How to evaluate diagnostic tests?

Prof. Margus Lember
Department of Internal Medicine
University of Tartu, Estonia

Why physicians use tests?

- Diagnosis- most common
- Monitoring- has someting changed?
- Treatment planning
- Prognosis
- Diminishing anxiety of patients (and doctors themselves)
- etc

Why to use diagnostic tests?

- Differentiate normal from abnormal
- An ideal test/measurement/assessment of a clinical symptom separates the diseased and disease-free groups
- In practice: almost always overlap

- **A test= measurement**

  Presence/absence of a symptom

  Presence/absence of a sign

  Laboratory or radiological measurement

  should increase probability of the presence or absence of a disease

Tests are used for

- **screening**: for asymptomatic disease (PSA, mammography, fasting blood sugar, PAP smear, chest x-rays) or case-finding while testing patients for diseases unrelated to their complaint
- **diagnostics**: to confirm or rule out a diagnosis in symptomatic patients or after positive screening test

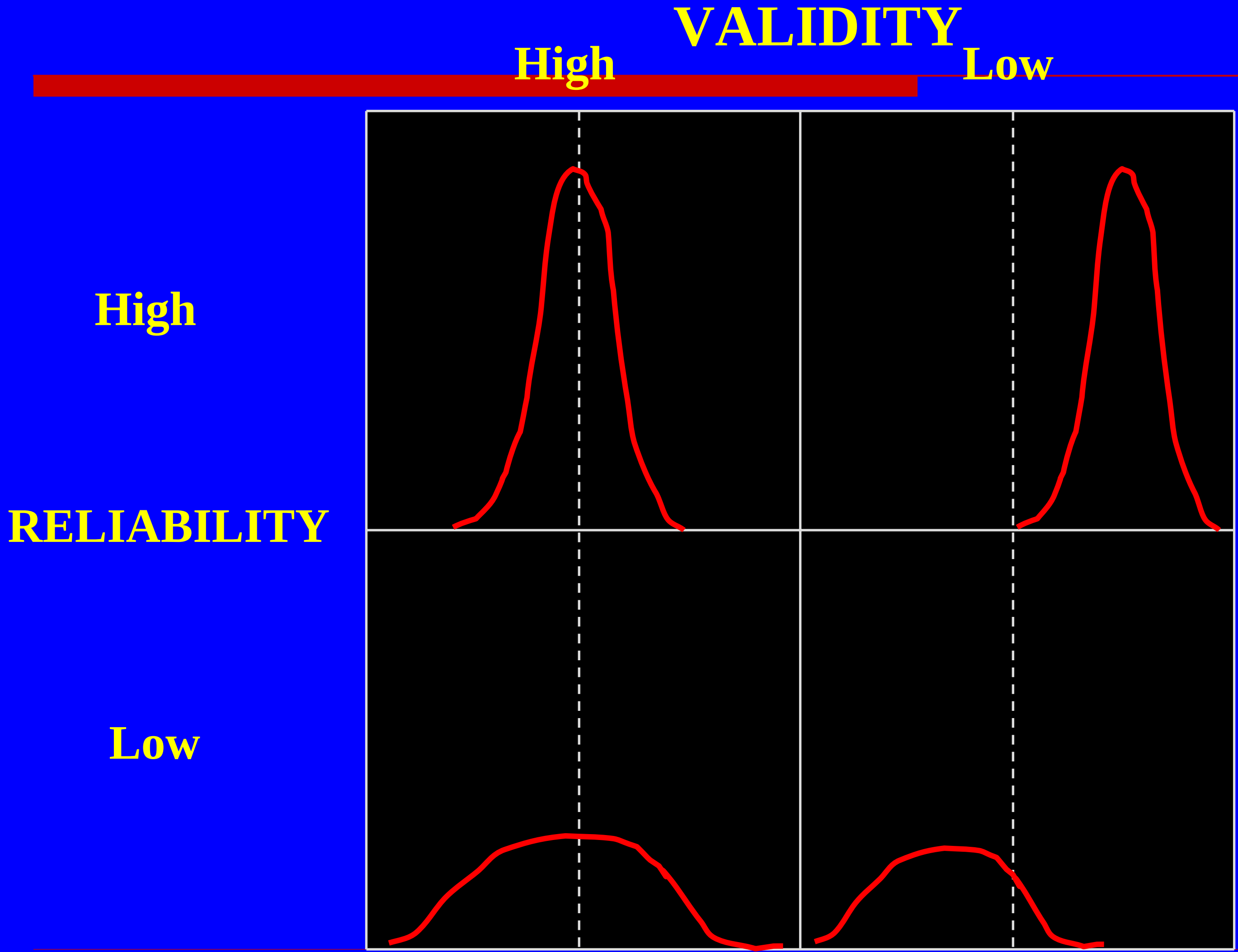
Quality of measurements is characterized by

