Diagnostic Test Accuracy Studies

Paschalis Paschos MD, MSc, PhD

Clinical Research & Evidence-Based Medicine Unit

Aristotle University of Thessaloniki

MSc in Health Statistics and Data Analytics

Medical School

Aristotle University of Thessaloniki

"…before we can treat **patients** appropriately we must get the **diagnosis** right…"

Definition of *diagnosis*

Merriam-Webster SINCE 1828

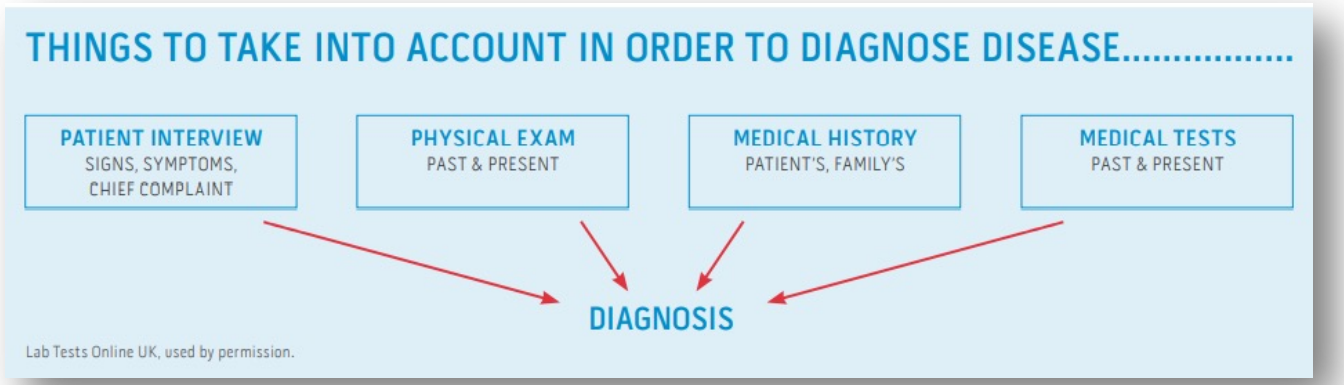
1 a : the art or act of identifying a disease from its signs and symptoms

"Absolute certainty in diagnosis is unattainable, no matter how much information we gather, how many observations we make, or how many tests we perform…"

Kassirer, M.D. N Engl J Med 1989

What is Diagnosis?

Testing during diagnostic process refers to any method for collecting additional information about the current health status of a patient.

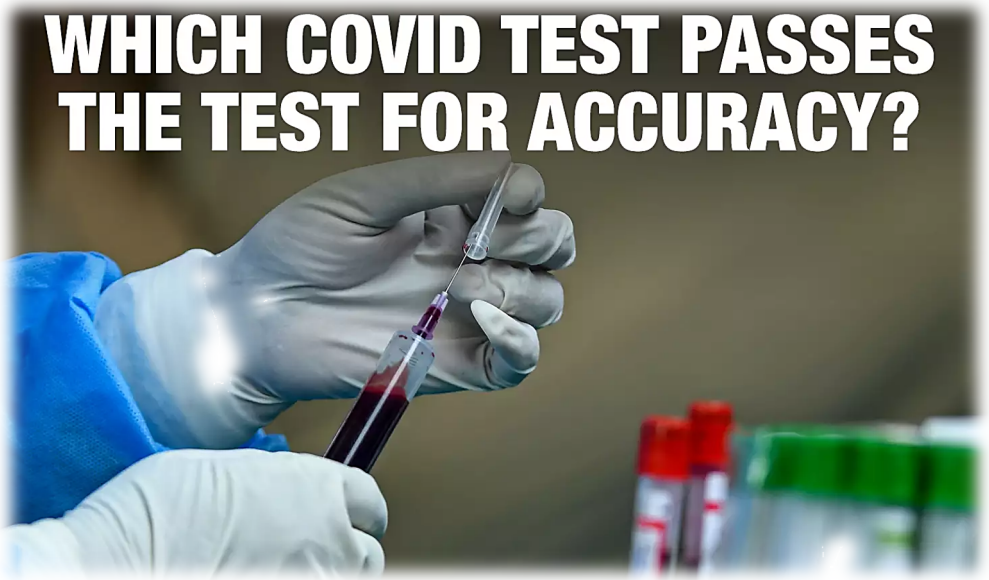


THERE ARE SEVERAL REASONS WHY DOCTORS ORDER TESTS.....

- For **diagnosis**, to determine the likelihood that someone has the disease their symptoms suggest;
- If someone is considered to be at significant risk, but doesn't have the disease, to help **classify the risk and manage it;**
- For **screening**, in order to evaluate risk in a defined group of people with no recognised symptoms;
- When someone does have a disease, to know how it is likely to progress (**prognosis**);
- To determine which **treatment** regime is best;
- To **monitor** a therapy that someone has been put on, to check progress.

