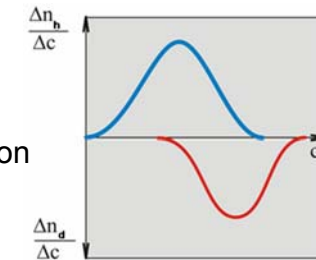
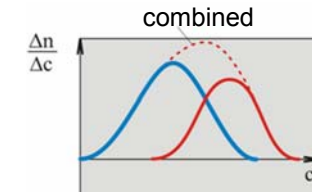
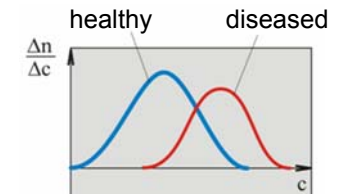


KAD 2015.11.23

Overlapping distributions

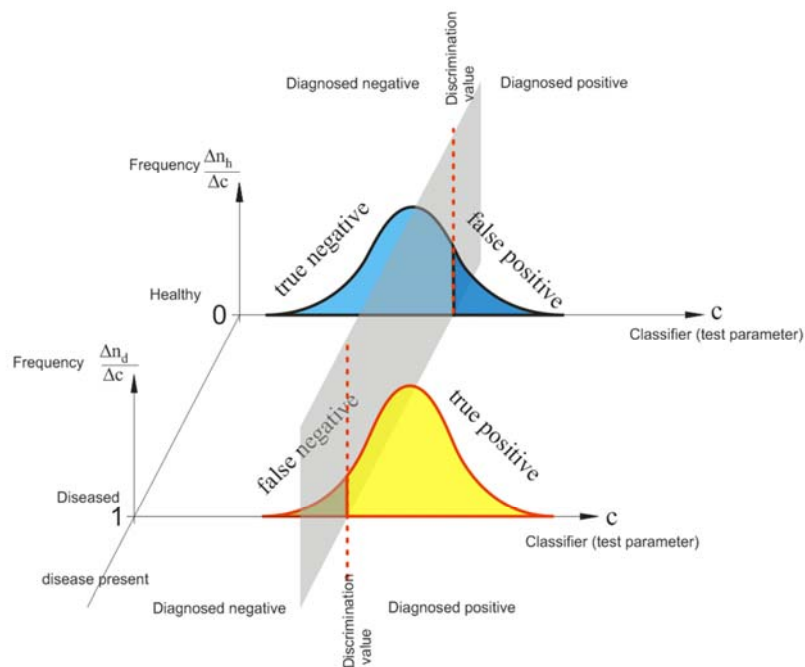
assumption:

a classifier value
(e.g. serum concentration)
changes (e.g. increases)
in diseased subpopulation



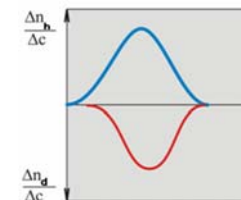
novel
representation

2



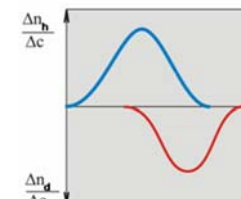
3

Based on overlap magnitude:



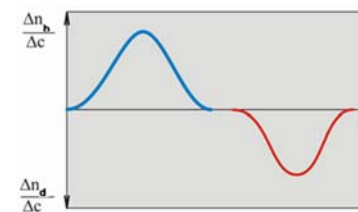
full
overlap

useless method



partial
overlap

real-life situation

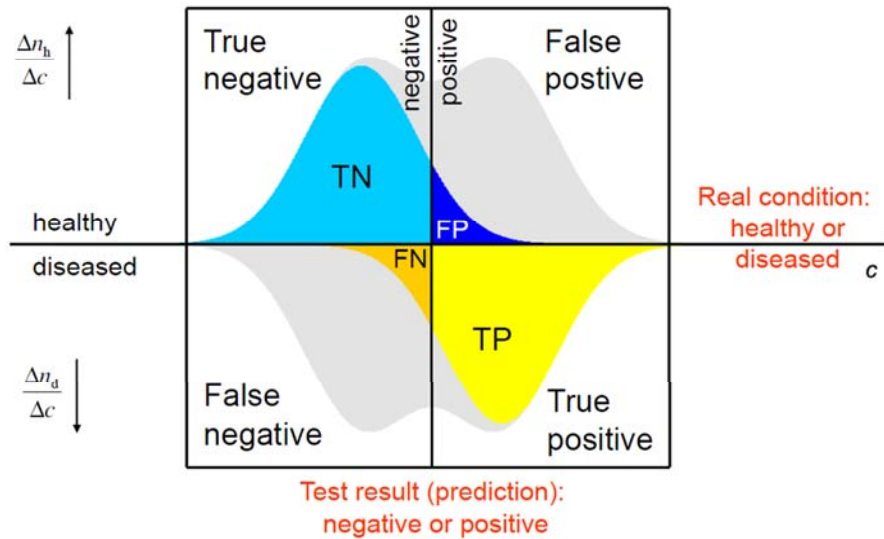


complete
separation

perfect method

4

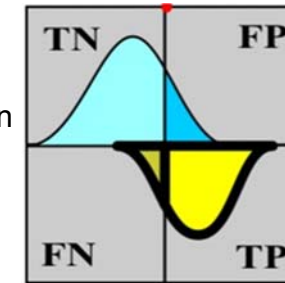
Contingency table: Confusion matrix (binary classification)



5

Prevalence

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

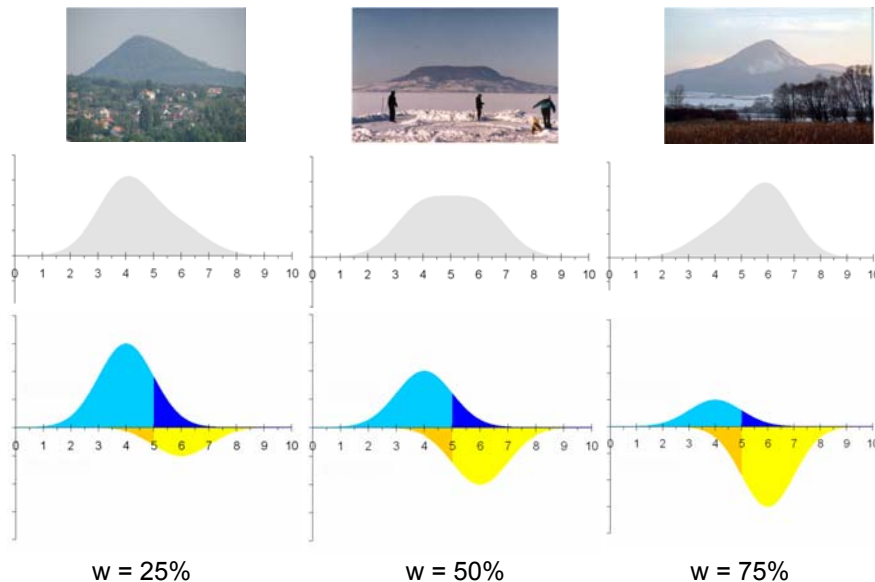


measure of
how common
the disease is

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

6

Shape of combined distributions



7

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)

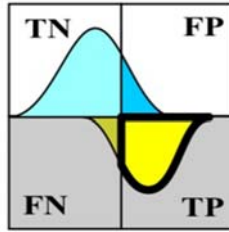
Gold standard: method known to
always work; often autopsy