- Validity (accuracy)
  - Does it measure what it is intended to?
  - Lack of bias
- Reliability (reproducibility, precision, consistency) of measurements

Validity (accuracy) of a test

- Test is valid if it detects most people with the target disorder and excludes most people without the disorder

- What is the “golden standard” against which to validate a test? (biopsy, not always possible)


VALIDITY
High
Low
High
RELIABILITY
Low

Sources of variation: biologic or true variation

- between individuals
- within individuals (e.g., diurnal variation in BP)
- distinguish between individuals with normal & abnormal test results

Biologic Variation of Human Populations

- **<u>Bimodal curve</u>:** distribution with 2 peaks
  - Relatively easy to separate most of the population into 2 groups (e.g., ill & not ill; have condition or abnormality & do NOT have condition or abnormality)
  - Some fall into "gray zone" – may belong to either curve
  - Most human characteristics are NOT distributed bimodally

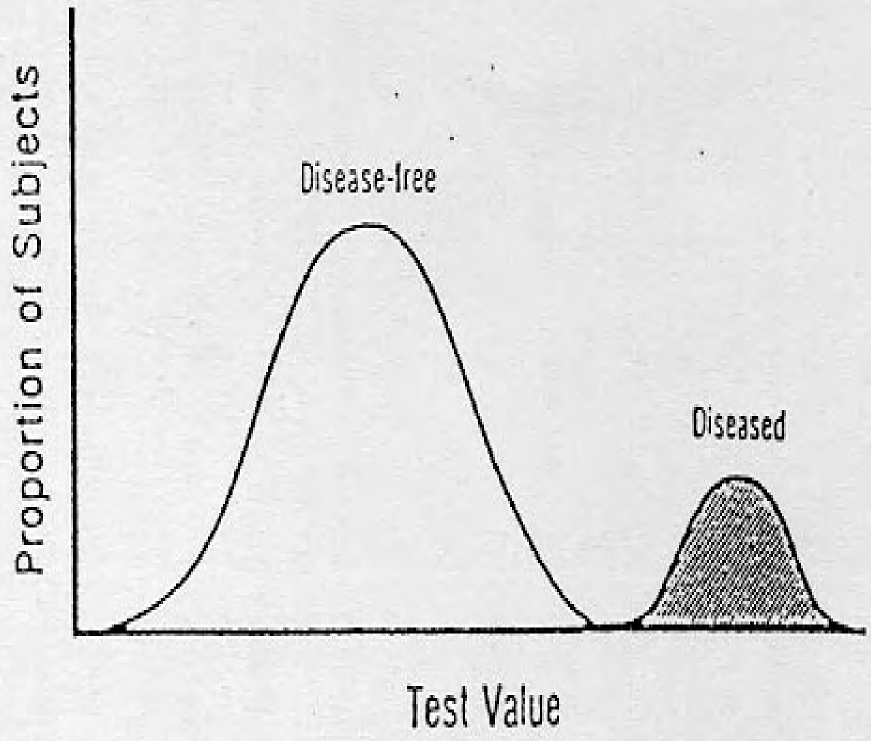


Fig. 16.2. Two nonoverlapping distributions

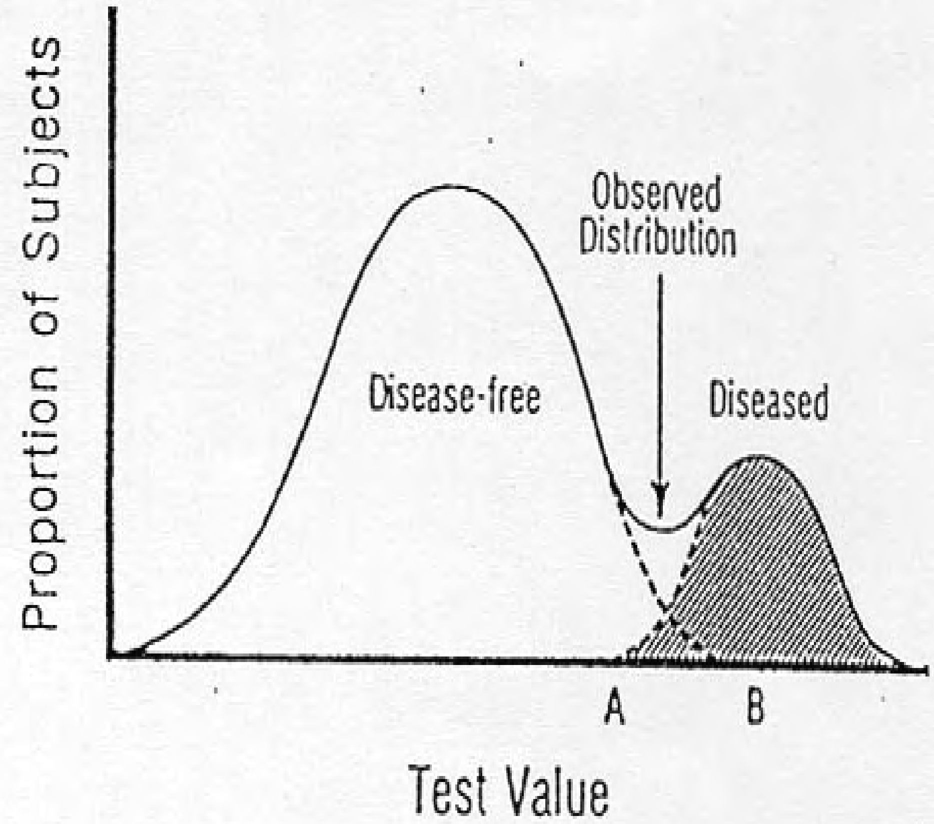


Fig. 16.5. Bimodal distribution

Biologic Variation of Human Populations

- **<u>Unimodal curve</u>:** distribution with 1 peak
  - Must set a cutoff level to distinguish those with condition & those without condition
  - Relatively easy to distinguish the extreme values of abnormal & normal
  - Uncertainty about those cases in the gray zone

Sources of variation: Measurement error

- random error vs systematic error (bias)
- method (measuring instrument)
- Observer (even “harder” methods- x-ray, biopsies- have “soft” aspects)

