Prof **François Bochud**

Institute of Radiation physics (IRA)

UNIL - CHUV

Master of Science EPF-ETH degree in **Nuclear Engineering**
Medical Radiation Physics

Receiver operating characteristics (ROC)

Learning objectives

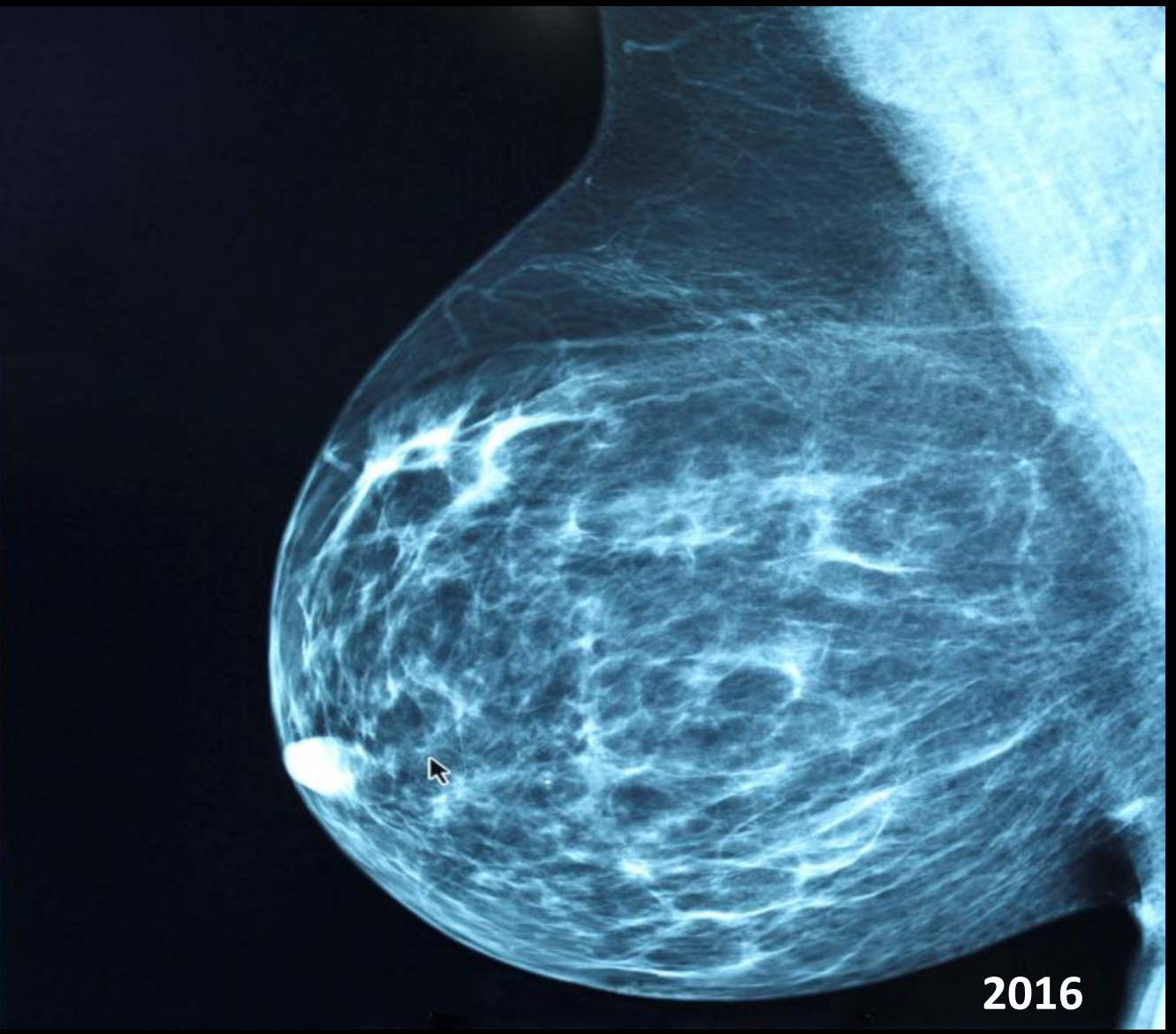
- Explain the **meaning of a ROC curve**
- Calculate a posteriori probabilities with **Bayes' theorem** and other **simple methods**
- Know the basic principles of a **detection experiment**
- Be able **to build a ROC curve** from a detection experiment
- **Communicate** efficiently the meaning of a ROC analysis

ROC theory

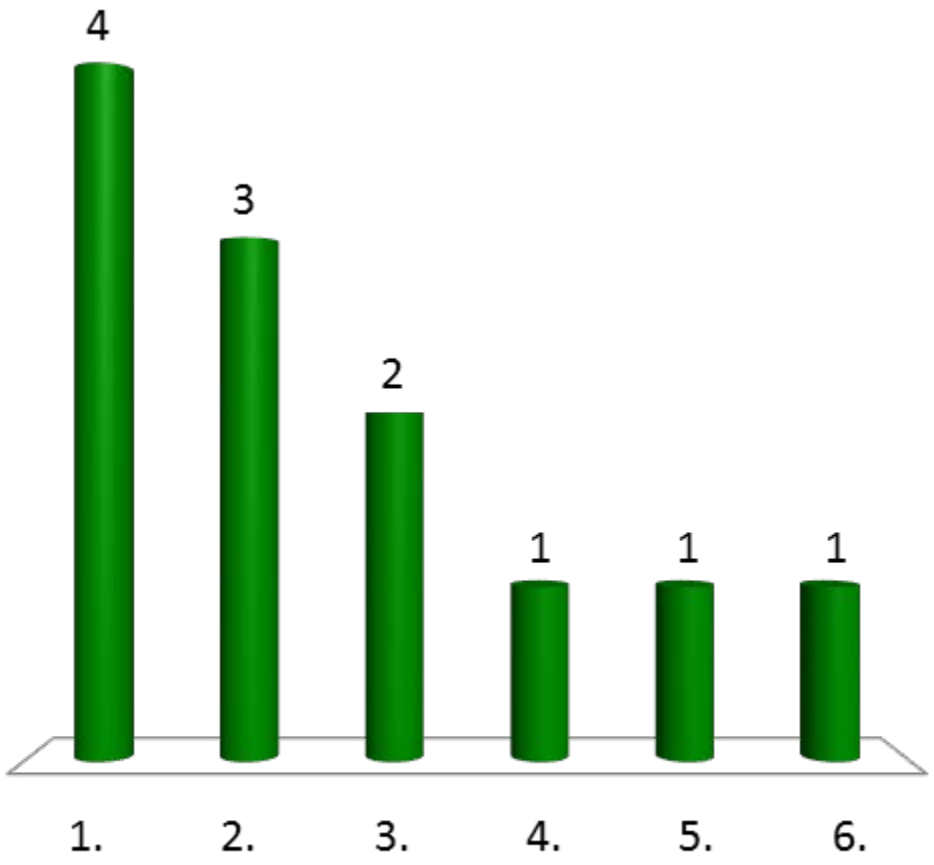
1.

Quantification of image quality

Which **image** has the **best quality**?

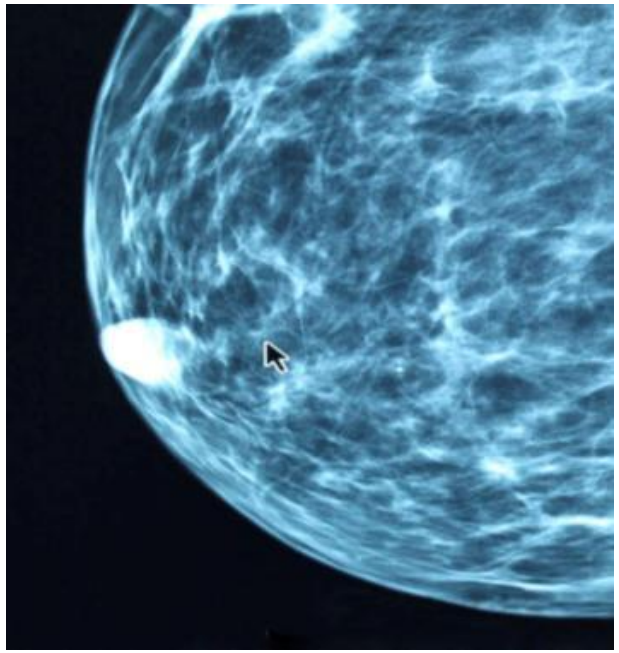


How should we measure **image quality**?



6 sur 6

1. **Technical**
   Physical measurements *(MTF, SNR, etc.)*
2. **Diagnostic accuracy**
   Agreement between diagnoses and truth
3. **Diagnostic thinking**
   Physician certainty about the diagnostic
4. **Therapeutic**
   Treatment of management of the patient
5. **Outcome**
   Result of treatment
6. **Societal**
   Value to society