How to evaluate a diagnostic test

"The development of an ….. diagnostic test from a good idea to a clinically relevant tool takes several step"



Leeflang et al. Clin Microbiol Infect. 2019
Ruffano et al. BMJ 2012

Frameworks for diagnostic tests

Shared components of frameworks which propose a chain of phases used to evaluate the evidence for new tests:

Analytic validity:
How well new assay performs in laboratory setting (technical efficacy etc)

Clinical Validity:
Accuracy of test to diagnose outcome of interest in clinical settings

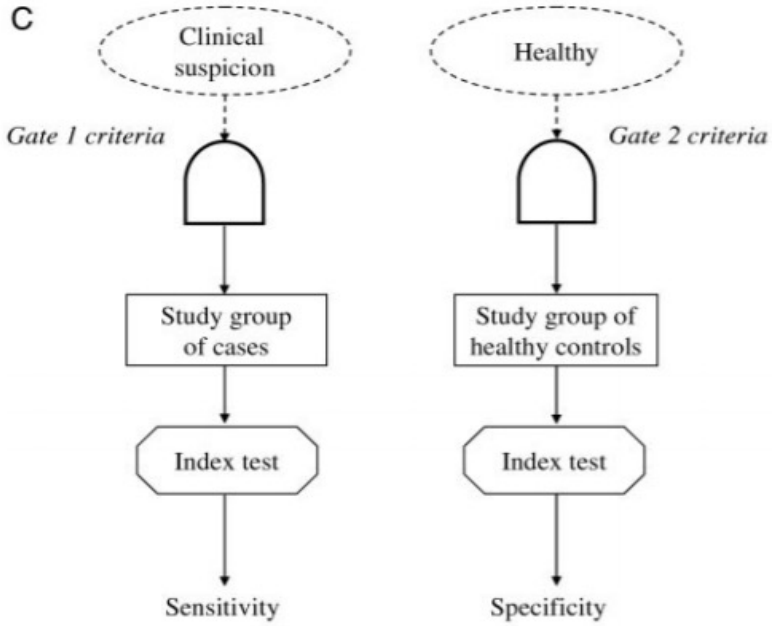
Clinical Utility:
Effects on clinical decision making, and patient outcomes

Other effects:
Cost effectiveness
Ethical, legal or societal implications

How to evaluate a diagnostic test

Phase 1: discrimination between two groups in an ideal situation

- ✓ Case – control study / two-gate study
- ✓ Poor clinical applicability of sensitivity – specificity



