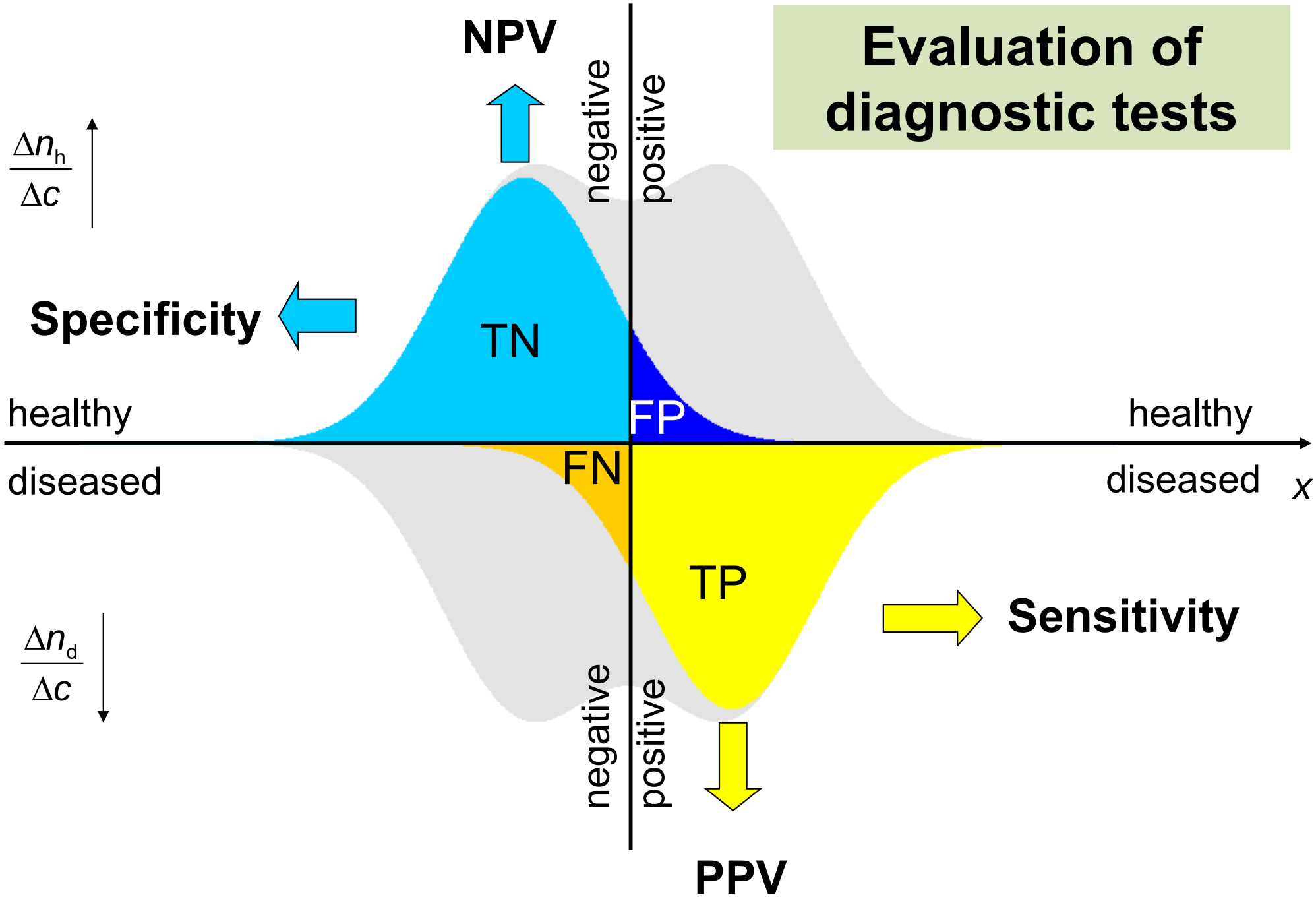
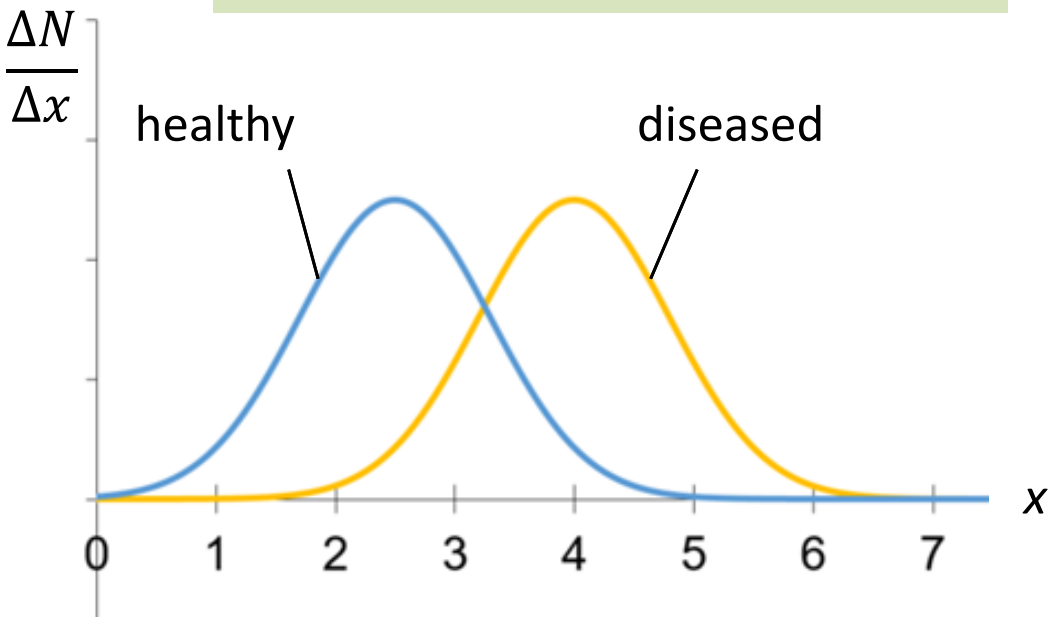
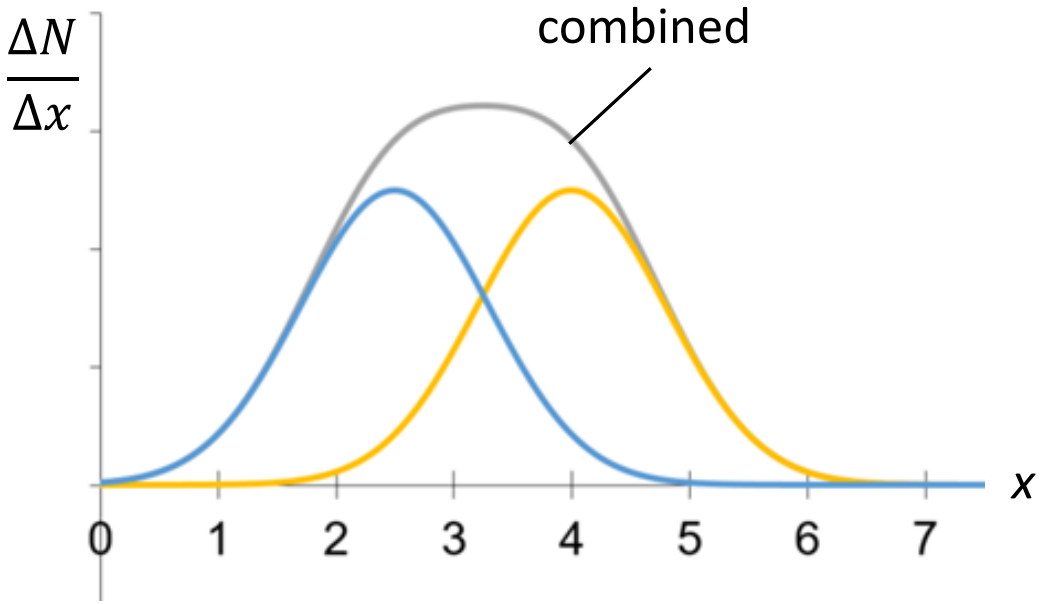


KAD 2023.11.07

Overlapping distributions





everything has a **distribution** (eye color, height, cholesterol level,…)

