

Statistical Tools for Medical Diagnostics

Evaluation of Diagnostics Tests

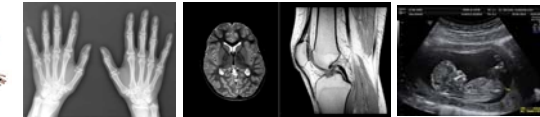
DIF, Semmelweis University, Budapest, November 28, 2018.



Medical diagnostics

Aim: obtaining simple, discriminating information
on health/disease condition

Imaging techniques:



X-Ray, MRI, CT, PET, Ultrasound, ... => Image

Physiological, biochemical measurements, evaluations:



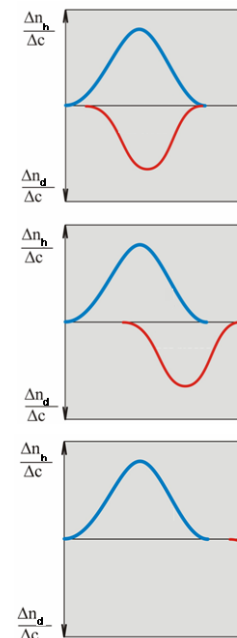
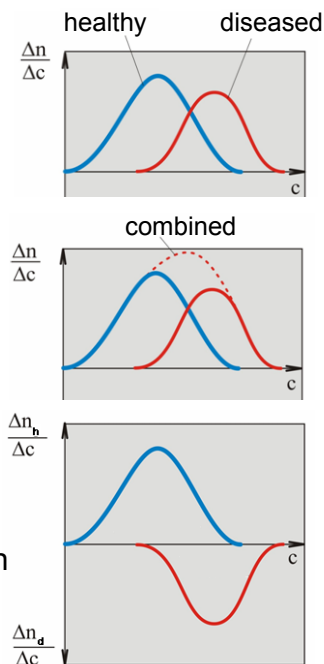
Pressure, Frequency, Density,
Concentration
Measurement => **SINGLE NUMBER:**
CLASSIFIER VALUE (c)

Discrimination of health and disease condition

Based on overlapping distributions

assumption:
c: classifier value
(e.g. serum concentration)
changes in healthy and
diseased subpopulations