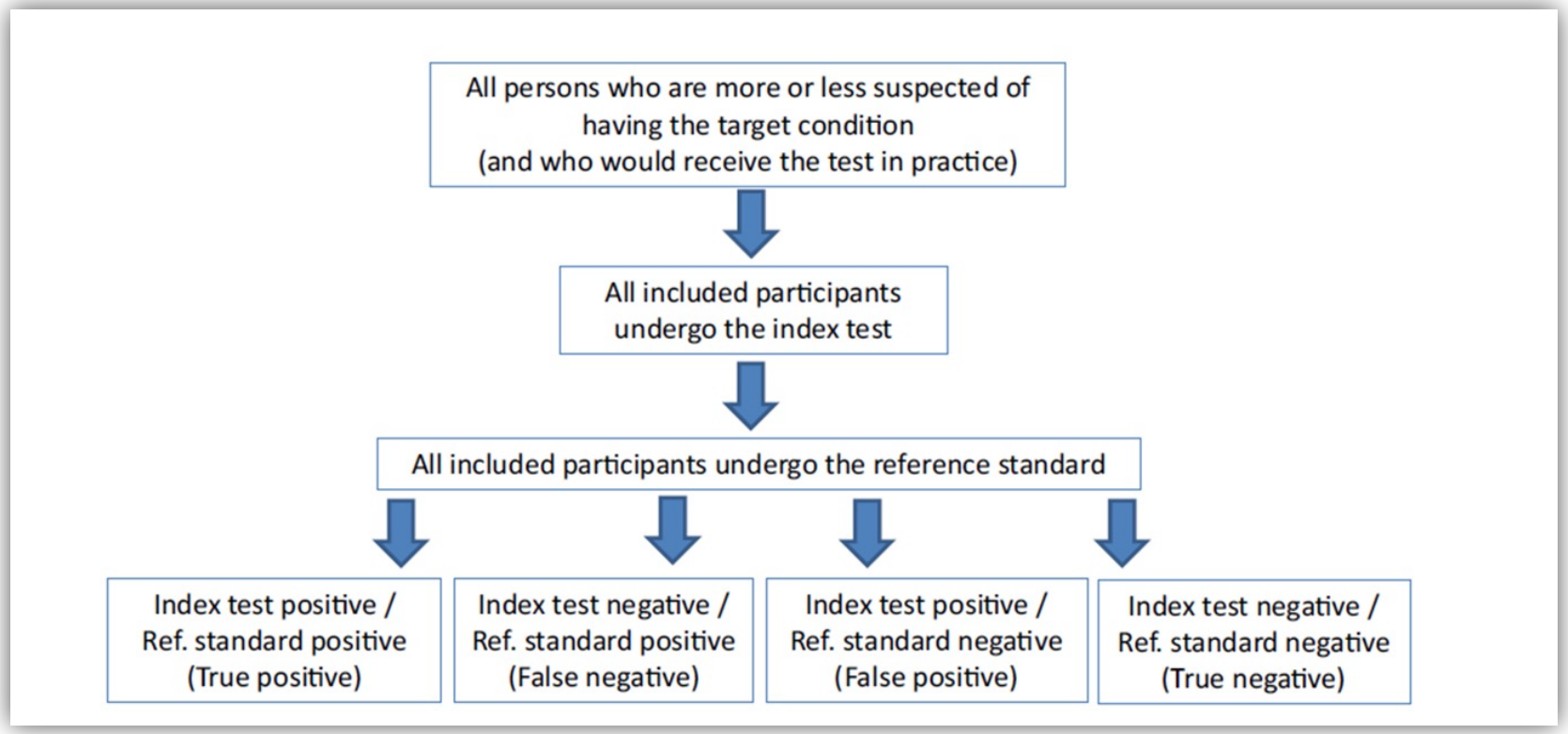
How to evaluate a diagnostic test

Phase 2: Technical validity

- ✓ if it is repeated several times, does it produce the same result?
- ✓ Repeatability and reproducibility
- ✓ Analytical sensitivity (minimally detectable levels).

How to evaluate a diagnostic test

Phase 3: clinical validity and accuracy in a clinically relevant situation



How to evaluate a diagnostic test

Who will be tested?

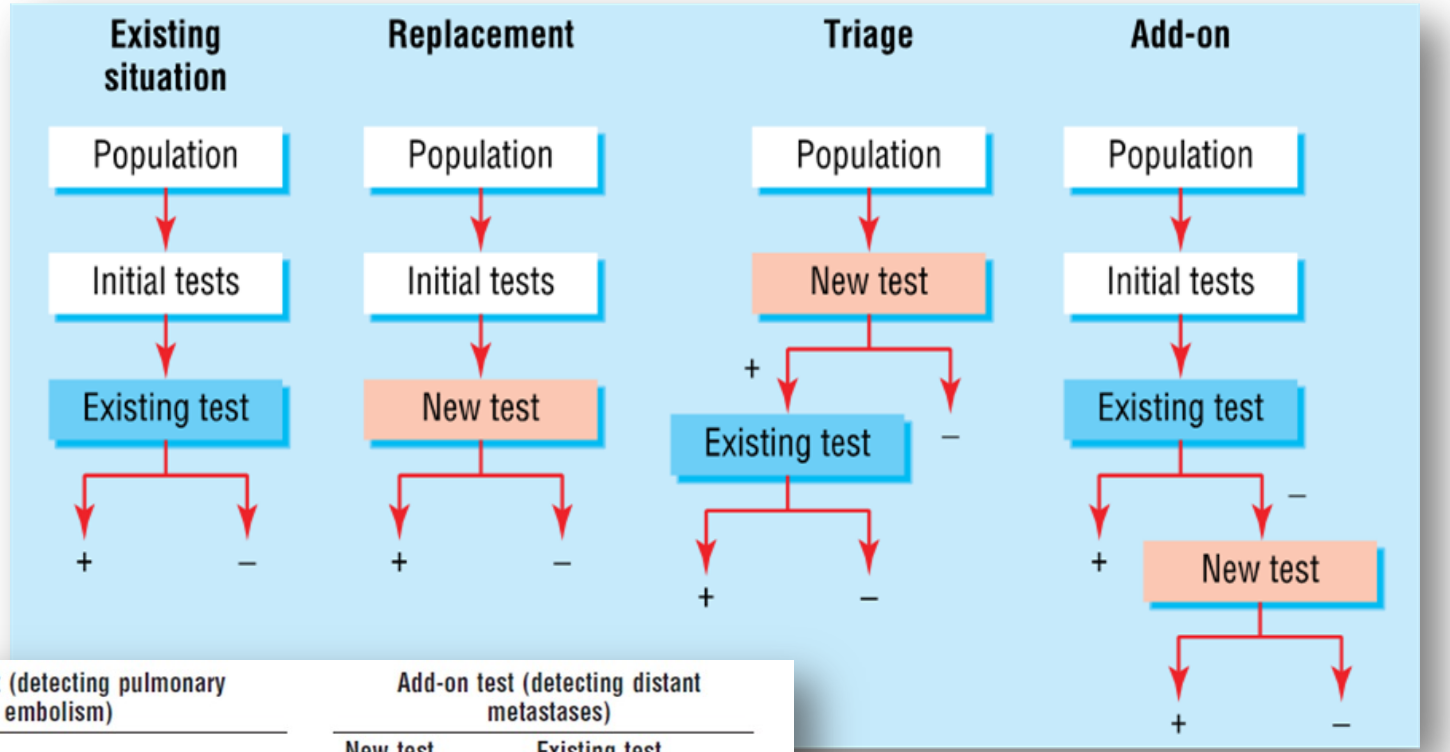
The population for which it is intended to be used in clinical practice.

Individuals are selected randomly or consecutively on the presence of predefined symptoms or signs of the target condition (**cross-sectional study**).

