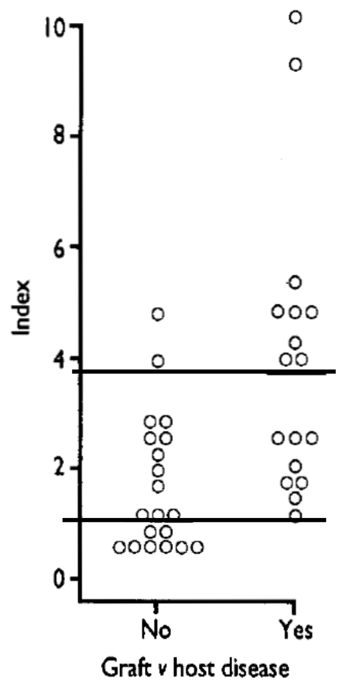
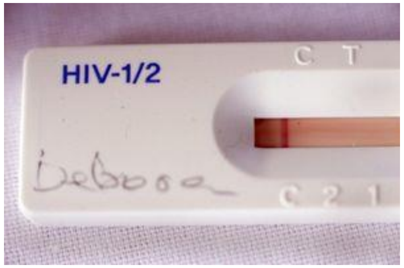
Biostatistics
Week 7

- **Diagnostic tests as "patient classifier"**
  - How can we describe the quality of a diagnostic test with binary outcome:
  → Sensitivity, Specificity

  - How can we describe the predictive value of a binary diagnostic test:
  → PPV, NPV or positive and negative predictive value

  - How to evaluate a diagnostic test with continuous score outcome:
  → ROC curve analysis and its AUC





How to quantify the performance of a test?

1. Performance characteristics of a diagnostic test in a lab setting

   Sensitivity

   Specificity

   Choice of a threshold

2. Performance of a diagnostic test in a population application

   Positive predictive value of a test (PPV)

   Negative predictive value of a test (NPV)

   Impact of disease prevalence, sensitivity and specificity on predictive values

Binary test ore binary classification rule

Explanatory variable **X**
(e.g.blood sample)

Classification

predicted class
$\hat{Y}$

Target Variable Y

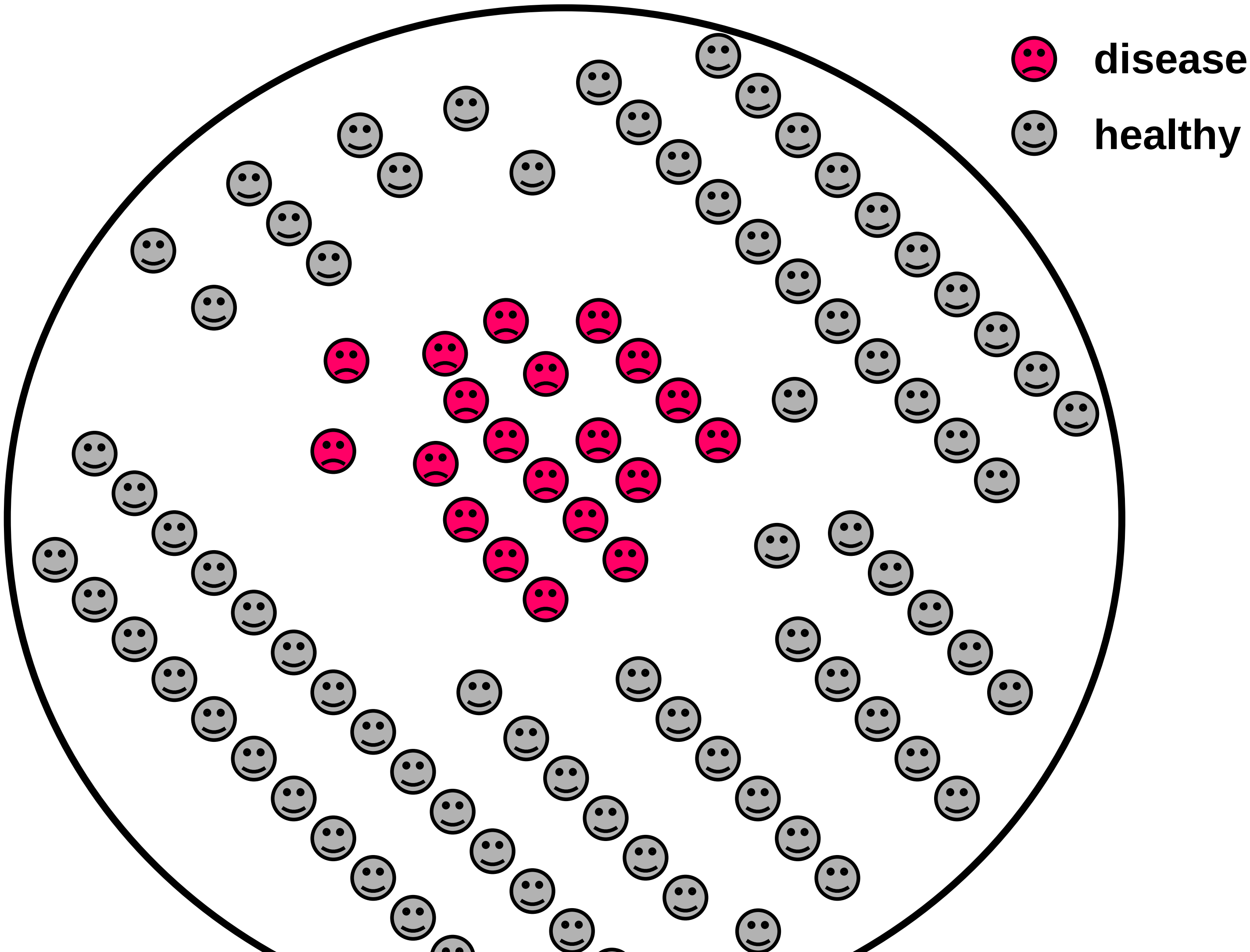
2 classes:
Postive or Negative
1 or 0
Yes or No
Diseased or Healthy

Each observation unit described by input **x,** belongs to one of two classes.