8

Diagnostic sensitivity

= positive within diseased
= true positive rate
= recall rate



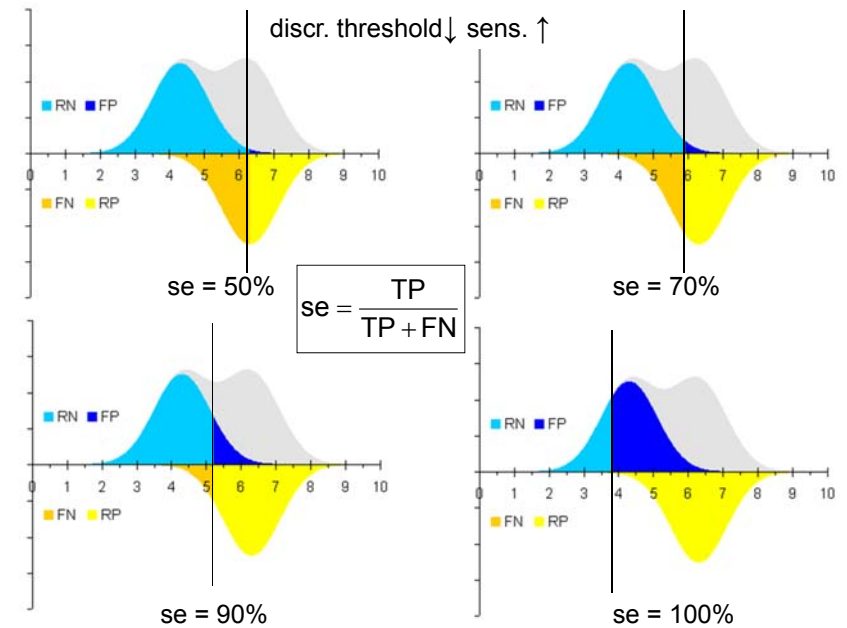
probability that
the test finds the
diseased positive

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

Large-sensitivity tests are required:

In early diagnosis (screening) so that few patients remain unrecognized.
If the risk of disease is greater than the risk of treatment.

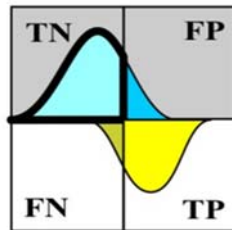
9



10

Diagnostic specificity

= negative among
healthy
= true negative rate



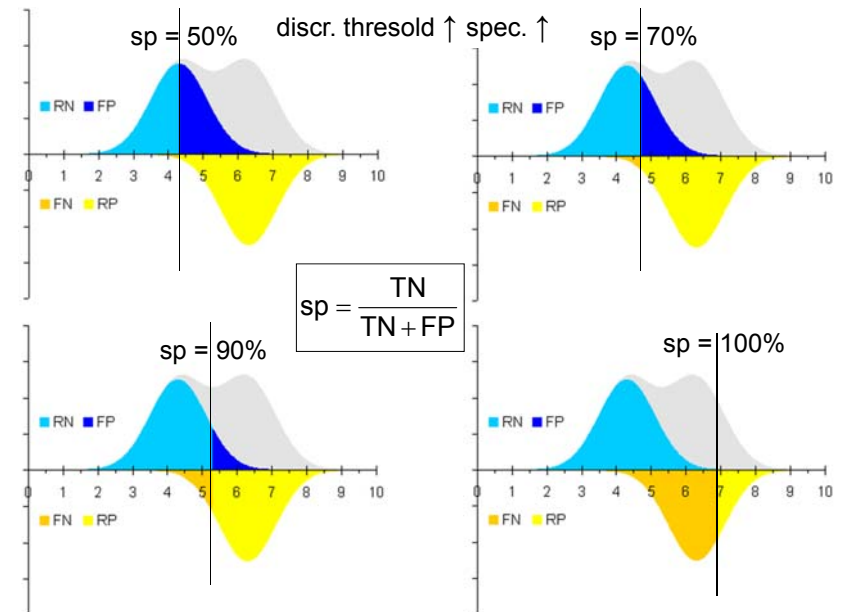
probability that
the test finds a
healthy negative

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

High-specificity tests are important:

When the false positive values have severe consequences (e.g. surgery).
When the risk of treatment is greater than the risk of disease.