novel
representation



full
overlap

partial
overlap

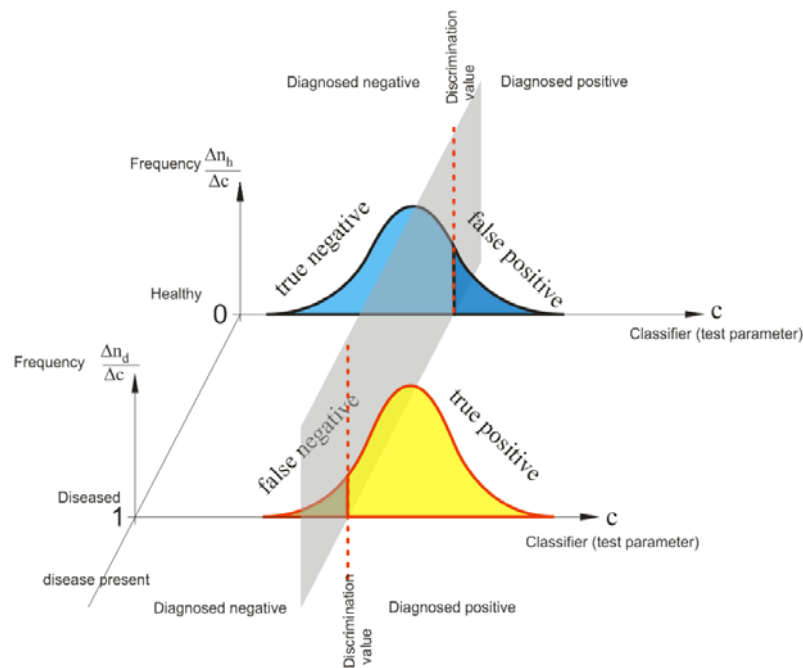
complete
separation

**Discrimination based on
overlap magnitude:**

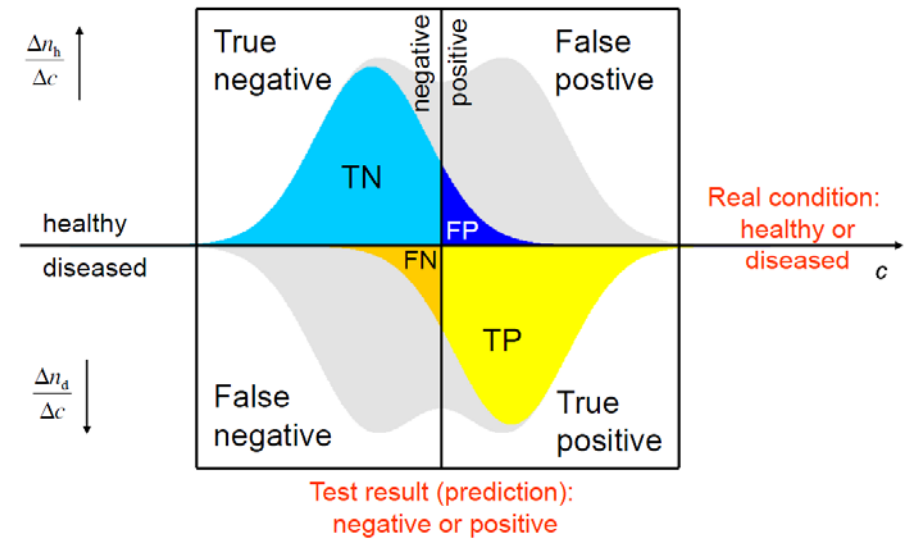
useless method

real-life situation

perfect method



Contingency table:
Confusion matrix (binary classification)



Parameters of diagnostic „goodness”

The goodness of a test can be described in terms of the following diagnostic parameters

Sensitivity
Specificity
PPV, relevance
NPV, segregation

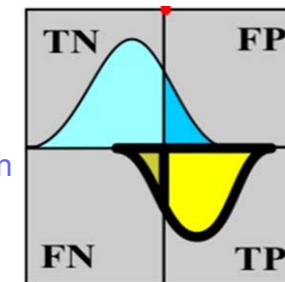
Every method must be compared with a reference-method (gold standard)

Gold standard: method known to always work; often autopsy



Prevalence

= frequency of diseased in examined population
= probability prior to test
= a-priori-probability



measure of how common the disease is

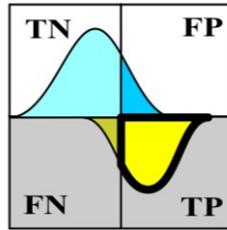
$$\frac{\text{Diseased}}{\text{Total}} = w = \frac{\text{diseased}}{\text{total}} = \frac{TP + FN}{TP + TN + FN + FP} = \frac{de - sp}{se - sp}$$

Incidence

Rate of new cases of disease/year/person in a given population
(e. g. per year, per 10 000 people)

Diagnostic sensitivity

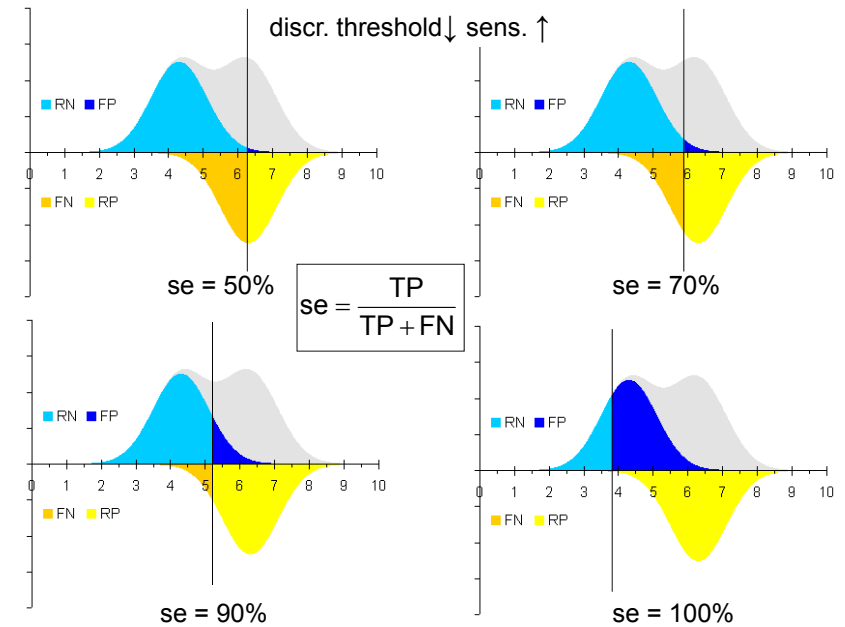
= positive within diseased
= true positive rate
= recall rate