Non-toxicity

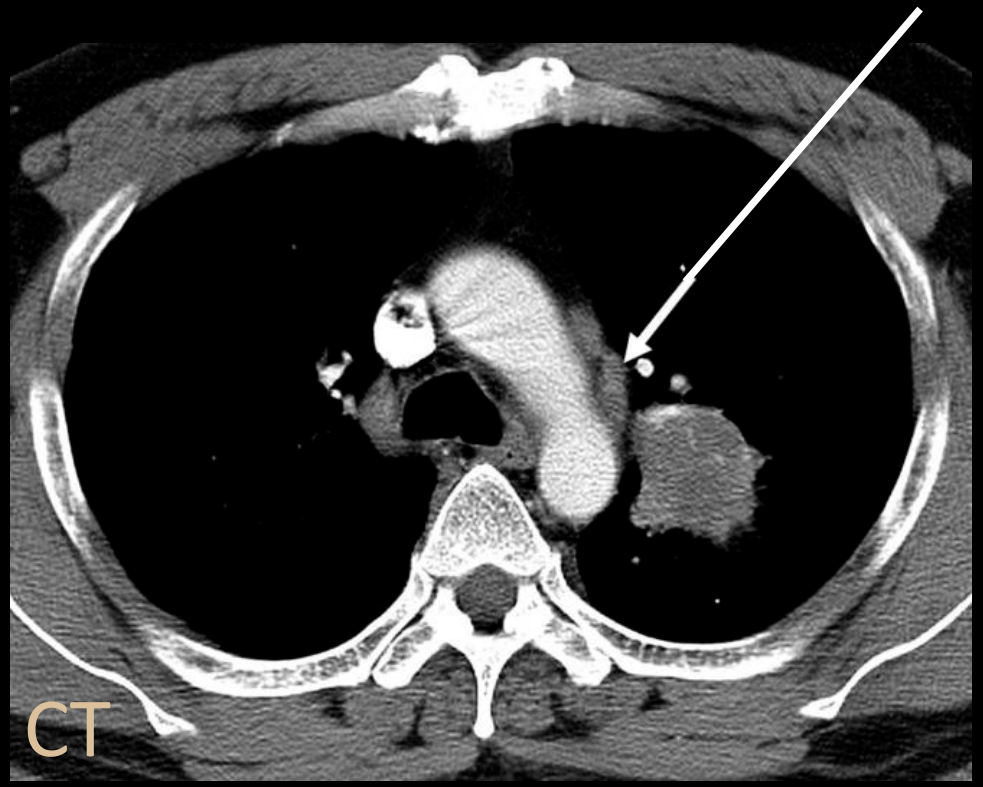
Efficiency

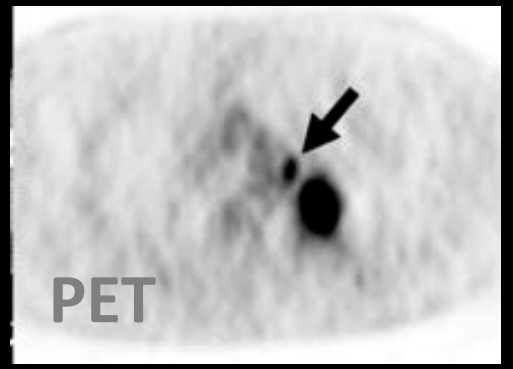
1. **Technical**
   Physical measurements *(MTF, SNR, etc.)*
2. **Diagnostic accuracy**
   Agreement between diagnoses and truth
3. **Diagnostic thinking**
   Physician certainty about the diagnostic
4. **Therapeutic**
   Treatment of management of the patient
5. **Outcome**
   Result of treatment
6. **Societal**
   Value to society

ICRU Report 79, Receiver Operating Characteristic Analysis in Medical Imaging (2008)

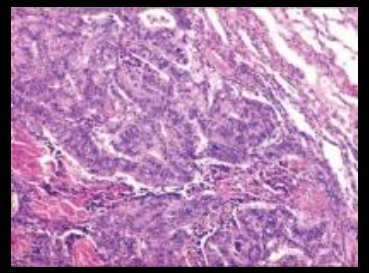
Can you
detect the bighorn?

TRUE **Positive** mediastinal lymph node metastasis





Additional PET/CT image **positive** again



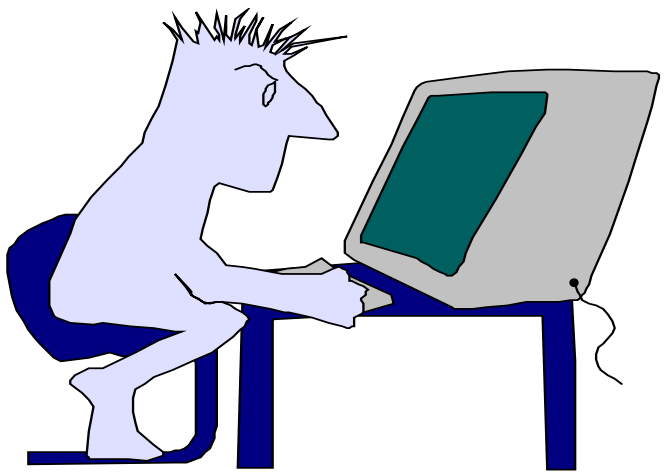
Pathology **positive** again

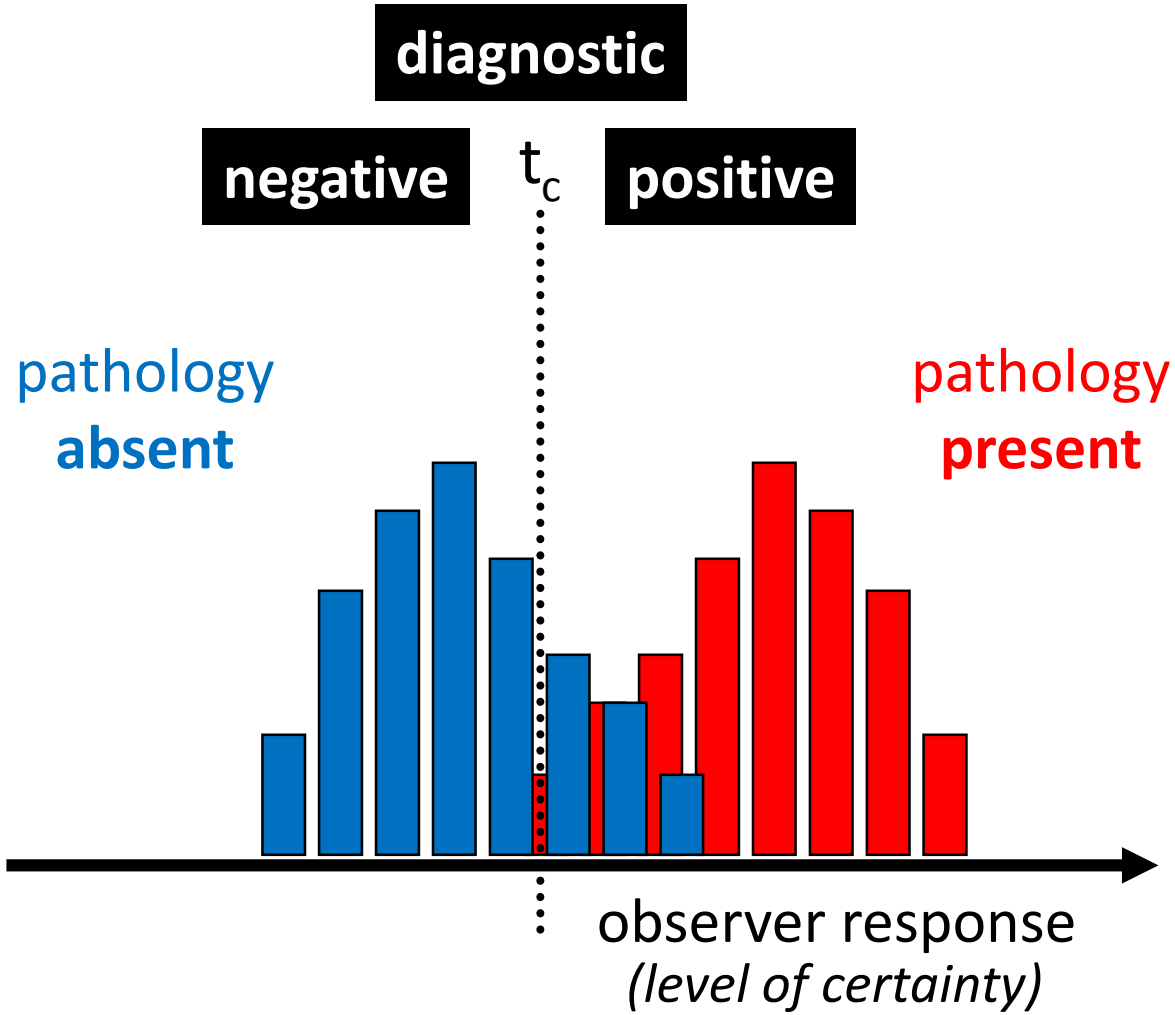
Kim et al, doi 10.1002/cncr.22518 (2008)

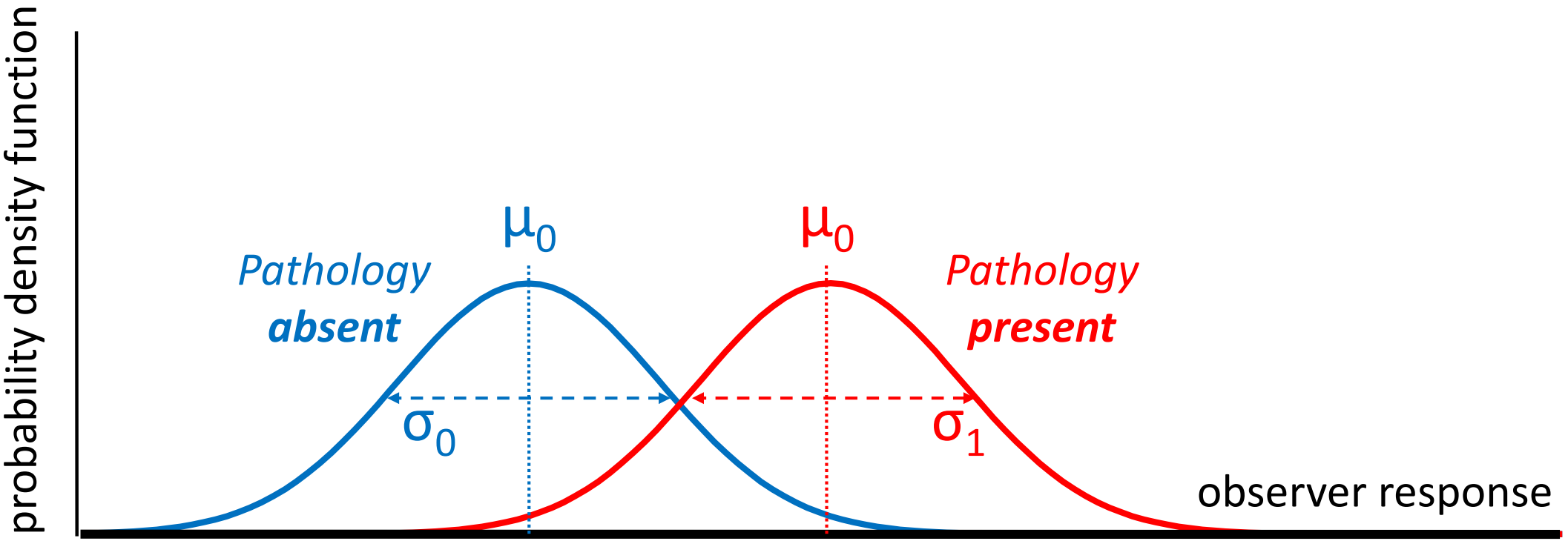
What would you propose to characterize the **quality of a detection**?

And if you were a **patient**?