11

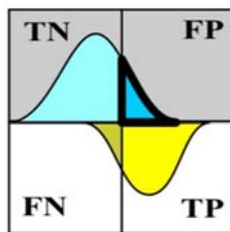


12

Diagnostic False Positive Rate

(Type-I error)

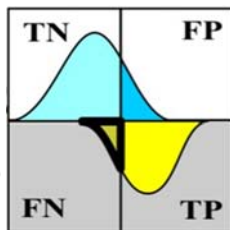
$$\frac{\text{FP}}{\text{TN} + \text{FP}} = 1 - \text{sp} = \frac{\text{FP}}{\text{healthy}} = p(\text{positive}|\text{healthy})$$



Diagnostic False Negative Rate

(Type-II error)

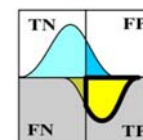
$$\frac{\text{FN}}{\text{FN} + \text{TP}} = 1 - \text{se} = \frac{\text{FN}}{\text{diseased}} = p(\text{negative}|\text{diseased})$$



13

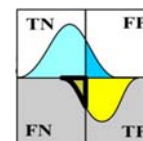
Horizontal rates are independent of prevalence

sensitivity
(se)



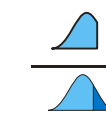
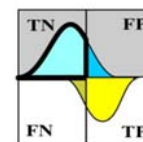
$$\text{se} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

false negative rate
(1-se)



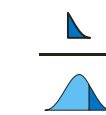
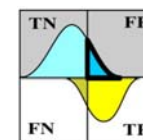
$$1 - \text{se} = \frac{\text{FN}}{\text{FN} + \text{TP}}$$

specificity
(sp)



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

false positive rate
(1-sp)



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

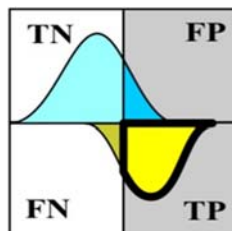
14

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



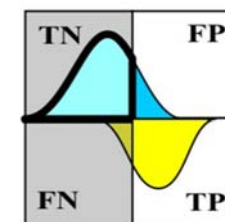
probability of
disease if test is
positive

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

15

Negative predictive value

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



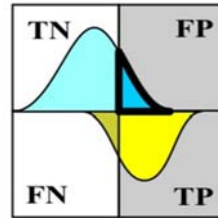
probability of
healthiness if test is
negative

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

16

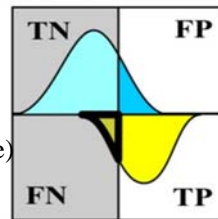
False alarm rate

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



False reassurance rate

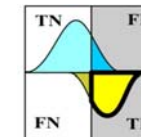
$$\frac{\text{FN}}{\text{FN} + \text{TN}} = 1 - \text{NPV} = \frac{\text{FN}}{\text{negative}} = \frac{\text{FN}}{\text{FN} + \text{TN}} = p(\text{diseased}|\text{negative})$$



17

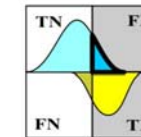
Vertical rates are dependent of prevalence

positive predictive value (PPV)



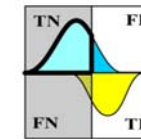
$$\text{PPV} = \frac{\text{TP}}{\text{FP} + \text{TP}}$$

false alarm rate (1-PPV)



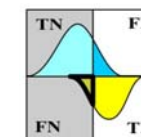
$$1 - \text{PPV} = \frac{\text{FP}}{\text{FP} + \text{TP}}$$

negative predictive value (NPV)



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

false reassurance rate (1-NPV)



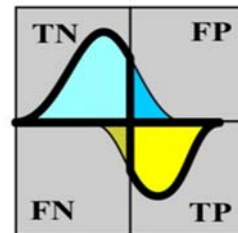
$$1 - \text{NPV} = \frac{\text{FN}}{\text{TN} + \text{FN}}$$

