



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

# Learning objectives

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

- *Defining the dimension of disparity (covered already)*
- 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

# 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**

# Measuring disparities

A difference in health outcomes between groups that is *unjust* (preventable).

Measuring disparities entails:

1. Choosing outcome and 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)

# Defining the dimension of inequality

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

# Choosing the measure for a dimension of disparity

- You have decided on the dimension of disparity you want to study...e.g. SES
  - Several different ways of measuring the dimension.
    - How do you choose a measure?

# How do you choose a measure?

## It depends on your research question

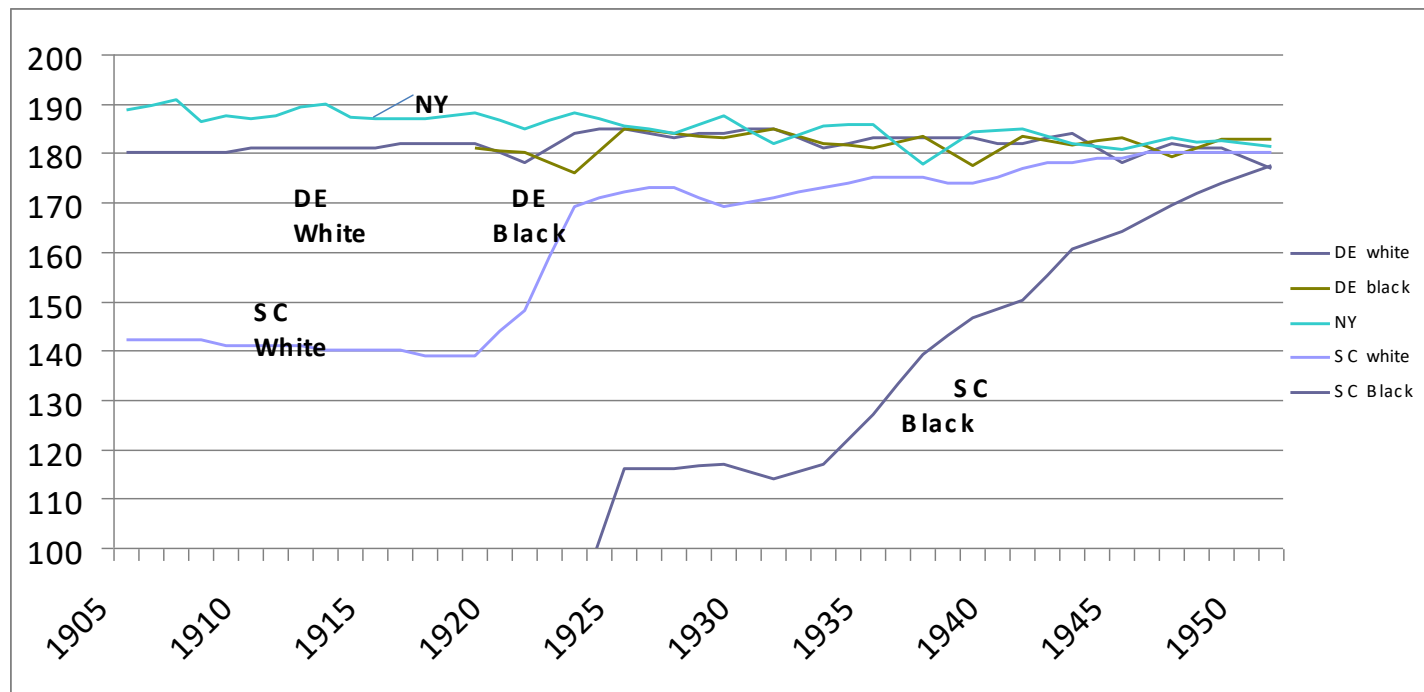
- What type of study you are conducting: etiologic/intervention research or surveillance/descriptive?
- If etiologic:
  - Theoretically, what is most likely to be relevant to your outcome?
  - Articulate mechanisms by which SES may influence your health outcome
- Describe why your measure is inadequate/incomplete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

# 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
    - $\approx 8.9$  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
    - $\approx 7.9$  years
  - A comparable black kid born in SC in 1925:
    - School was open 1,288 days
    - On average he attended 958 days
    - $\approx 5.3$  years (40% less than the NY kid)
- Other dimensions of quality differed as well, so this understates the magnitude of the problem

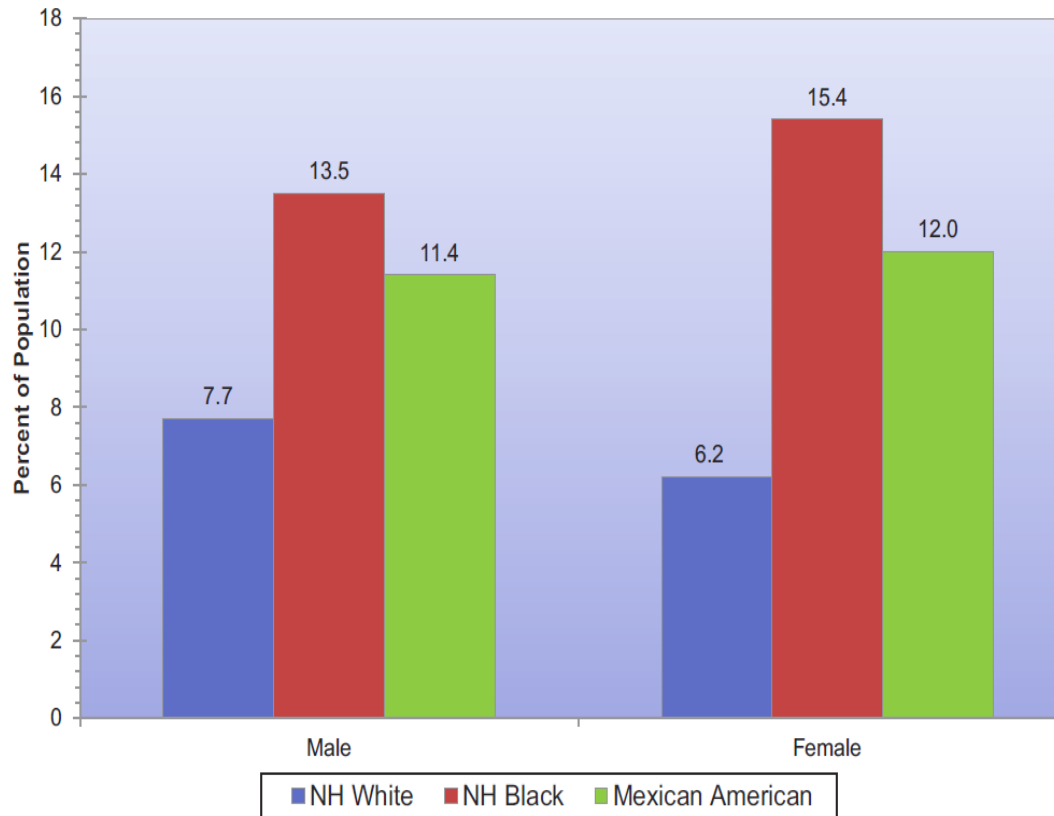
# Other considerations in defining measures of inequality for SES

- Define time in life (childhood, current, cumulative)
- Define level (own SES, household, neighborhood)
- Choose absolute vs relative position (income, income ratio to poverty threshold, 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)

# Choosing a comparison group

- How do you decide on your comparison group?

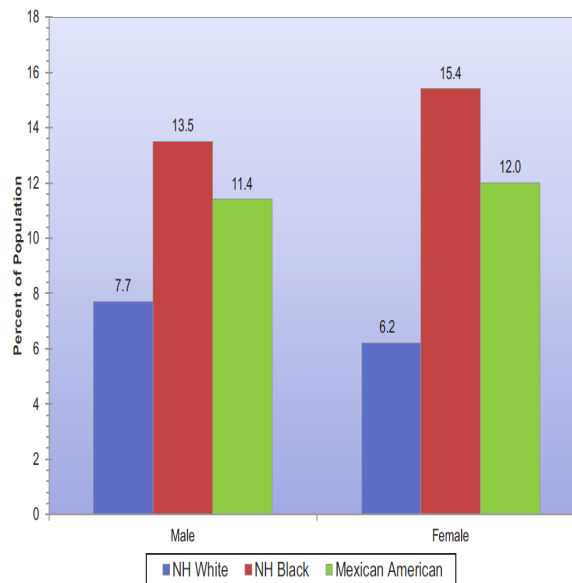
# Choosing a comparison group



Which will you choose as your comparison group and why?