		Diagnostic	
		Negative	**Positive**
Truth	pathology **absent**	**TN** true negative	**FP** false positive
	pathology **present**	**FN** false negative	**TP** true positive

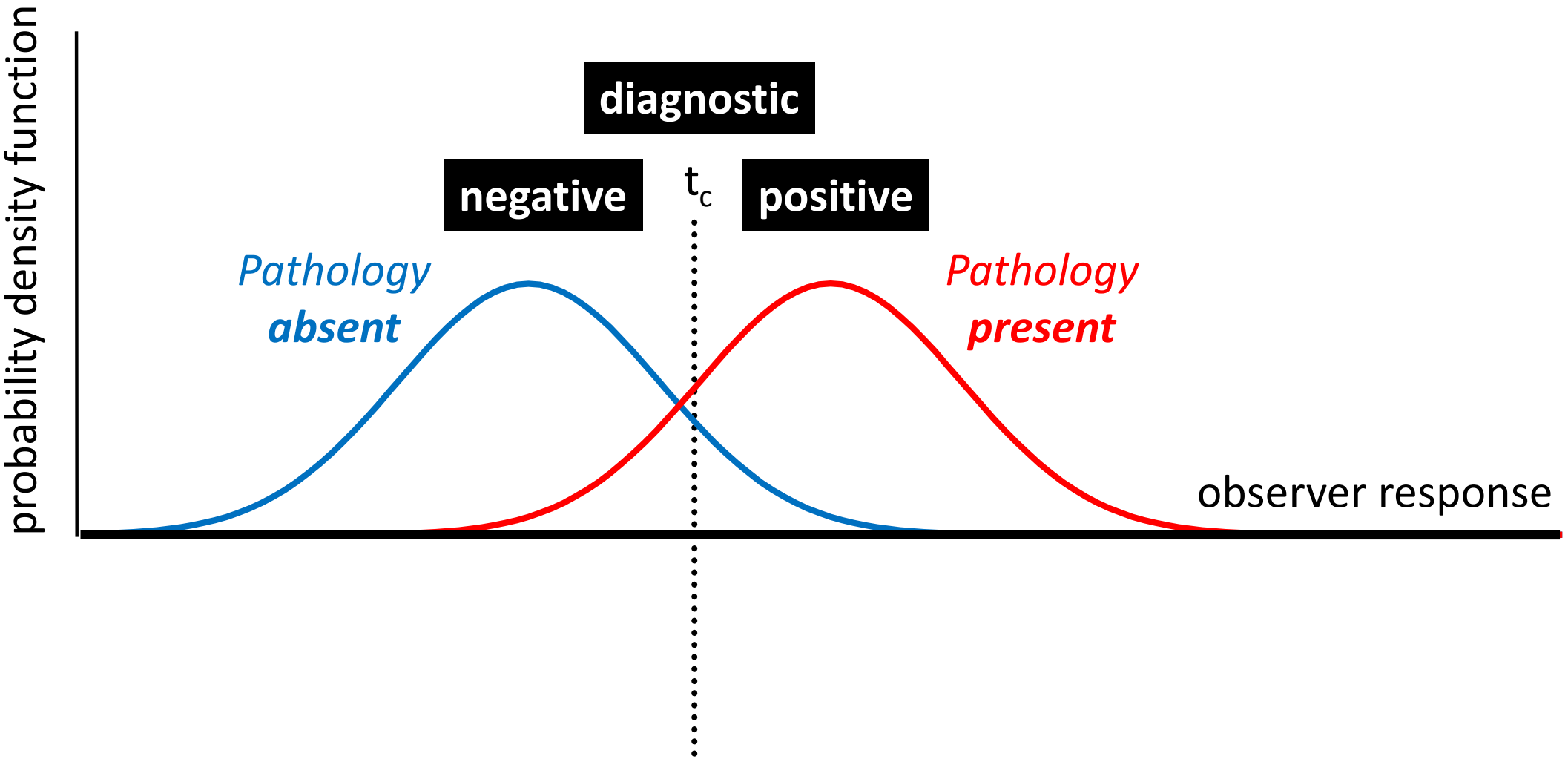


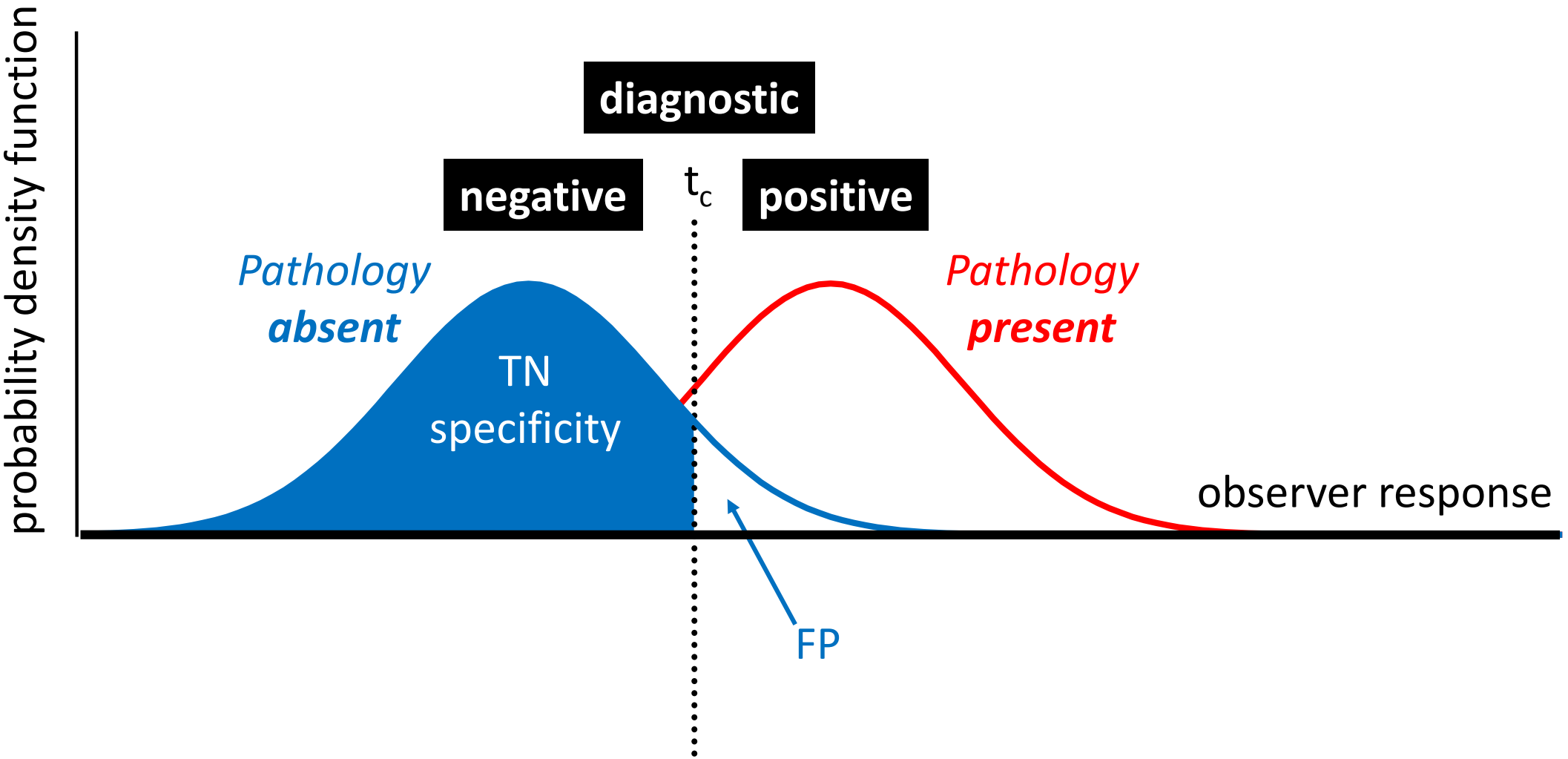

diagnostic
negative
t_c
positive
pathology
absent
pathology
present
observer response
(level of certainty)