18

Diagnostic efficacy/efficiency

= accuracy

= correct classification rate



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

often: discrimination threshold is chosen so that de is maximized

19

Effect of prevalence

case1: $w = 50\%$

NPV = 90%

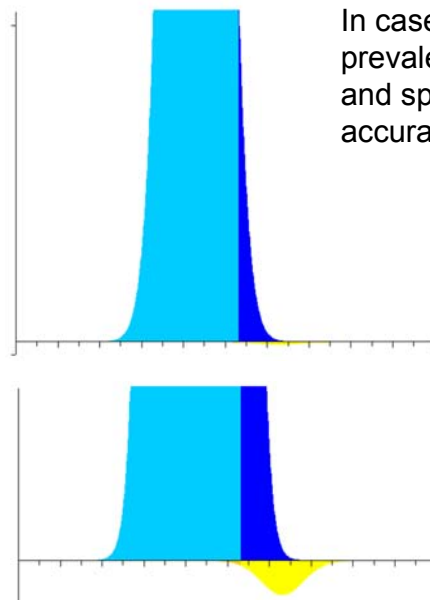
		Test		
		negative	positive	
sp = 90%	Gold-standard	healthy	90	10
		diseased	10	90
(de = 90%)		PPV = 90%		se = 90%

NPV = 99%

Case 2: $w = 10\%$

		Test		
		negative	positive	
sp = 90%	Gold-standard	healthy	810	90
		diseased	10	90
(de = 90%)		PPV = 50%		se = 90%