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

# 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.

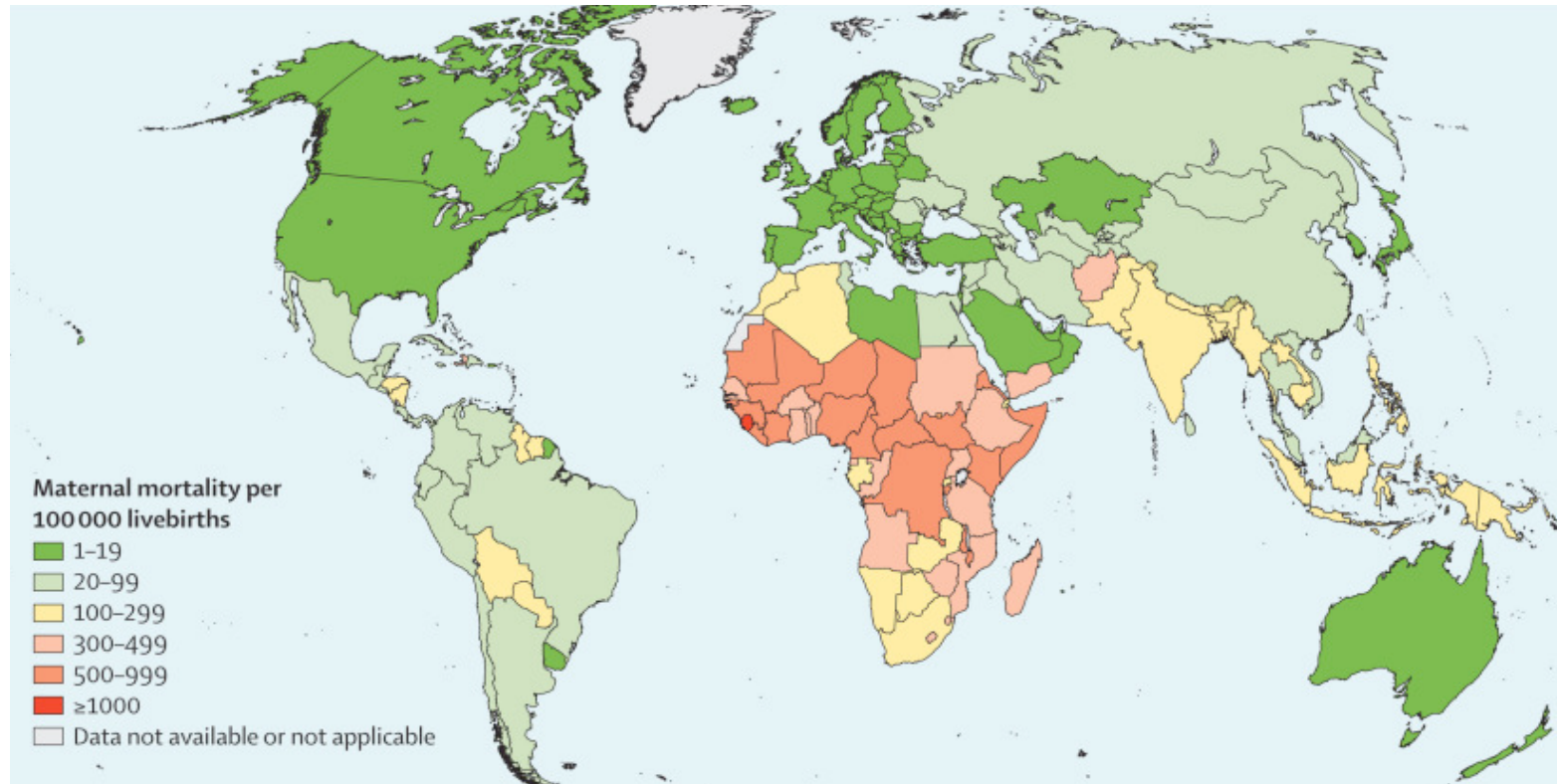
- Use whites as the comparator because they are the “default” or the most advantaged

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

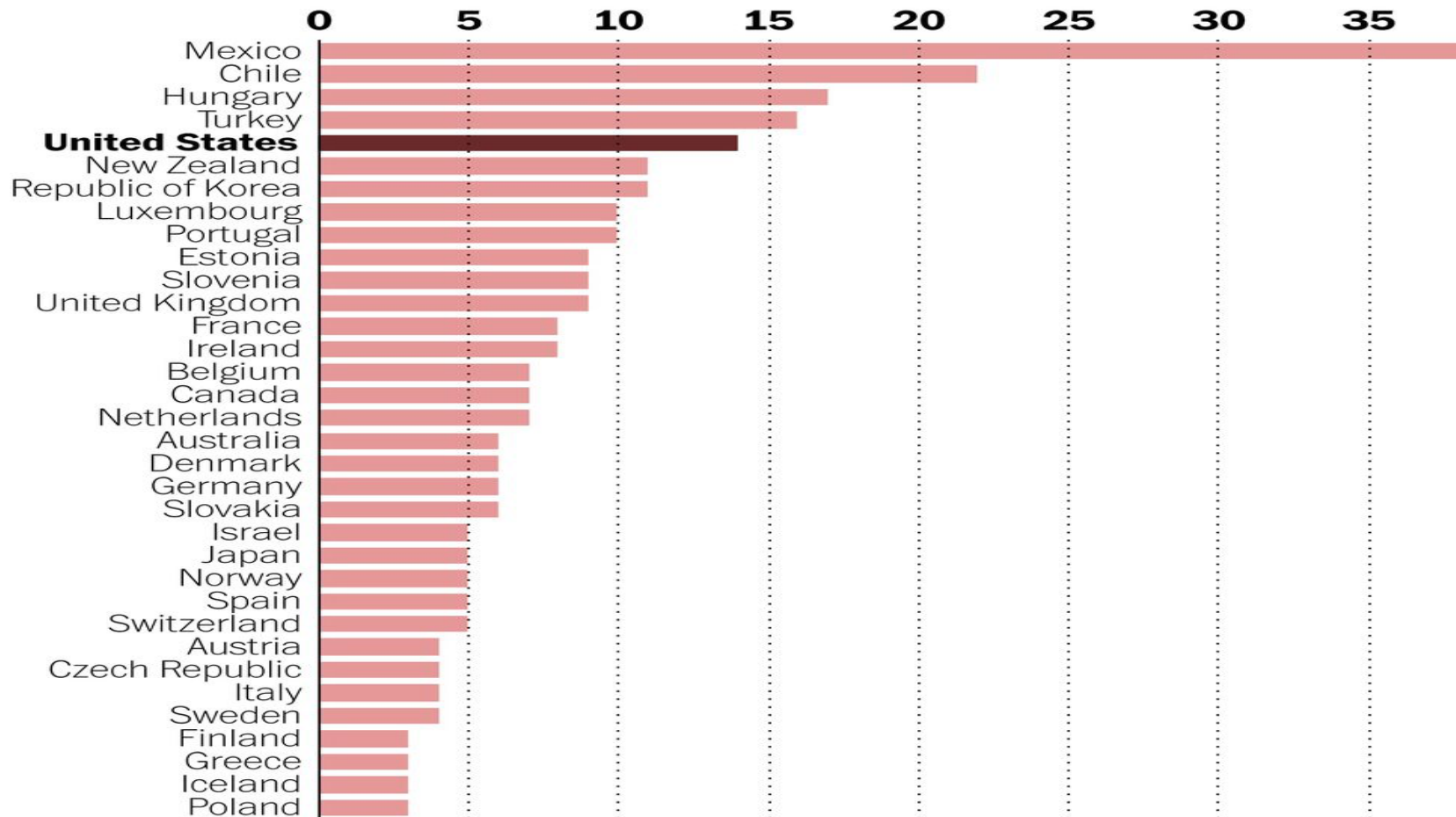
# Maternal mortality



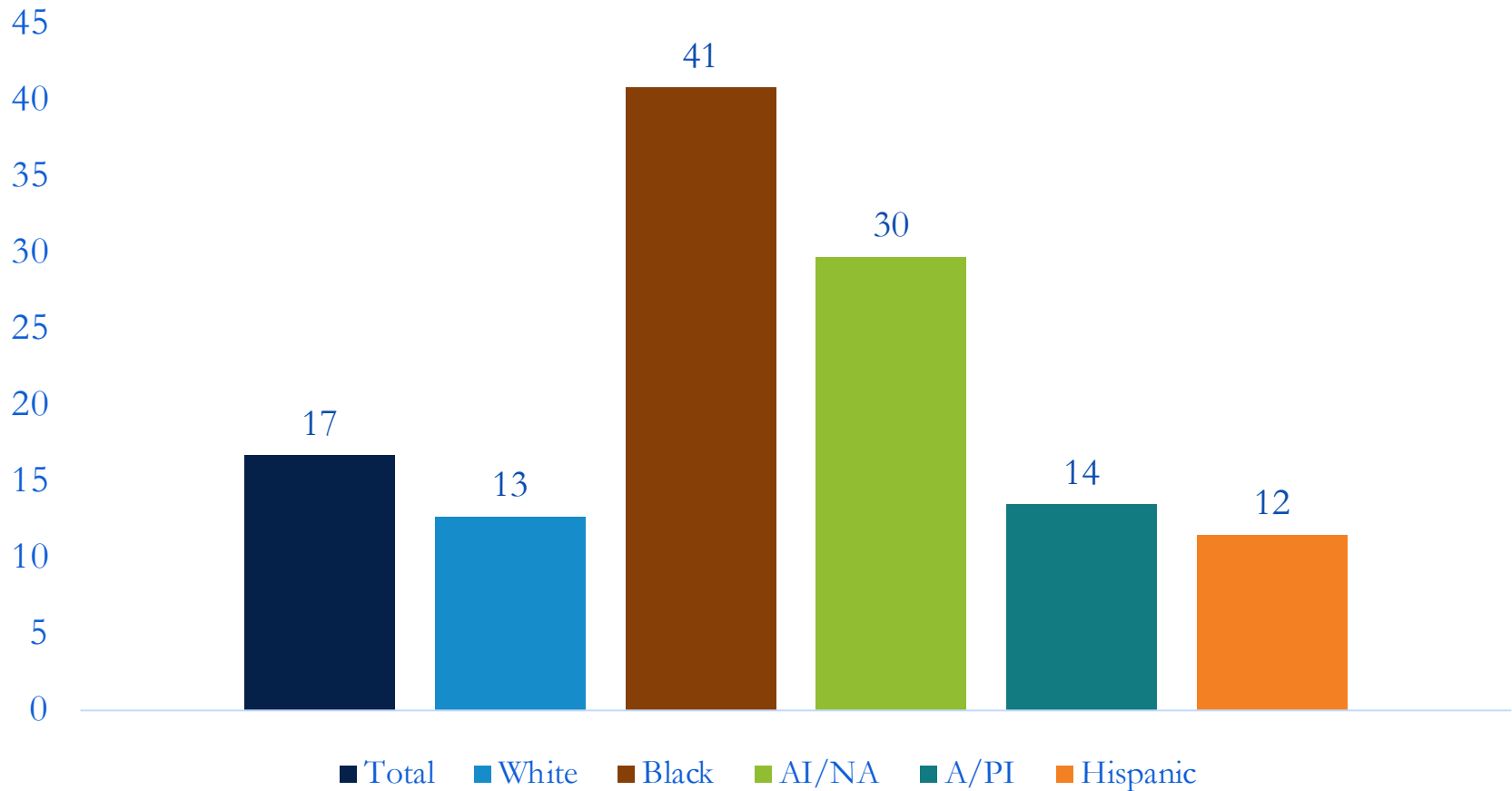
What does this hide?

