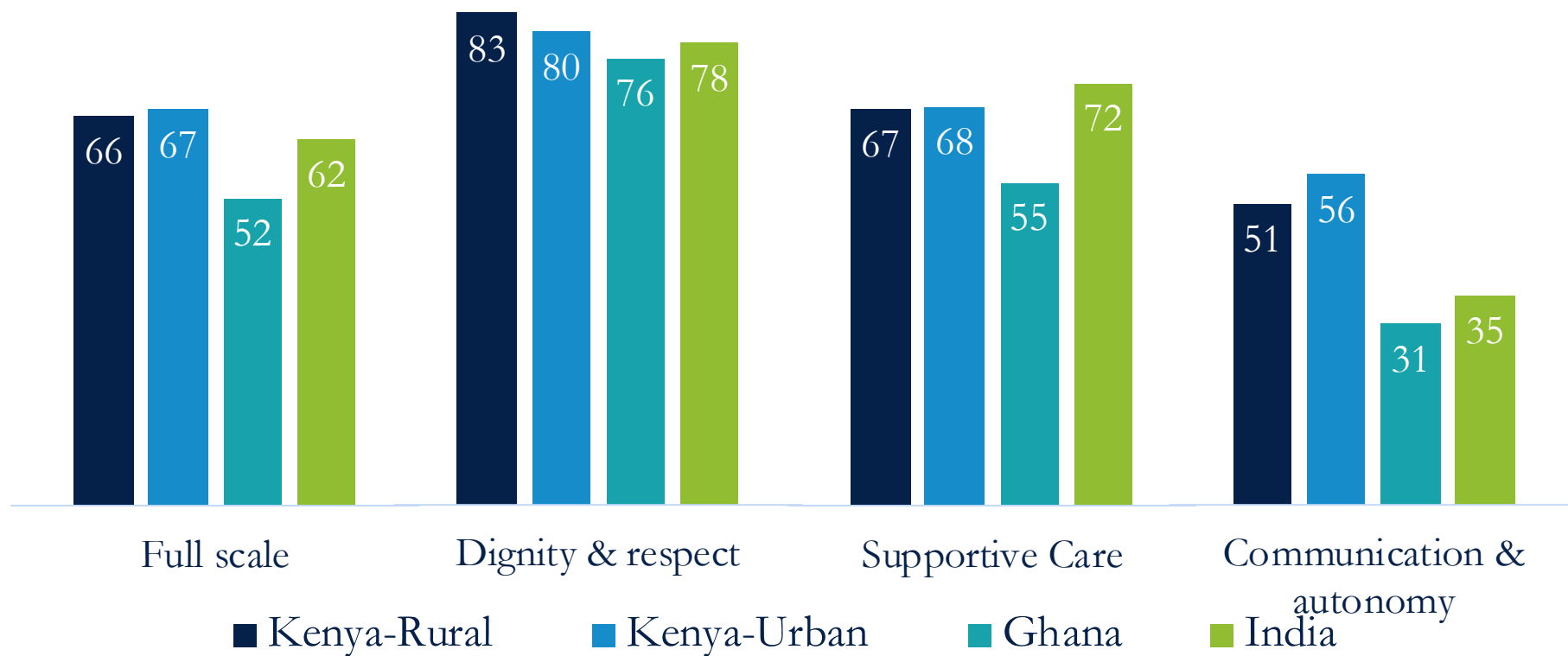
University of California
San Francisco

Questions? Comments?

Patience.Afulani@ucsf.edu

PCMC scores on full scale & by domains

Rescaled scores



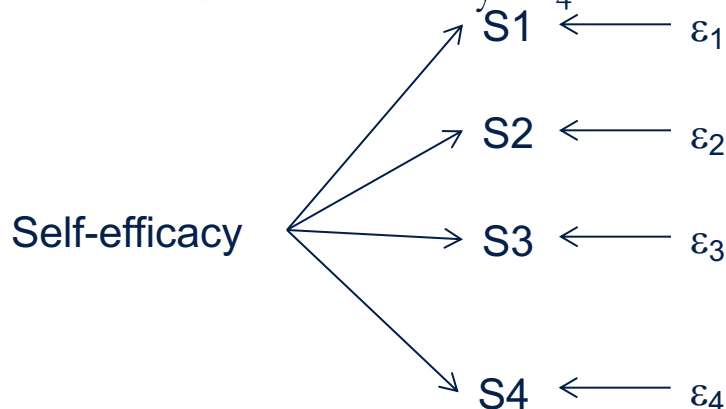
Group Activity

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.