20



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 %

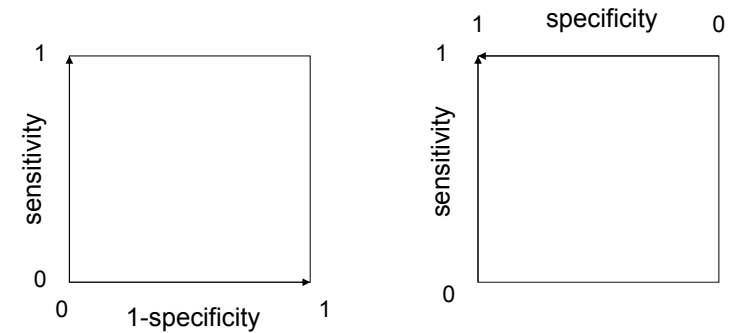
21

Comparison of diagnostic tests: the ROC space

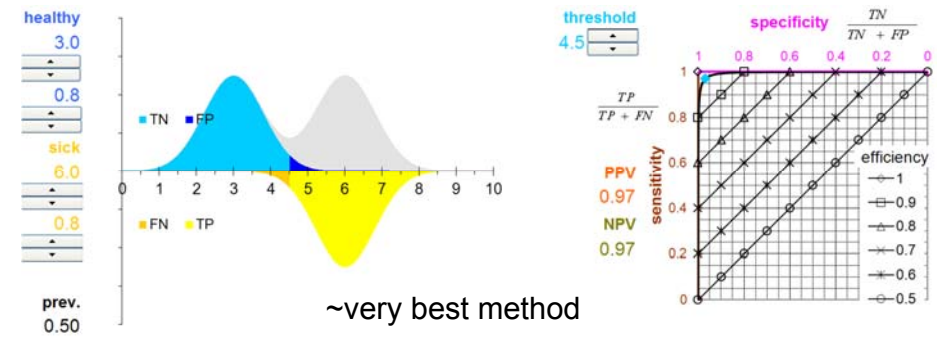
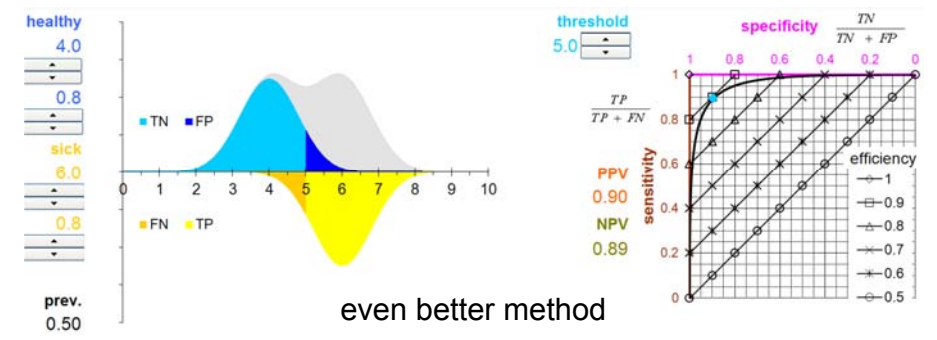
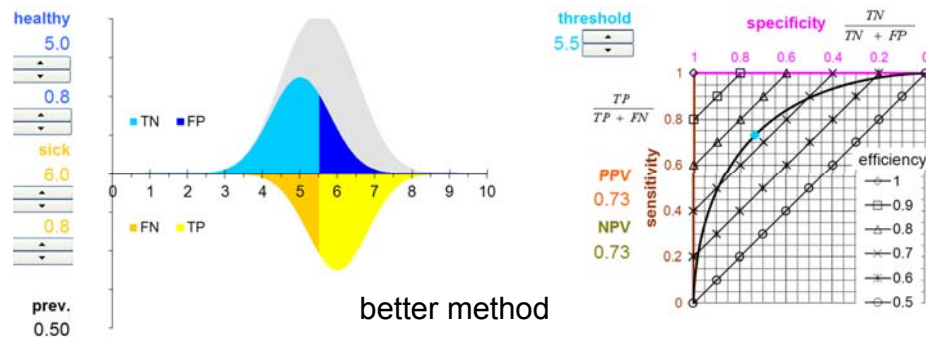
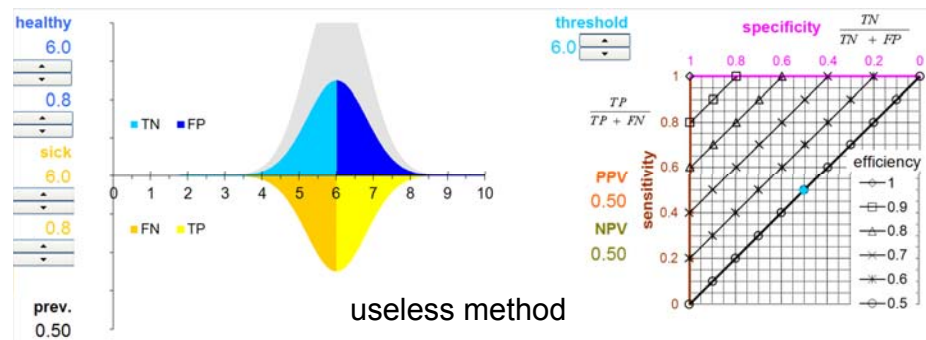
ROC: receiver-operator (operating) characteristic

~ 1950: first ROC Analysis (receiver: Radar)

~ 1970: first medical applications

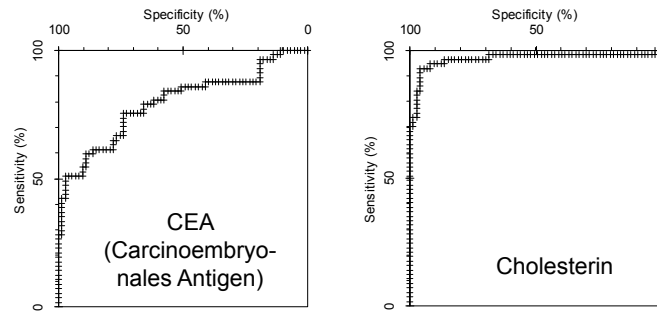


22



E.g.: Tumor markers in the ascites

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



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

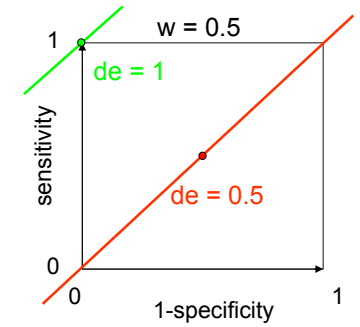
25

$$de = se \cdot w + sp \cdot (1 - w)$$

$$\frac{de}{1 - w} = \frac{w}{1 - w} se + (sp - 1) + 1$$

$$(1 - sp) + \frac{de}{1 - w} - 1 = \frac{w}{1 - w} se$$

$$se = \underbrace{\frac{1 - w}{w}}_{\text{slope}} (1 - sp) + \underbrace{\frac{1}{w} de + \frac{w - 1}{w}}_{\text{intercept}}$$



if $w = 0.5$: $se = 1 \cdot (1 - sp) + 2 \cdot de - 1$

The points have the same diagnostic efficiency belong to a line with a slope of 1.