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

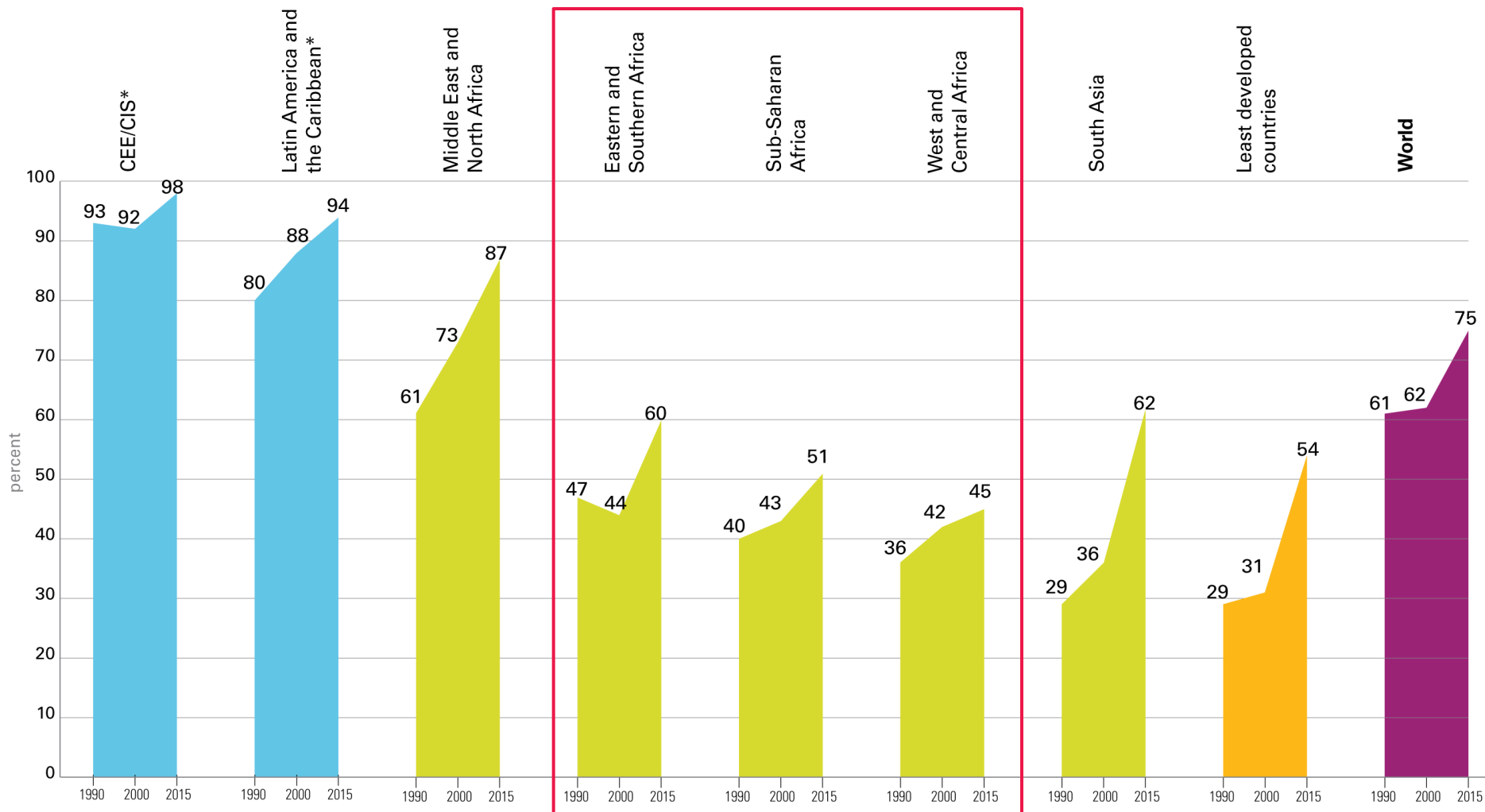
Maternal deaths due to pregnancy or labor complications per 100,000 live births, OECD countries



# Racial/ethnic disparities in US maternal mortality

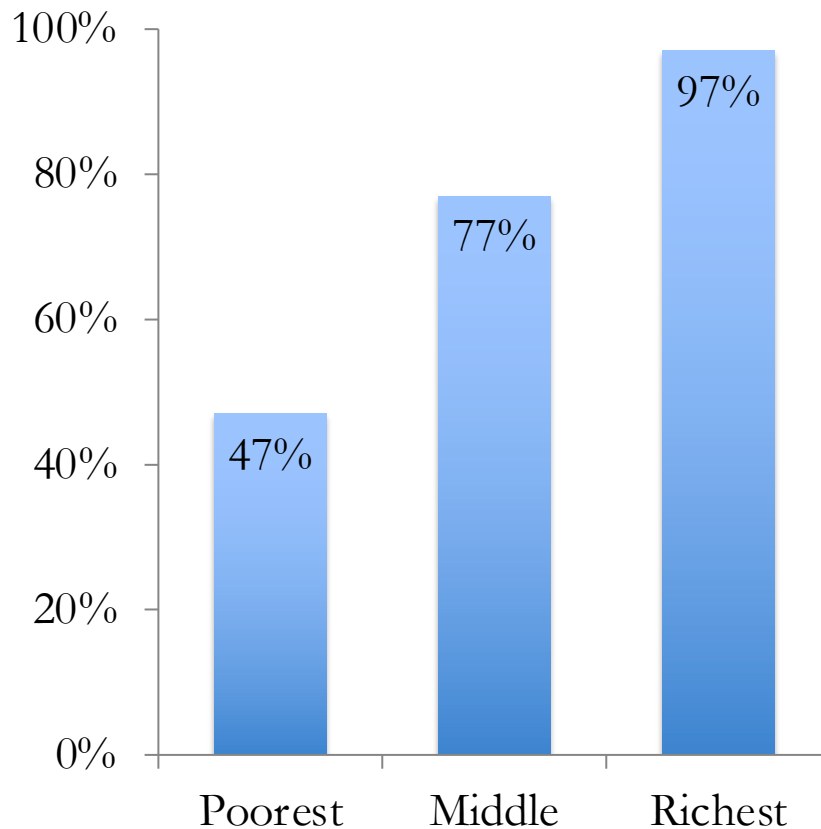


# Births by skilled attendants globally

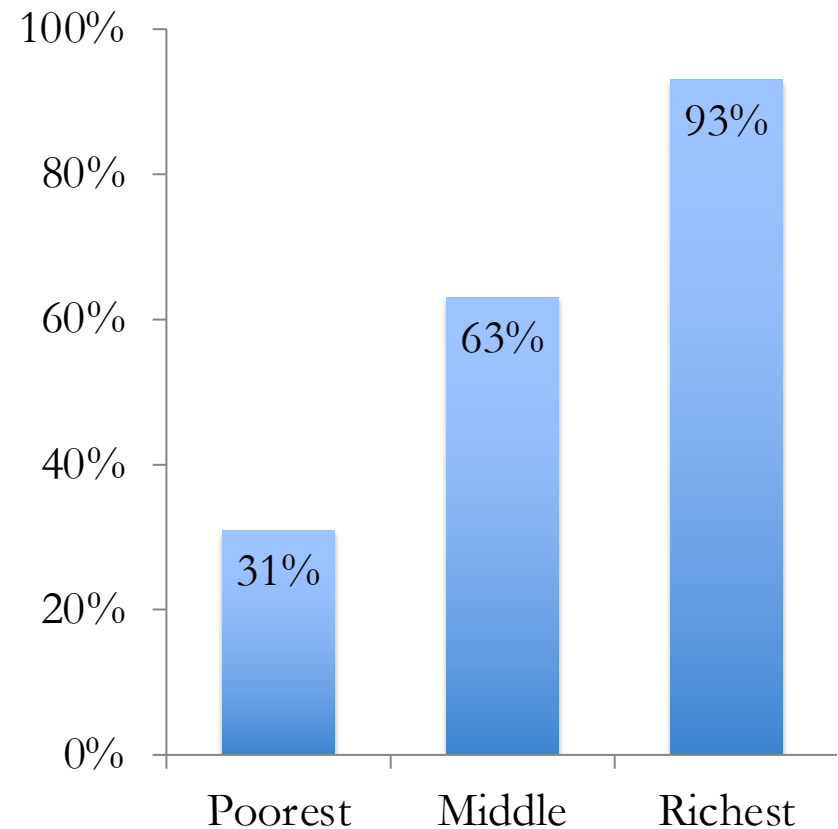


# Births by skilled attendants by wealth quintile

**Ghana**



**Kenya**



# Choosing the outcome

- What outcomes are you interested in?

# Choosing the outcome

- How do you express your outcomes?
  - Expression in terms of favorable events or adverse events
    - Mortality or survival, % that got a mammogram or not...
  - Favorable events or characteristics desirable and promoted through public health action.
  - Adverse events or characteristic undesirable, so public health goal is reduction or elimination.
  - Expressing relative disparities in terms of adverse events would enhance the consistency and comparability of relative measures of disparity across indicators and over time.