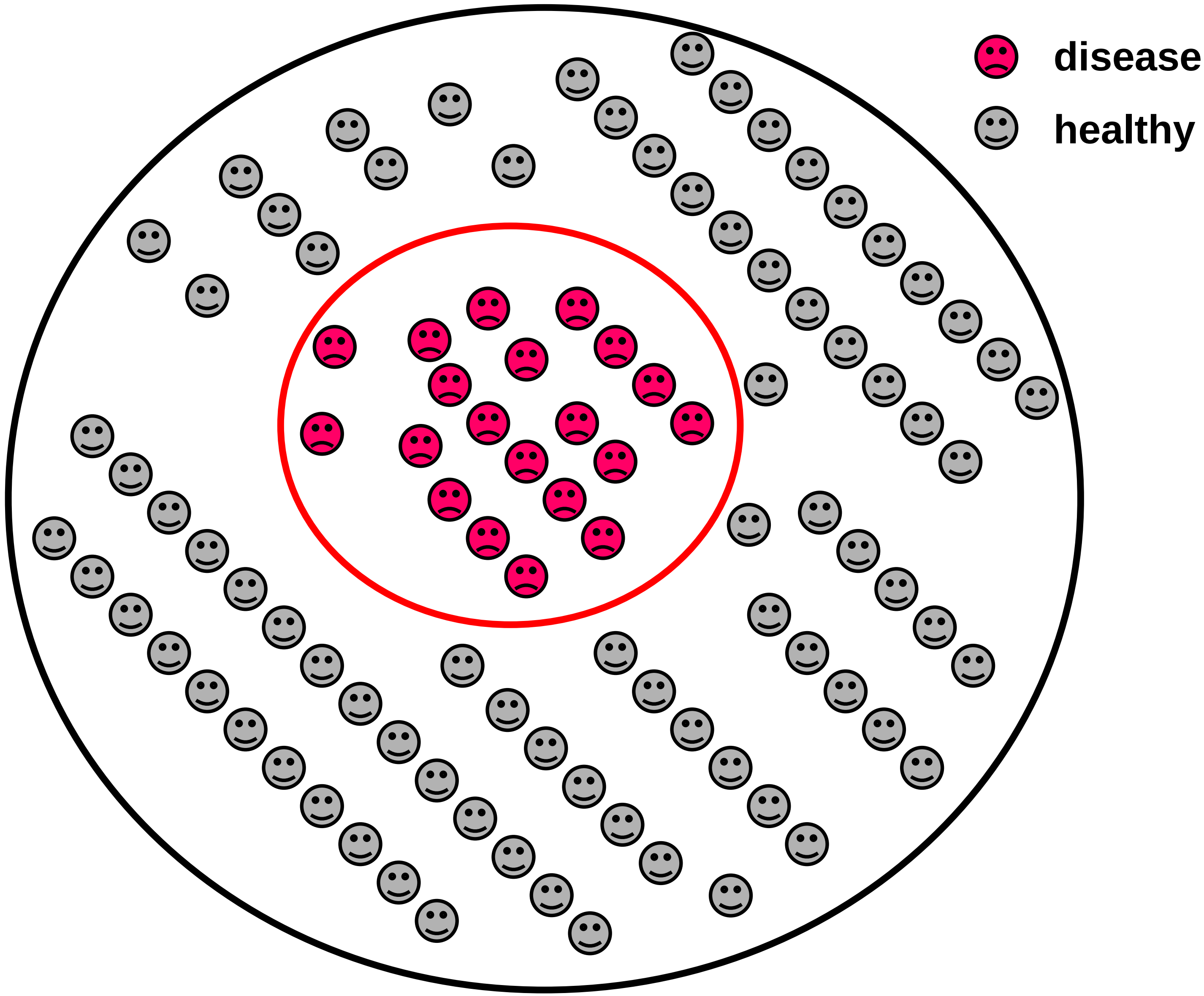
Y: true class

$\hat{Y}$: predicted class

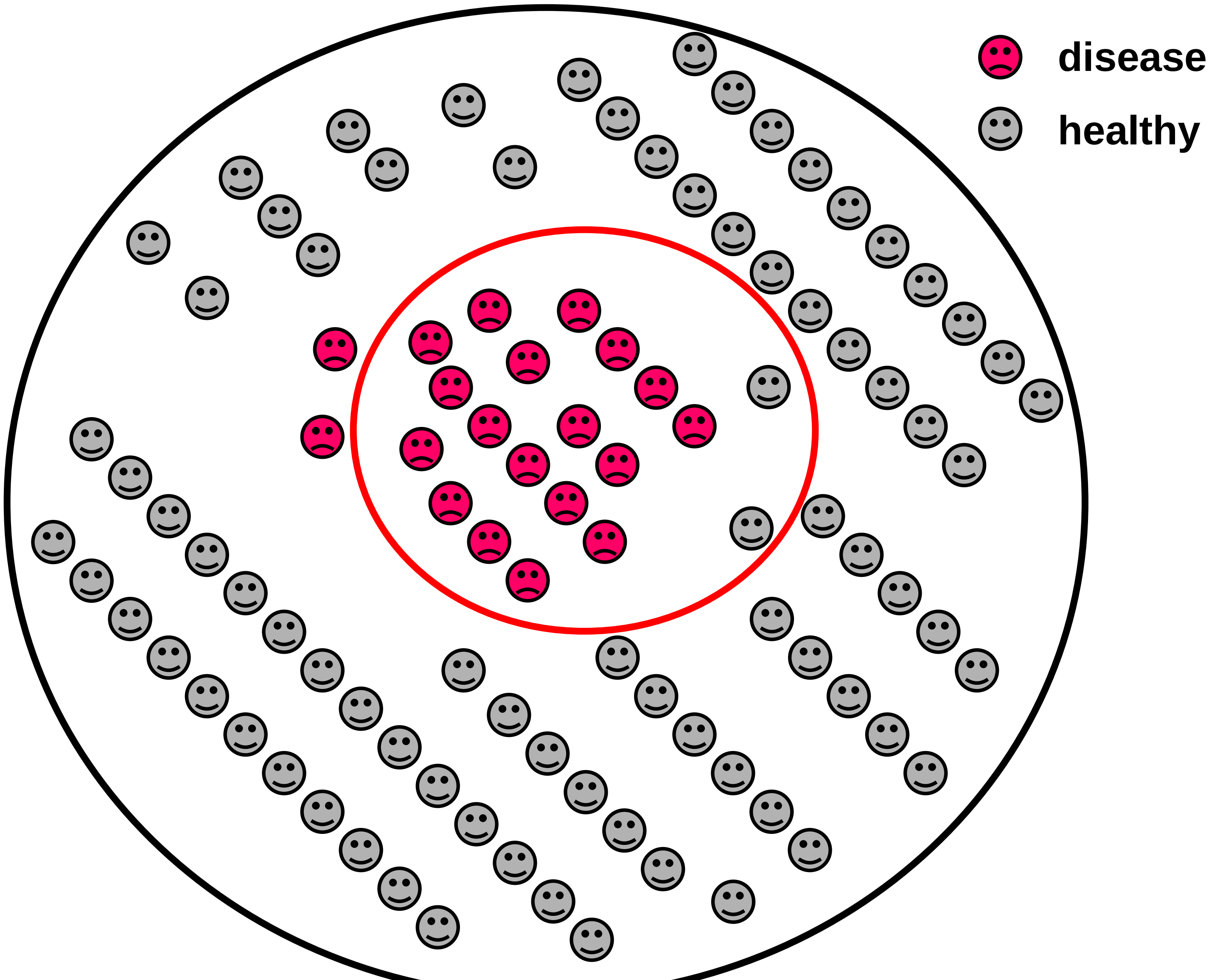
Population
with diseased and healthy individuals



A perfect diagnostic test turns out positive for the diseased individuals only

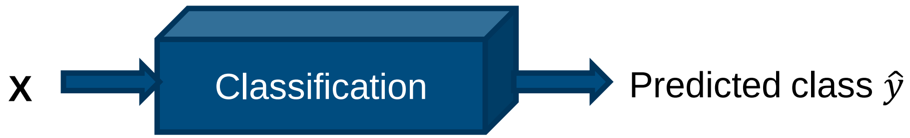


Real tests are not perfect



Confusion matrix: Evaluate a performed classification

Evaluation is done on a test set with known true class y and the predicted class $\hat{y}$.



id	true_class	pred_class
1	P	P
2	N	P
3	N	N
4	P	P
5	N	N
6	N	N

		True class	
		Positive	Negative
Predicted class	Positive	TP=2	FP=1
	Negative	FN=0	TN=3

Sensitivity and Specificity derived from a confusion matrix

Evaluation is done on a test set with known true class labels y and the predicted class label $\hat{y}$.

True class

Predicted class		**Positive**	**Negative**
	Positive	**TP**	**FP**
	Negative	**FN**	**TN**
		$sens = \frac{TP}{TP+FN}$	$spec = \frac{TN}{FP+TN}$

The sensitivity is derived from the positive examples and the specificity from the negative examples → both do not depend on the ratio of positive and negative classes in the test sample.

The sensitivity (recall) of a binary classifier is its ability to identify correctly the positive class.

Also called true positive rate (TPR) since it corresponds to the proportion of “Positive” instances that were classified as “Positive”

The specificity of a binary classifier is its ability to identify correctly the negative class.

Also called true negative rate (TNR) since it corresponds to the proportion of “Negative” instances that were classified as “Negative”

Positive predictive value (PPV) and negative predictive value (NPV)

Evaluation is done on a test set with known true class labels y and the predicted class label $\hat{y}$.

True class / **Predicted class**

	Positive	Negative	
Positive	TP	FP	$PPV = \frac{TP}{TP + FP}$
Negative	FN	TN	$NPV = \frac{TN}{TN + FN}$
	$sens = \frac{TP}{TP + FN}$	$spec = \frac{TN}{FP + TN}$	

The PPV gives the probability that a instance, that was as “positive” predicted, is indeed “positive”.

The NPV gives the probability that a instance, that was as “negative” predicted, is indeed “negative”

The PPV is derived from all as positive classified examples and the NPV from all as negative classified examples → both depend on the ratio of positive and negative classes in the two prediction groups.