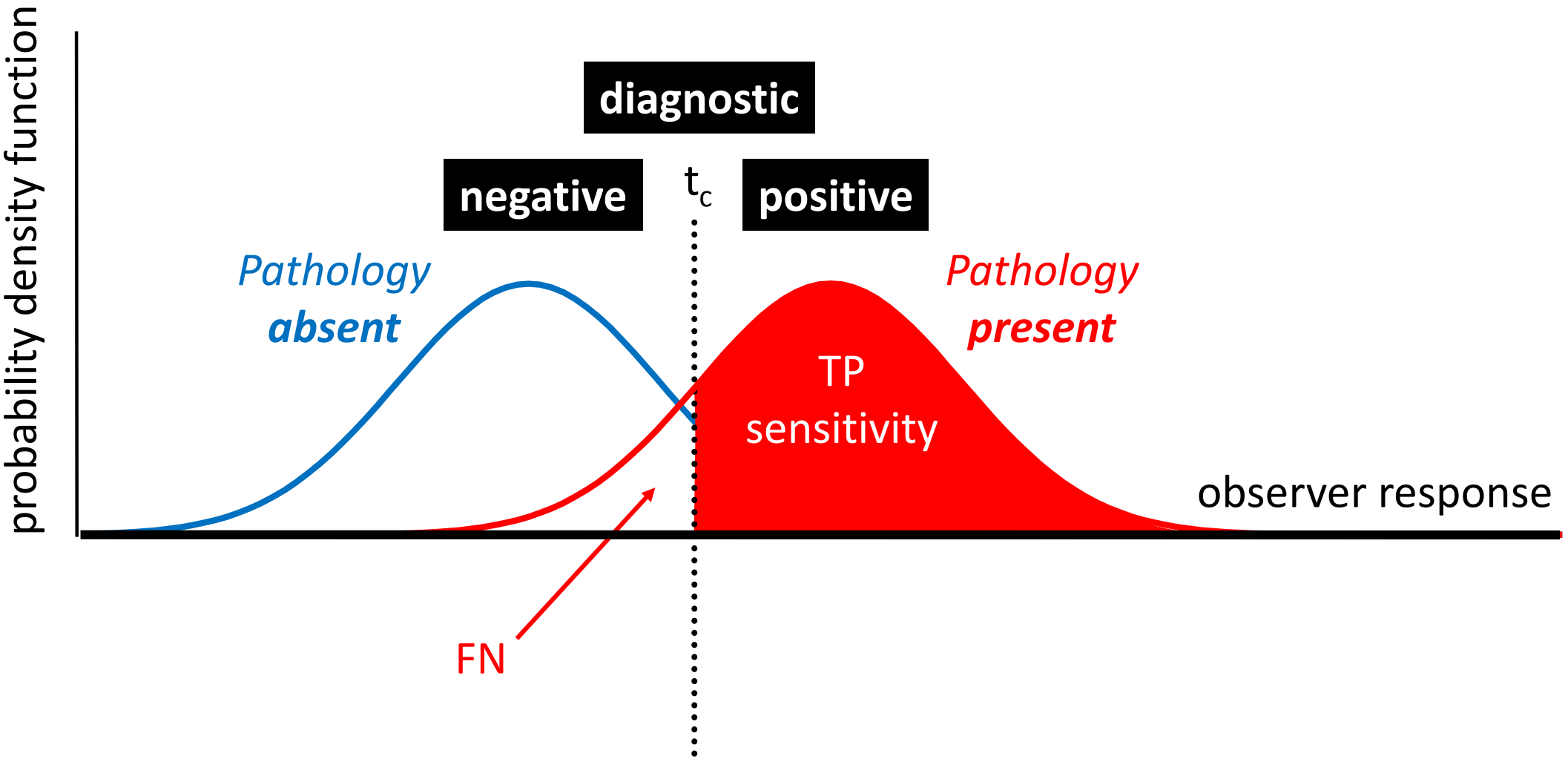


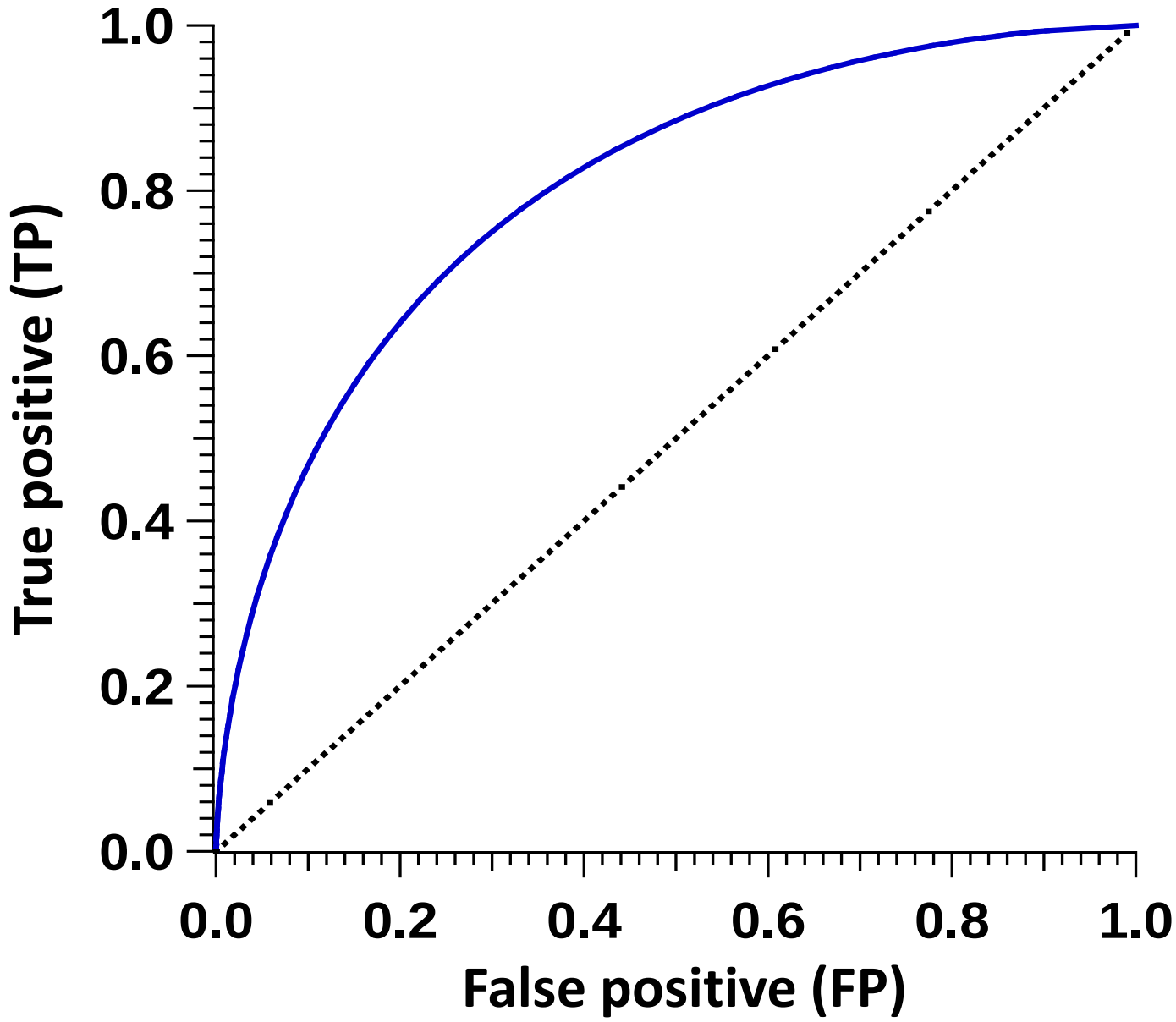
$$\mathrm{SNR} = \frac{\left|\mu_1 - \mu_0\right|}{\sqrt{\frac{1}{2}\left(\sigma_1^2 + \sigma_0^2\right)}}$$

this **signal-to-noise ratio**
is also known as
"**detectability index**"


probability density function
diagnostic
negative
t_c
positive
Pathology
absent
Pathology
present
observer response


probability density function
diagnostic
negative
t_c
positive
Pathology
absent
Pathology
present
TN
specificity
observer response
FP


probability density function
diagnostic
negative
t_c
positive
Pathology
absent
Pathology
present
TP
sensitivity
observer response
FN



Receiver

Operating

Characteristic

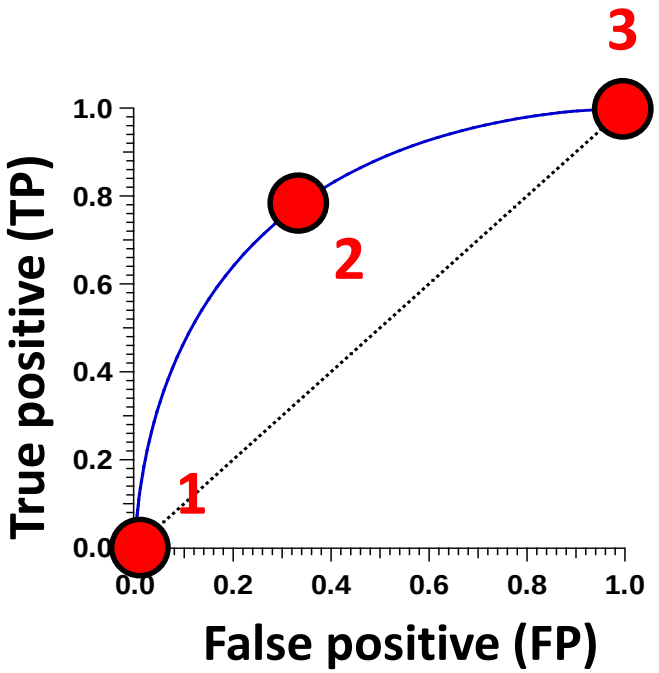
ROC-curve

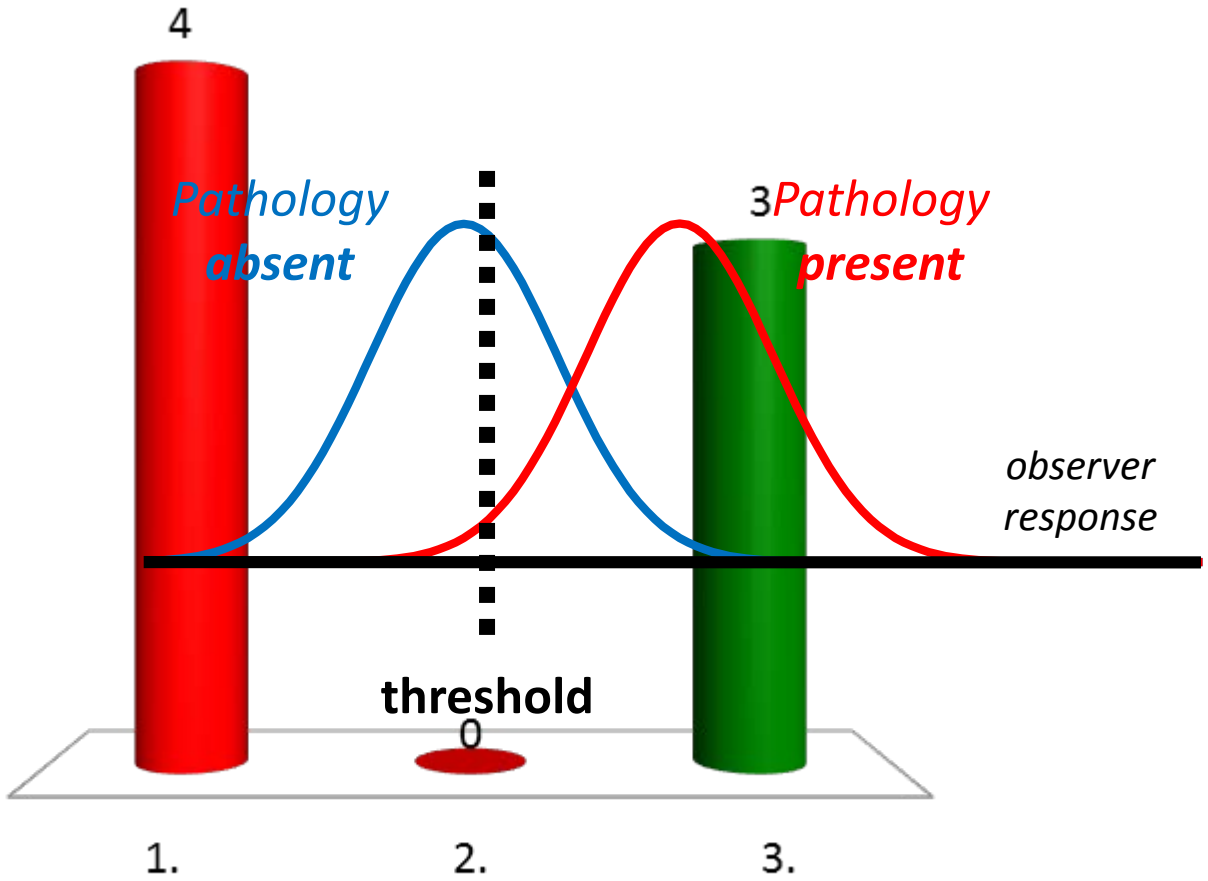
Which point of the ROC-curve corresponds to threshold = -∞?

1.

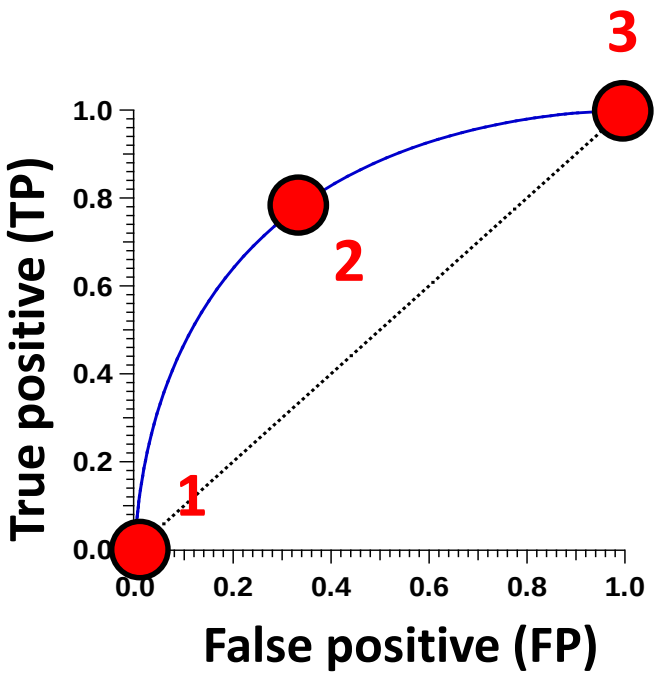
2.