probability that
the test finds the
diseased positive

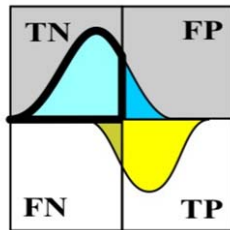
$$\frac{\text{TP}}{\text{TP} + \text{FN}} = \text{se} = \frac{\text{true positive}}{\text{diseased}} = p(\text{positive}|\text{diseased})$$

High-sensitivity tests are required:
In early diagnosis (screening) so that few patients remain unrecognized.
If the risk of disease is higher than the risk of treatment.



Diagnostic specificity

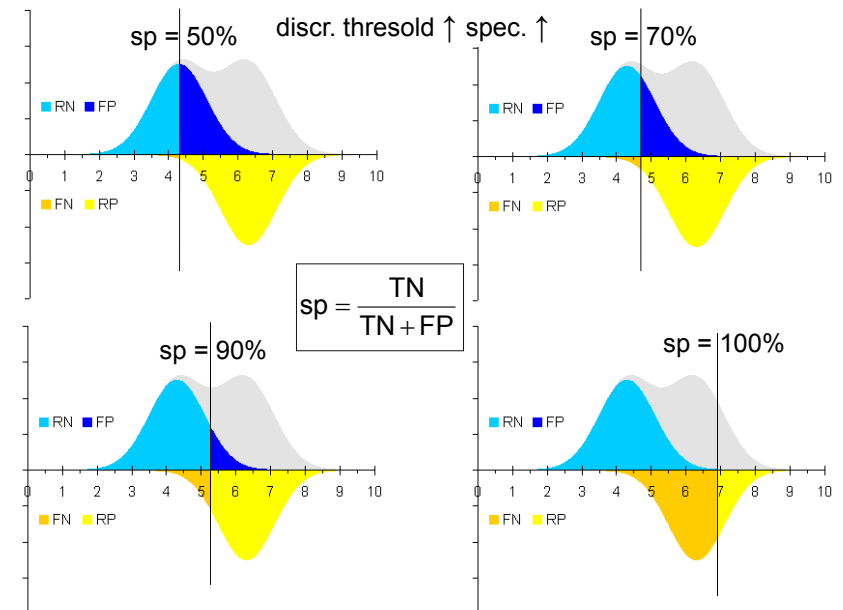
= negative among
healthy
= true negative rate



probability that
the test finds a
healthy negative

$$\frac{\text{TN}}{\text{TN} + \text{FP}} = \text{sp} = \frac{\text{true negative}}{\text{healthy}} = p(\text{negative}|\text{healthy})$$

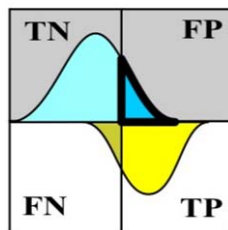
High-specificity tests are important:
When the false positive values have severe consequences (e.g. surgery).
When the risk of treatment is higher than the risk of disease.



Diagnostic False Positive Rate

(Type-I error)

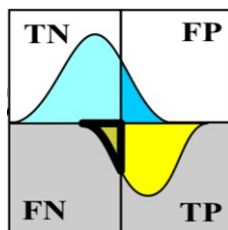
$$\frac{\text{FP}}{\text{TN} + \text{FP}} = 1 - \text{sp} = \frac{\text{FP}}{\text{healthy}} = p(\text{positive}|\text{healthy})$$



Diagnostic False Negative Rate

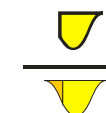
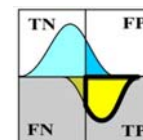
(Type-II error)

$$\frac{\text{FN}}{\text{FN} + \text{TP}} = 1 - \text{se} = \frac{\text{FN}}{\text{diseased}} = p(\text{negative}|\text{diseased})$$