# Choosing the outcome

- Distinguish between the construct (e.g., health) and how you will measure that construct (e.g., self-rated health, stroke, etc.)
- More on this later

# Choosing the way to compare outcomes

- Absolute differences and relative differences in health
    - Absolute: how many fewer cases per person
    - Relative: ratio of rates or prevalence per person
  
  - 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)
-

# Choosing the way to compare outcomes

## Relative measures

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

## Absolute measures

- Risk difference
- Rate difference
- Years of life lost

# Choosing the way to compare outcomes

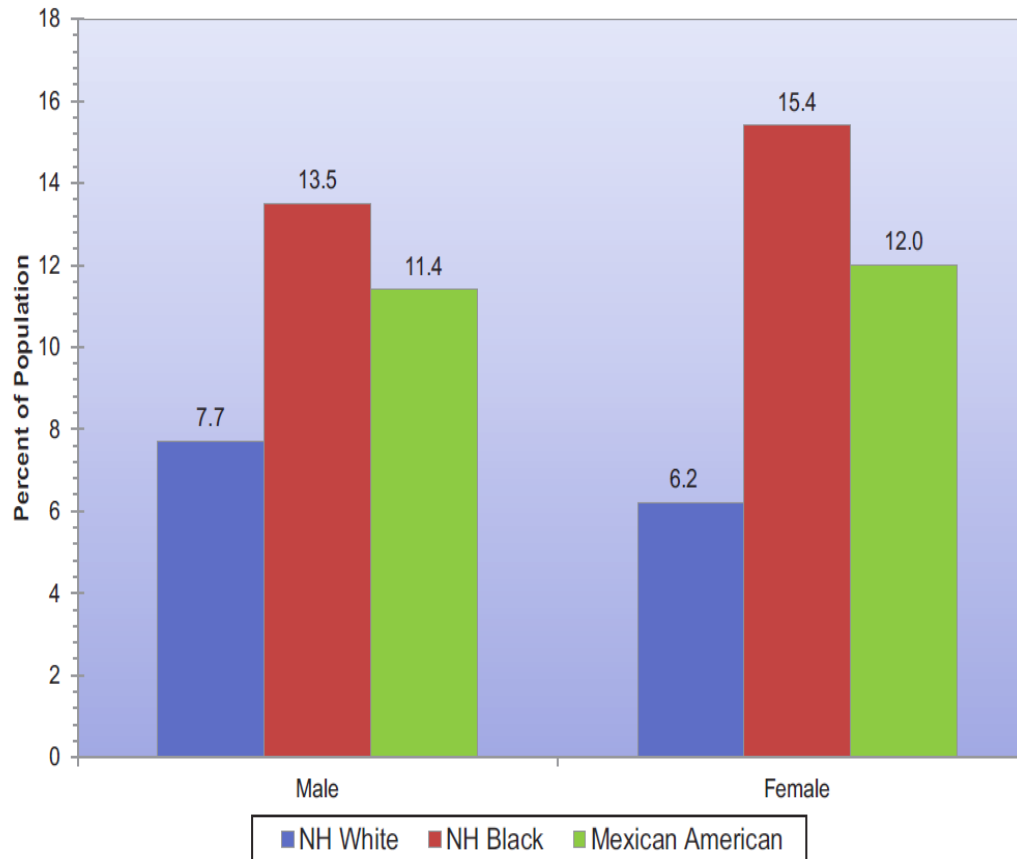
## ■ *Absolute or relative differences*

- An absolute measure of disparity is a simple arithmetic difference between a group rate and a specified reference point. An absolute measure of disparity is expressed in the same units as the rates themselves.
  - Simple difference = rate of interest – reference point =  $R_i - R_r$
  - Can be expressed as a risk or rate difference
- A relative measure of disparity expresses the difference between rates in terms of the chosen reference point. The percentage difference expresses the simple difference from the reference point as a percentage of the reference point.

$$\text{Percentage difference} = \frac{\text{rate of interest} - \text{reference point}}{\text{reference point}} \bullet 100 = \frac{R_i - R_r}{R_r} \bullet 100$$

Keppel et al. 2005;

# Relative vs absolute comparisons of the groups?



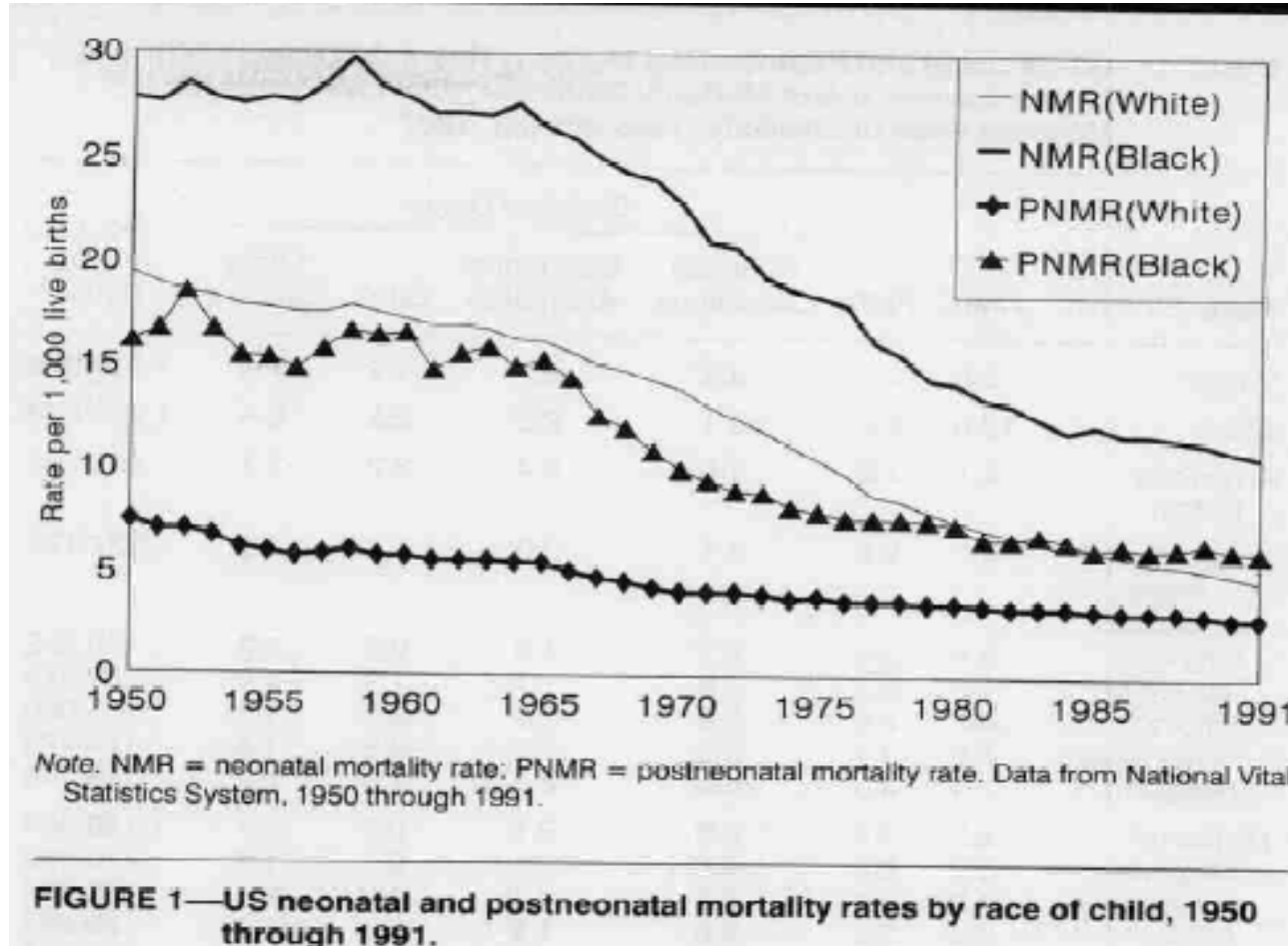
**Relative** effects approach: black men have  $13.5/7.7=1.75$  **times** the prevalence of diabetes as white men, or 75% higher prevalence. Women: 2.48 times

**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?

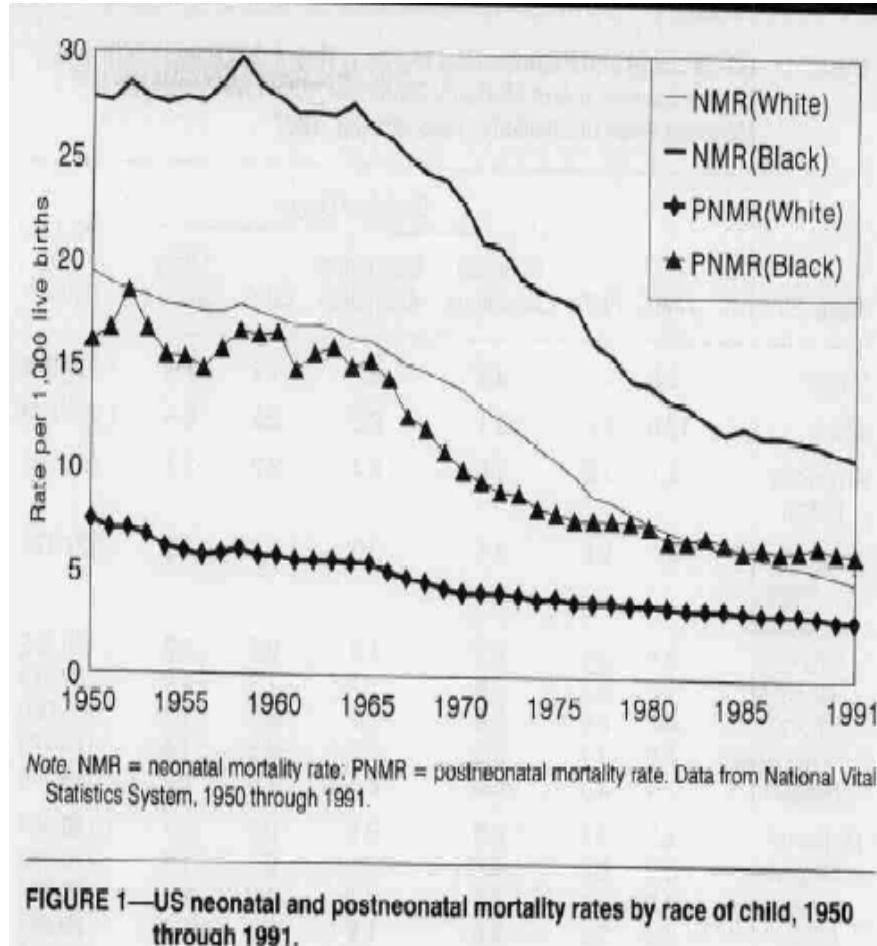
Age adjusted prevalence of physician diagnosed diabetes mellitus, NHANES: 2007-2010