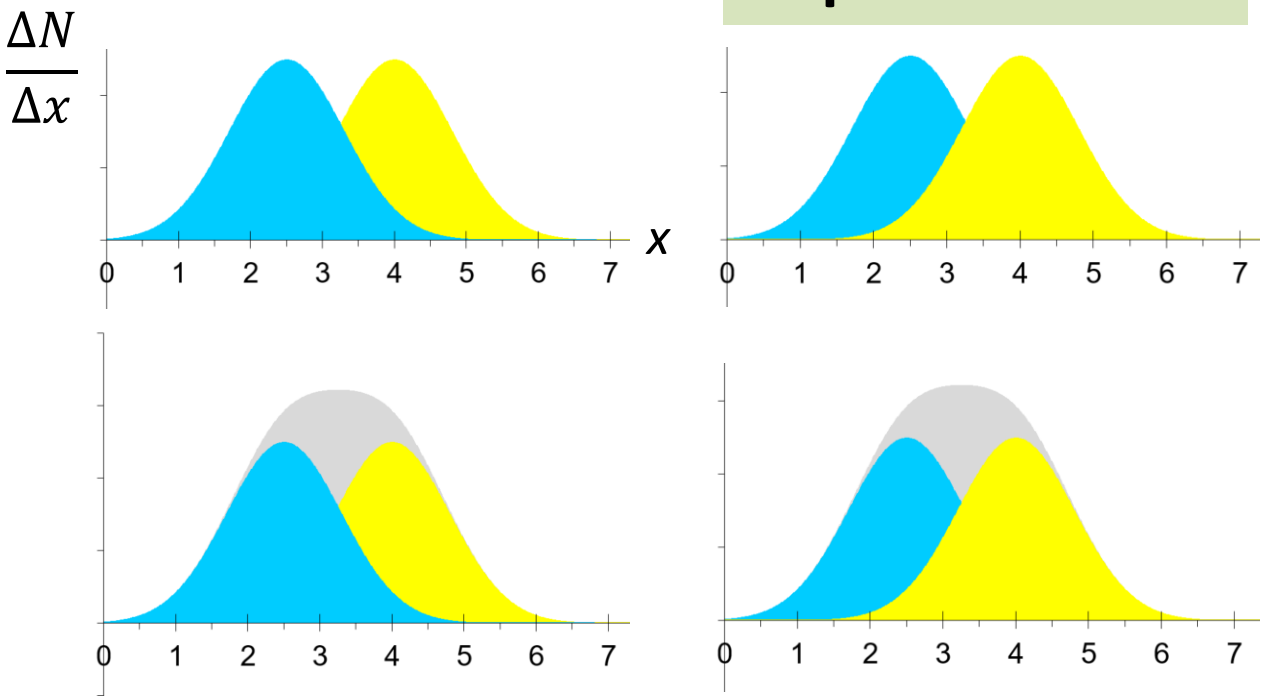
consider a continuous variable with a different distribution in the diseased population than in the healthy one

suppose that the measurable parameter is typically **larger** in the **diseased** population than in the healthy one (if smaller, the reciprocal of the original parameter can be used)

the figure shows two such density functions; the area under the curve corresponds to the number of individuals

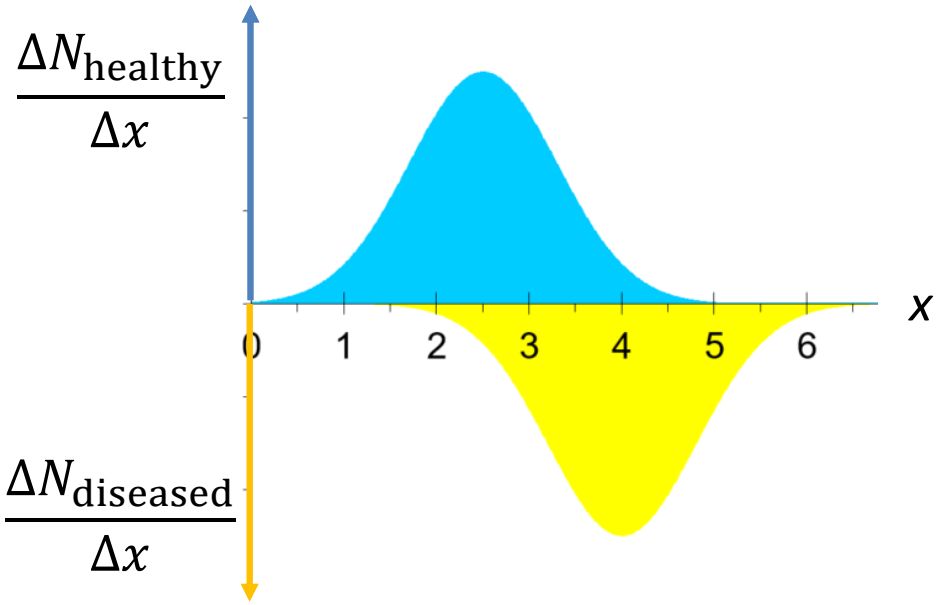
in the present example the number of healthy and diseased is the same (the standard deviation of the parameter is the same)

Representations



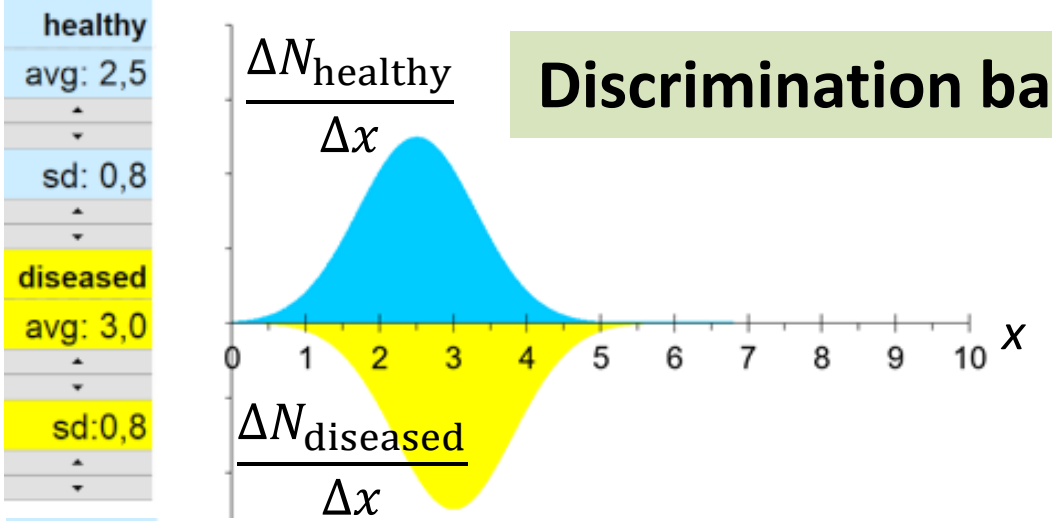
due to the great importance of the areas under the curve, we prefer an image that colors the areas instead of a line drawing

due to the overlaps, correct coloring is difficult (or impossible) in the usual representation

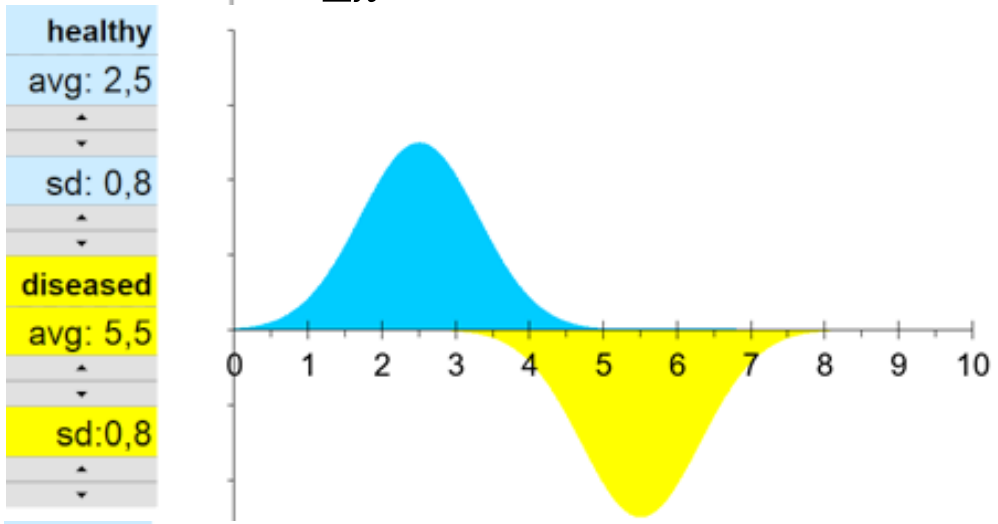


proposed new representation: instead of the negative axis, another positive axis, for the diseased

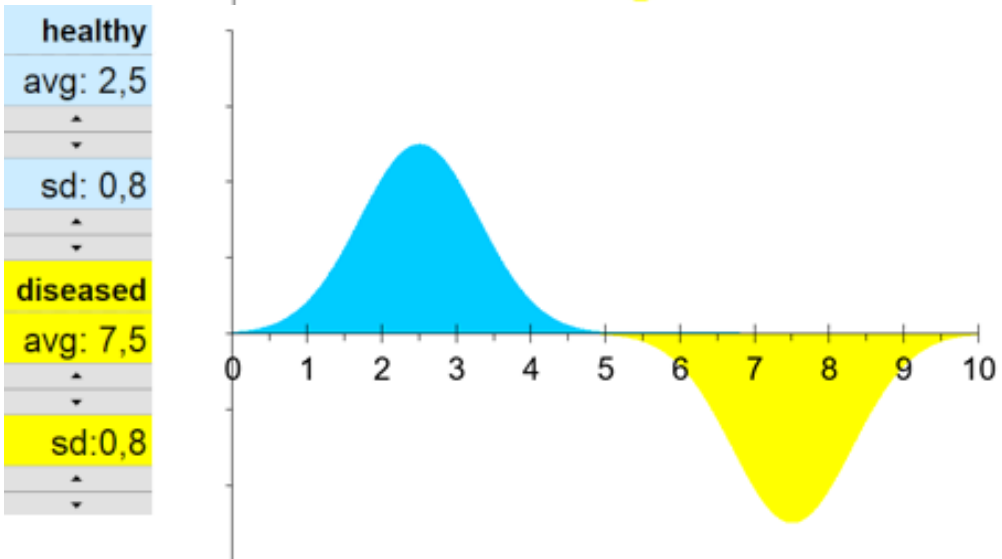
Discrimination based on overlap magnitude