Performance measures expressed as (conditional) probabilities

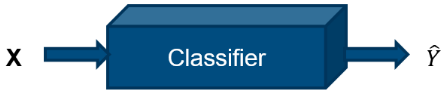
- **$P(\hat{Y} = Y)$ = acc** : accuracy

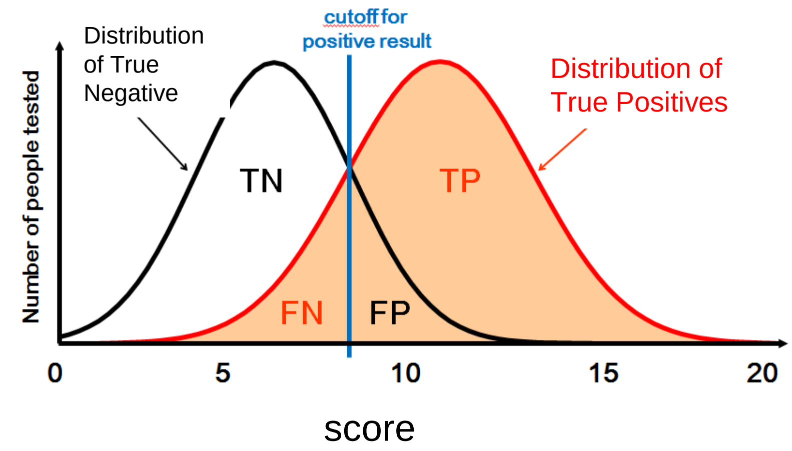
- **$P(\hat{Y} = 1 \mid Y = 1)$ = Sens** : true positive rate or sensitivity or recall
- **$P(\hat{Y} = 0 \mid Y = 0)$ = Spec** : true negative rate or specificity

- **$P(Y = 1 \mid \hat{Y} = 1)$ = PPV** : positive predictive value or precision
- **$P(Y = 0 \mid \hat{Y} = 0)$ = NPV** : negative predictive value

		True class		
		Positive	**Negative**	
Predicted class	**Positive**	**TP**	**FP**	$PPV = \frac{TP}{TP+FP}$
	Negative	**FN**	**TN**	$NPV = \frac{TN}{TN+FN}$
		$sens = \frac{TP}{TP+FN}$	$spec = \frac{TN}{FP+TN}$	

Score based classifier



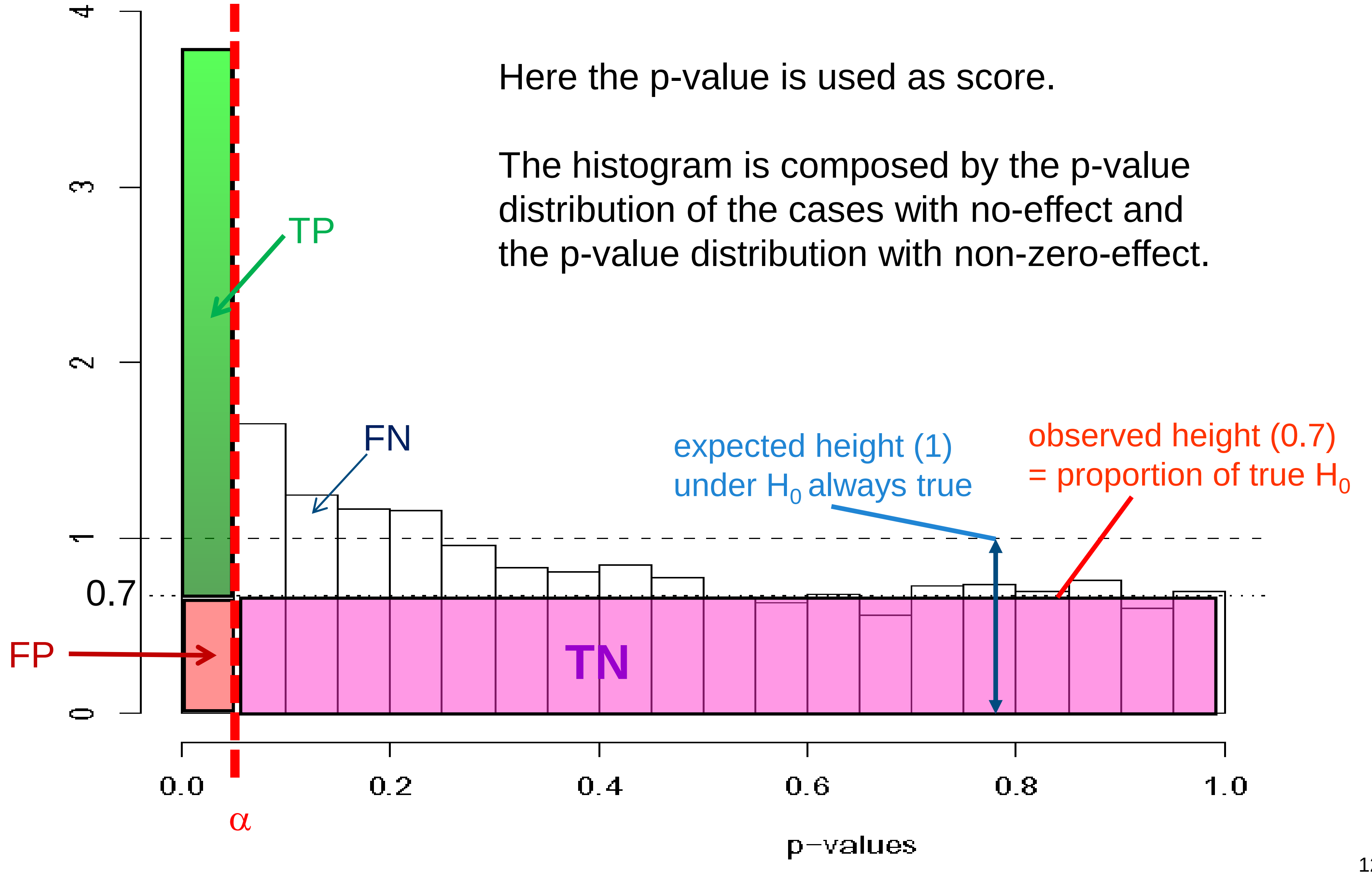


- Output: continuous score $\hat{Y}(x)$ (instead of actual class prediction)
- Discretized by choosing a cut-off
  - score ≥ c ➔ class «positive» or 1
  - score < c ➔ class «negative» or 0