# Trends in infant mortality



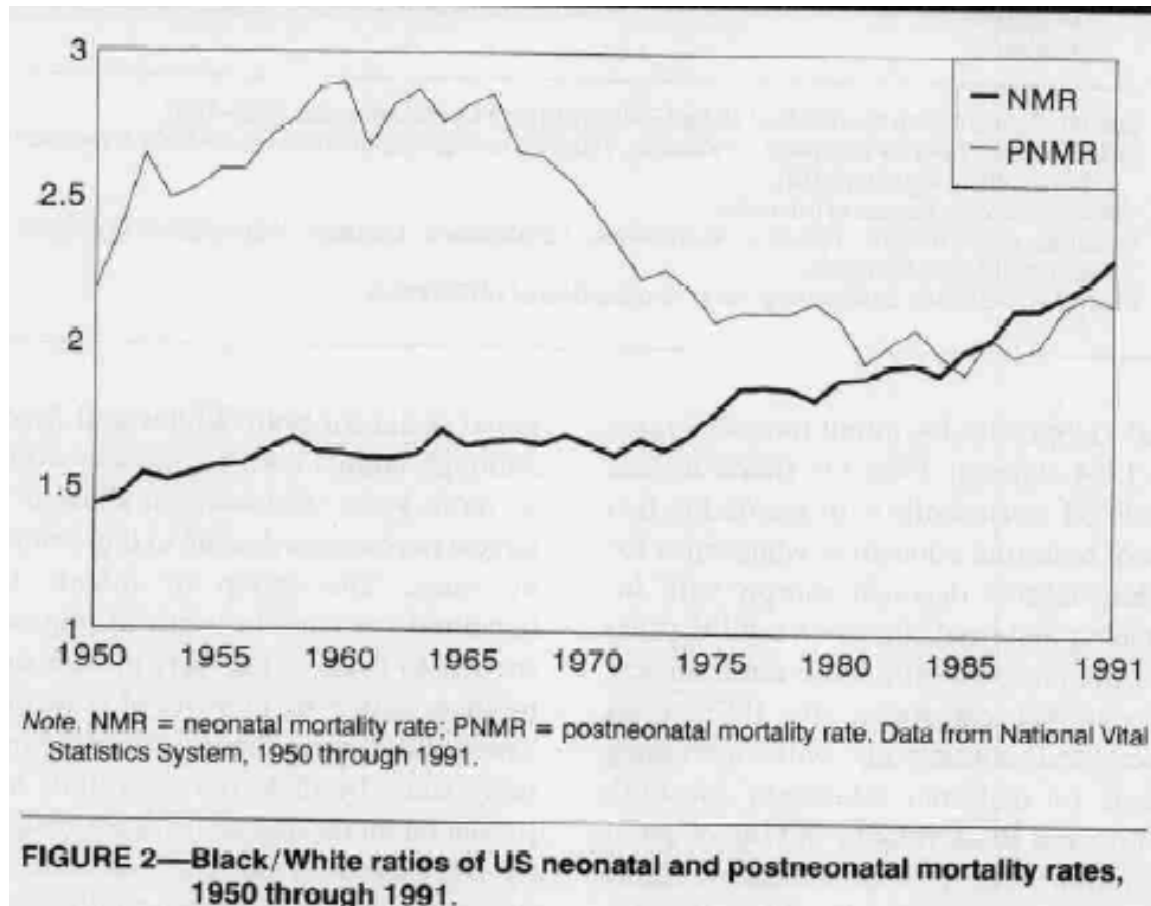
Did the black white disparity in infant mortality increase or decrease?

# 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

# Absolute vs relative Changes

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

# Effect measure modification

- 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.

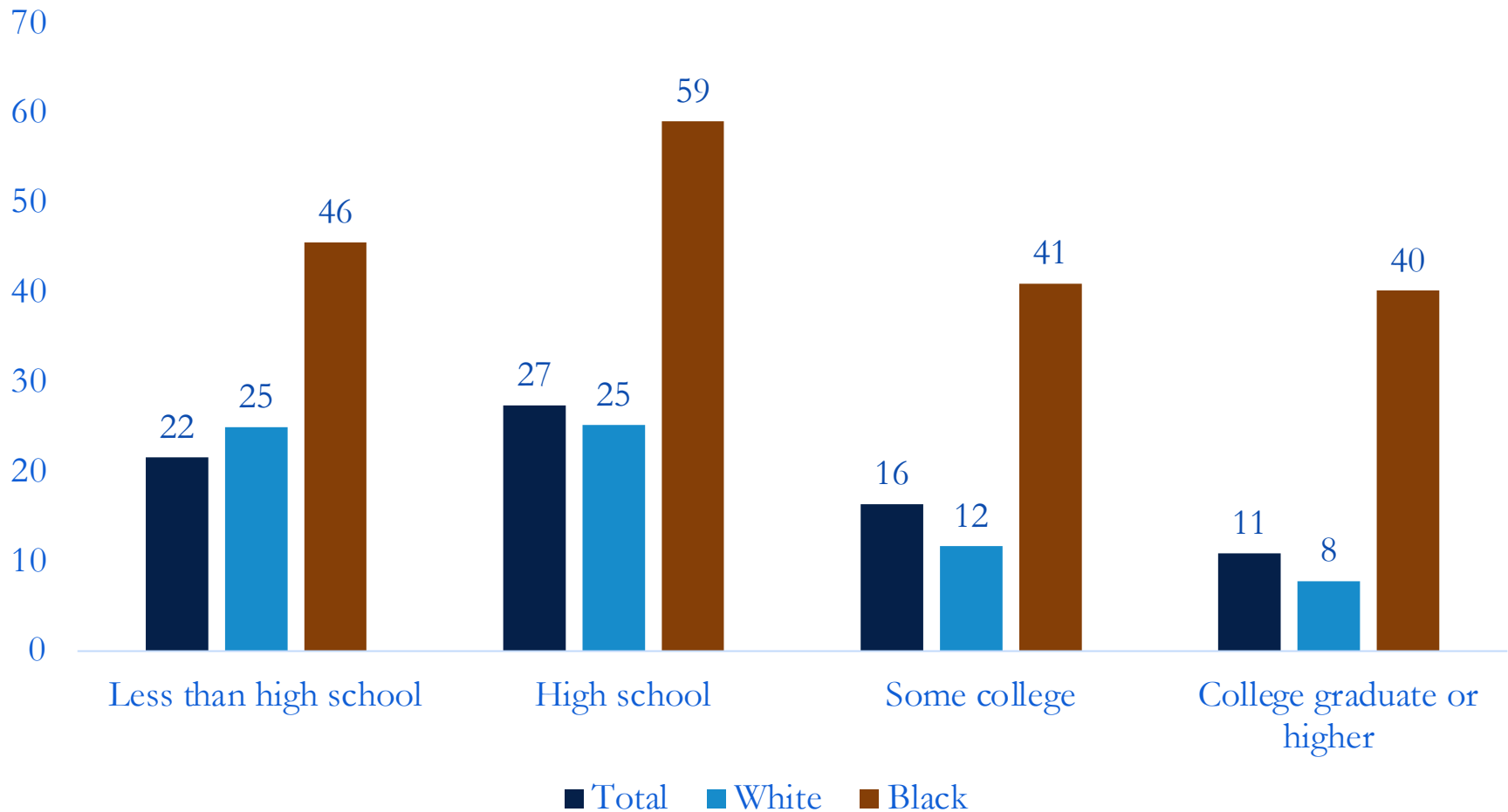
# Effect modification vs interaction

- Effect measure modification occurs when the measure of an effect changes over values of some other variable (which does not necessarily need to be a causal factor). Used to identify populations that are particularly susceptible to the exposure of interest.
- Interaction occurs when two exposures contribute to the causal effect of interest, and they are both causal factors. Generally studied in order to clarify etiology.
- “Interaction is defined in terms of the effects of 2 interventions whereas effect modification is defined in terms of the effect of one intervention varying across strata of a second variable. Effect modification can be present with no interaction; interaction can be present with no effect modification. “

VanderWeele TJ. On the Distinction Between Interaction and Effect Modification. *Epidemiology*. 2009;20(6):863-871

---

# Racial/ethnic disparities in US maternal mortality by education



# “Guidelines” for Measuring Disparities

- When disparities are measured, the reference point should be explicitly identified and the rationale for choosing a particular reference point should be provided.
- If comparisons are made between two groups, the more favorable group rate should be used as the reference point.
- Disparities should be measured in both absolute and relative terms in order to understand their magnitude.
- When relative measures of disparity are employed to compare disparities across different indicators of health, all indicators should be expressed in terms of adverse events.
- .....