- Spectrum bias- study population has a different clinical spectrum (more advanced cases etc)
- Selection bias- if preselected patients are used for calibration, but the test is used e.g. in primary care

Tests with Dichotomous Results (+ or -)

- Results of a dichotomous test with actual disease status:

Population

Test Results	With Disease	Without Disease
Positive	True +	False +
Negative	False -	True -

Sensitivity = TP / (TP + FN)   Specificity = TN / (TN + FP)

Sensitivity

- Proportion of the truly diseased persons who are identified as diseased by the test (true-positives) among all the diseased people

Specificity

- Proportion of truly non-diseased by the test (true-negatives) among all non-diseased

Lactose Test U-gal

- **Golden standard for hypolactasia: activity of lactase in small intestinal mucosa**

Test Results	Hypolactasia	Normolactasia
Positive	TP 22	FP 2
Negative	FN 3	TN 30

Sensitivity = TP / (TP + FN)   Specificity = TN / (TN + FP)

Lactose Test U-gal

- Sensitivity 88%   Specificity 94%

Test Results	Hypolactasia	Normolactasia
Positive	TP 22	FP 2
Negative	FN 3	TN 30

Sensitivity = TP / (TP + FN)   Specificity = TN / (TN + FP)

Sensitivity and specificity

Assess correct classification of:

- People with the disease (sensitivity)
- People without the disease (specificity)

- **Sensitivity and specificity are characteristics of a test**

- For screening- sensitivity is more important

- For confirming diagnosis- specificity is more important

Predictive value (PV)

- **More relevant to clinicians and patients**

- **Positive PV**: probability of the test positive person truly having the disease
- (what proportion of patients who test positive actually have the disease in question?)