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.
- $S1 = b * \text{self-efficacy} + \epsilon_1$
- $S2 = b * \text{self-efficacy} + \epsilon_2$
- $S3 = b * \text{self-efficacy} + \epsilon_3$
- $S4 = b * \text{self-efficacy} + \epsilon_4$



Classic Measurement Model assumes:

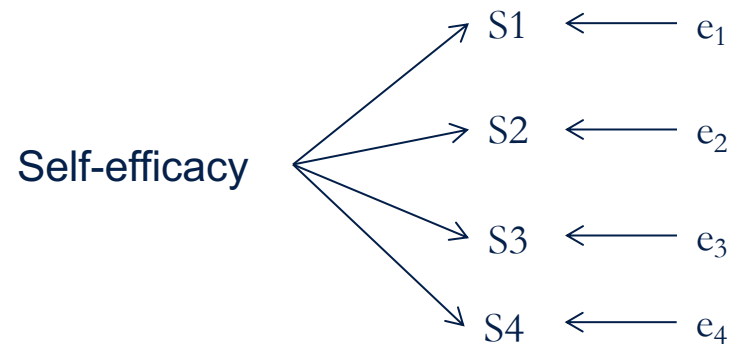
- Errors are random
- Errors are not correlated with each other
- Errors are not correlated with the true score
- Latent variable affects all items equally
- Variance of each error is the same

Why repeat basically the same question in different ways?

■ Classical measurement theory:

- $S1 = b \cdot \text{self-efficacy} + e_1$
- $S2 = b \cdot \text{self-efficacy} + e_2$
- $S3 = b \cdot \text{self-efficacy} + e_3$
- $S4 = b \cdot \text{self-efficacy} + e_4$
- $\text{self-efficacy} \sim N(0, s^2)$
- $e \sim N(0, d^2)$
- Assume $b=1$
- What % of the variance in S1 is due to self-efficacy?

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



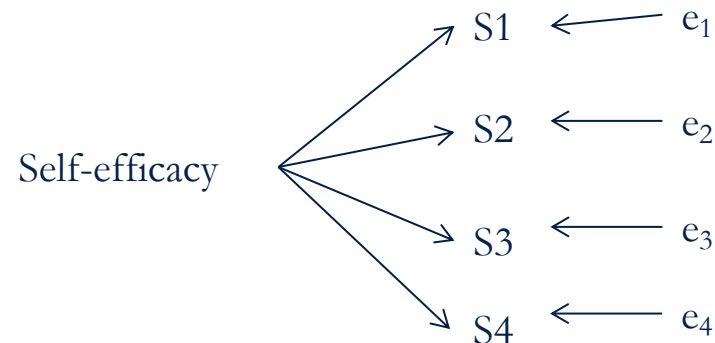
Adding items reduces measurement error and improves reliability

■ Classical measurement theory:

- $S1 = b \cdot \text{self-efficacy} + e_1$
- $S2 = b \cdot \text{self-efficacy} + e_2$
- $S3 = b \cdot \text{self-efficacy} + e_3$
- $S4 = b \cdot \text{self-efficacy} + e_4$
- $\text{self-efficacy} \sim N(0, s^2)$
- $e \sim N(0, d^2)$
- 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%

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

- Now calculate $(S1 + S2) / 2$
 - What % of the variance in $(S1 + S2) / 2$ is due to self-efficacy?
 - $S1 + S2 = (2 \cdot b \cdot \text{self-efficacy} + e_1 + e_2) / 2$
 - $\text{Var}(S1 + S2) = \text{Var}(S1) + \text{Var}(S2) + 2 \cdot \text{Cov}(S1, S2)$
 - The $\text{Cov}(S1, S2)$ is entirely due to self-efficacy
- So self-efficacy contributes disproportionately.



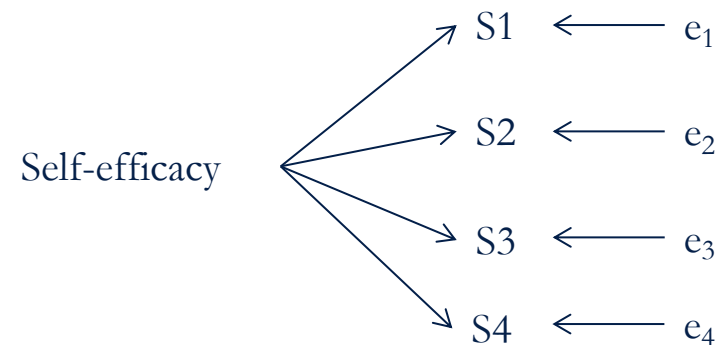
Adding items reduces measurement error and improves reliability

■ Classical measurement theory:

- $S1 = b * \text{self-efficacy} + e_1$
- $S2 = b * \text{self-efficacy} + e_2$
- $S3 = b * \text{self-efficacy} + e_3$
- $S4 = b * \text{self-efficacy} + e_4$
- $\text{self-efficacy} \sim N(0, s^2)$
- $e \sim N(0, d^2)$
- Variance of $S1 = s^2 + d^2$ (because e_1 is independent of self-efficacy)
- What % of the variance in $S1$ is due to self-efficacy? 50%

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

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



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())