---

Keppel K, Pamuk E, Lynch J, et al. Methodological issues in measuring health disparities. *Vital Health Stat* 2. 2005;(141):1-16.

# 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—constructs
  - 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
- Discrimination
- Segregation
- 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

# 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.

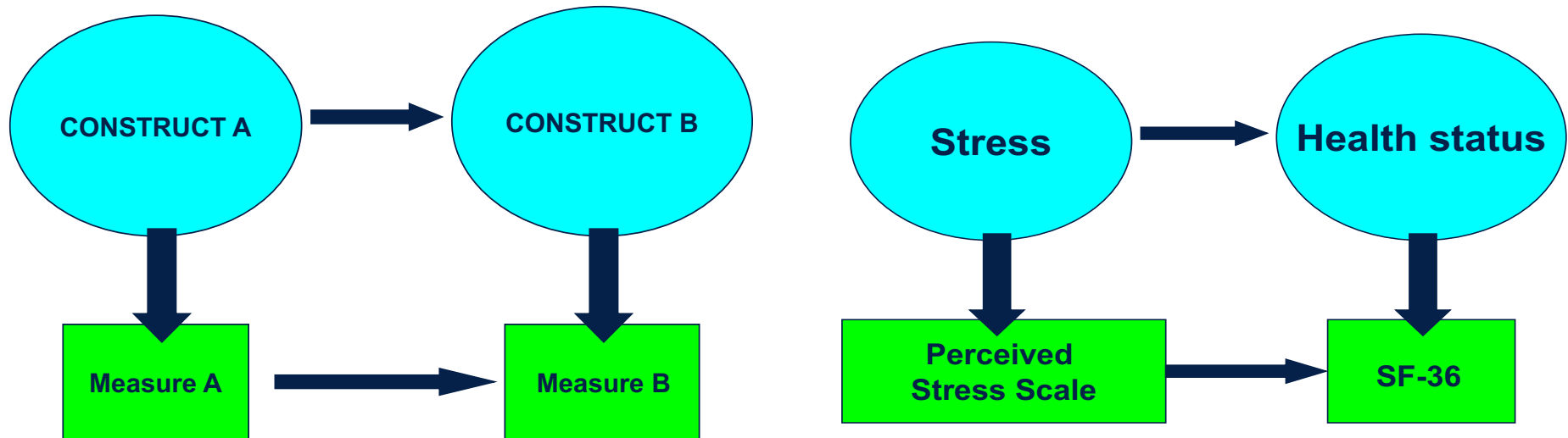
# Concepts/constructs/latent variables

- 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.

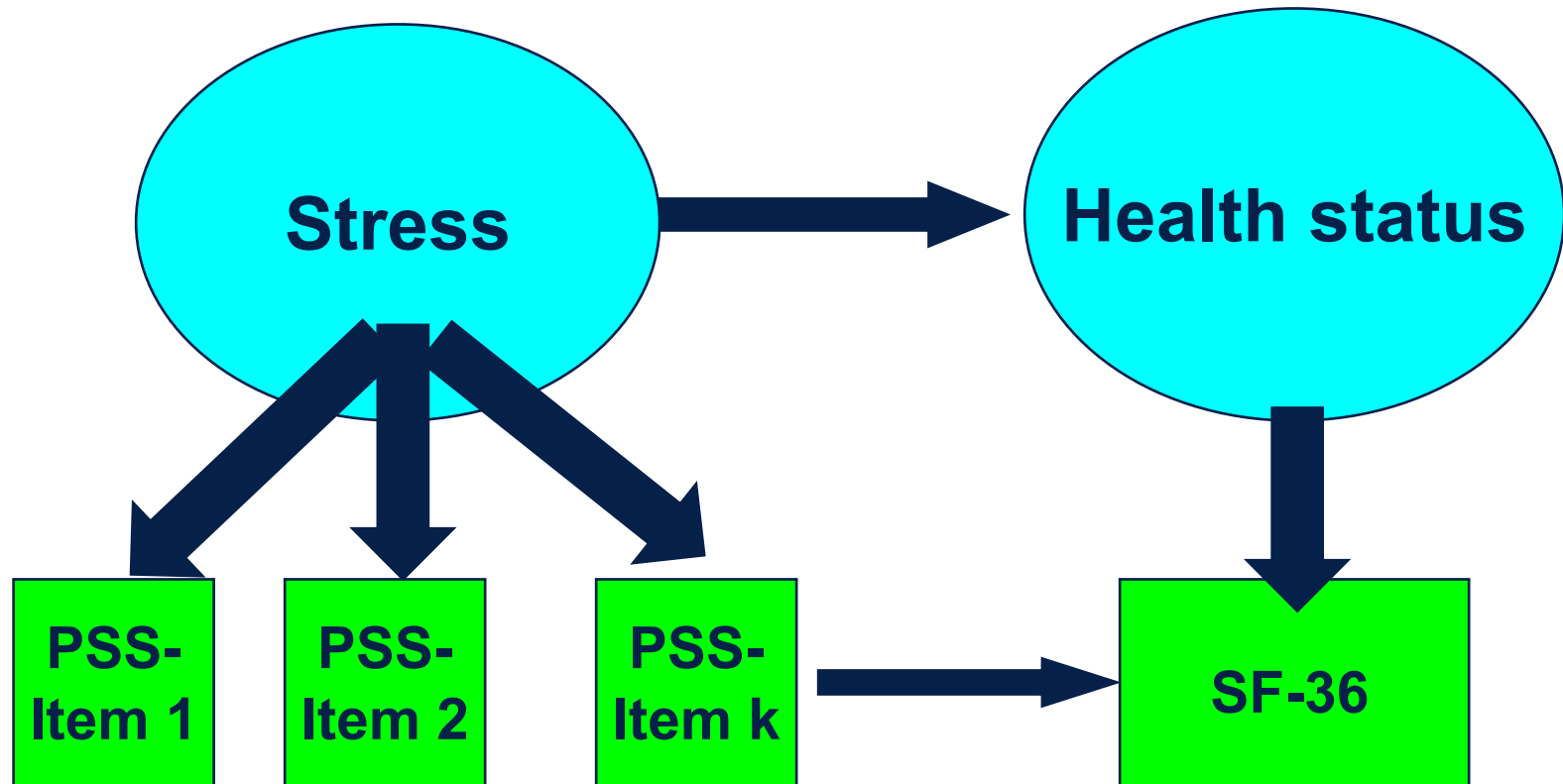
*(USDHHS Theory at a glance)*

- 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

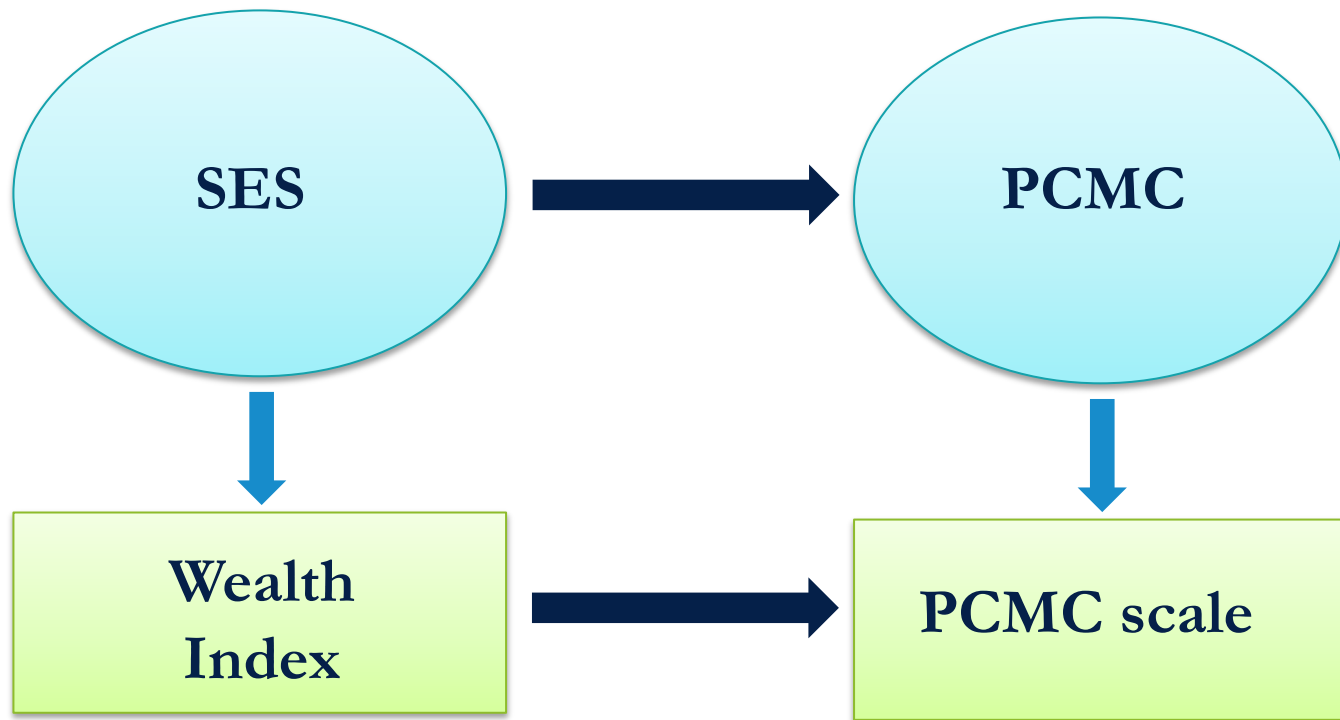
# Depicting latent variables and measures