full overlap — useless method



partial overlap — (possibly) usable method



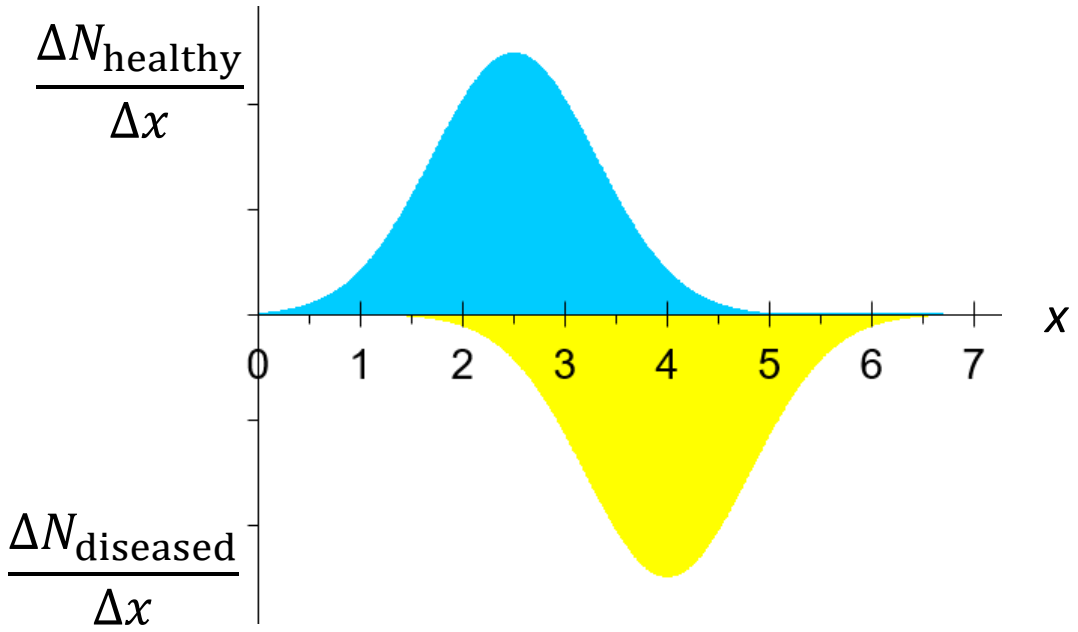
no overlap — perfect method

Prevalence

frequency of diseased in examined population

measure of how common the disease is

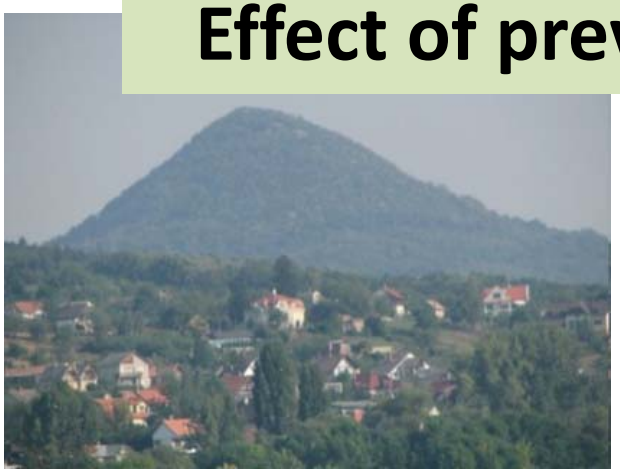
= probability prior to test
= a-priori-probablity

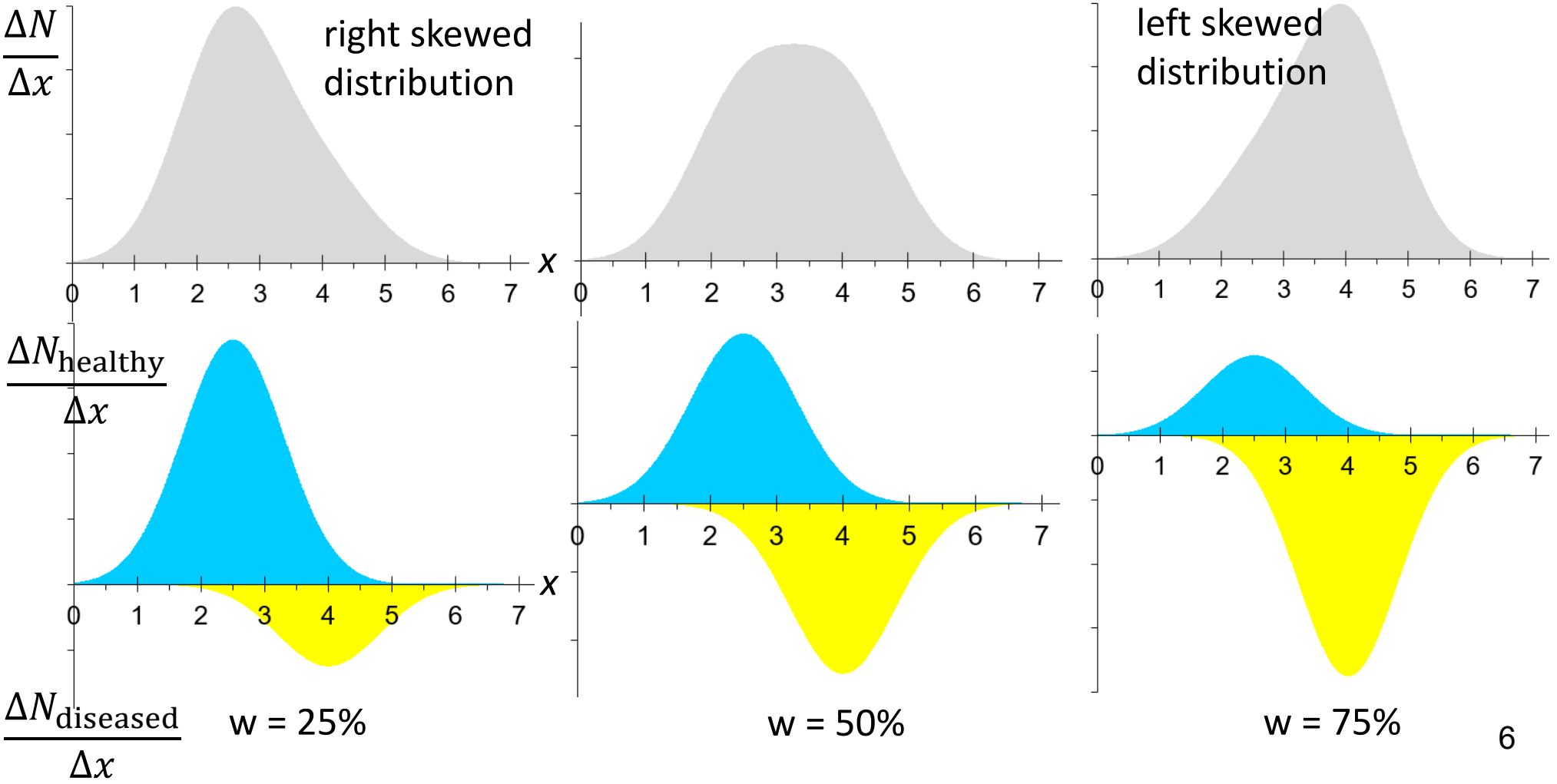


$$w = \frac{\text{diseased}}{\text{total}} = \frac{\text{diseased}}{\text{diseased} + \text{healthy}} = \frac{\text{de} - \text{sp}}{\text{se} - \text{sp}}$$

cf: incidence = the number of new cases in a given period and in a given number of population, e.g. 29 per year, per 10 000 people

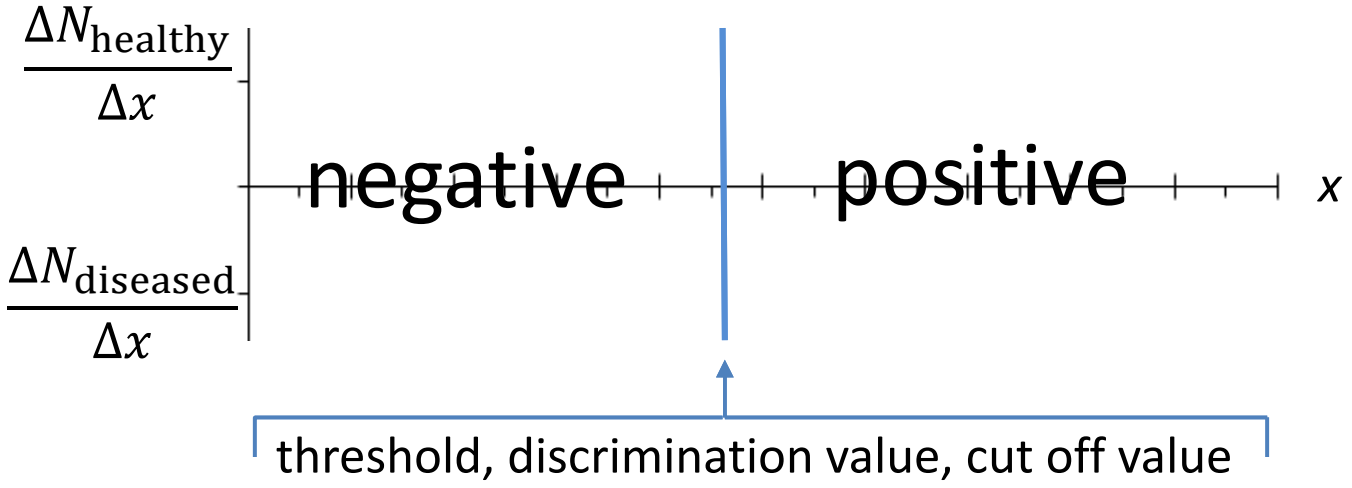
Effect of prevalence on combined distributions