Important to define:

- ✓ Clinical setting
- ✓ Spectrum of disease
- ✓ Role of the test in the clinical pathway

Evaluation of Test – Role of new Test



Features	Replacement test (detecting herniated discs)		Triage test (detecting pulmonary embolism)		Add-on test (detecting distant metastases)	
	New test (magnetic resonance imaging)	Existing test (myelography)	New test (D-dimer)	Existing test (spiral computed tomography)	New test (positron emission tomography)	Existing test (computed tomography and ultrasound)
Accuracy	High	High	Low	High	High	High
Invasiveness	Non-invasive	Invasive	Non-invasive	Non-invasive	Non-invasive	Non-invasive
Waiting time	Yes	Yes	No	Yes	Yes	No
Knowledge and skills needed	Moderate	Moderate	Low	Moderate	Moderate	Moderate
Interpretable	Most tests	All tests	All tests	Most tests	Most tests	Most tests
Cost	High	High	Low	Higher	High	Medium

Reference Standard

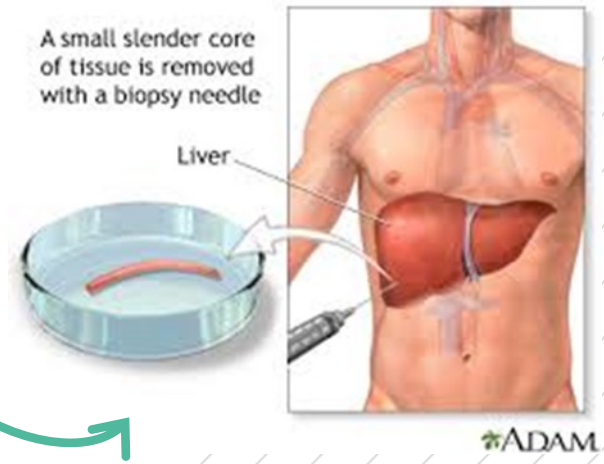
Test, series of tests, or set of procedures that is used to determine the presence or absence of the target condition

- ✓ "Gold standards" with 100% sensitivity and specificity are rare
- ✓ The best available, clinically accepted, error-free procedure

Principles

- to be applied on all included subjects
- independent assessment of test and reference standard
- standardized protocol

"Even biopsies are not infallible, as they rely on subjective assessment from human observers with varying expertise. "



Reference Standard

"For many diseases, a perfect diagnostic test may not exist or cannot be applied owing to costs or ethical concerns."

Type of bias	When does it occur?	Impact on accuracy
Misclassification bias	When the reference test does not correctly classify patients with the target condition	Depends on whether both the reference and index test make the same mistakes
Partial verification bias	When a non-random set of patients does not undergo the reference test	Usually leads to overestimation of sensitivity, effect on specificity varies
Differential verification bias	When a non-random set of patients is verified with a second or third reference test	overestimation of accuracy
Information bias	When the reference test data is interpreted with the knowledge of the index test results	overestimation of accuracy
Incorporation bias	When the index test is incorporated in a (composite) reference test	overestimation of accuracy

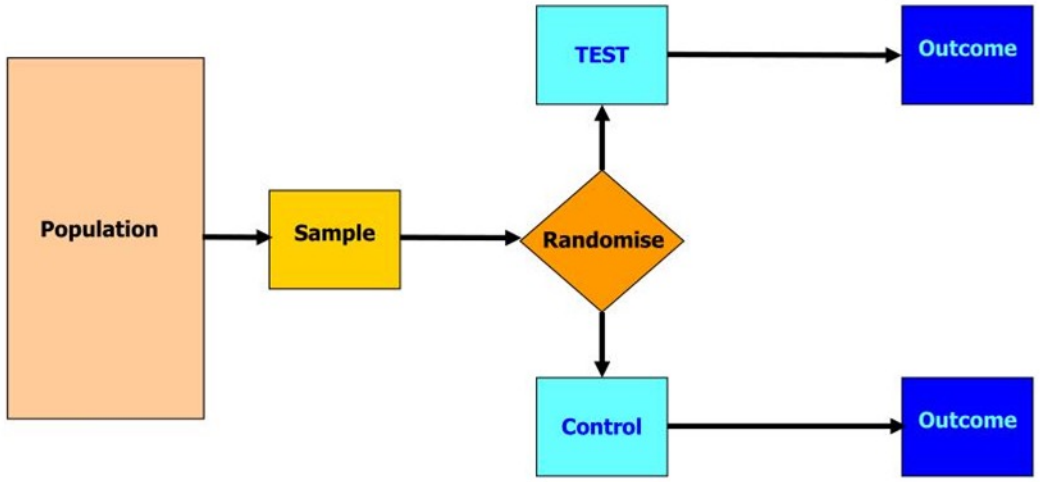
How to evaluate a diagnostic test

Phase 4: Effect on outcomes important for the patient and on society

"Does the patient's situation improve after testing?"