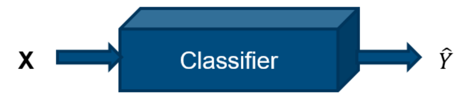
True class

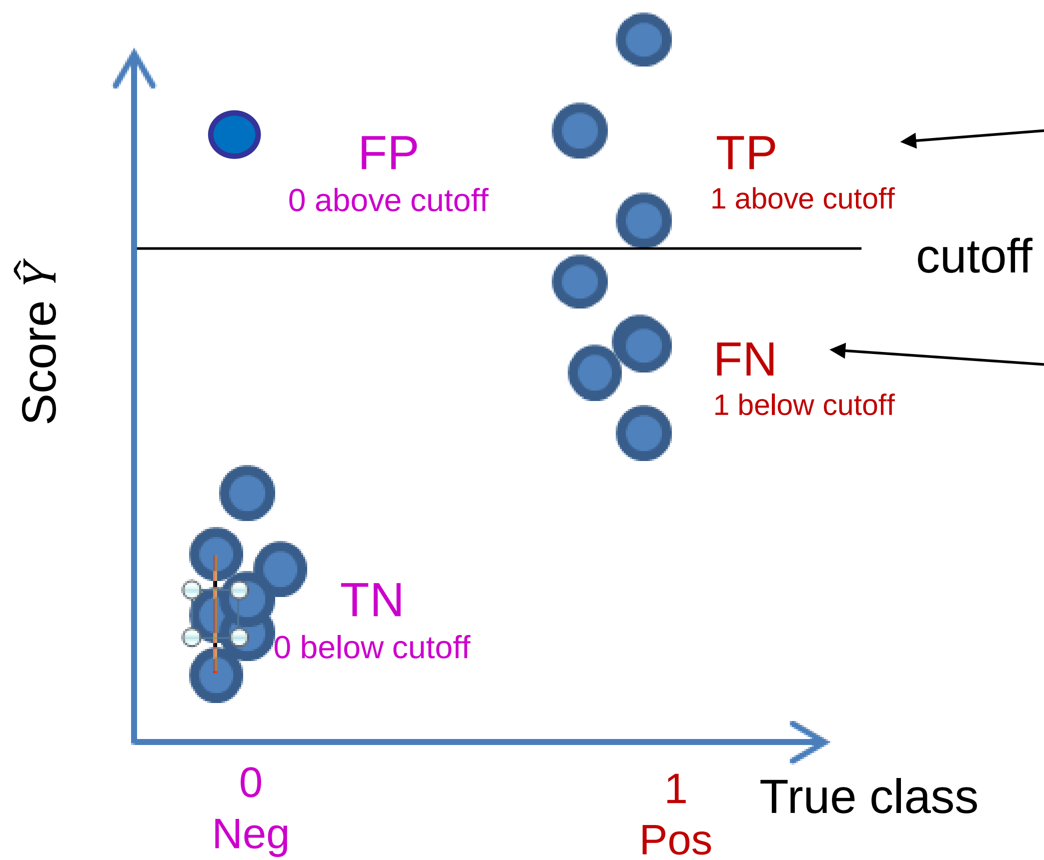
Predicted class		**Positive**	**Negative**	
	Positive	**TP**	**FP**	$PPV = \frac{TP}{TP+FP}$
	Negative	**FN**	**TN**	$NPV = \frac{TN}{TN+FN}$
		$sens = \frac{TP}{TP+FN}$	$spec = \frac{TN}{FP+TN}$	

Recall the p-value histogram
we can read off the content of the confusion matrix



Score based classifier





All instances with a score above the cutoff get prediction 1/Pos.

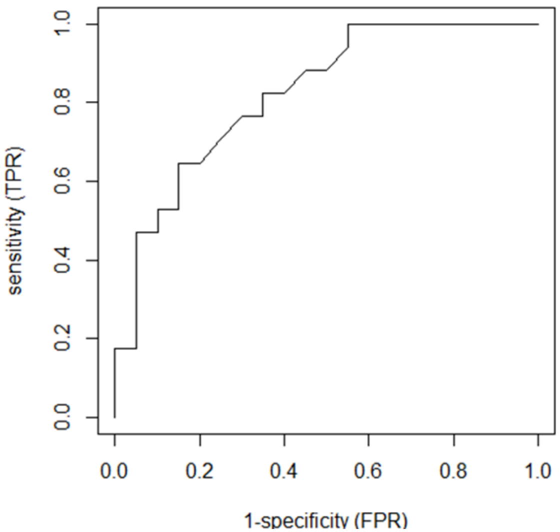
All instances with a score below the cutoff get prediction 0/Neg.

		True class	
		Positive	**Negative**
Predicted class	**Positive**	**TP**	**FP**
	Negative	**FN**	**TN**

We can use a continuous score such as probability to construct a ROC curve

For each cutoff we get a classification rule (classify each observation with score>cutoff as class 1) and a corresponding confusion matrix and can determine sensitivity and specificity

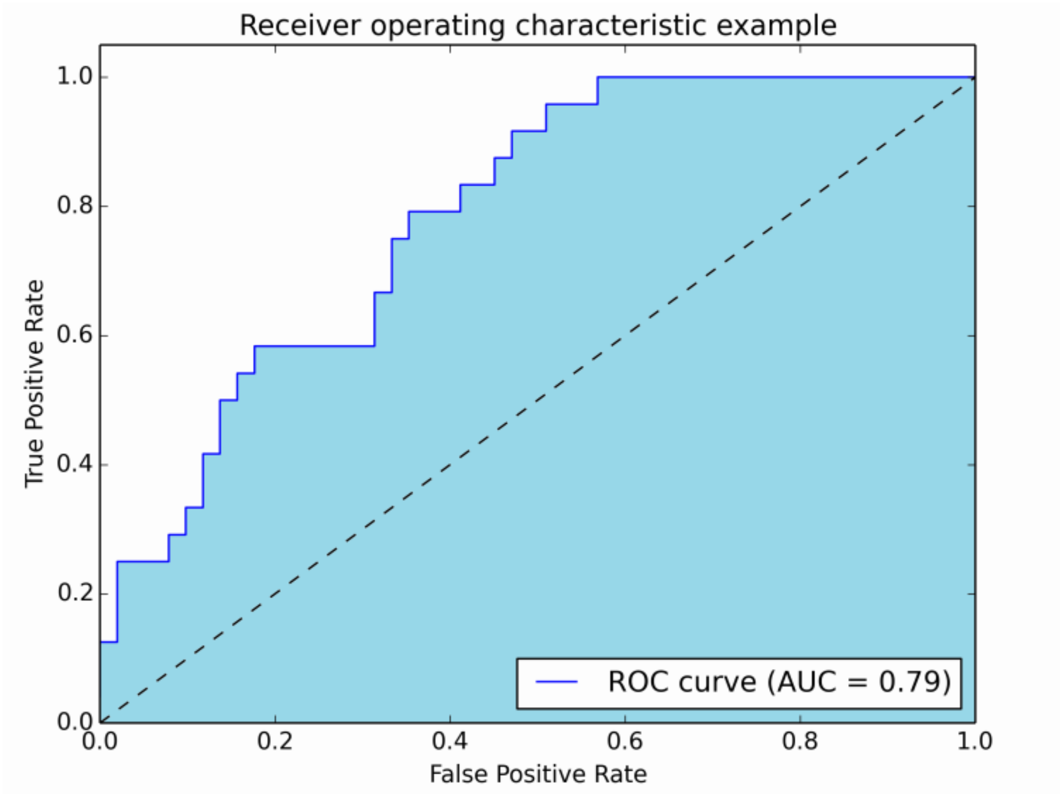
Determine the Sensitivity (true positive rate) and Specificity (true negative rate) for the indicated 2 cut-offs.

