# PHC 6010
## EPIDEMIOLOGY METHODS I

## MEASUREMENT AND CLASSIFICATION OF DISEASE

Hamisu Salihu, MD, PhD

## Disease misclassification

- **Definition**: incorrect assignment of disease status either in terms of a value along a continuum or categorical class.
- Categorical: misclassification bias
- Continuous: measurement error
- The concept of disease as a continuum

## Validity of a measure

- **Definition**: Is the extent to which an instrument (e.g., a test) measures what it claims to measure (e.g., IQ test).
- The best and most obvious way of appraising validity is to find a criterion (Gold standard) that we know is close to the truth.
- e.g., PSA versus biopsy in diagnosis of prostate cancer

## Validity of a measure

**Types of validity**

Criterion validity: how valid the test is in comparison to the gold standard. How sensitive and specific.

Predictive validity: Association between results of the test and subsequent events. Is a positive result associated with disease or response? E.g. GRE score at entry to graduate program and probability of graduating successfully. How valid is GRE in the prediction.

Concurrent Validity occurs when the criterion measures are obtained at the same time as the test scores. It is demonstrated where a test correlates well with a measure that has previously been validated. E.g., A new Job satisfaction instrument tested against a previously validated questionnaire at the same time.

## Validity of a measure

Construct validity: The extent to which the measurement corresponds to theoretical concepts (constructs) concerning the phenomenon under study. Epidemiologically speaking, you have a variable that is measured by a number of other variables, each purportedly measuring the same thing (construct). Correlations that fit the expected pattern contribute evidence to construct validity

Convergent validity: measures that should be related are in fact related (high coefficient of correlation). Shows that the assessment is related to what it should theoretically be related to (e.g., High GPA in math and high GRE quantitative score is evidence of convergent validity

## Validity of a measure

**Types of validity**

Convergent validity (illustration): Depression versus symptom query (Item I = feeling sad; Item II= feeling guilty; Item III = my life is worthless; Item IV = Insomnia)

| | | | |
|---|---|---|---|
| **1.00** | **.83** | **.89** | **.91** |
| **.83** | **1.00** | **.85** | **.90** |
| **.89** | **.85** | **1.00** | **.86** |
| **.91** | **.90** | **.86** | **1.00** |

**Verdict: High correlation, hence converging on the same construct (depression).**

## Validity of a measure

**Types of validity**

Discriminant validity: Measures that should not be related are in reality not related.

E.g., items on depression measures and locus of mood control.

| | $LOC_1$ | $LOC_2$ |
|---|---|---|
| $r_1$ | 0.12 | 0.09 |
| $r_2$ | 0.04 | 0.11 |

Discriminant correlations are always lower than convergent correlations. The correlations provide evidence that the two sets of items discriminate.

The two sets of measures each seem to be related to different constructs

## Discriminant validity

- Consider a researcher developing a new scale designed to measure Narcissism. They may want to show discriminant validity with a scale measuring Self-esteem.
- Narcissism and Self-esteem are theoretically different concepts, and therefore it is important that the researcher show that their new scale measures Narcissism and not simply Self-esteem.

## Validity of a measure

Combining convergent validity and discriminant validity together gives us a more comprehensive picture termed construct validity.

| | $r_1$ | $r_2$ | $r_3$ | $LOC_1$ | $LOC_2$ | $LOC_3$ |
|---|---|---|---|---|---|---|
| $r_1$ | 1.0 | .83 | .89 | .02 | .12 | .09 |
| $r_2$ | .83 | 1.0 | .85 | .05 | .11 | .03 |
| $r_3$ | .89 | .85 | 1.0 | .04 | .00 | .06 |
| $LOC_1$ | **.02** | **.05** | **.04** | 1.0 | .84 | .93 |
| $LOC_2$ | **.12** | **.11** | **.00** | .84 | 1.0 | .91 |
| $LOC_3$ | **.09** | **.03** | **.06** | .93 | .91 | 1.0 |

## Validity of a measure

Face validity: Examine (usu by observation) and then deciding "on face value" whether it seems like a good translation of the construct.