Patient-relevant outcomes:

- Risk of disease,
- Risk of death,
- Patient's quality of life



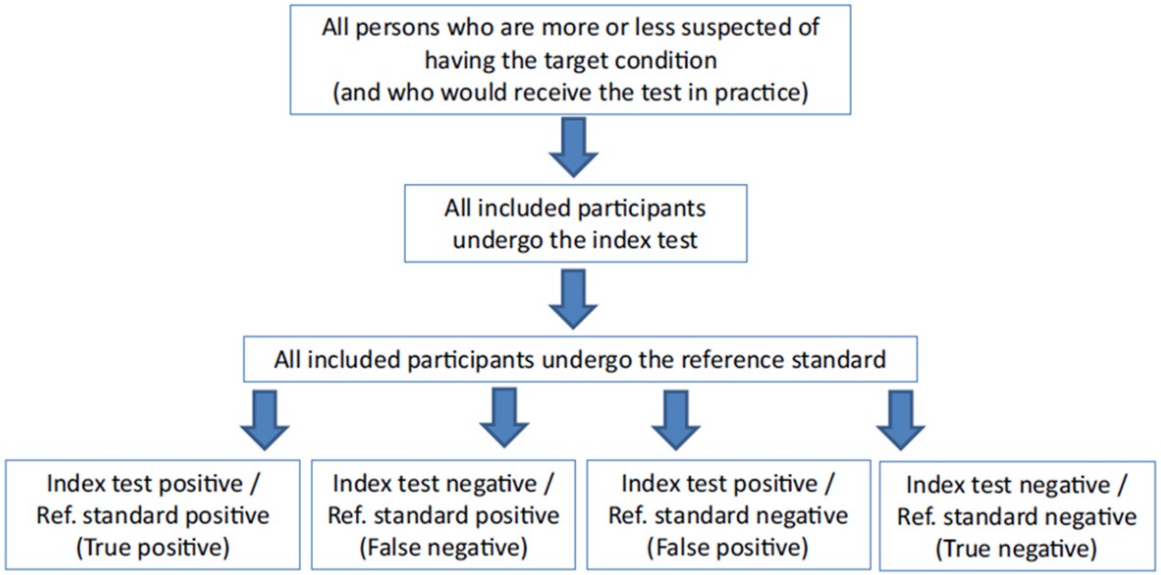
Metrics of Diagnostic Test Accuracy Studies

Measures of Diagnostic Accuracy

Diagnostic accuracy: the ability of a test to discriminate between the target condition and health.

- Sensitivity
- Specificity
- Predictive values
- Likelihood ratios
- Area under ROC curve

- Not fixed indicators of test performance
- Depend on: disease prevalence
  disease spectrum
  disease definition
  design of the study



2X2 Table

	Reference standard positive	**Reference standard negative**
Index test positive	*True positive (tp)*	*False positive (fp)*
Index test negative	*False negative (fn)*	*True negative (tn)*

True positive (TP): Diseased individuals with positive test results.

True negatives (TN): Non-diseased individuals with negative test results.

False positive (FP): Non-diseased individuals with positive test results.

False negative (FN): Diseased individuals with negative test results.

Sensitivity (Sens)

The proportion of people who have the disease who have a positive test result.

"So, a test with 90% sensitivity....means that the test identifies 90 out of 100 people WITH the disease"

	Reference standard positive	Reference standard negative
Index test positive	*True positive (tp)*	*False positive (fp)*
Index test negative	*False negative (fn)*	*True negative (tn)*

$$\frac{\text{True Positive}}{\text{True Positive+ False Negative}}$$

Specificity (Sp)

The proportion of patients who don't have the disease who have a negative test.

"So, a test with 75% specificity will be NEGATIVE in 75 out of 100 people WITHOUT the disease"

	Reference standard positive	Reference standard negative
Index test positive	*True positive (tp)*	*False positive (fp)*
Index test negative	*False negative (fn)*	*True negative (tn)*

$$\frac{\text{True Negative}}{\text{True Negative+ False Positive}}$$

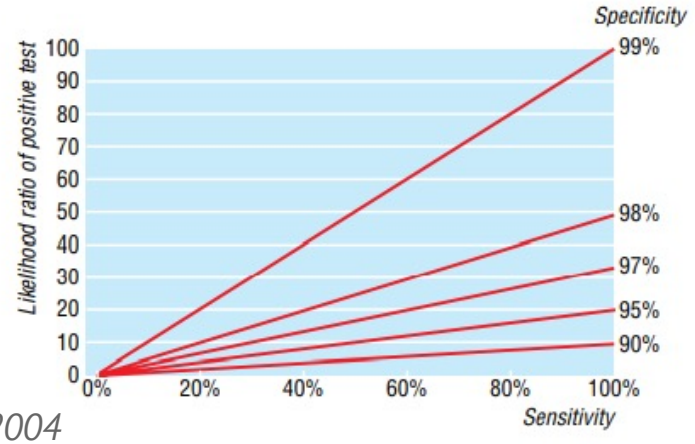
Ruling a diagnosis in or out with "SpPIn" and "SnNOut"