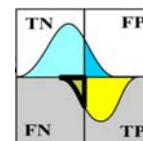
Horizontal rates are independent of prevalence

sensitivity
(se)



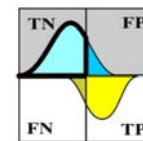
$$\text{se} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

false negative rate
(1-se)



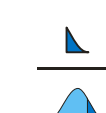
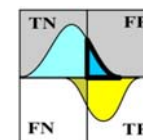
$$1 - \text{se} = \frac{\text{FN}}{\text{FN} + \text{TP}}$$

specificity
(sp)



$$\text{sp} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

false positive rate
(1-sp)



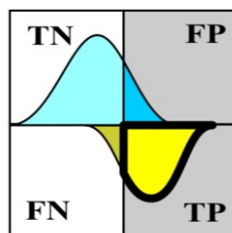
$$1 - \text{sp} = \frac{\text{FP}}{\text{TN} + \text{FP}}$$

Predictive values (vertical rates)

a-posteriori-probabilities; they depend strongly on prevalence

Positive predictive value

= PPV
= predictive value positive
= PVP
= diagnostic **relevance**
= diseased among positive

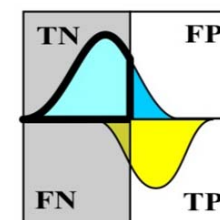


probability of
disease if test is
positive

$$\frac{\text{TP}}{\text{TP} + \text{FP}} = \text{PPV} = \frac{\text{TP}}{\text{positive}} = p(\text{diseased}|\text{positive}) = \frac{\text{se} \cdot w}{\text{se} \cdot w + (1 - \text{sp}) \cdot (1 - w)}$$

Negative predictive value

= NPV
= predictive value negative
= PVN
= diagnostic segregation
= healthy among negatives

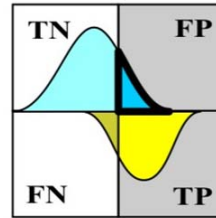


probability of health
if test is negative

$$\frac{\text{TN}}{\text{TN} + \text{FP}} = \text{NPV} = \frac{\text{TN}}{\text{negative}} = p(\text{healthy}|\text{negative}) = \frac{\text{sp} \cdot (1 - w)}{\text{sp} \cdot (1 - w) + (1 - \text{se}) \cdot w}$$

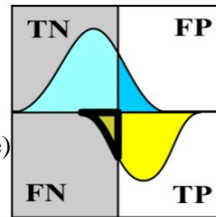
False alarm rate

$$\frac{\text{FN}}{\text{FN} + \text{TP}} = 1 - \text{PPV} = \frac{\text{FP}}{\text{positive}} = \frac{\text{FP}}{\text{FP} + \text{TP}} = p(\text{healthy}|\text{positive})$$



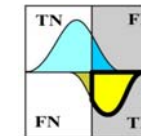
False reassurance rate

$$\frac{\text{FP}}{\text{FP} + \text{TP}} = 1 - \text{NPV} = \frac{\text{FN}}{\text{negative}} = \frac{\text{FN}}{\text{FN} + \text{TN}} = p(\text{diseased}|\text{negative})$$



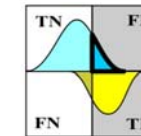
Vertical rates are dependent on prevalence

positive predictive value (PPV)



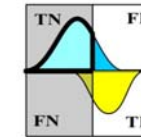
$$\text{PPV} = \frac{\text{TP}}{\text{FP} + \text{TP}}$$

false alarm rate (1-PPV)



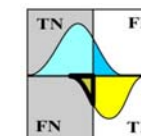
$$1 - \text{PPV} = \frac{\text{FP}}{\text{FP} + \text{TP}}$$

negative predictive value (NPV)



$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}}$$

false reassurance rate (1-NPV)