- E.g., math tests to measure math ability.
- Limitation: Weak
- Reinforcement: send it to a sample of experts and read their report

## Validity of a measure

Consensual validity:

the use of agreement between 2 or more subject-matter authorities as providing evidence for the validity of the phenomenon.

Example: If numerous experts accept the indicator as valid then the measurement satisfies consensual validity requirements.

## Validity of measurements

- The ideal diagnostic test should discriminate unerringly between diseased and healthy individuals (perfectly accurate)
- Such a test is rare
- Hence, the search for a perfect test yields a series of tests that achieve partial accuracy
- The need therefore, arises to rate diagnostic tests in an objective manner
- In other words, how do we decide between two tests based on their relative accuracy

## Validity measurement

Sensitivity and specificity: the validity of a test is oftentimes used to determine whether it should be used to classify individuals as diseased or non-diseased.

Sensitivity: proportion of correct results among people who actually have the disease (or condition).

Specificity: proportion of correct results among people who are actually free of the disease (or condition).

False negative: proportion of incorrect results among people who actually have the disease (or condition).

False positive: proportion of incorrect results among people who are actually free of the disease (or condition).

## Calculating Prevalence or Incidence

| | Disease present | Disease absent | |
|---|---|---|---|
| Positive test | | | a + b (all +) |
| Negative test | | | c + d (all -) |
| | a + c (all diseased) | b + d (all non-diseased) | |

Prevalence = a + c/a + b + c + d

= those with disease/those at risk for disease

## Validity of screening tests

- Purpose of a *screening test* is to identify individuals or groups who have a high probability of having a particular disease or other attribute.
- Definition of a screening test (1951; US Commission on Chronic Illness): *the presumptive identification of unrecognized disease or defect by the application of tests, examinations or other procedures which can be applied rapidly. Screening tests sort out apparently well persons who probably have a disease from those who probably do not. A screening test is not intended to be diagnostic.*

## Screening for Diagnosing Disease

**How do we judge the usefulness of a test?**

1. How often is the test result correct for persons in whom the disease is known to be present? SENSITIVITY
2. How often is the test result correct for persons in whom the disease is known to be absent? SPECIFICITY

**Sensitivity & Specificity are measures of test function.**

- Independent of disease prevalence.
- If a test is not sensitive, it will fail to detect disease in some diseased subjects. FALSE-NEGATIVE
- If a test is not specific, it will falsely identify disease in whom the disease is not present. FALSE-POSITIVE

## Calculating Sensitivity & Specificity

| | Disease present | Disease absent | |
|---|---|---|---|
| Positive test | | | a + b (all +) |
| Negative test | | | c + d (all -) |
| | a + c (all diseased) | b + d (all non-diseased) | |

Sensitivity = a/a + c OR tp/tp + fn

- Positive in disease;
- Negative test 'rules out' the disease.

False-negative rate = c/a + c

- Type II error (beta)

Specificity = d/b + d OR tn/tn + fp

- Negative in health
- Positive test 'rules in' the disease.

False-positive rate = b/b + d

- Type I error (alpha)

## Sensitivity and Specificity

- Important notes:

**False-positive rate**

- Proportion of disease-free cases that test positive.
  = number of false positives/total disease-free
- = 1-specificity of the test
- In statistical hypothesis testing, this fraction is given the symbol $\alpha$, and 1- $\alpha$ = specificity.
- Hence, as the specificity increases, the probability of type 1 errors is lowered.
- But this could raise the probability of type II errors (false negatives that reject the alternative hypotheses when it is true)

## Sensitivity and Specificity

**False-negative rate**

- Proportion of diseased cases that test negative.
  = number of false negatives/total diseased cases
- = 1-sensitivity of the test
- In statistical hypothesis testing, this fraction is given the symbol β, and 1- β = power.
- Hence, as the sensitivity increases, the probability of type II errors is lowered.
- But this could raise the probability of type I errors (false positives that reject the null hypothesis when it is true)

## Applying Sensitivity & Specificity

- A screening test used to 'rule out' a diagnosis should have a high degree of sensitivity.
- A confirmatory test used to 'rule in' a diagnosis should have a high degree of specificity. (SpIn). In order to reduce false-positives (α), specificity should be high (because α = 1- specificity).