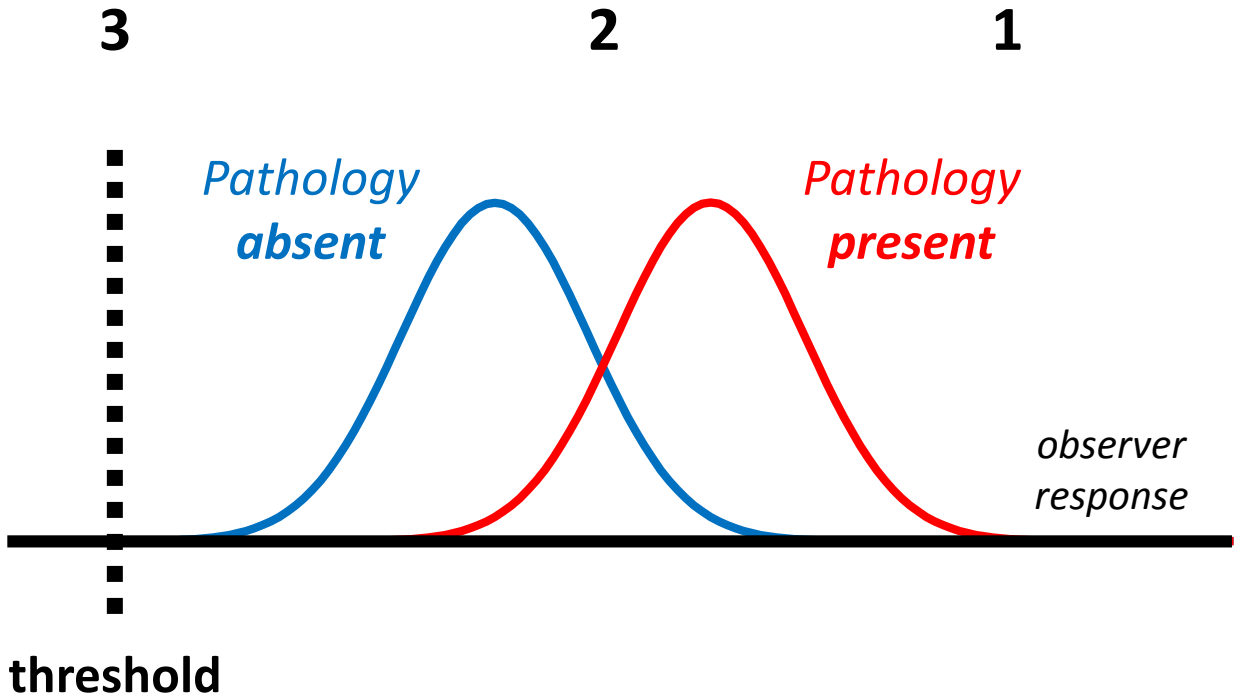
3.



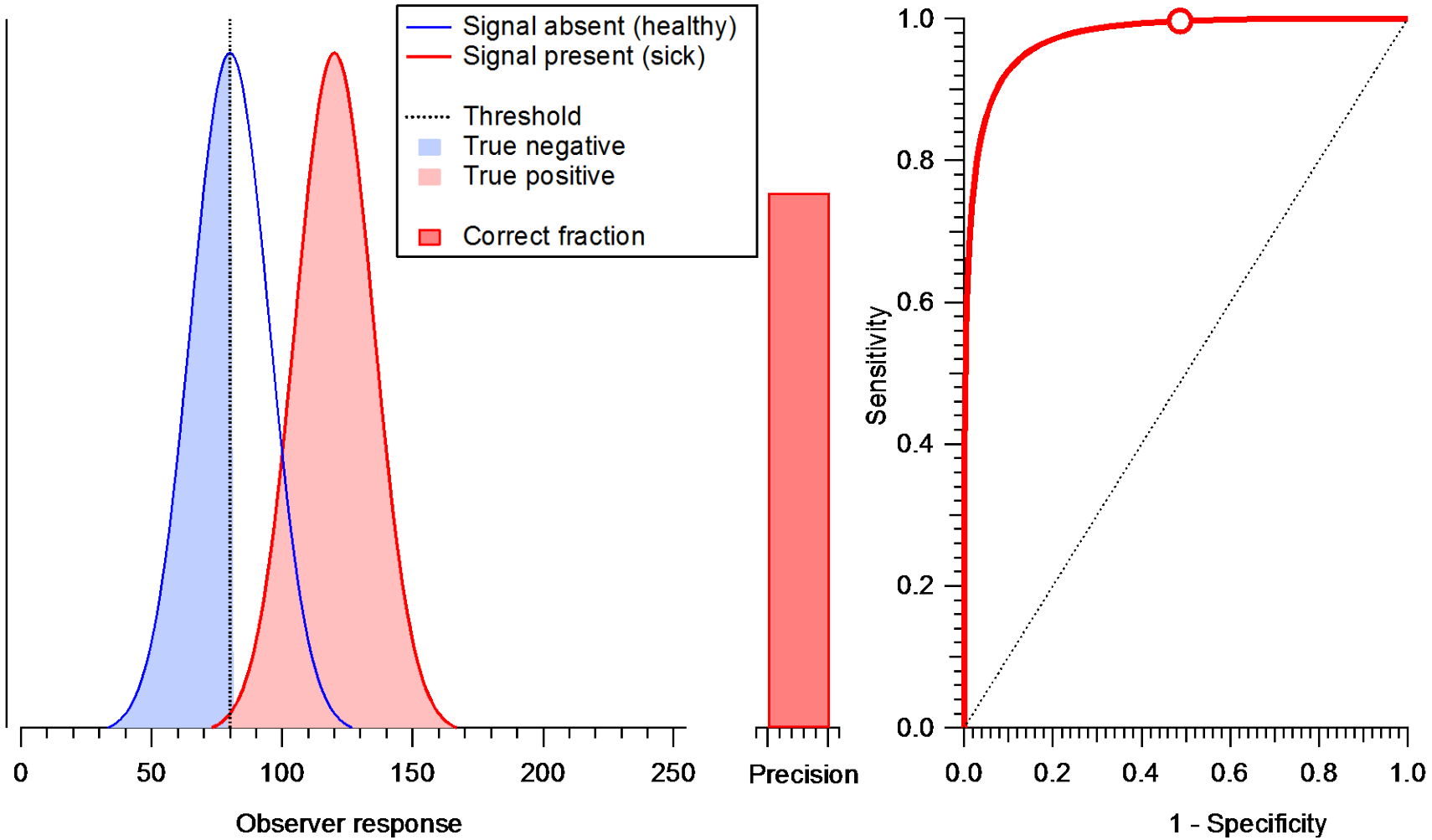


7 sur 7

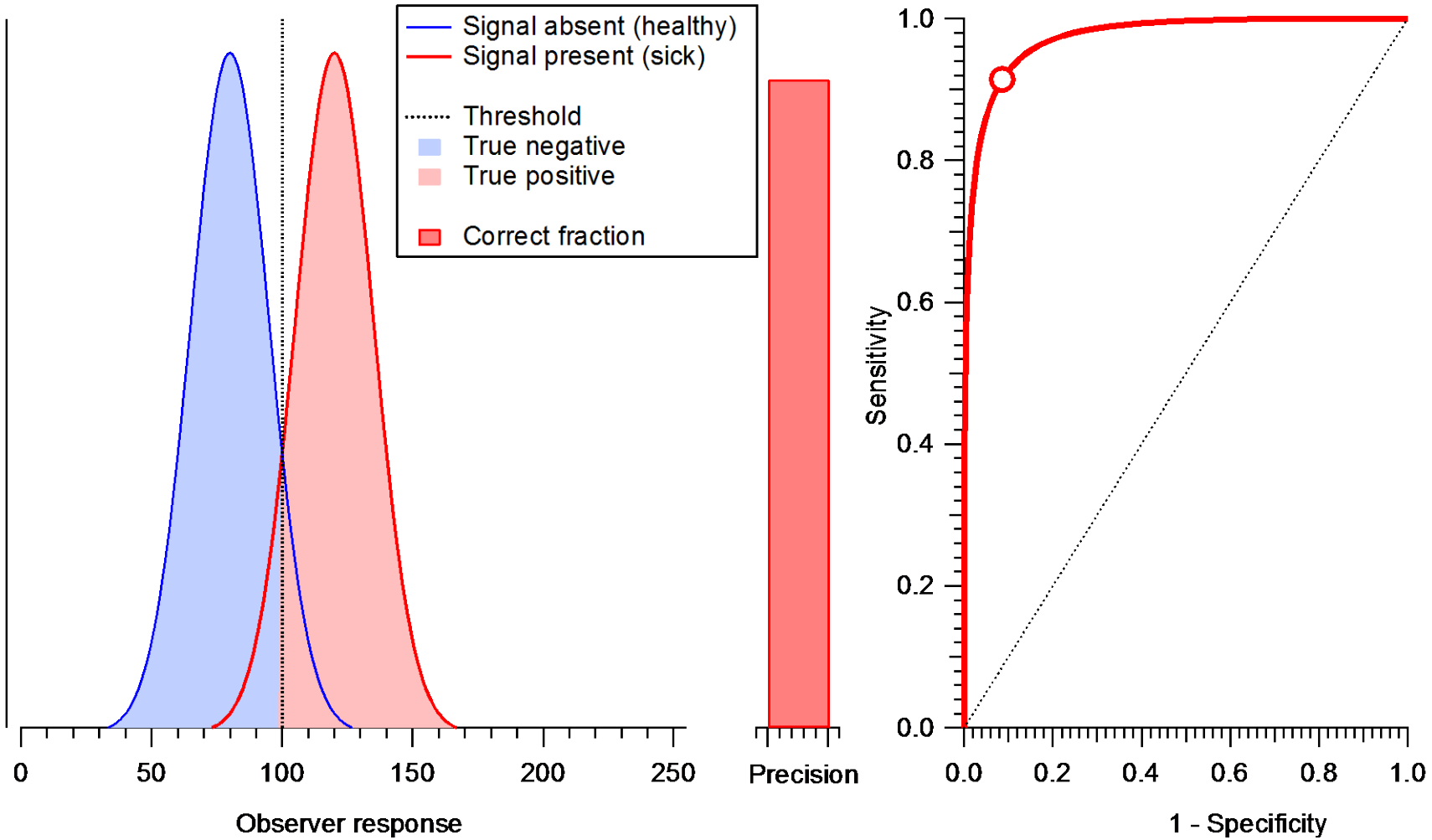

3
2
1
1.0
0.8
0.6
0.4
0.2
0.0
True positive (TP)
0.0
0.2
0.4
0.6
0.8
1.0
False positive (FP)