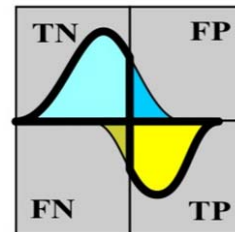


$$1 - \text{NPV} = \frac{\text{FN}}{\text{TN} + \text{FN}}$$

Diagnostic efficacy/efficiency

also called as **accuracy**

= probability of correct diagnosis



$$\frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \text{de} = \frac{\text{TP} + \text{TN}}{\text{total}} = \text{se} \cdot w + \text{sp} \cdot (1 - w)$$

often: discrimination threshold is chosen so that **de** is maximized

Effect of prevalence

case1: **w = 50%**

NPV = 90%

sp = 90%

Gold-standard	healthy	Test	
		negative	positive
	healthy	90	10
	diseased	10	90

se = 90%

(de = 90%)

PPV = 90%

NPV = 99%

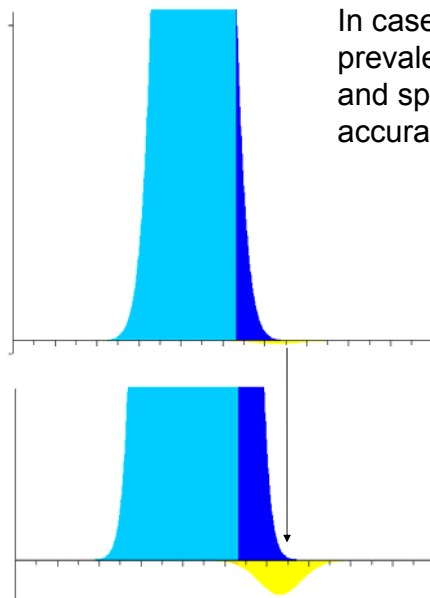
Case 2: **w = 10%**

Gold-standard	healthy	Test	
		negative	positive
	healthy	810	90
	diseased	10	90

se = 90%

(de = 90%)

PPV = 50%



In case of very small prevalence a highly sensitive and specific test have low accuracy (PPV).