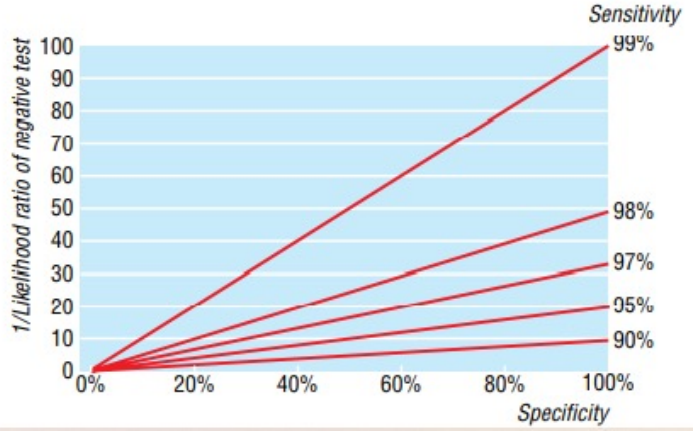
"**SnNout**" When a test has a very high **Se**nsitivity, a **N**egative result effectively rules **out** the diagnosis

"**SpPin**" When a test has a very high **Sp**ecificity, a **P**ositive result effectively rules **in** the diagnosis

Note of caution

"The power of a test to rule a diagnosis out does not depend exclusively on its sensitivity, but is reduced by low specificity. Similarly, the power to rule in depends on both specificity and sensitivity."





Pewsner D, et al. BMJ. 2004

Predictive Values

Positive Predictive Value (PPV)

"What is the likelihood that this patient has the disease when the test result is positive?"

$$\frac{\text{True Positive}}{\text{True Positive+ False Positive}}$$

Negative Predictive Value (NPV)

"What is the likelihood that this patient does not have the disease when the test result is negative?"

$$\frac{\text{True Negative}}{\text{True Negative+ False Negative}}$$

	Reference standard positive	**Reference standard negative**
Index test positive	*True positive (tp)*	*False positive (fp)*
Index test negative	*False negative (fn)*	*True negative (tn)*

Sensitivity – Specificity

- ✓ NOT influenced by the disease prevalence
- ✓ Depend on the spectrum of the disease in the studied group.

Predictive values

- ✓ Dependent on the disease prevalence.
- ✓ The same diagnostic test will have varying predictive values in different populations.
- ✓ A higher prevalence will increase the PPV and decrease the NPV.

Predictive Values

Your father went to his doctor and was told that his test for a disease was positive. He is really worried, and comes to ask you for help!

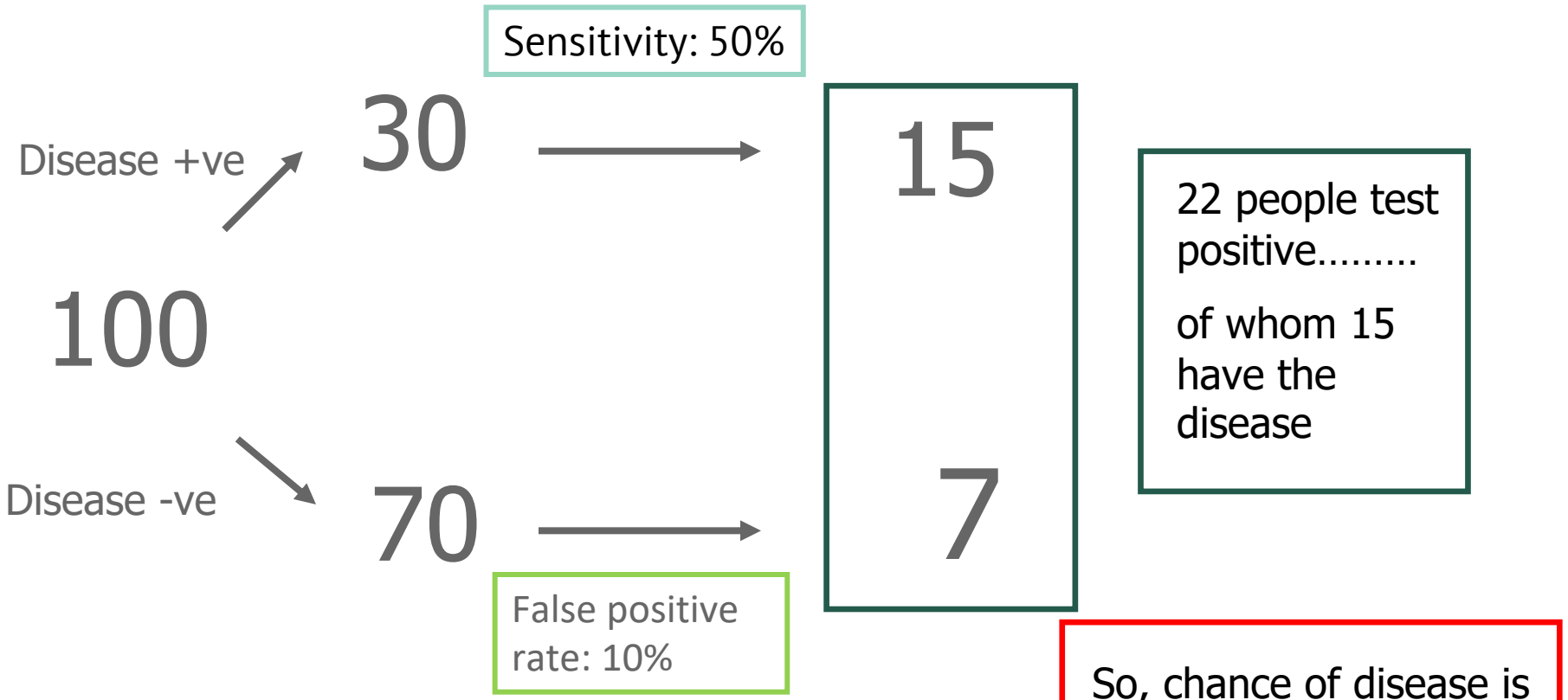
"Tell me what's the chance I have this disease?"