# Depicting latent variables and measures



# Depicting Latent Variables and Measures

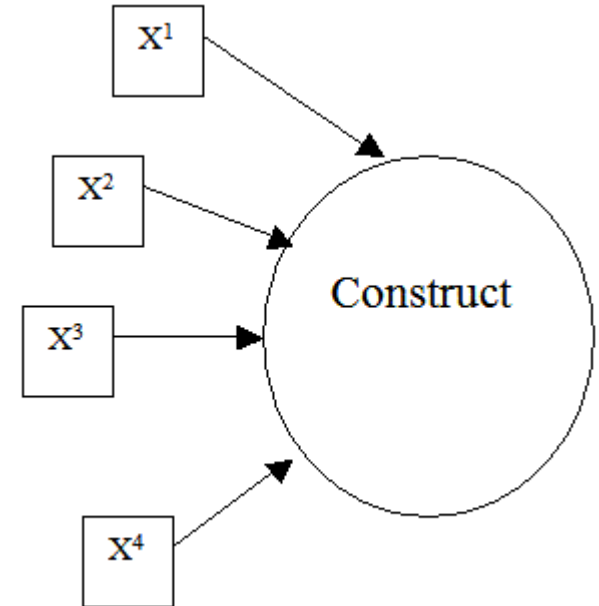
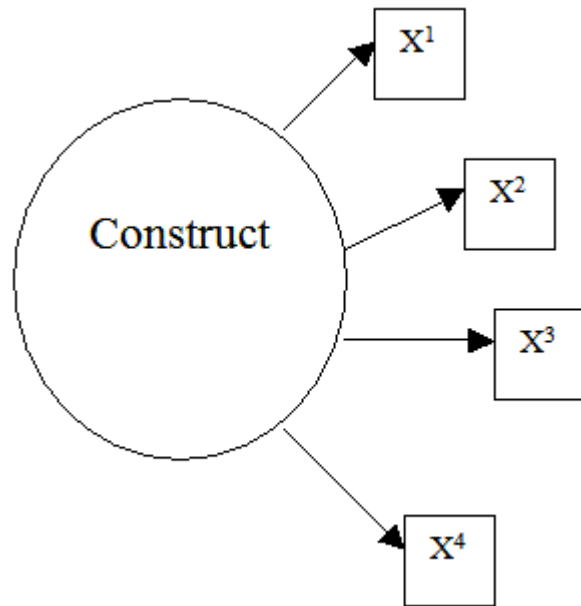


# Scales vs indexes (DeVellis)

- 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

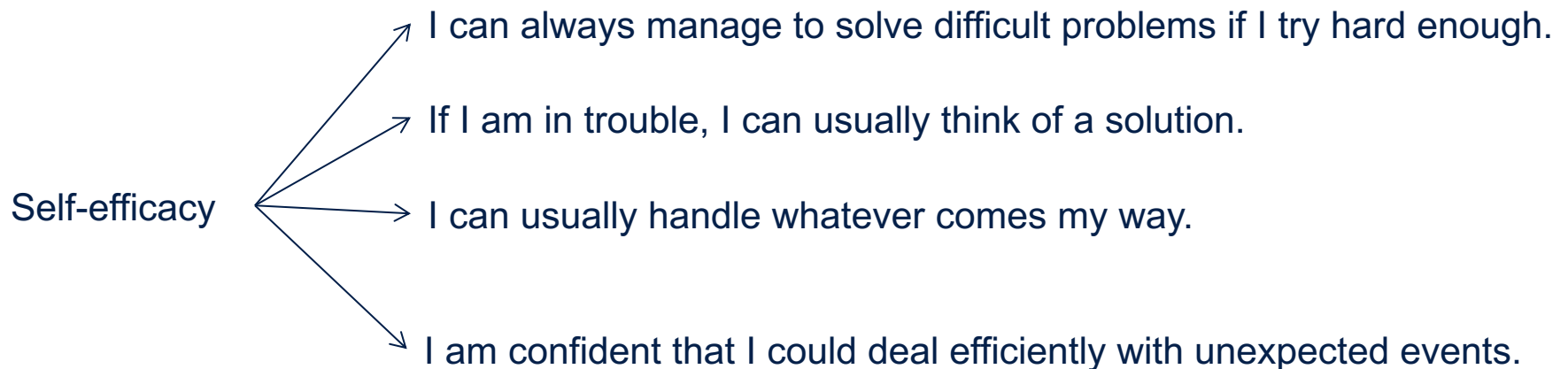
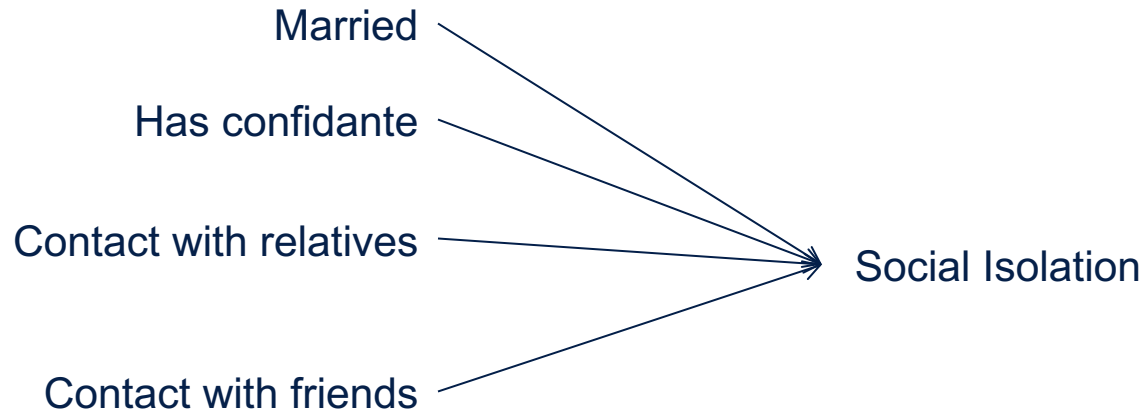
- If you encounter the SEM or psych literature, they often use the terms formative and reflective, corresponding to scale vs index



Scales- Reflective: items reflect the underlying construct

Index-Formative: items form the construct

# Scales vs indexes (per the Devellis framework): Common effect and cause



# How do you select measures for your construct?

# Process of selecting good measures for your studies

- Define your construct

# 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

# Process of selecting good measures for your studies

- Define your construct
- Identify potential measures
-

# 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
-

# Examples of measures of discrimination

- “Experiences of Discrimination” (EOD)
  - “conceptualized as measuring “self-reported experiences of discrimination,” not “perceived discrimination,” because while self-reported experiences must be perceived, not all perceived experiences are necessarily reported, depending upon individuals’ willingness or ability to report them(Krieger,2000).
- The Williams Major and Everyday discrimination measures.
- Single-item discrimination questions??

Krieger N, Smith K, Naishadham D, Hartman C, Barbeau EM. Experiences of discrimination: validity and reliability of a self-report measure for population health research on racism and health. *Soc Sci Med*. 2005;61(7):1576-1596.

- Chronic Work Discrimination and Harassment

# 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

# Developing/Adapting 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/Adapting items

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

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

# Importance of cognitive interviewing

- “Answering questions about exposure to racial discrimination is complex and can be distressing for respondents who may prefer not to recall such memories....
- Cognitive interviews using open-ended probes could yield new information on important content areas in discrimination, and optimal item wording, and framing to develop comprehensive measures that are less distressing and have more power to explain variations in health.”

Stewart AL, Nápoles-Springer AM. Advancing health disparities research: can we afford to ignore measurement issues? *Med Care*. 2003;41(11):1207-1220.