prevalence = 0.1 %
sensitivity = 98 %
specificity = 98 %

PPV = 4 %

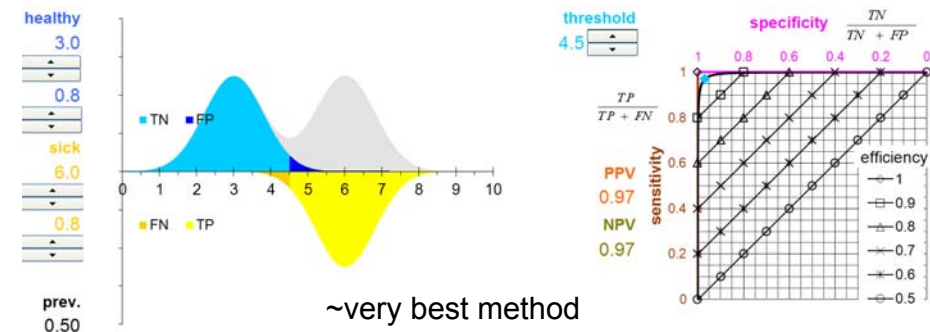
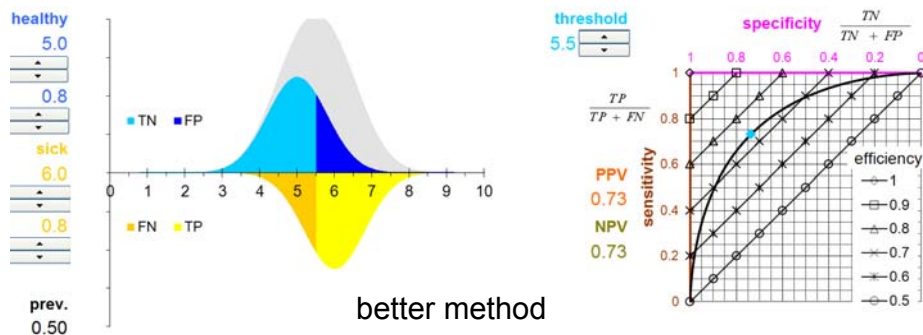
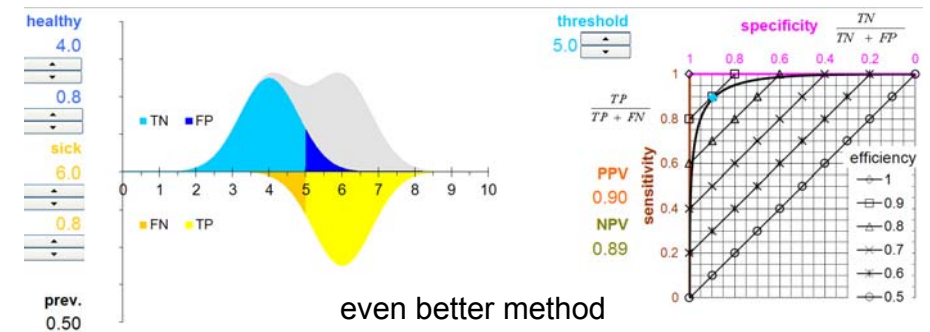
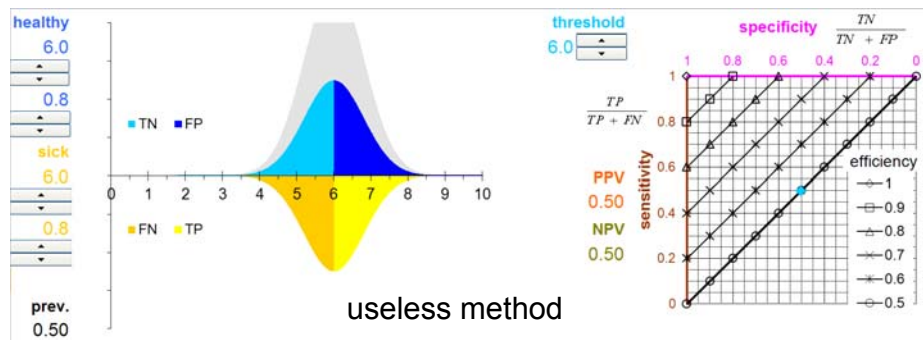
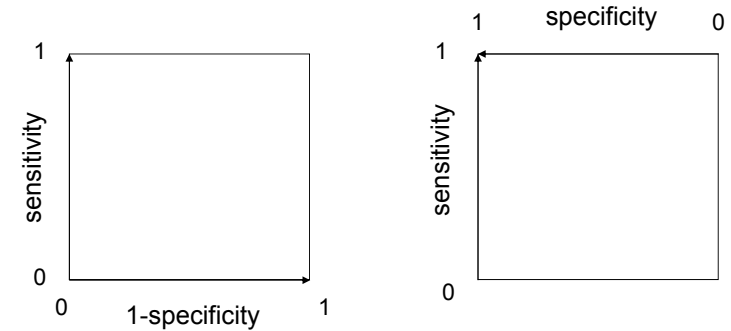
Comparison of diagnostic tests: the ROC curves

[NOT EXAM MATERIAL]

ROC: receiver-operator (operating) characteristic

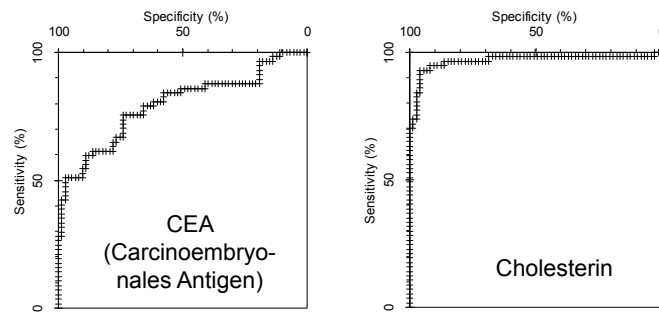
~ 1950: first ROC Analysis (receiver: Radar)

~ 1970: first medical applications



Example: Tumor markers in the ascites

increased CEA and/or cholesterol concentrations in ascites are diagnostic markers for carcinomatosis



Main use of ROC curves: comparison of diagnostic methods

Which method is better? What discrimination threshold should be used? =>

Gulyás M, Kaposi AD, Elek G, Szollár LG, Hjerpe A, Value of carcinoembryonic antigen (CEA) and cholesterol assays of ascitic fluid in cases of inconclusive cytology, J Clinical Pathology 2001 (54) 831-835

THANK YOU for your attention
...and PATIENCE!!

!! One can use the FORMULA COLLECTION at Exams