A negative test result below the threshold and a positive test result above it

among the possible measurement parameter values, by designating a **threshold** value, we decide which will be the positive values and which are the negative ones according to the test method

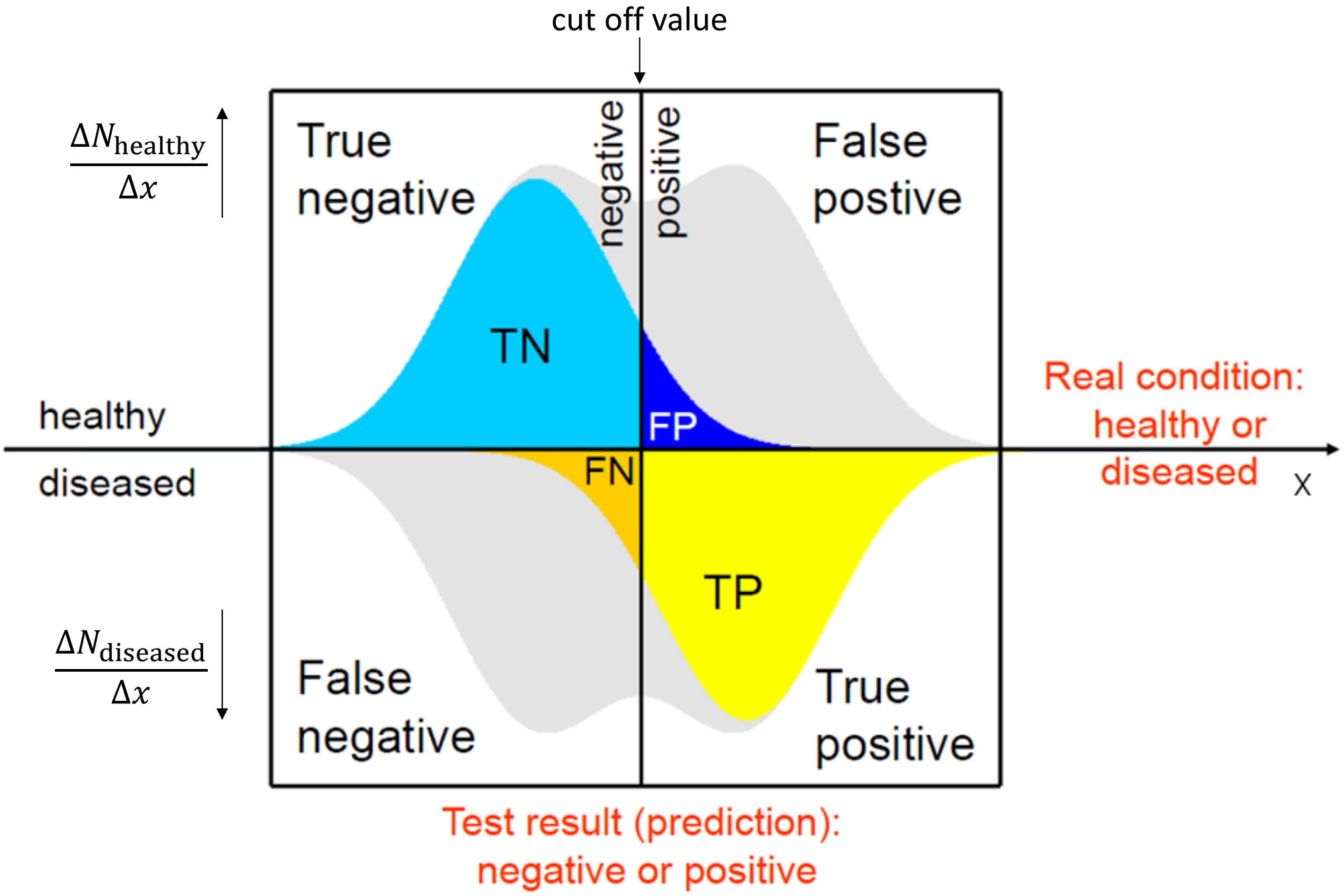


the wish/**desire**/request that the diseased and the positive, respectively healthy and negative match each other as much as possible

however, the **classification** is almost never perfect:

there will be diseased who are positive: true positive, TP ✓
there will be diseased who are negative: false negative, FN 💣
there will be healthy who are negative: true negative, TN ✓
there will be healthy who are positive : false positive, FP 💣

Confusion matrix



Parameters of diagnostic **„goodness”**

based on one (or more) measured parameters diagnostic tests divide the examined into (test) **positive** and (test) **negative** groups

the **“goodness”** of grouping **cannot** be characterized by a single number

(a) how well does it catch those **to be caught**?
e.g. the probability of a COVID infected stating/determining to be positive

(b) how well does it leave those **to be left alone**?
e.g. the probability of claiming to be negative for a person not infected with a COVID

(c) how reliable is a **positive test result**?
in the case of a positive test result, how certain the patient is diseased
e.g. in the case of a positive COVID test, how certain it is that the person is infected with COVID

(d) how reliable the **negative test result** is?
in the case of a negative test result, how certain the person is healthy
e.g. in the case of a negative COVID test, how certain it is that the person is not infected with COVID

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

Sensitivity

Specificity

PPV, relevance

NPV, segregation

only 3 independent!

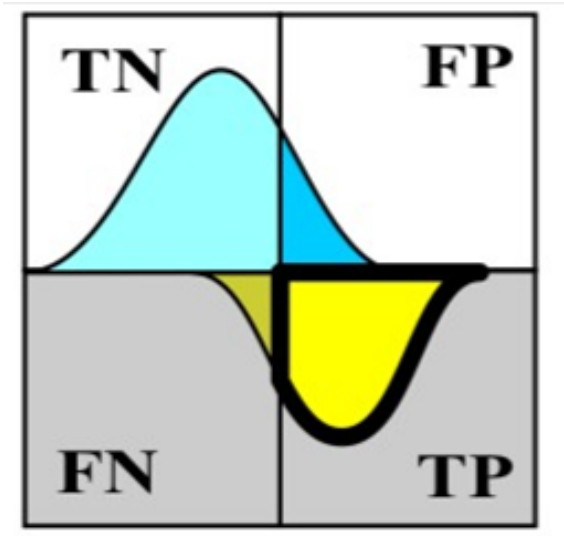
Every method must be compared with a reference-method: **gold standard**
method known to always work
(sometimes only the result of an autopsy)

Diagnostic **sensitivity**

= positive within diseased

= true postive rate

= recall rate



probability that the test finds the diseased positive

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

discr. threshold↓ sens. ↑

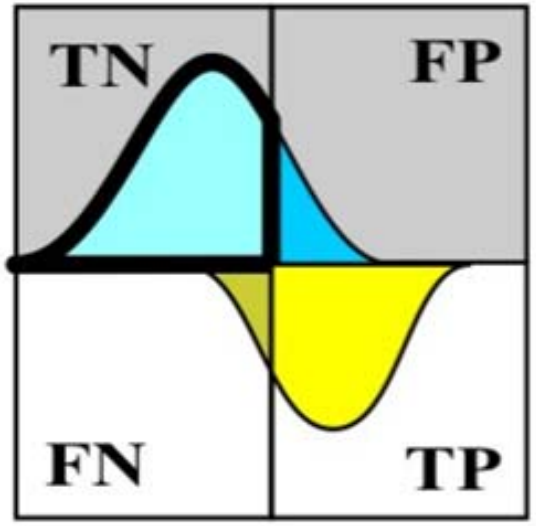
Large-sensitivity tests are required:

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

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

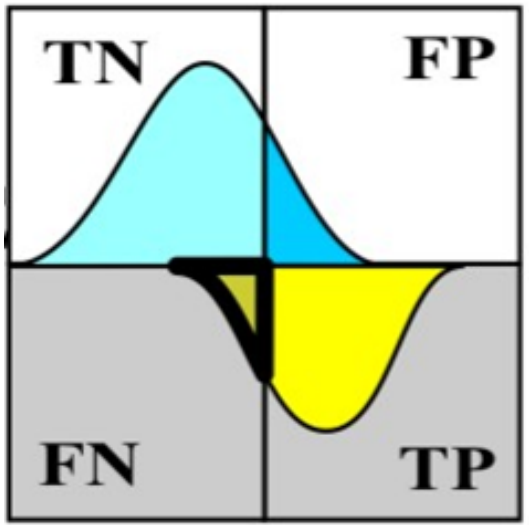
discr. thresold ↑ spec. ↑

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 Negative Rate

Type-II error



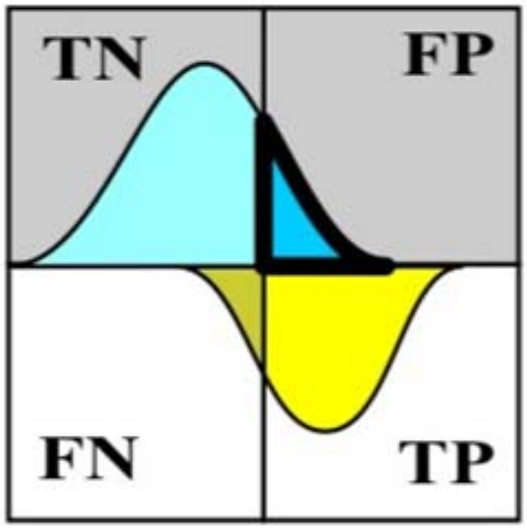
the probability that the test will find a diseased negative

negative among diseased

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

Diagnostic False Positive Rate

Type-I error



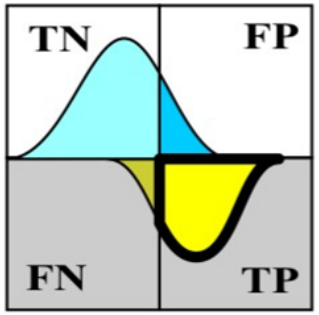
the probability that the test will find a healthy positive

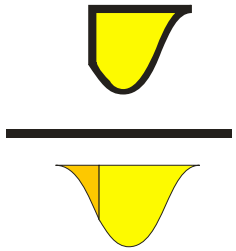
positive among the healthy

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

Horizontal rates are independent of prevalence

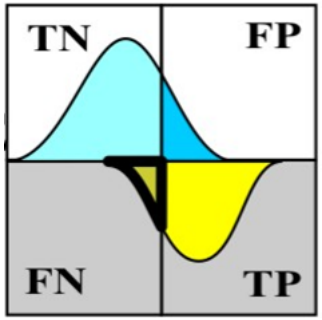
sensitivity
(se)

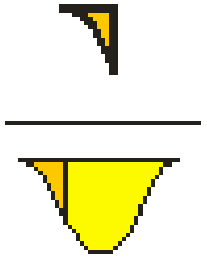




$$se = \frac{TP}{TP + FN}$$

false negative rate
(1-se)

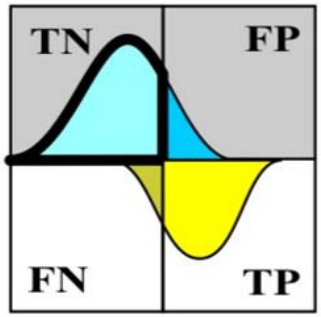


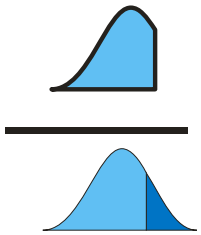


$$1 - se = \frac{FN}{FN + TP}$$

---

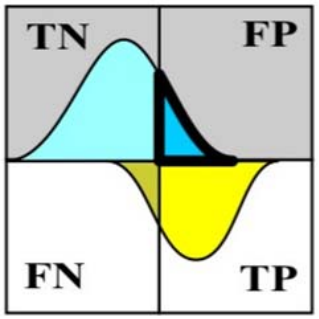
specificity
(sp)

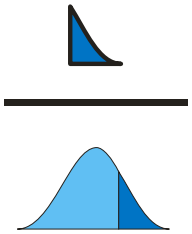




$$sp = \frac{TN}{TN + FP}$$

false positive rate
(1-sp)





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

Predictive values (vertical rates)

a-posteriori-probabilities; they depend strongly on prevalence

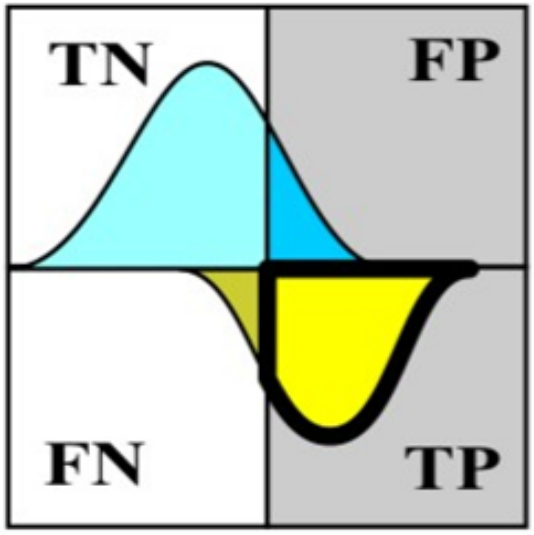
Positive predictive value

= **PPV**

= predictive value positive

= PVP

= diagnostic **relevance**



probablity of disease if test is positive

diseased among positives

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

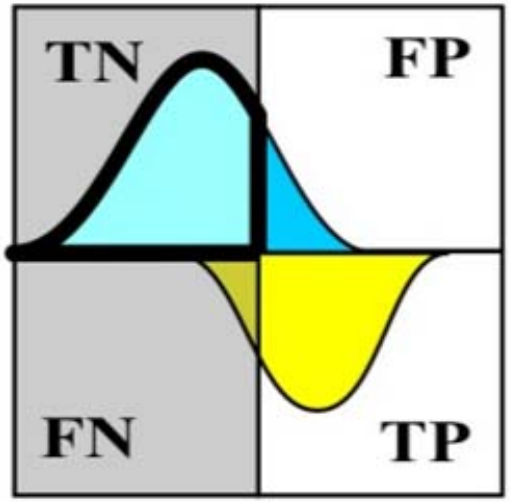
Negative predictive value

= **NPV**

= predictive value negative

= PVN

= diagnostic **segregation**



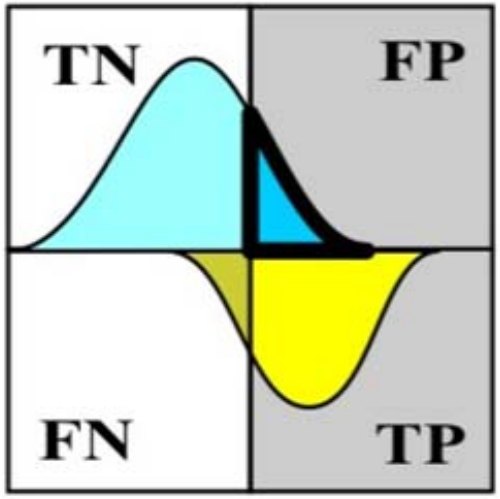
probabilty of health if test is negative

healthy among negatives

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

False alarm rate

=1-PPV



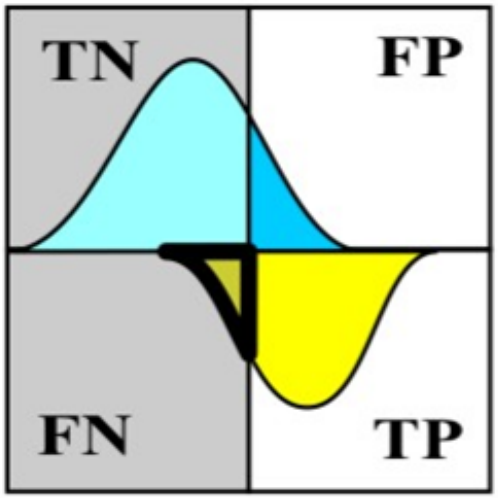
the probability of the absence of the disease if the test is positive

healthy among positives

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

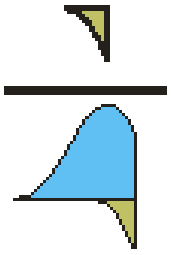
False reassurance rate

=1-NPV



the probability of the presence of the disease if the test is negative

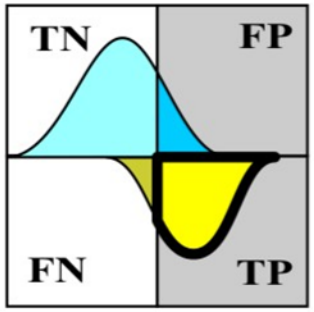
diseased among negatives



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

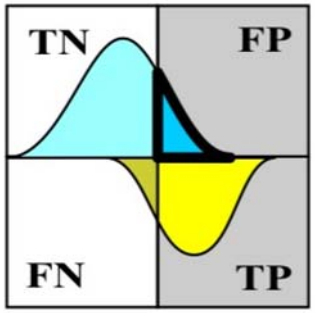
Vertical rates are dependent of prevalence

positive predictive value (PPV)



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

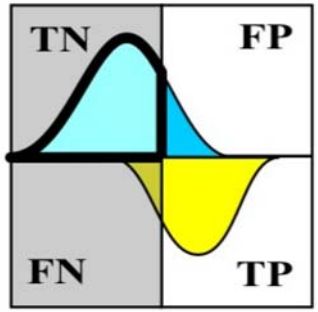
false alarm rate (1-PPV)