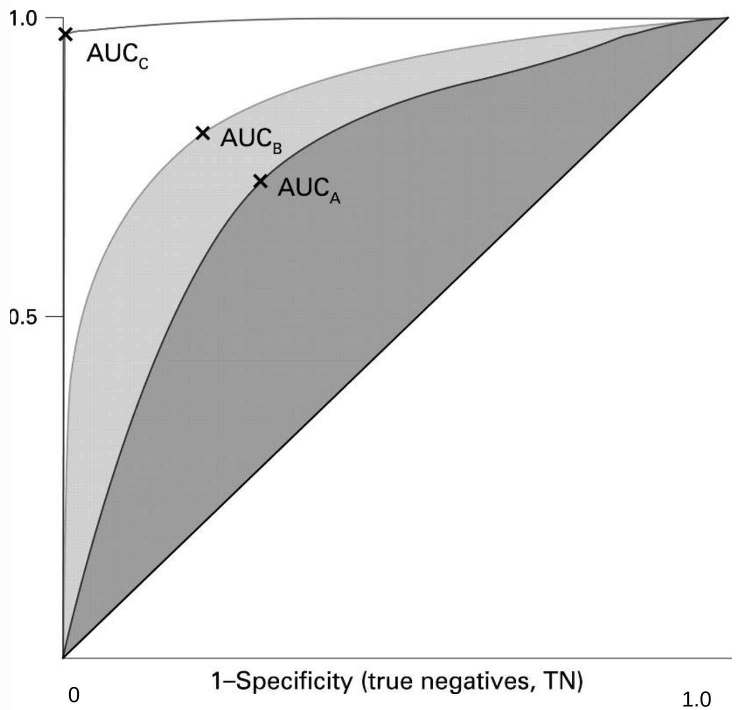


Do inn-class exercise

BMJ 1994; 309:188

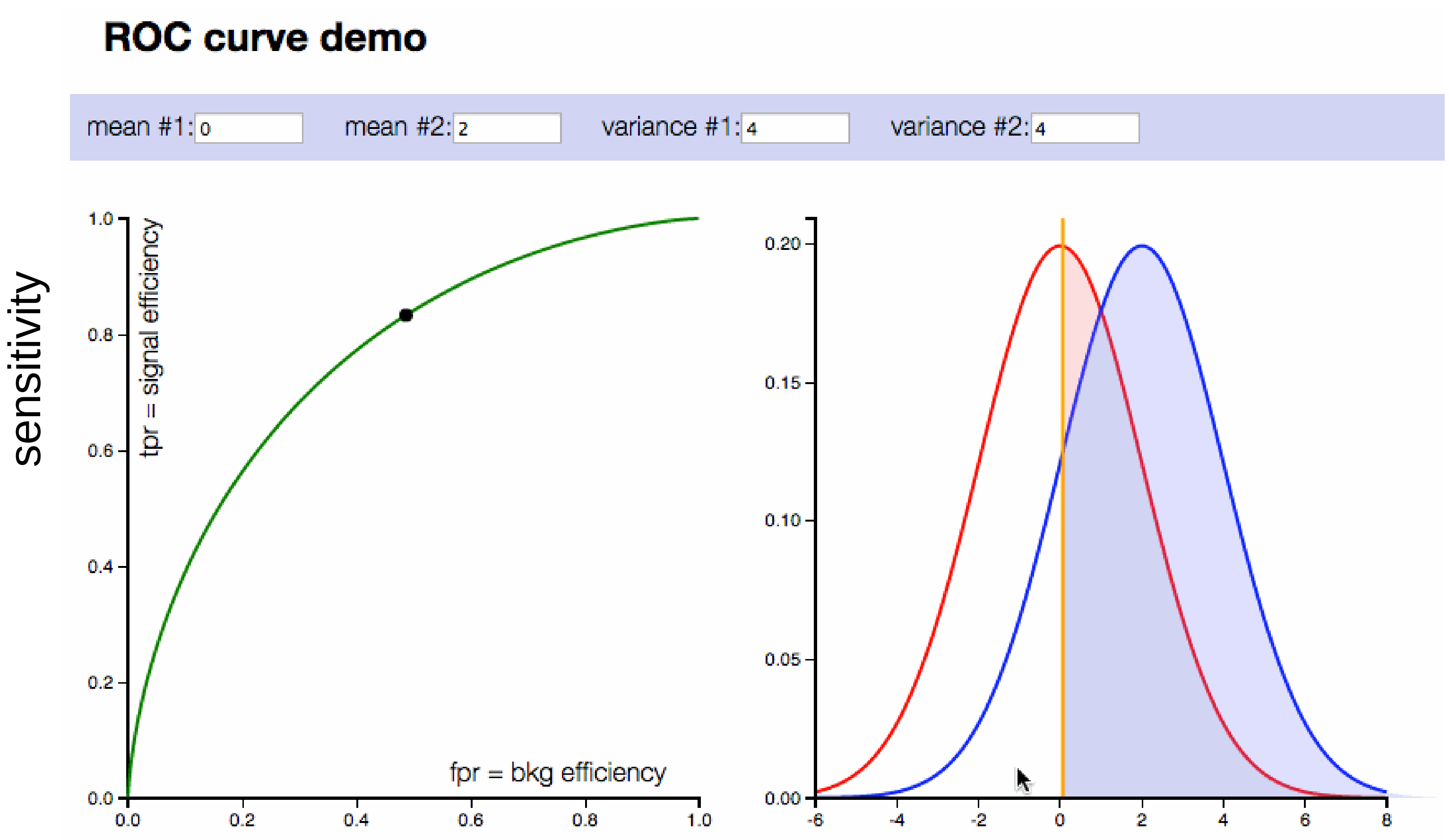
Use the ROC curve as performance measure by quantifying the area under the curve (AUC)





The larger the AUC the better is the performance of the diagnostic test.
A useless test has an AUC = 0.5.
A perfect test has an AUC= 1.

Nice online demos

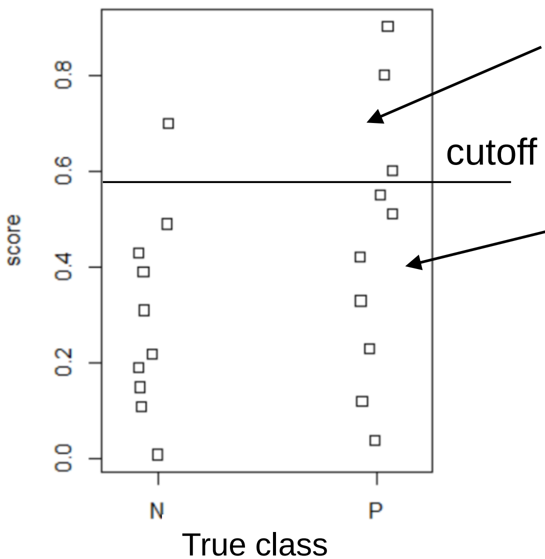


http://arogozhnikov.github.io/2015/10/05/roc-curve.html

Check out: http://www.navan.name/roc/

http://mlwiki.org/index.php/ROC_Analysis

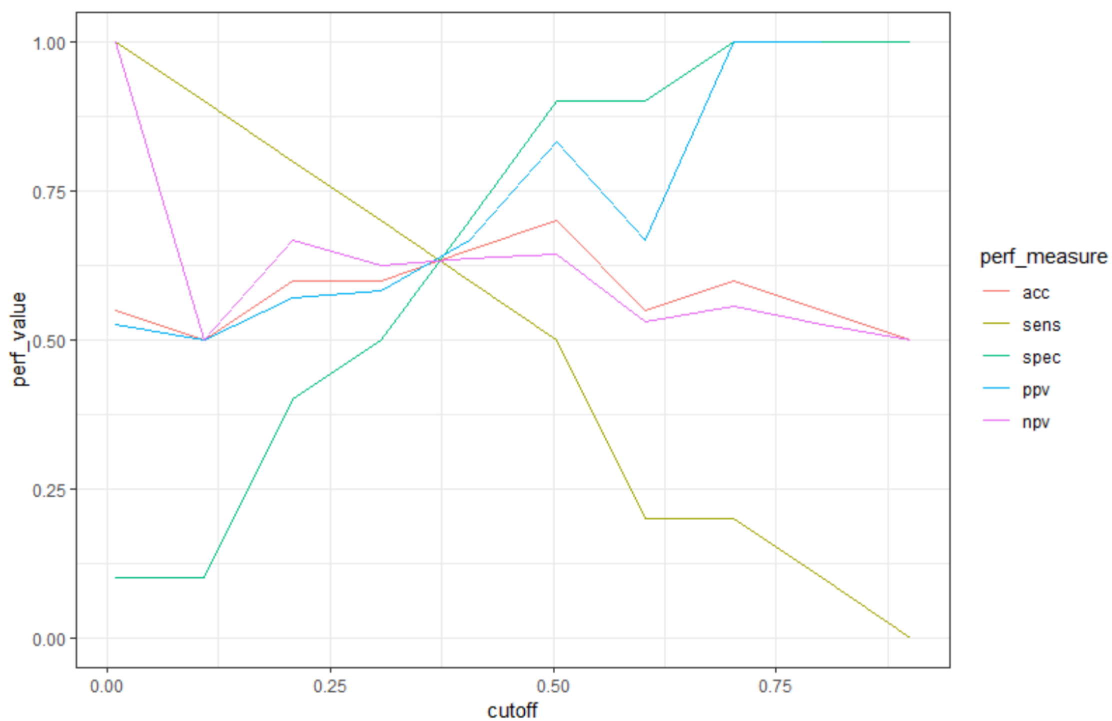
Let's move the cutoff in scoring classifier and determine performance of resulting classification rule



All instances with a score above the cutoff get prediction 1 (Pos).

All instances with a score below the cutoff get prediction 0 (Neg).

Predicted class		True class: Positive	True class: Negative	
	Positive	TP	FP	$PPV = \frac{TP}{TP+FP}$
	Negative	FN	TN	$NPV = \frac{TN}{TN+FN}$
		$sens = \frac{TP}{TP+FN}$	$spec = \frac{TN}{FP+TN}$	



How reliable is the result of a Aids-Test?