Predictive value

- **Negative PV**: probablity of the test negative person not truly having the disease
- (What is the probability that this patient does not have the disease?)

Predictive value

- **Affected by the prevalence** of the disease (higher the prevalence of the disease, greater the predicitive values)
- Thus, usefulness depends on the context in which the test is used
- Differences while using in community and in clinical setting

Lactose Test U-gal

Positive Predictive Value **(PPV) = TP : (TP+FP)**

Negative Predictive Value **(NPV) = TN : (TN+FN)**

Pre-test probability (prevalence) (TP + FN):(TP+FN+FP+TN)

Test Results	Hypolactasia	Normolactasia
Positive	TP 22	FP 2
Negative	FN 3	TN 30

Lactose Test U-gal

Positive Predictive Value (PPV) = 22 / 24 = 92%

Negative Predictive Value (NPV) = 30/33 = 91%

Pre-test probability (prevalence) 25/57=44%

Test Results	Hypolactasia	Normolactasia
Positive	TP 22	FP 2
Negative	FN 3	TN 30

What if prevalence 16%?

Positive Predictive Value (PPV) = 22 / 30 = 73%

Negative Predictive Value (NPV) = 120/123 = 98%

Pre-test probability (prevalence) 25/153=16%

Test Results	Hypolactasia	Normolactasia
Positive	TP 22	FP 8
Negative	FN 3	TN 120

What if prevalence 1%?

Positive Predictive Value (PPV) = 22 / 177 = 12%

Negative Predictive Value (NPV) = 2320/2323 = 99.9%

Pre-test probability (prevalence) 25/2500=1%

Test Results	Hypolactasia	Normolactasia
Positive	TP 22	FP 155
Negative	FN 3	TN 2320

- A major factor in improving the positive predictive value of a test is to limit the use of the test to the patients who are likely to have the disease in question

Predictive Value of a Test

- PV is affected by 2 factors:
  - **(1) Prevalence of disease in population tested**
  - **(2) Specificity of test when disease is infrequent**

- The difference in the PV among different patient samples is the main reason that:

- Hospital doctors believe GPs miss "easy" diagnoses
- GPs believe hospital doctors overinvestigate
- Misunderstandings with oncologists

Relation of Predictive Value to Disease Prevalence

- A screening program is most productive & efficient if it is directed to a high-risk target population
  - If screen a total population for a relatively infrequent disease $\Rightarrow$ waste resources & yield few previously undetected cases for the amount of effort involved

- Demonstrates that test results must be interpreted in context of prevalence of the disease in the population from which the subject originates

Relation of Predictive Value to Disease Prevalence

- Same test can have a VERY different PV when administered to a high-risk (high prevalence) population or to a low-risk (low prevalence) population
  - Clinical implications

Choice of cut-point

If higher score increases probability of disease

- Lower cut-point:
  - increases sensitivity, reduces specificity
- Higher cut-point:
  - reduces sensitivity, increases specificity

Considerations in selection of cut-point

Implications of false positive results

- burden on follow-up services
- labelling effect

Implications of false negative results

- Failure to intervene

Likelihood ratio

- Takes into account both the pre-test probability (prevalence) and post-test probability (positive and negative predictive values).
- The post-test odds divided by the pre-test odds is the likelihood ratio

Likelihood ratio

- Likelihood ratio (LR) = $\frac{\text{sensitivity}}{\text{1-specificity}}$
- = $\frac{\text{true positive rate}}{\text{false positive rate}}$

- More accurate tests do not necessarily improve management!