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

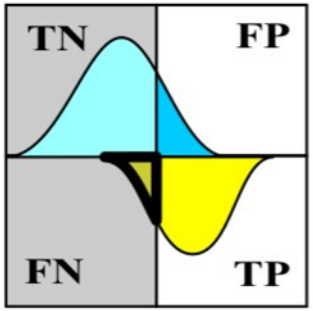
---

negative predictive value (NPV)



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

false reassurance rate (1-NPV)



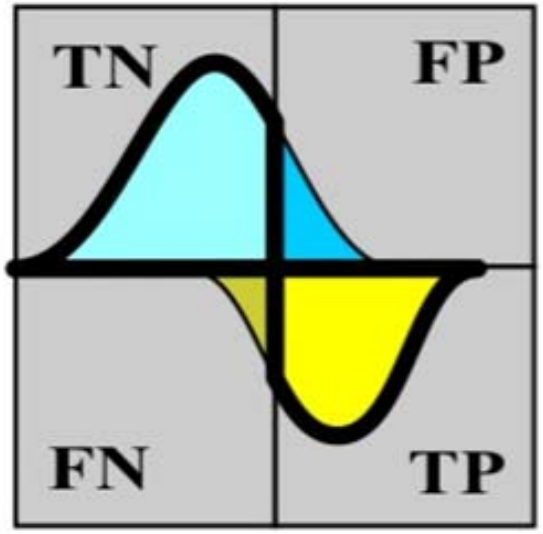
$$1 - NPV = \frac{FN}{TN + FN}$$

Diagnostic **accuracy**

= da=de

= efficacy/efficiency

= correct classification rate



probability of correct diagnosis

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

often: discrimination thresold is chosen so that accuracy is maximized

Effect of prevalence

NPV = 94.7%

case1: ***w* = 50%**

		Test	
		negative	positive
Gold-standard	healthy	90	10
	diseased	5	95

sp = 90%

se = 95%

(de = 92.5%)

PPV = 90.5%

NPV = 99.4%

Case 2: ***w* = 10%**

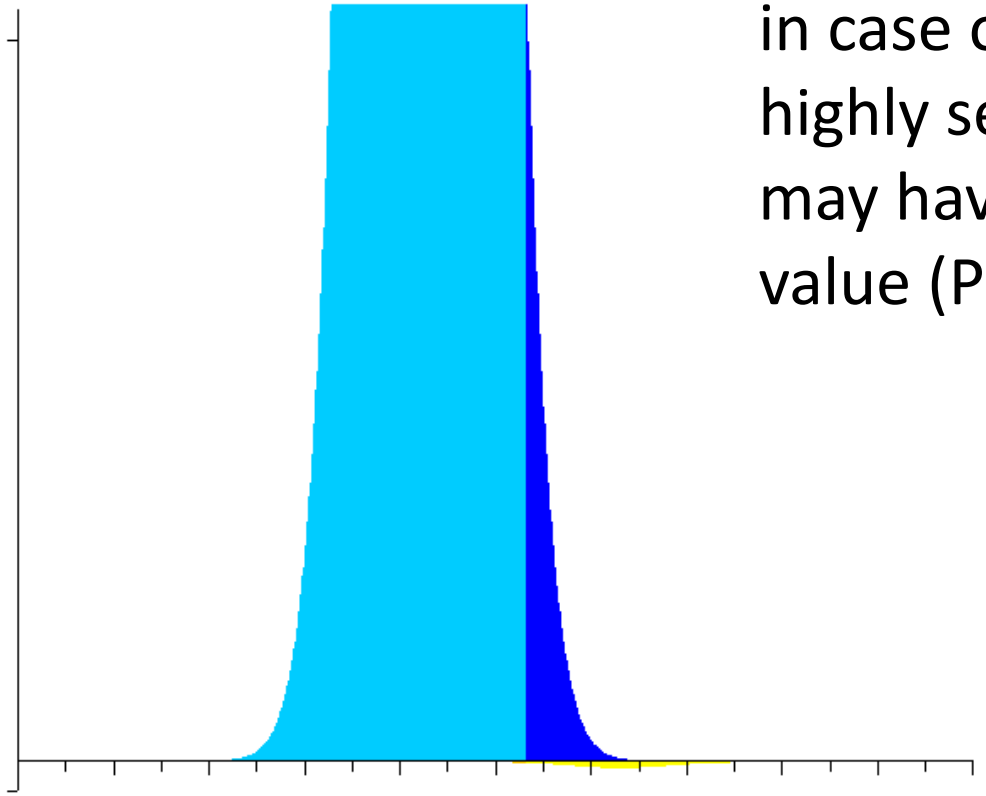
		Test	
		negative	positive
Gold-standard	healthy	810	90
	diseased	5	95

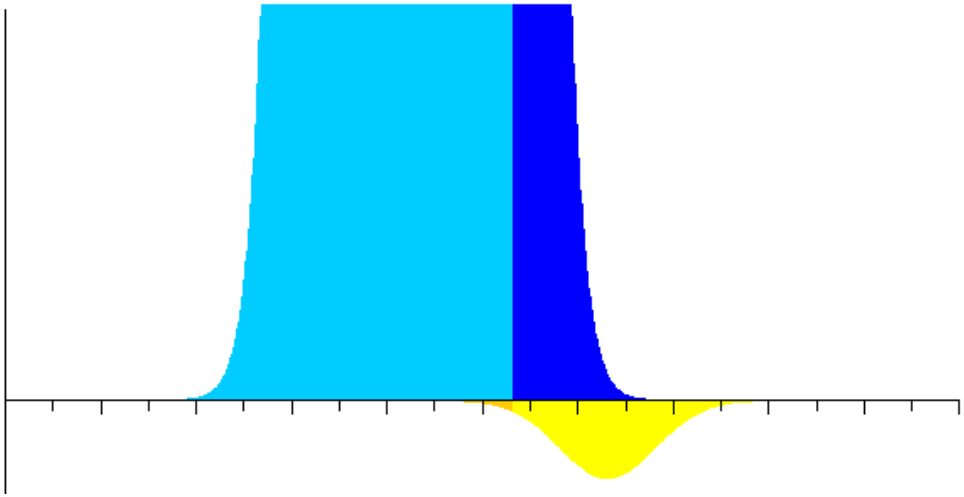
sp = 90%

se = 95%

(de = 90.5%)

PPV = 51.4%





in case of very small prevalence a highly sensitive and specific test may have low positive predictive value (PPV)

prevalence = 0.1 %

sensitivity = 98 %

specificity = 98 %

↓

PPV = 4 %

A patient comes to your office frantic over the results of a home HIV test. The test touts 99% sensitivity and 99% specificity. On questioning, you determine that this patient is at low risk for HIV; given your assessment of his risk factors, you believe he comes from a population group that has a baseline prevalence of HIV of 1 in 100,000. He now presents to you with a positive result on his home HIV test. Given his baseline risk and the positive home test, what are the chances that this patient is actually HIV positive?

(PPV = $9.89 \times 10^{-4} \cong 0.001$)

Stuart Spitalnic: Test properties I: Sensitivity, specificity, and predictive values; Hospital Physician, September 2004, 27-31

Overview

Sensitivity	se	$\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}$	$p(P\|D)$	positive within diseased	True Positive Rate
Specificity	sp	$\frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}}$	$p(N\|H)$	negative among healthy	True Negative Rate
False Negative Rate	1-se	$\frac{\mathrm{FN}}{\mathrm{TP}+\mathrm{FN}}$	$p(N\|D)$	negative among diseased	
False Positive Rate	1-sp	$\frac{\mathrm{FP}}{\mathrm{TN}+\mathrm{FP}}$	$p(P\|H)$	positive among the healthy	
Positive Predictive Value	PPV	$\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}$	$p(D\|P)$	diseased among positives	Relevance
Negative Predictive Value	NPV	$\frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FN}}$	$p(H\|N)$	healthy among negatives	Segregation
False alarm rate	1-PPV	$\frac{\mathrm{FP}}{\mathrm{TP}+\mathrm{FP}}$	$p(H\|P)$	healthy among positives	
False reassurance rate	1-NPV	$\frac{\mathrm{FN}}{\mathrm{TN}+\mathrm{FN}}$	$p(D\|N)$	diseased among negatives	