If $de = 0.5$, the intercept is 0.

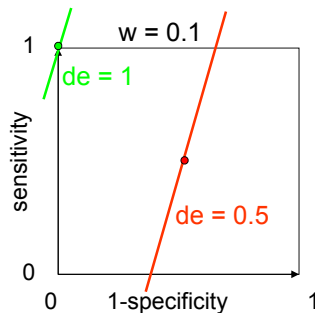
26

$$se = \underbrace{\frac{1 - w}{w}}_{\text{slope}} (1 - sp) + \underbrace{\frac{1}{w} de + \frac{w - 1}{w}}_{\text{intercept}}$$

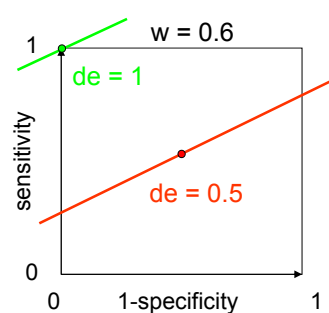
if $w < 0.5$: the slope of lines at identical diagnostic efficiencies is greater than 1.

if $w > 0.5$: the slope of lines at identical diagnostic efficiencies is smaller than 1.

e.g. $w = 0.1$, slope: 9

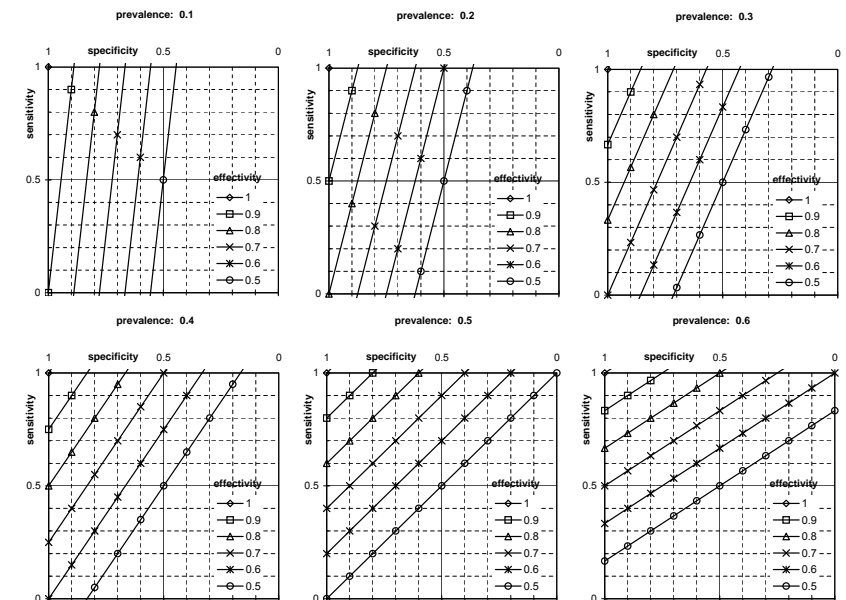