Ozzy Osbourne 'was told he could be HIV positive by doctors'

Rocker Ozzy Osbourne has revealed he was once told by doctors he could be HIV positive before a second test for the disease came back negative.

Ozzy Osbourne 'was told by doctors he could be HIV positive' Photo: AP

Prevalence, Sensitivity and Specificity

The probability that a randomly selected person has AIDS in Switzerland:

0.004

This is the prevalence of AIDS in Switzerland

Sensitivity of the ELISA-Test to detect a HIV+ blood sample:

0.999

Specificity of the ELISA-Test to identivy a HIV- blood sample correctly: :

0.997

-> in-class exercise with topic screening with the Aids test:

HIV$^+$/AIDS proportions in different countries

Rank	Land	HIV/AIDS Rate der Erwachsenen (%)
1	Swasiland	38.8
2	Botsuana	37.3
3	Lesotho	28.9
4	Simbabwe	24.6
5	Südafrika	21.5
6	Namibia	21.3
7	Sambia	16.5
8	Malawi	14.2
9	Zentralafrikanische Republik	13.5
10	Mosambik	12.2
11	Guinea-Bissau	10
12	Tansania	8.8
13	Gabun	8.1
14	Sierra Leone	7
15	Côte d'Ivoire	7
16	Kamerun	6.9
17	Kenia	6.7
18	Burundi	6
19	Liberia	5.9
20	Haiti	5.6
21	Nigeria	5.4
22	Ruanda	5.1
23	Kongo	4.9
24	Tschad	4.8
25	Äthiopien	4.4
26	Demokratische Republik Kongo	4.2
27	Burkina Faso	4.2
28	Uganda	4.1
29	Togo	4.1
30	Angola	3.9
31	Äquatorialguinea	3.4
32	Trinidad und Tobago	3.2
33	Guinea	3.2
34	Ghana	3.1
35	Bahamas	3
43	Benin	1.9
44	Honduras	1.8
45	Dominikanische Republik	1.7
46	Madagaskar	1.7
47	Suriname	1.7
48	Thailand	1.5
49	Barbados	1.5
50	Ukraine	1.4
51	Myanmar	1.2
52	Gambia	1.2
53	Niger	1.2
54	Jamaika	1.2
55	Russische Föderation	1.1
56	Guatemala	1.1
57	Estland	1.1
58	Somalia	1
59	Panama	0.9
60	Indien	0.9
61	Senegal	0.8
62	Spanien	0.7
63	El Salvador	0.7
64	Brasilien	0.7
65	Kolumbien	0.7
66	Argentinien	0.7
67	Venezuela	0.7
68	Vereinigte Staaten	0.6
69	Costa Rica	0.6
70	Papua-Neuguinea	0.6
71	Lettland	0.6
72	Mauretanien	0.6
73	Italien	0.5
74	Nepal	0.5
75	Paraguay	0.5
76	Peru	0.5
77	Malaysia	0.4
78	Portugal	0.4
79	Schweiz	0.4

Numbers from 2008

Confusion Matrix

From the tree diagram given in the in-class exercise we can read of the content of the corresponding confusion matrix.

	T +	T -	Summe
HIV +	30‘769	31	30‘800
HIV -	23‘008	7‘646‘192	7‘669‘200
sum	53‘777	7‘646‘223	7‘700‘000

Prevalence

$$P(HIV^{+}) = \frac{30800}{7700000} = 0.004$$

Sensitivity

$$P(T+ \mid HIV^{+}) = \frac{30769}{30800} = 0.999$$

Specificity

$$P(T- \mid HIV^{-}) = \frac{7646192}{7669200} = 0.997$$

Definition of the conditional probability

The conditional probability of an event (e.g. A or D+) given that some other event (e.g. B or T+) has already occurred is written as P(A|B) and defined as the quotient of the probability of the joint of events A and B, and the probability of B. Der vertical dash means „given that“ or „under the condition“ B has already occurred.

A and B are two events and P(B) ≠ 0. The <u>conditional probability of A given B</u> is defined as:

$$P(A|B) := \frac{P(A \cap B)}{P(B)}$$

<u>Remark:</u> If A and B are ***independent*** , we get:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \cdot P(B)}{P(B)} = P(A)$$

Bayes's theorem
Inversion of a conditional probability

Bayes's theorem gives the rule how to invert a conditional probability,

and how to update the probability by using some additional information:

Bayes‘ s theorem:

$$P(B \mid A) = \frac{P(A \mid B) \cdot P(B)}{P(A)} = \frac{P(A \mid B)}{P(A)} \cdot P(B)$$

posteriori probability for B
or updated probability
or predictive value

info

a-priori probability for B
or prevalence of B

proof:

$$P(B \mid A) := \frac{P(A \cap B)}{P(A)} = \frac{\frac{P(A \cap B)}{P(B)} \cdot P(B)}{P(A)} = \frac{P(A \mid B) \cdot P(B)}{P(A)}$$

$$P(A \mid B) := \frac{P(A \cap B)}{P(B)}$$

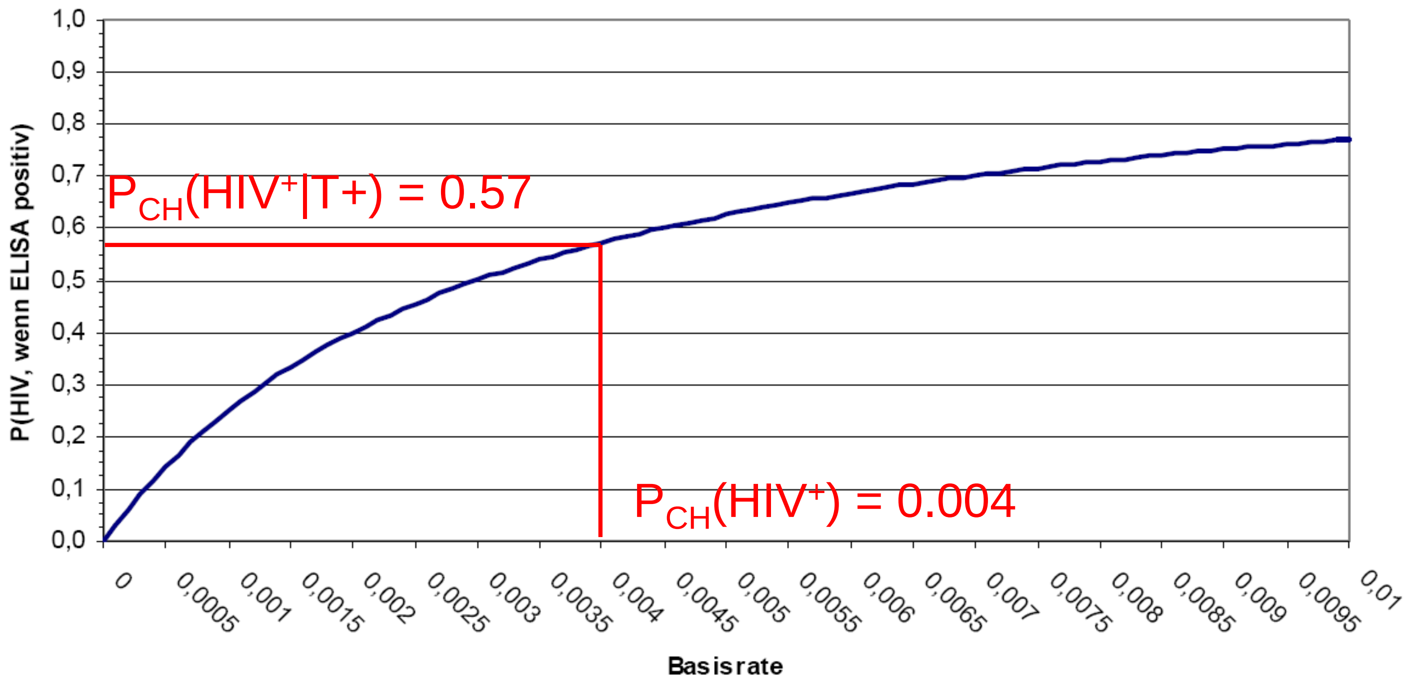
Wahrscheinlichkeiten neu bewerten:
A-priori- und a-posteriori-Wahrscheinlichkeit