3
2
1
Pathology
absent
Pathology
present
observer
response
threshold

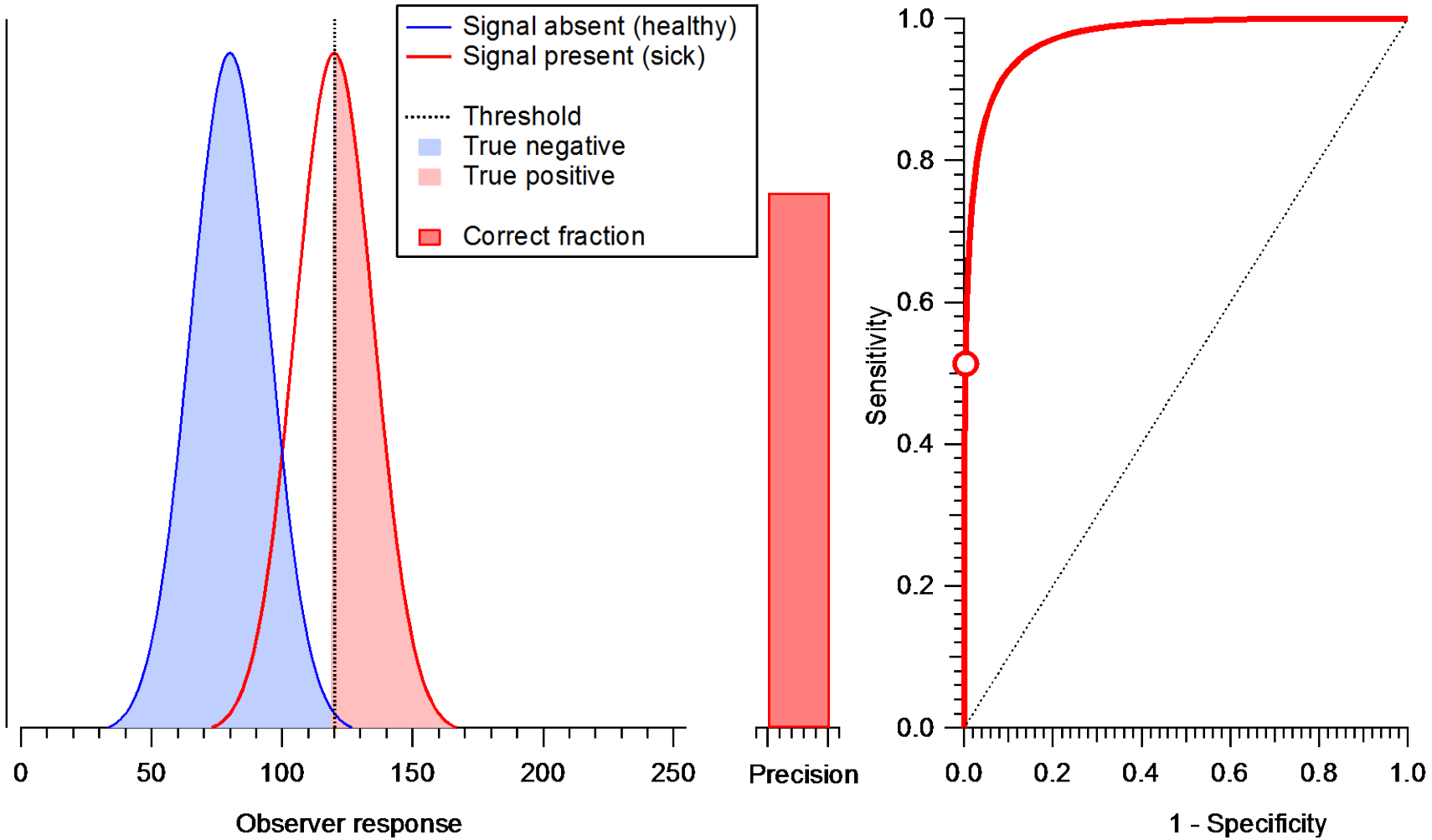
ROC curve: **low** threshold



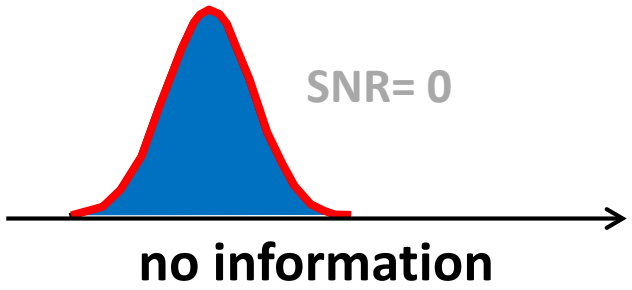
ROC curve: **medium** threshold

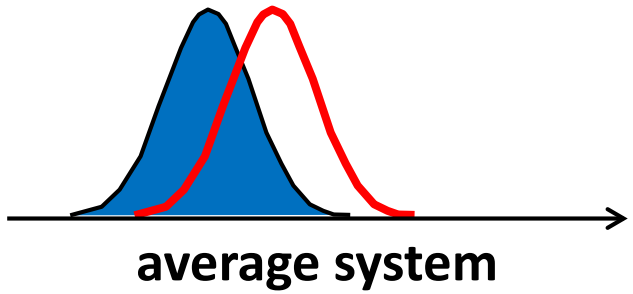


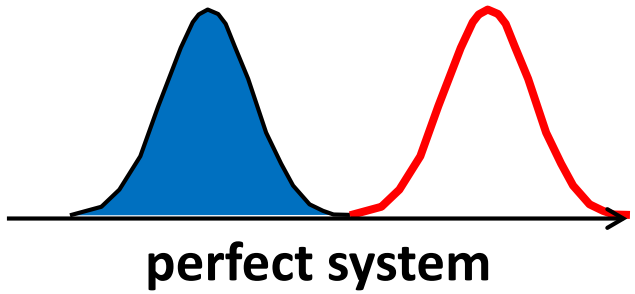
ROC curve: **high** threshold

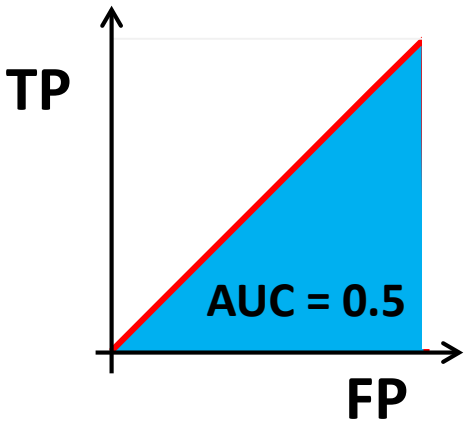


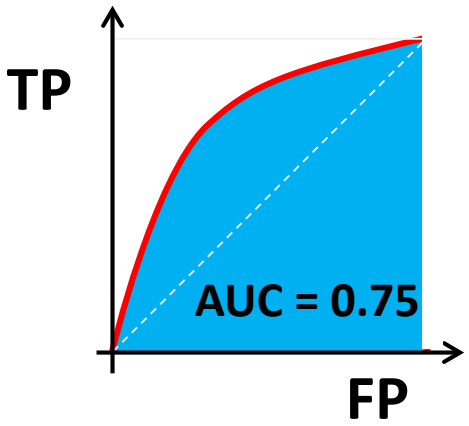
In general, the **larger** the area under the curve **(AUC)** the **better** the diagnostic **system**