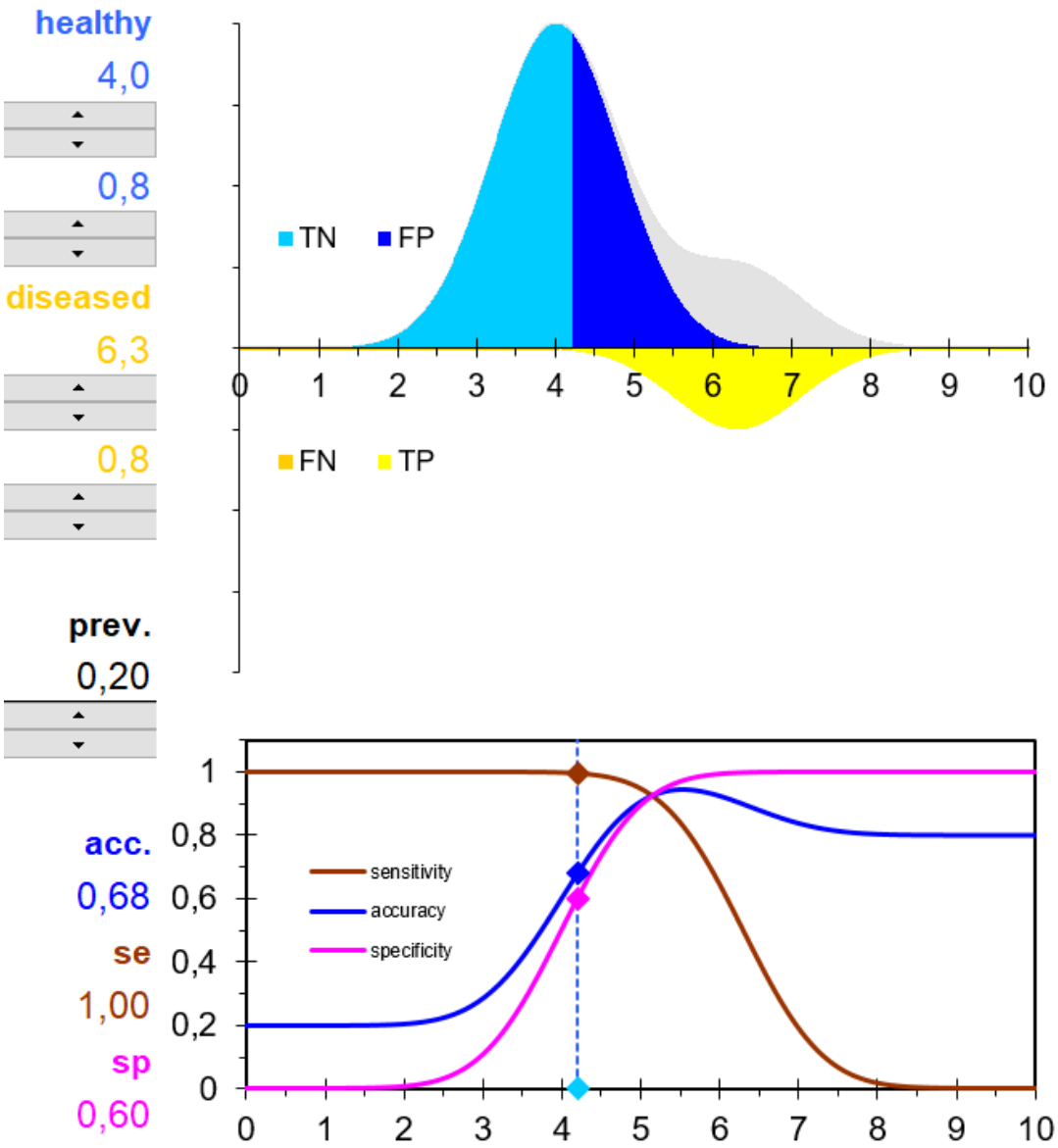
prevalence-independent

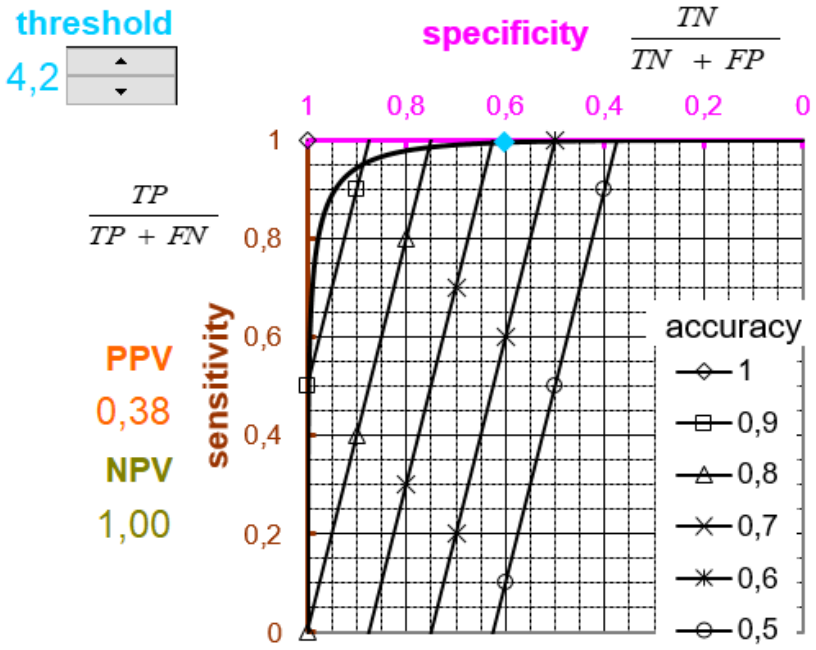
prevalence-dependent

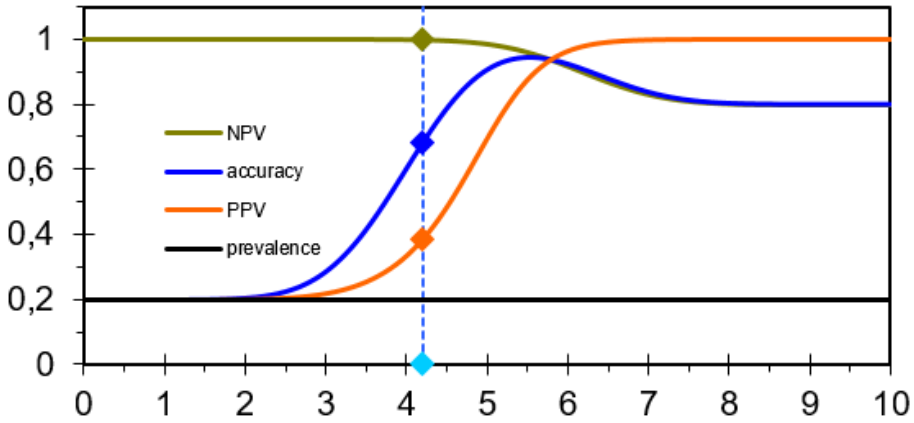
conditional probability (Bayes)

how well does it catch those **to be caught**?