After doing some reading, you find that for men of his age:

- ✓ The prevalence of the disease is 30%
- ✓ The test has sensitivity of 50% and specificity of 90%

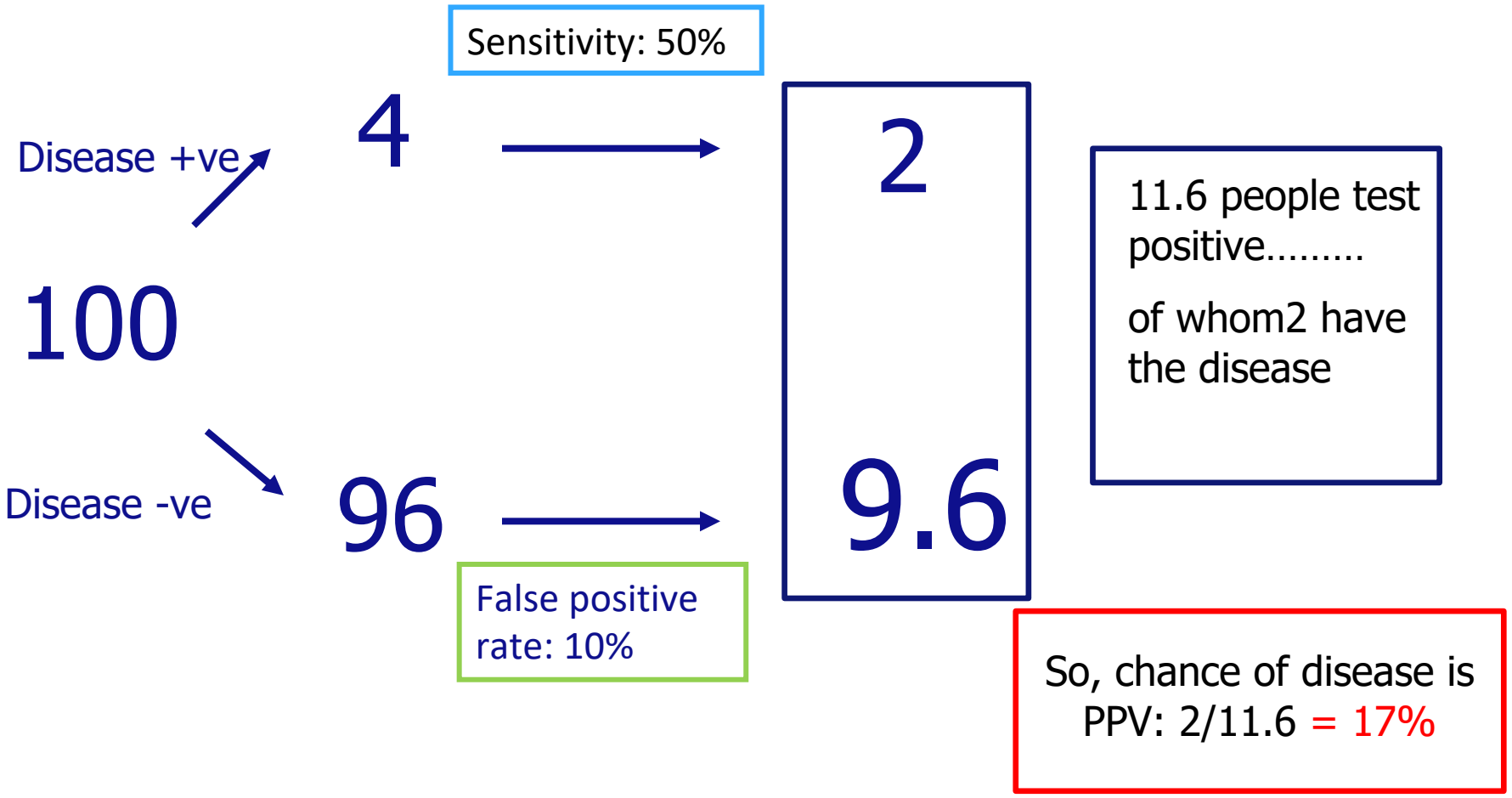
Predictive Values

Prevalence of 30%, Sensitivity of 50%, Specificity of 90%



Predictive Values

Prevalence of 4%, Sensitivity of 50%, Specificity of 90%



Likelihood Ratios

$$LR = \frac{\text{Probability of the specific test result in subjects with the disease}}{\text{Probability of the same result in subjects without the disease}}$$

LR provides a sense of how "powerful" a test is in influencing our pretest probability of disease.