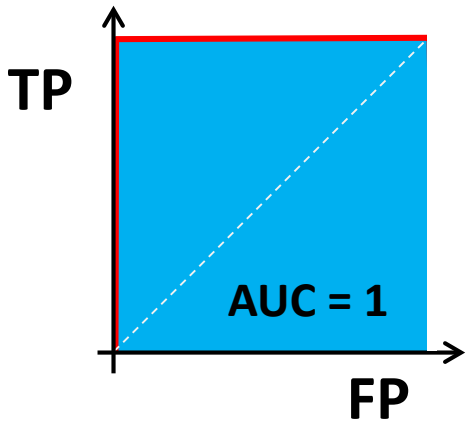












AUC: **A**rea **U**nder the (ROC) **C**urve

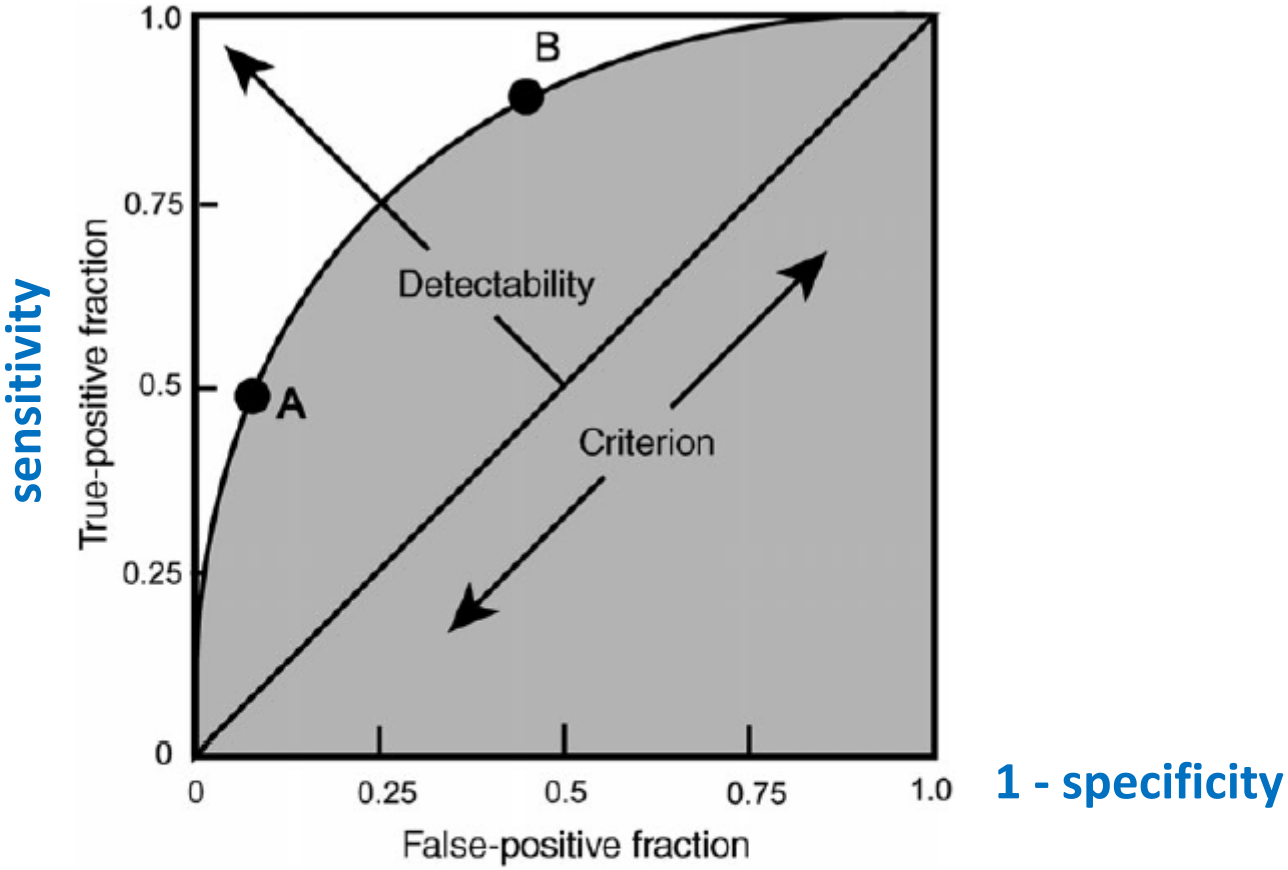


Figure 1.1. An ROC curve. The ROC curve is a plot of TPF against FPF. It describes the inherent detectability of the signal as the decision criterion varies from strict (point A on the curve) to lenient (point B on the curve). The area under the curve (AUC) shown by the shaded region is one of the single valued parameter that is used to describe the curve. The area below the diagonal line, the "guessing line," is 0.5 and is the location of points that occur due to chance.

ICRU Report 79, Receiver Operating Characteristic Analysis in Medical Imaging (2008)

Which physician do you prefer?

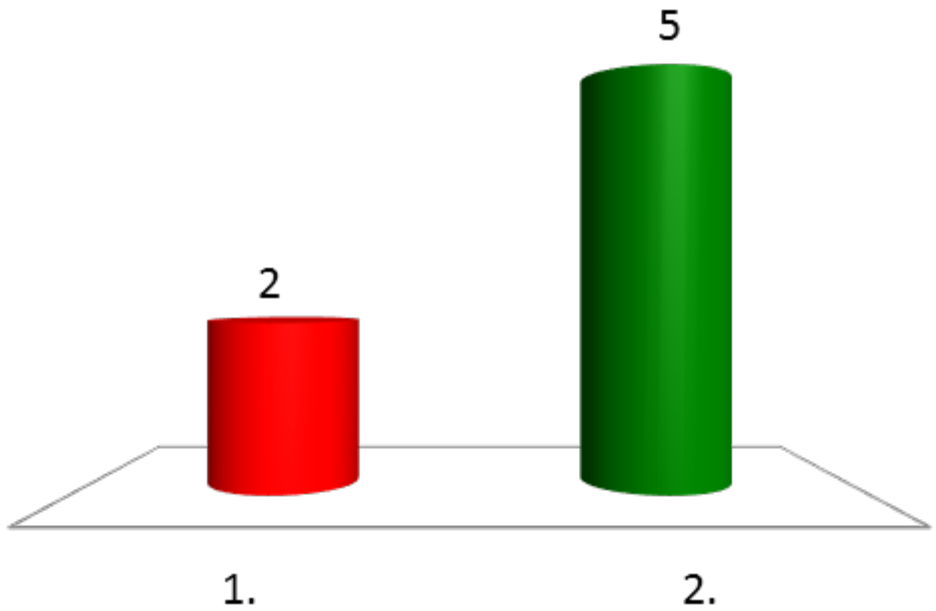
physician1

- if you are **sick**, **positive** diagnostic with probability **95%**
- if you are **healthy**, **positive** diagnostic with probability **95%**

physician 2

- if you are **sick**, **positive** diagnostic with probability **50%**
- if you are **healthy**, **positive** diagnostic with probability **5%**

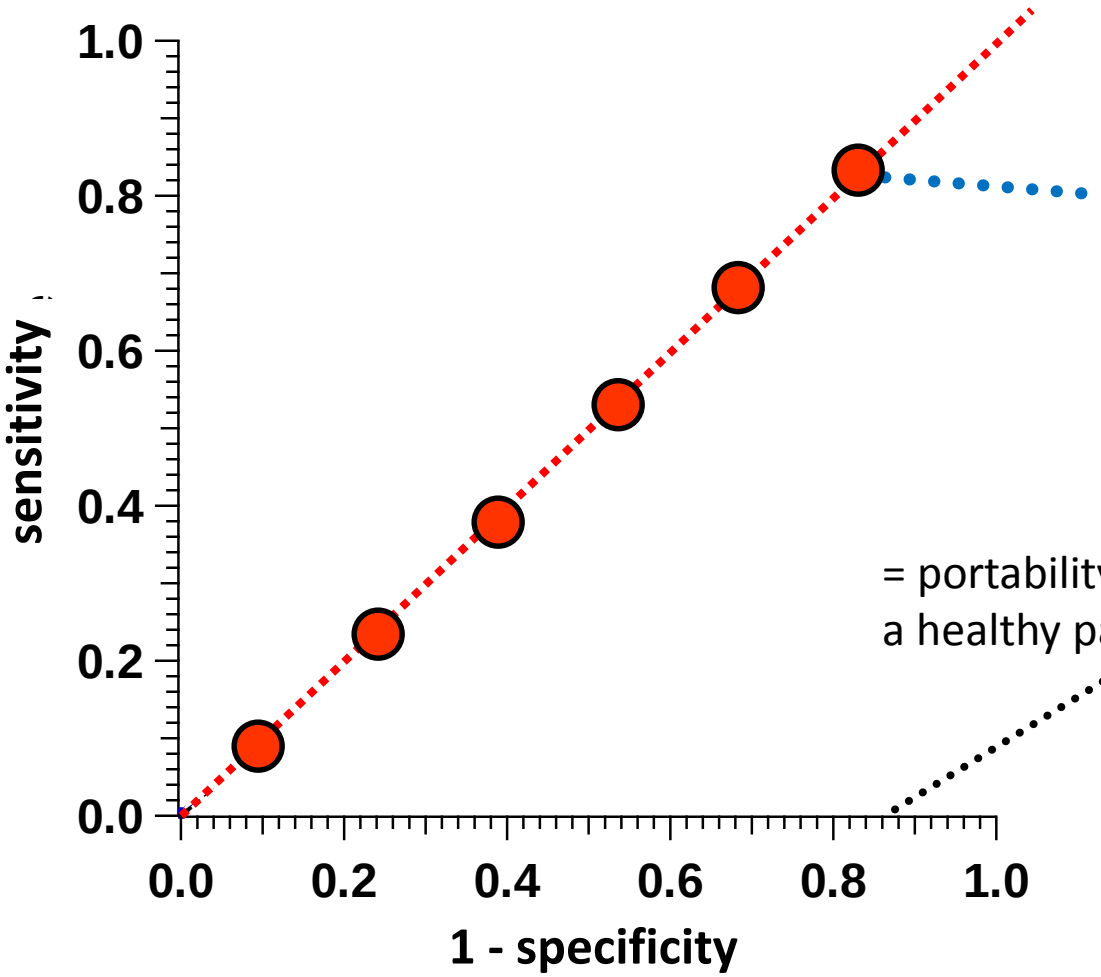
1. physician 1
2. physician 2



7 sur 7

ROC curve

worst system



sick or healthy patients have the same probability to be diagnosed positive

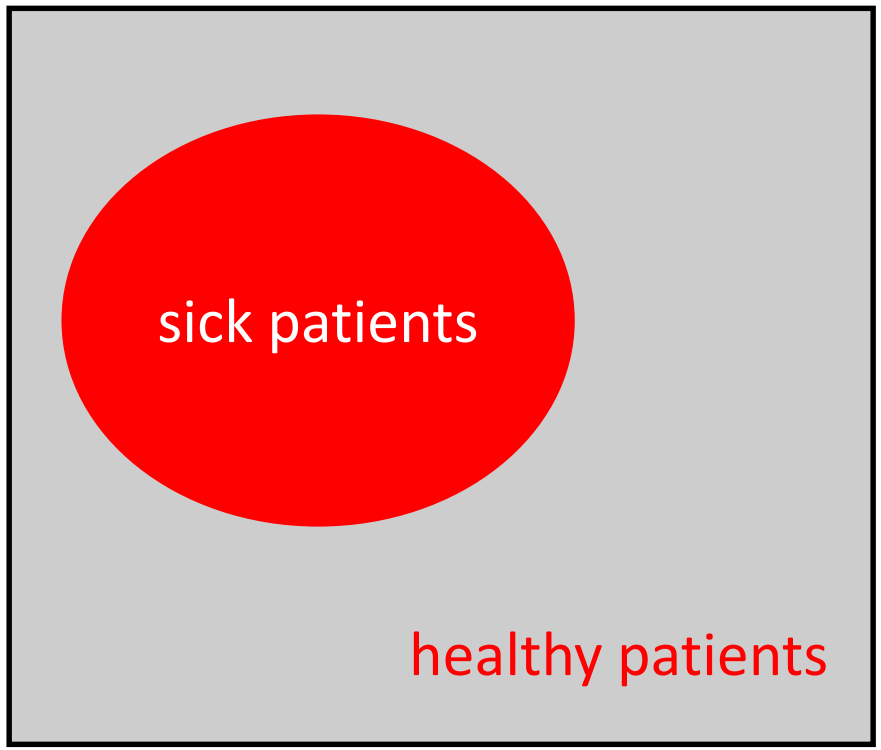
= portability of a positive diagnostic for a healthy patient

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

ROC theory

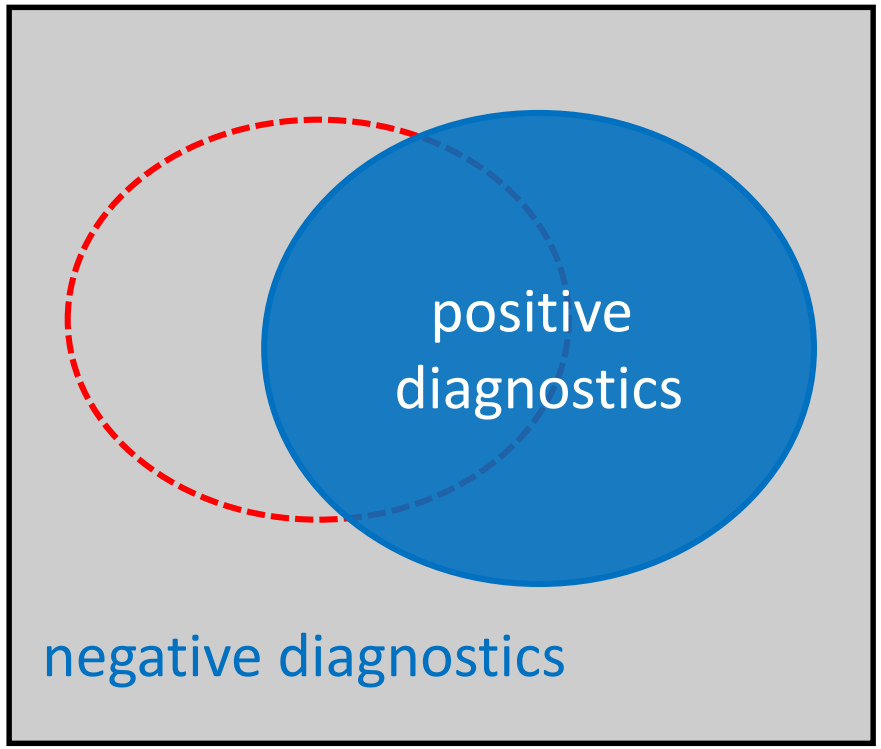
2.