# 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

# 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

# Assessing psychometric adequacy

## ■ 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

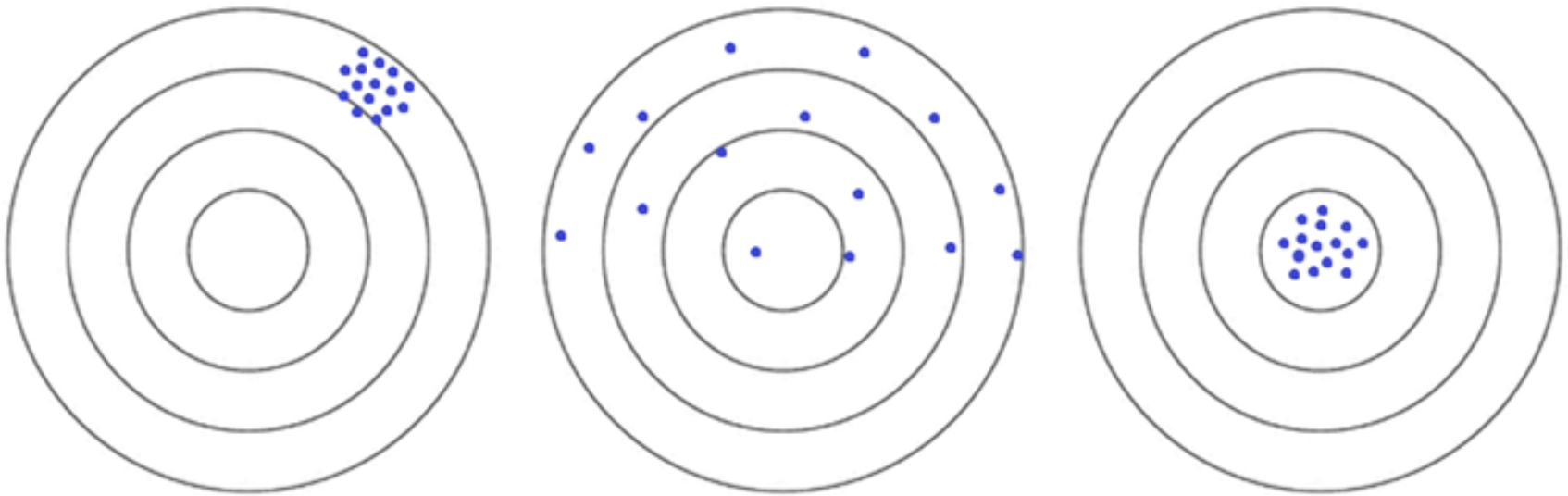
## **Validity:**

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

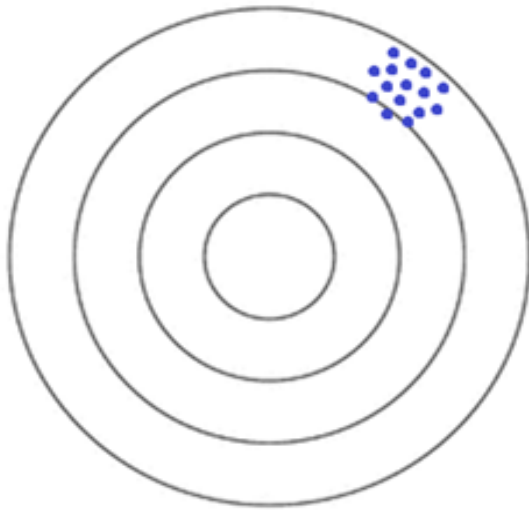
## **Reliability:**

- 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

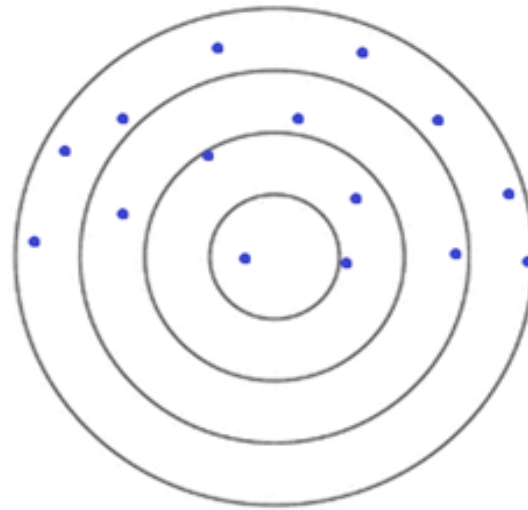
# Valid or reliable?



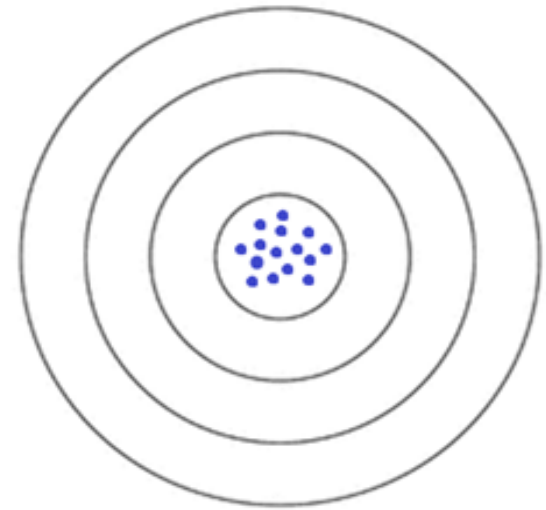
# Valid or reliable?



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 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

# 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?

# Assessing Adequacy of the Experiences of discrimination scale

- In the Krieger paper, what type of adequacy was assessed?
  - What else could they have done to assess the adequacy of the measure of discrimination?
  - The revised version, used in this study, newly asked about the frequency of occurrence, plus added two situations (“getting service in a store or restaurant,” and “getting credit, bank loans, or a mortgage”).
    - How do you think they get to this?

# Summary: Conceptual and psychometric adequacy within and across groups

|                          |  |                                                                       |  |                                                                    |                              |  |  |
|--------------------------|--|-----------------------------------------------------------------------|--|--------------------------------------------------------------------|------------------------------|--|--|
|                          |  | Conceptual                                                            |  |                                                                    |                              |  |  |
|                          |  | Concept is meaningful<br>within one group                             |  | Concept is equivalent<br>across groups                             |                              |  |  |
| Adequacy<br>in One Group |  |                                                                       |  |                                                                    | Equivalence<br>Across Groups |  |  |
|                          |  |                                                                       |  |                                                                    |                              |  |  |
|                          |  | Psychometric properties<br>meet minimal standards<br>within one group |  | Psychometric properties<br>invariant (equivalent)<br>across groups |                              |  |  |
|                          |  | Psychometric                                                          |  |                                                                    |                              |  |  |

\*Group(s) can refer to any sociodemographic strata being compared

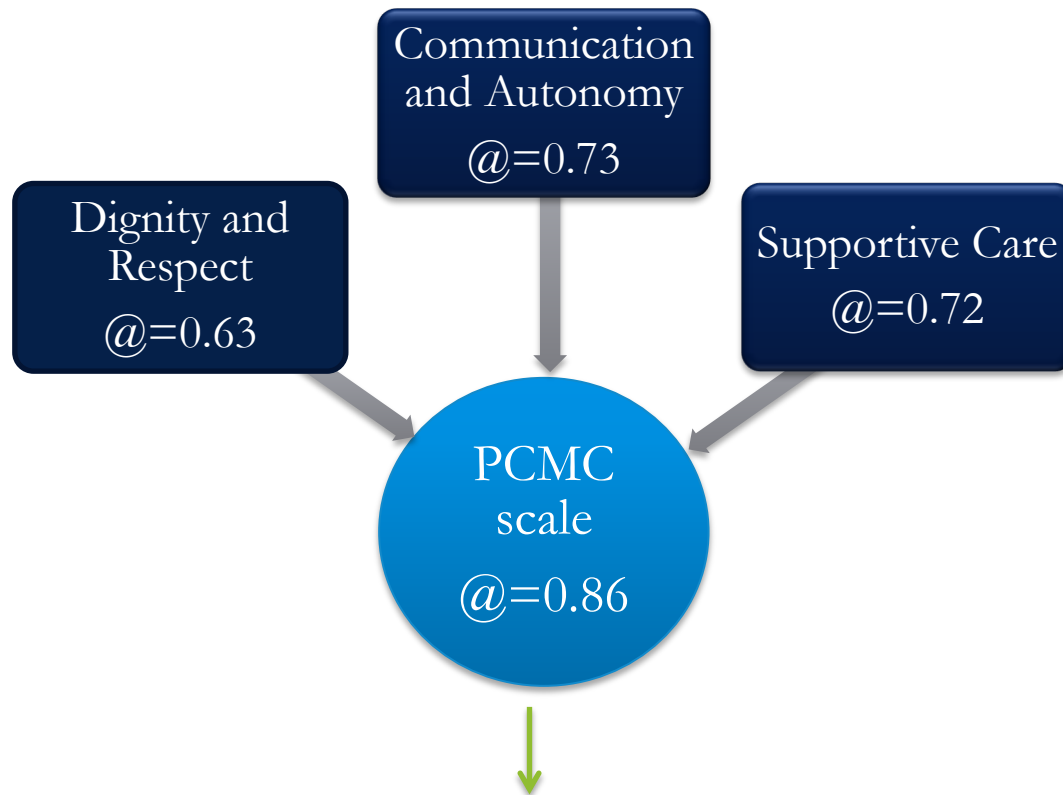
FIG. 1. Conceptual and psychometric adequacy and equivalence within one group and across groups.\*

Stewart AL, Nápoles-Springer AM. Advancing health disparities research: can we afford to ignore measurement issues? *Med Care*. 2003;41(11):1207-1220.

# Summary steps for developing/adapting measures and assessing their adequacy

- 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

# Person-centered maternity care scale developed in Kenya



| PCMC subscales and items              |                           |
|---------------------------------------|---------------------------|
| Sub-scale                             | Items                     |
| <b>Dignity and respect (6)</b>        |                           |
|                                       | Treated with respect      |
|                                       | Friendly                  |
|                                       | Verbal abuse              |
|                                       | Physical abuse            |
|                                       | Visual privacy            |
|                                       | Record confidentiality    |
| <b>Communication and autonomy (9)</b> |                           |
|                                       | Introduce self            |
|                                       | Called by name            |
|                                       | Involvement in care       |
|                                       | Consent to procedures     |
|                                       | Delivery position choice  |
|                                       | Language                  |
|                                       | Explain exams/ procedures |
|                                       | Explain medicines         |
|                                       | Able to ask questions     |
| <b>Supportive Care (15)</b>           |                           |
|                                       | 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                      |

Adaptation of the  
**PERSON-CENTERED  
MATERNITY CARE  
(PCMC) SCALE**  
for women of color in the U.S.

**INTERNAL TEAM REVIEW**

Met with broader internal team to review the already existing scale and adapt questions to better fit the needs of women in the US.



**EXPERT REVIEW SESSION 1**

After our initial review, we joined a CAB meeting in January 2019 and received feedback on what questions to remove/keep from the original scale.



**EXPERT REVIEW SESSION 2**

Met with several CAB members and other experts in the field and added new questions to the scale.



**COGNITIVE INTERVIEWS**

We conducted fifteen 1:1 interviews with pregnant and postpartum women to review the tool and ensure the questions on the survey were reflective of their experiences and the questions were worded clearly.



**PRETESTING**



Eight women tested the survey on iPads to identify any final issues with questions and survey flow.

**ONLINE SURVEY LAUNCH**



Over 300 women have completed the survey online.

**ANALYSIS**



**YOUR FEEDBACK!**

Today, we want to present some of the questions from the survey that we are considering removing.

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?

—

Thank you!



University of California  
San Francisco

Questions? Comments?

[Patience.Afulani@ucsf.edu](mailto:Patience.Afulani@ucsf.edu)

# Discussion

- What dimension of inequality are you interested in?
- What outcome?
- How do you plan to measure your exposure and outcomes
- How will you assess the adequacy of your measure?

# 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.