LR+ : how much the odds of the disease **increase** when a test is **positive**

LR(+) = Sensitivity / (1 - Specificity)

LR-: how much the odds of the disease **decrease** when a test is **negative**

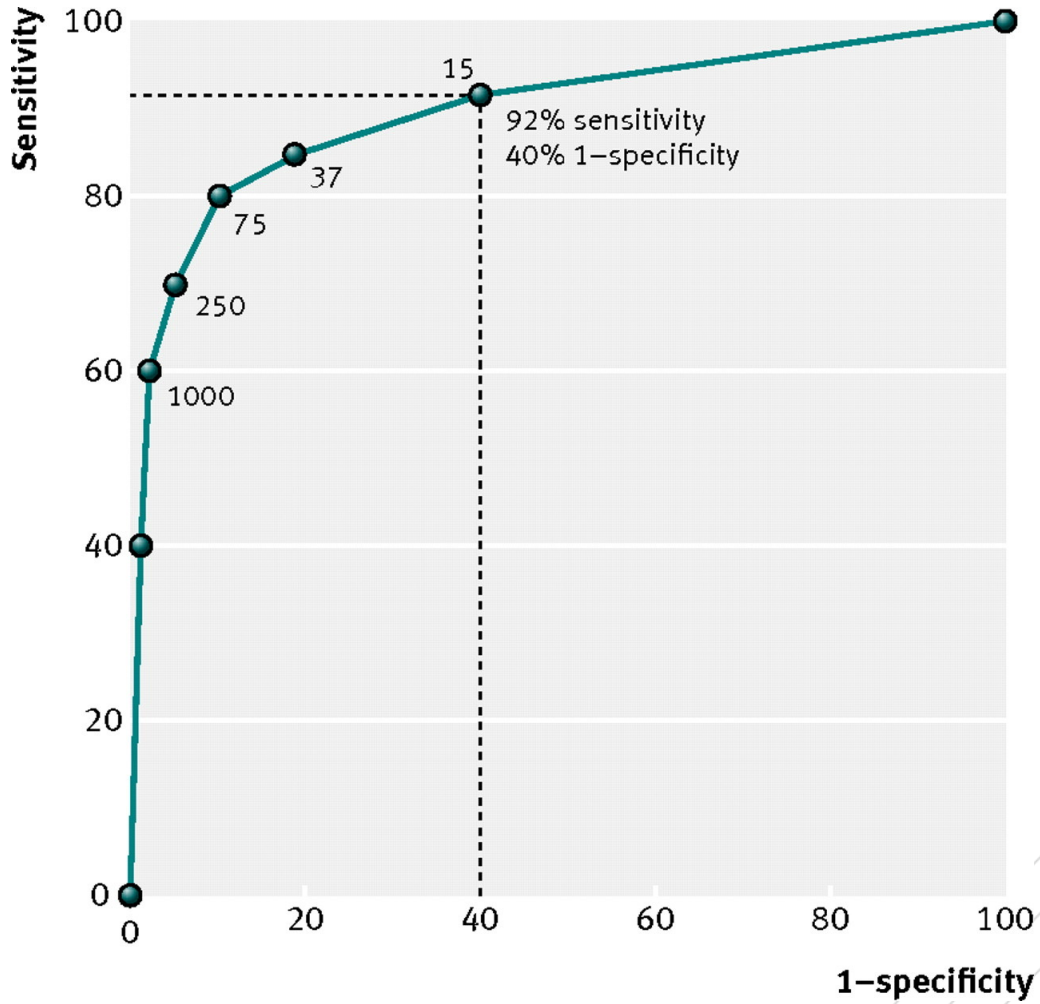
LR(-) = (1 -Specificity) / Sensitivity.

LR	Interpretation
> 10	Large and often conclusive increase in the likelihood of disease
5 –10	Moderate increase in the likelihood of disease
2 –5	Small increase in the likelihood of disease
1 –2	Minimal increase in the likelihood of disease
1	No change in the likelihood of disease
0.5 -1.0	Minimal decrease in the likelihood of disease
0.2 -0.5	Small decrease in the likelihood of disease
0.1 -0.2	Moderate decrease in the likelihood of disease
< 0.1	Large and often conclusive decrease in the likelihood of disease

ROC curve

Table 2. Characteristics of the PSA Test after Adjustment for Verification Bias, According to Age.*

<60 Yr			≥60 Yr		
Threshold for Biopsy Recommendation (ng/ml)	Sensitivity	Specificity	Threshold for Biopsy Recommendation (ng/ml)	Sensitivity	Specificity
0.9	1.00	0.56	1.1	0.84	0.43
1.4	0.74	0.79	2.1	0.68	0.70
2.6	0.36	0.94	4.1	0.35	0.88
4.1	0.18	0.98	6.1	0.19	0.94
6.1	0.08	0.99	10.1	0.08	0.99




Спасибо
Gracias
شكرا
Obrigado
Спасибо
Dank
Grazie
Ευχαριστω
Merci
Thank You
Ngiyabonga
Dank U
Diolch
Thank
You
Obrigado
Ngiyabonga
Tack
Merci
Dziękuję
Ευχαριστω
Danke
Grazie
תודה
Dank U
Diolch
Terima Kasih
Grazie
Tack
Diolch
Gracias
Merci
תודה