Prevalence

A-priori probability

Test result

a-posteriori probability



Inversion of a conditional probability

In general: $P(\text{ T+ } | \text{ HIV}^{+} \text{ }) \neq P(\text{ HIV}^{+} \text{ } | \text{T+ })$

Often we know a conditional probability as e.g.:

Sensitivity: $$P(T+|HIV^{+}) = \frac{P(T+\bigcap HIV^{+})}{P(HIV^{+})}$$

Specificity: $$P(T-|HIV^{-}) = \frac{P(T-\bigcap HIV^{-})}{P(HIV^{-})}$$

But we are interested in the predictive value of the diagnostic test which are the inversed conditional probabilities:

postive predictive Value $$\text{PPV}=P(HIV+|T+) = \frac{P(T_p | HIV^{+}) \cdot P(HIV^{+})}{P(T_p)} = \frac{TP}{TP+FP}$$

negative predictive Value $$\text{NPV}=P(HIV-|T-) = \frac{P(T-|HIV-) \cdot P(HIV-)}{P(T-)} = \frac{TN}{TN+FN}$$

Review: Power and level of significance, sensitivity and specificity of a test

A worked example

The fecal occult blood (FOB) screen test was used in 2030 people to look for bowel cancer:

		Patients with bowel cancer (as confirmed on endoscopy)		
		Condition Positive	Condition Negative	
Fecal Occult Blood Screen Test Outcome	Test Outcome Positive	**True Positive** (TP) = 20	**False Positive** (FP) = 180	Positive predictive value = TP / (TP + FP) = 20 / (20 + 180) = **10%**
	Test Outcome Negative	**False Negative** (FN) = 10	**True Negative** (TN) = 1820	Negative predictive value = TN / (FN + TN) = 1820 / (10 + 1820) ≈ **99.5%**
		Sensitivity = TP / (TP + FN) = 20 / (20 + 10) ≈ **67%**	**Specificity** = TN / (FP + TN) = 1820 / (180 + 1820) = **91%**	

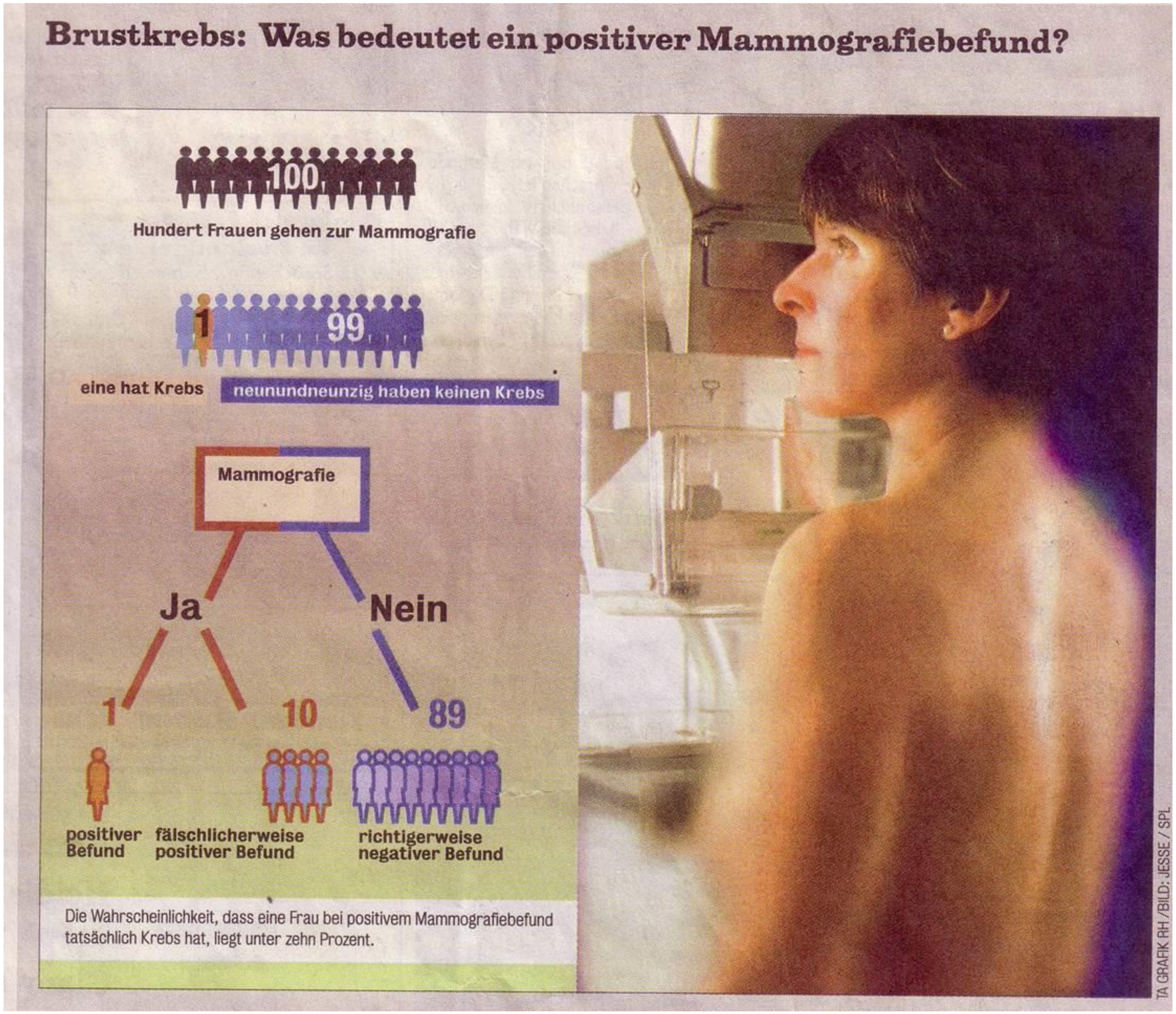
Characterizes quality of test

Posteriori probability
Depends on population e.g. the prevalence

fecal occult blood test (FOBT) checks for hidden (occult) blood in the stool (feces, excrements)

source: http://en.wikipedia.org/wiki/Positive_predictive_value

«Tagesanzeiger» explains
a-priori und a-posteriori probabilities



How to interpret a Mammography result

We can use the Bayes's theorem to determine the PPV and NPV of a Mammography result dependent on the prevalence.

prevalence	sensitivity	specificity	PPV	NPV
1.0%	86.6%	96.8%	21.5%	99.9%
4.5%	86.6%	96.8%	56.4%	99.3%
10.0%	86.6%	96.8%	75.1%	98.5%
50.0%	86.6%	96.8%	96.4%	87.9%

The breast cancer prevalence among British women aged 59 is 4.5%. (http://www.cancerresearchuk.org/cancer-info/)
The negative predictive value (NPV) is with 99.3% much higher than the PPV of 56%

In the "One Million Women Study" (Banks et al. 2004) 122'355 50- 64 year old women who had a Mammography were followed for one year and the histological confirmed breast cander incidences were determined.