Maximize diagnostic **sensitivity**

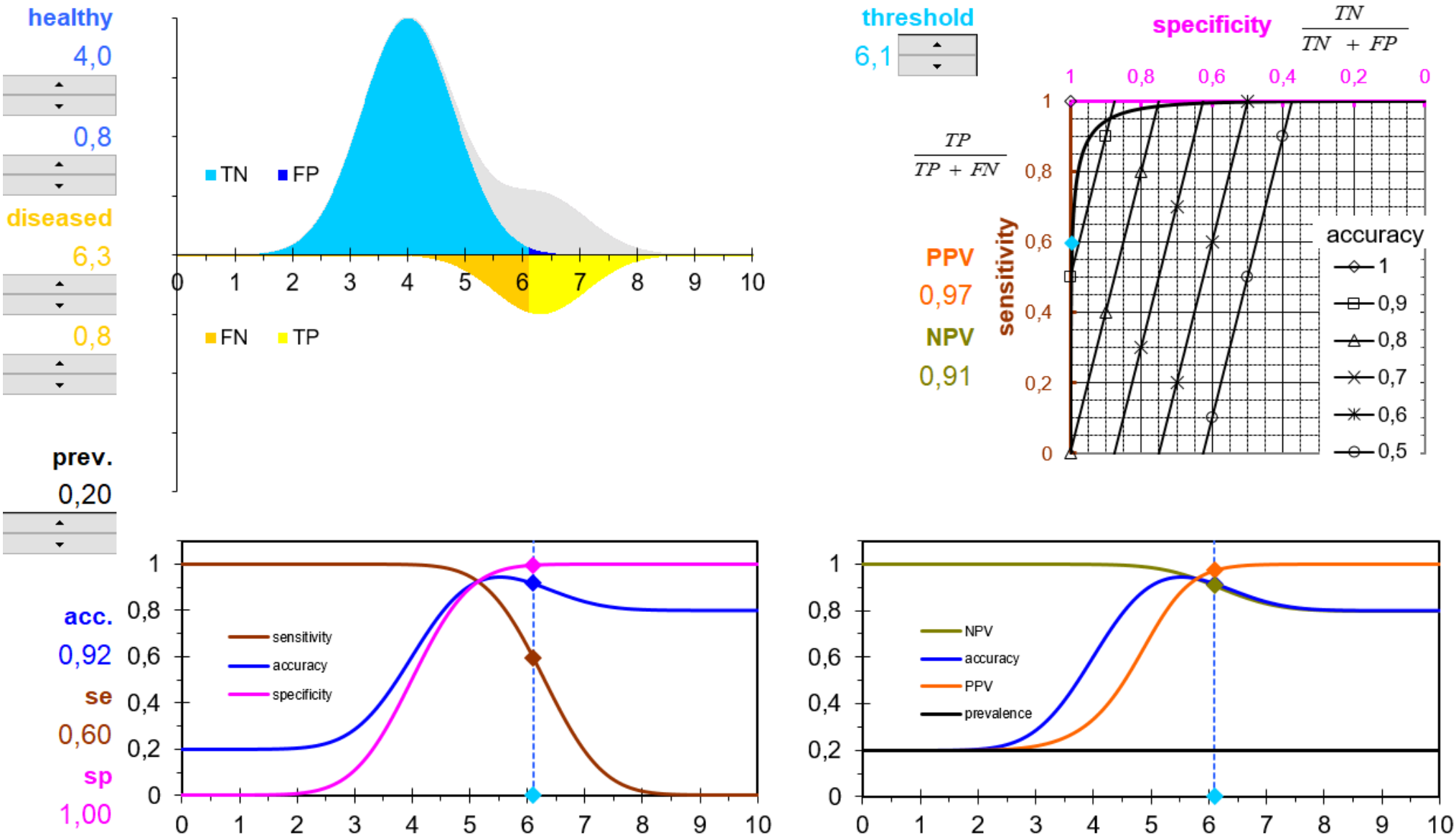






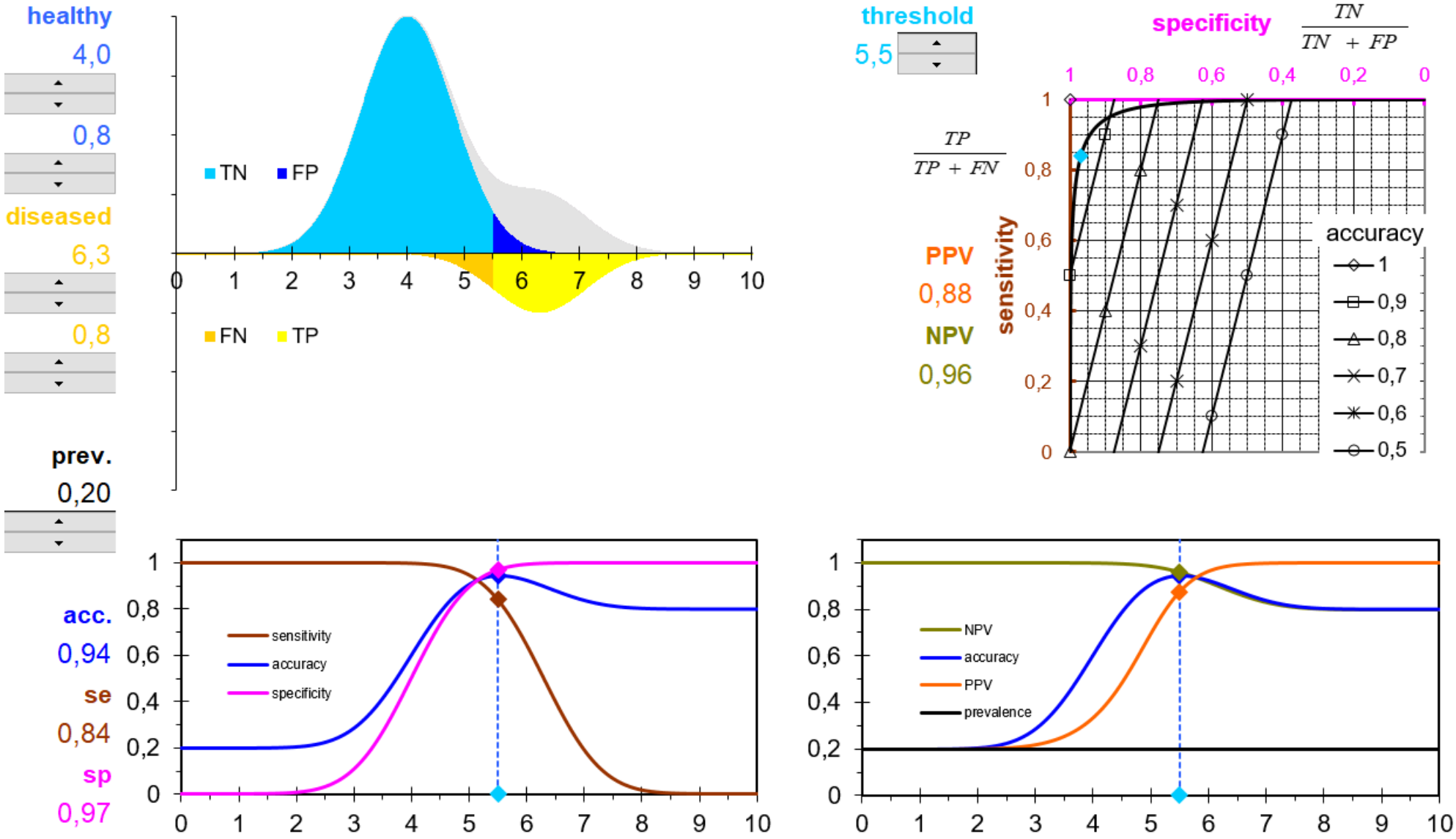
how well does it leave those **to be left alone**?

Maximize diagnostic **specificity**



it is equally important to catch those **to be caught** and to leave those **to be left alone**?

Maximize diagnostic **accuracy**



Take-home message

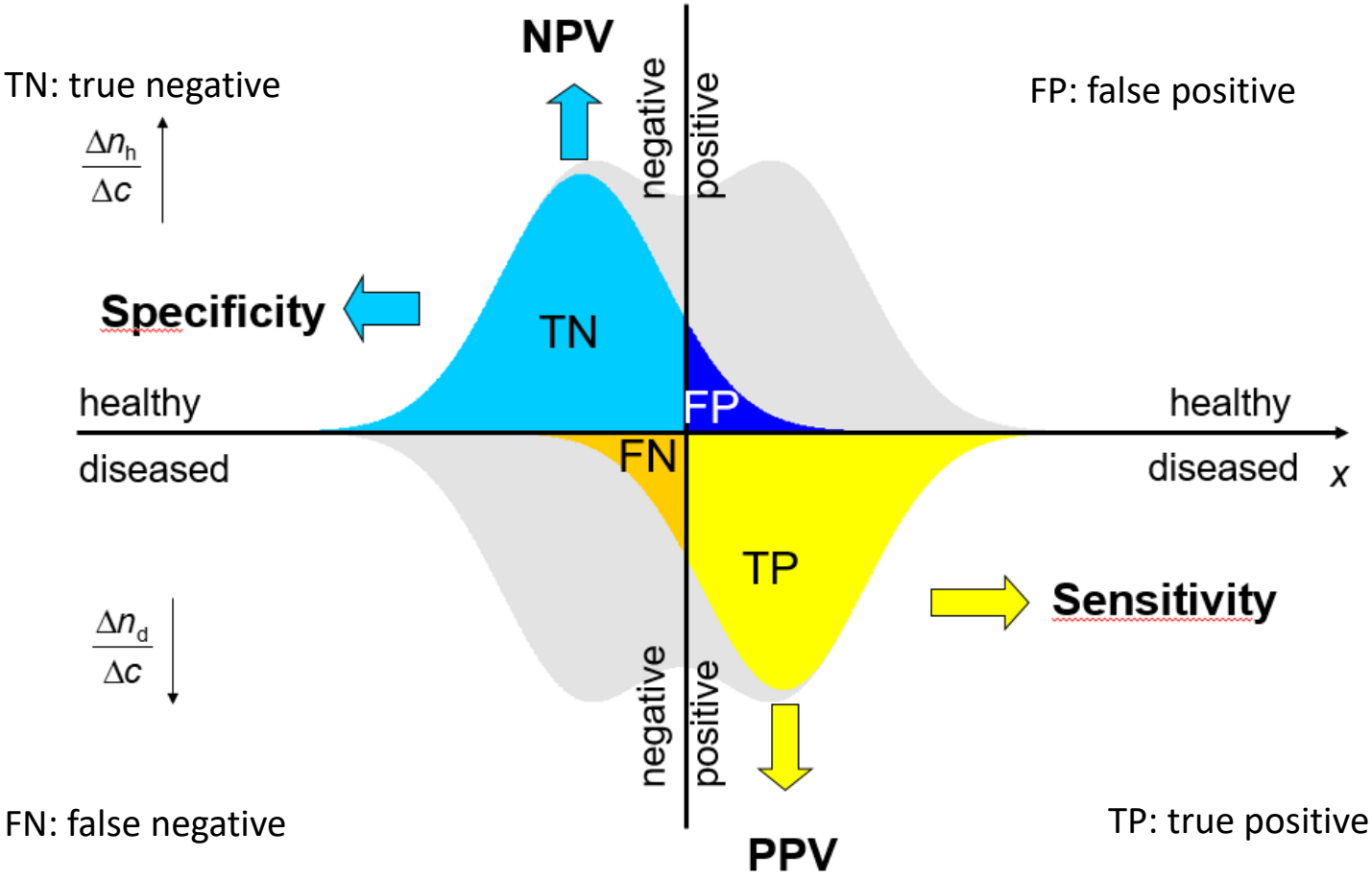
everything has a **distribution**; the distribution of sick and healthy values **overlap**

whether it is possible to decide which is more important :

to detect the disease in as many patients as possible in order to receive treatment (**maximizing sensitivity**), or

to assume a false positive value (minimizing false-positive ratio or **maximizing specificity**) in as few healthy people as possible so that they do not receive unnecessary therapy

if they cannot be decided, they are equally important: **maximizing accuracy**



TN: true negative

FP: false positive

diagnostic accuracy:

$$de = \frac{TP + TN}{total} = \frac{TP + TN}{TN + FP + FN + TP}$$

in case of several possible methods: ROC

FN: false negative

TP: true positive