e.g. $w = 0.6$, slope: 0.66



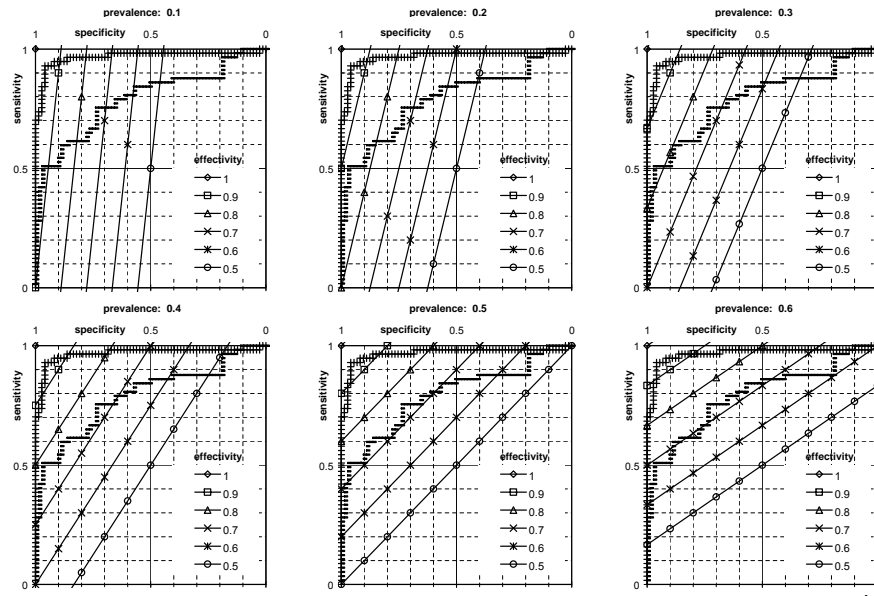
27

Isoeffective curves on the ROC

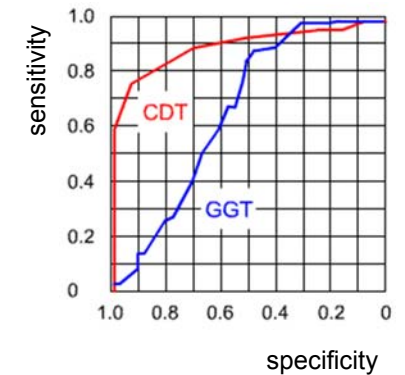
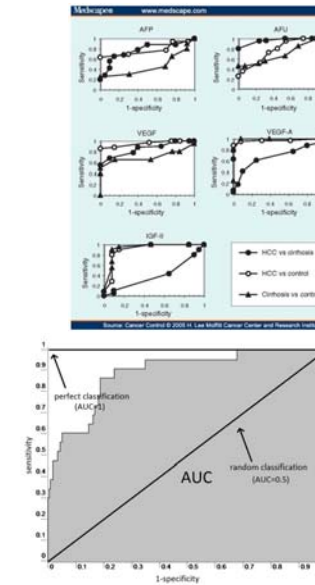


28

Ascites (+ Cholesterin, – CEA)



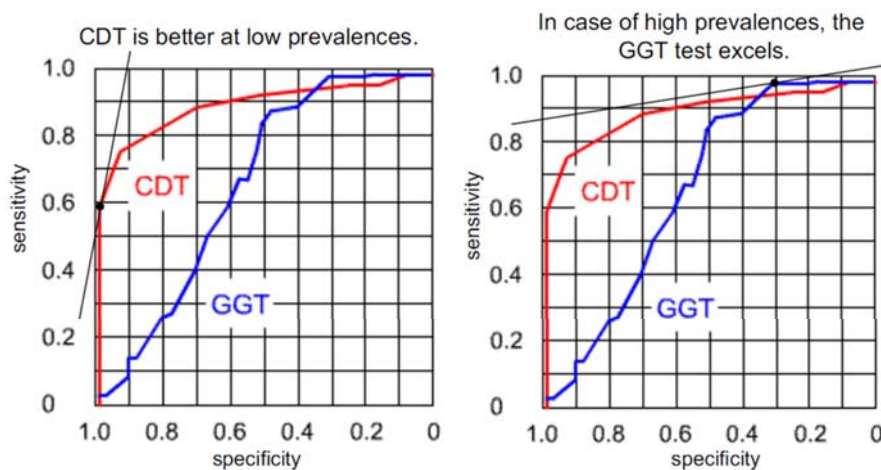
Additional examples



Alcoholism diagnostics with CDT (carbohydrate deficient transferrin) and GGT (gamma-glutamyltransferase). AUC of CDT is larger than of GGT. Is it a better method?

30

If we maximize the diagnostic accuracy...



31