28

The Fagan-Nomogram allows to graphically determine the posteriori probability

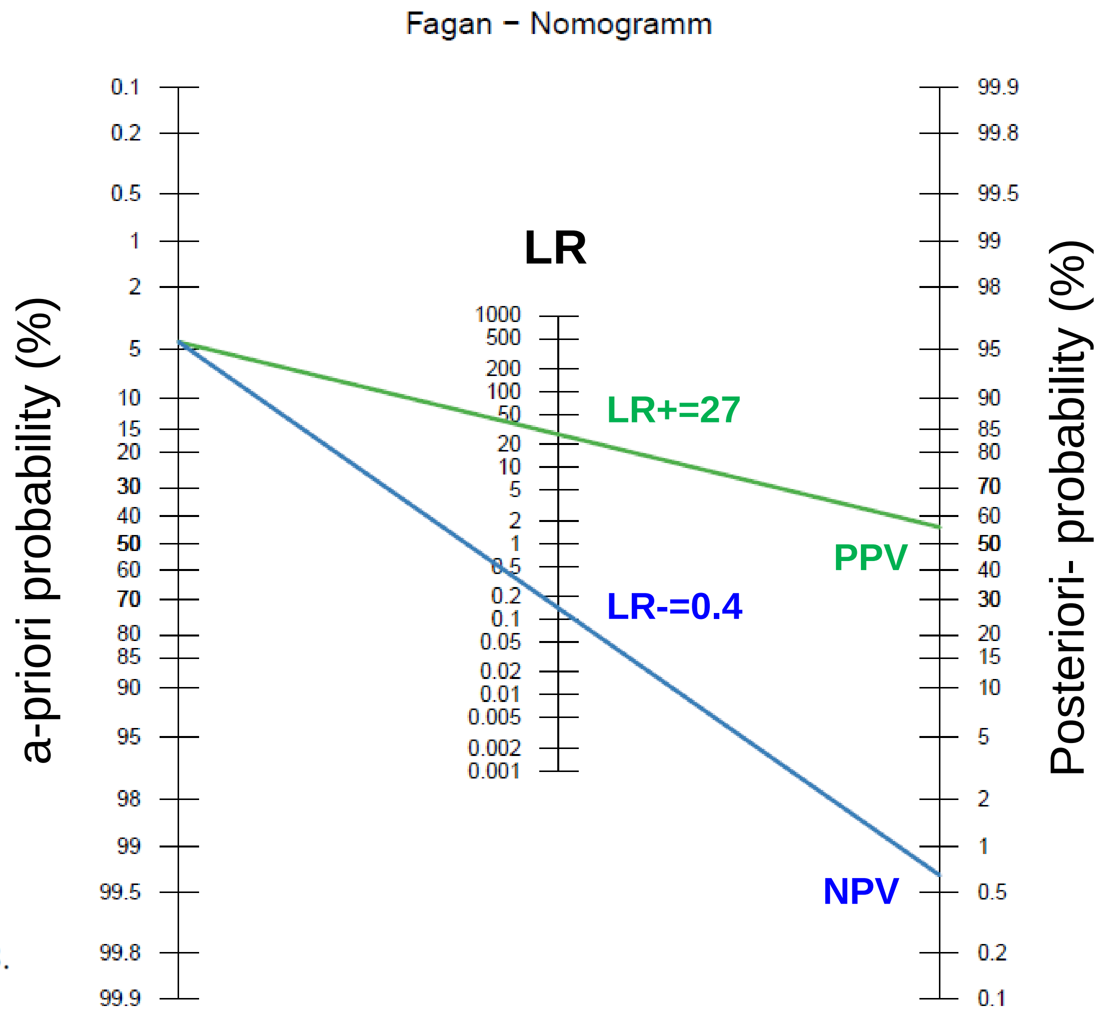
We can use likelihood ratios LR+ and LR- to get graphically from the prevalence to the predictive value of a test:

$$LR+ = \frac{\text{sensitivity}}{1-\text{specificity}}$$

$$LR- = \frac{1-\text{sensitivity}}{\text{specificity}}$$

Mammography:
sensitivity=86.6%
specificity=96.8%

$$\mathrm{LR}^{+} = \frac{86.6\%}{3.2\%} \approx 27.1 \qquad \mathrm{LR}^{-} = \frac{13.4\%}{96.8\%} \approx 0.138.$$

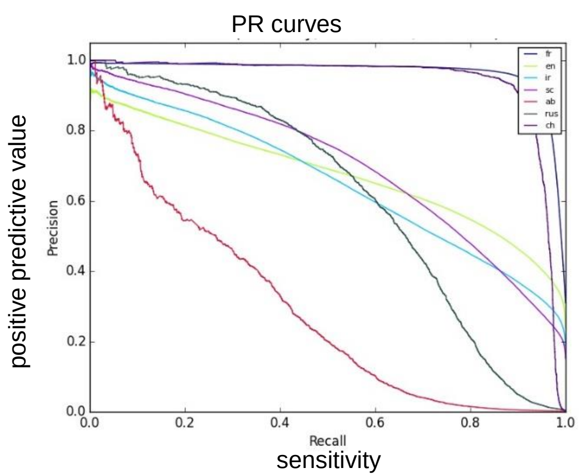


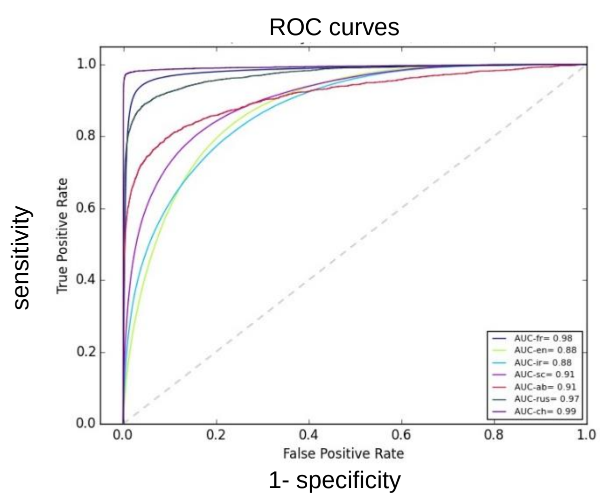
Published in NEJM, July 1975

29

Summary as extended confusion table & ROC and PR curves

		predicted condition			
	total population	prediction positive	prediction negative	Prevalence $= \frac{\Sigma \text{ condition positive}}{\Sigma \text{ total population}}$	
true condition	condition positive	**True Positive (TP)**	**False Negative (FN)** (type II error)	True Positive Rate (TPR), Sensitivity, Recall, Probability of Detection $= \frac{\Sigma \text{ TP}}{\Sigma \text{ condition positive}}$	False Negative Rate (FNR), Miss Rate $= \frac{\Sigma \text{ FN}}{\Sigma \text{ condition positive}}$
	condition negative	**False Positive (FP)** (Type I error)	**True Negative (TN)**	False Positive Rate (FPR), Fall-out, Probability of False Alarm $= \frac{\Sigma \text{ FP}}{\Sigma \text{ condition negative}}$	True Negative Rate (TNR), Specificity (SPC) $= \frac{\Sigma \text{ TN}}{\Sigma \text{ condition negative}}$
	Accuracy $= \frac{\Sigma \text{ TP} + \Sigma \text{ TN}}{\Sigma \text{ total population}}$	Positive Predictive Value (PPV), Precision $= \frac{\Sigma \text{ TP}}{\Sigma \text{ prediction positive}}$	False Omission Rate (FOR) $= \frac{\Sigma \text{ FN}}{\Sigma \text{ prediction negative}}$	Positive Likelihood Ratio (LR+) $= \frac{\text{TPR}}{\text{FPR}}$	Diagnostic Odds Ratio (DOR) $= \frac{\text{LR+}}{\text{LR−}}$
		False Discovery Rate (FDR) $= \frac{\Sigma \text{ FP}}{\Sigma \text{ prediction positive}}$	Negative Predictive Value (NPV) $= \frac{\Sigma \text{ TN}}{\Sigma \text{ prediction negative}}$	Negative Likelihood Ratio (LR−) $= \frac{\text{FNR}}{\text{TNR}}$	





RemarK: Unlike the ROC curve, PR curves are very sensitive to imbalance. A classifier that is optimized for good AUC, might yield poor precision-recall results on an unbalanced data.

Summary

- We need a (new) test set with known true binary outcome to evaluate the performance of a diagnostic test (or classifier)

- A binary diagnostic test (classifier) can be evaluated based on the
  - confusion matrix (determined in real world conditions) that allows to compute
    - test specific performance measures that do not depend on the disease prevalence
      - sensitivity: Probability that the test classifies a positive case as positive
      - specificity: Probability that the test classifies a negative case as negative
      - accuracy: overall classification rate
    - predictive performance measures that depend on the disease prevalence
      - positive predictive value: probability that a positive tested subject is sick
      - negative predictive value: probability that a negative tested subject is healthy

- A diagnostic scoring test with continuous score as outcome can be evaluated by using different score-cutoffs to define positive and negative predictions
  - by moving the cutoff we can determine a
    - ROC curve (sensitivity vs 1-specificity) and use the AUC (area under the curve) as performance measure
    - PR curve (Precision=positive-predictive-value vs Recall=sensitivity) and its AUC