Bayes' theorem



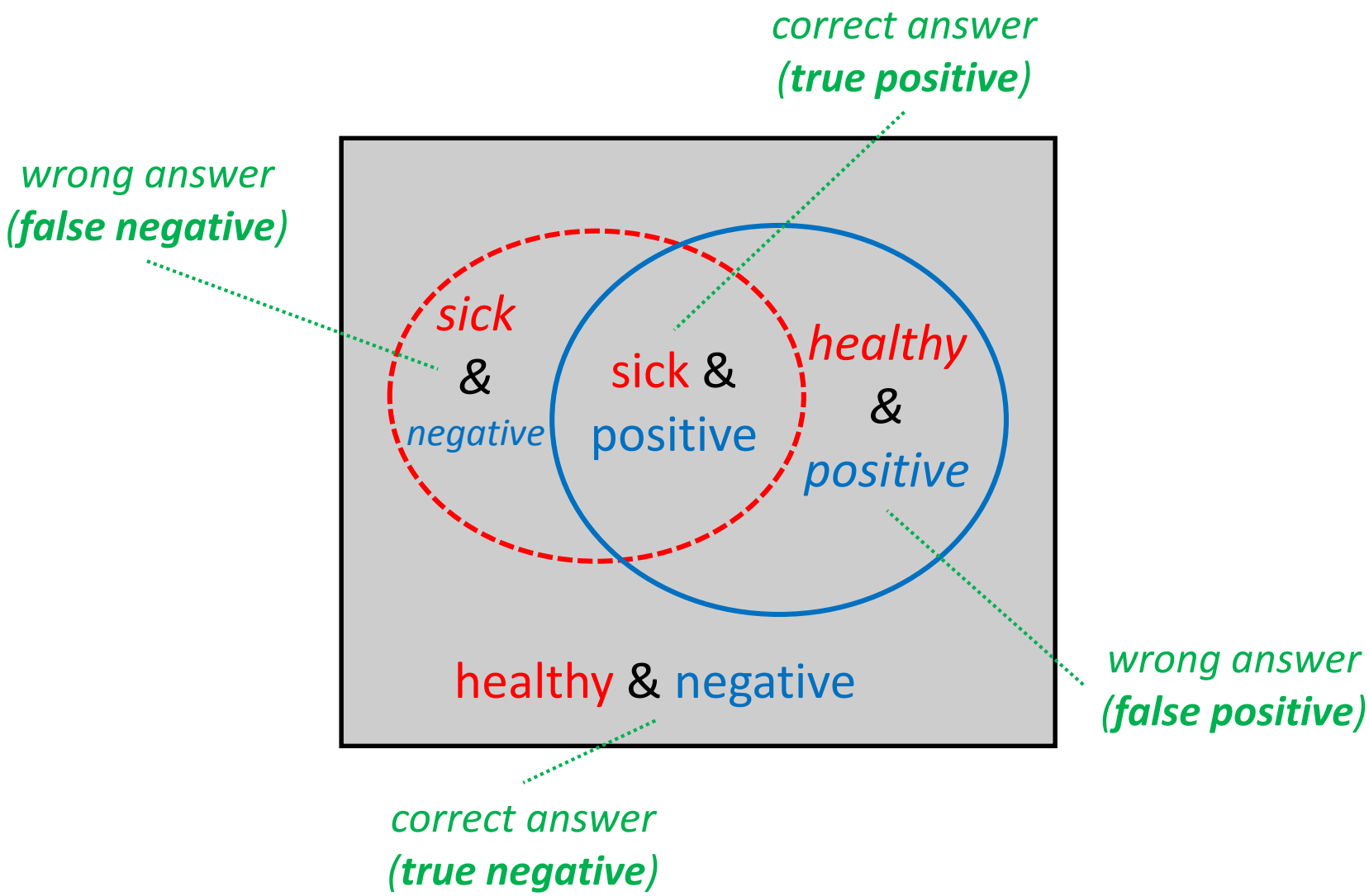
whole population

(for instance, ensemble of possible CT images of the liver)



whole population

(for instance, ensemble of possible CT images of the liver)


correct answer
(true positive)
wrong answer
(false negative)
sick
&
negative
sick &
positive
healthy
&
positive
healthy & negative
wrong answer
(false positive)
correct answer
(true negative)

Exercise

Derive the Bayes' theorem

$$P(\text{sick}|+) = \frac{P(+|\text{sick})P(\text{sick})}{P(+)}$$

Hint: start by expressing $P(\text{sick} \cap +)$

1. Press "1" when you are finished

3 sur 7

Exercise

What is the probability to be sick with the following data

Prevalence: 50 %

Sensitivity: 50 %

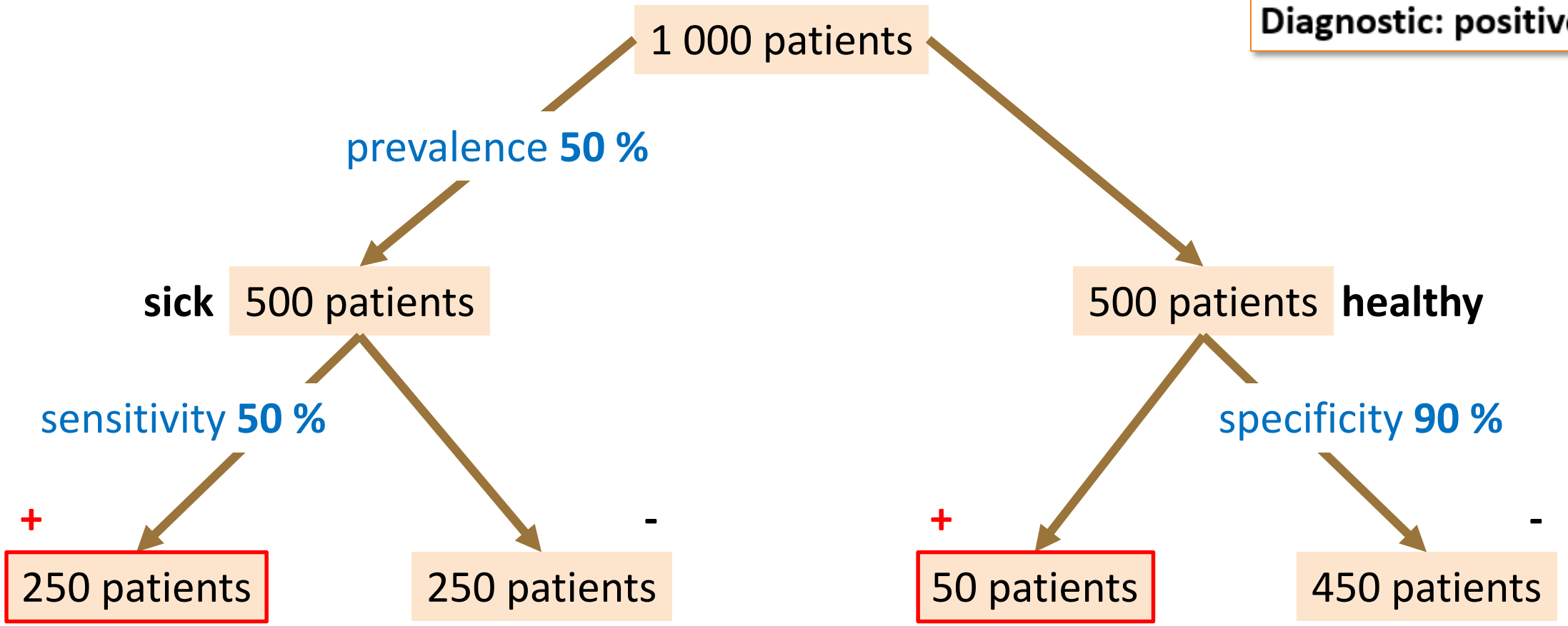
Specificity: 90 %

Diagnostic: Positive

1. Press "1" when you are finished

Same problem: **tree method**

Prevalence: 50 %
Sensitivity: 50 %
Specificity: 90 %
Diagnostic: positive



probability to be sick after a **positive** diagnostic: 250 / (250+50) = 0.83 = **83 %**

ROC theory

3.

Gold standard

The **gold standard**:

Achilles tendon of the ROC theory

- **Clinical and pathological** verification
  - **Highest level** of verification
  - **Biopsy** is the standard method for verifying the presence of cancer and many other disease processes
    (performance of pathology is tested by ROC studies…)
  - Patient **clinical course** may also provide information about the image content
- Verification using **imaging procedure**
  - Previous or subsequent imaging procedure
    (better to use **another procedure** that the one being tested)
  - Lack of change in a mass after imaging observation
    over a number of years may be evidence that it is benign

The **gold standard**:
Achilles tendon of the ROC theory

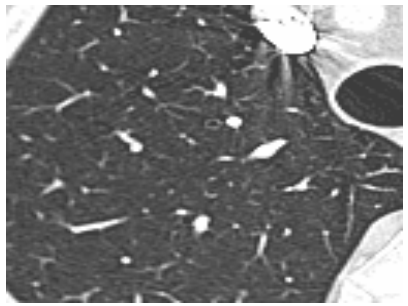
- Verification by **consensus among experts**
  - Consensus **≠ truth**
  - Even **experts disagree** about subtle cases (and such cases may be important in the study)
  - When expert panel is the only way to establish a diagnostic, an appropriate method to **combine the judgments** has to be used

ROC theory

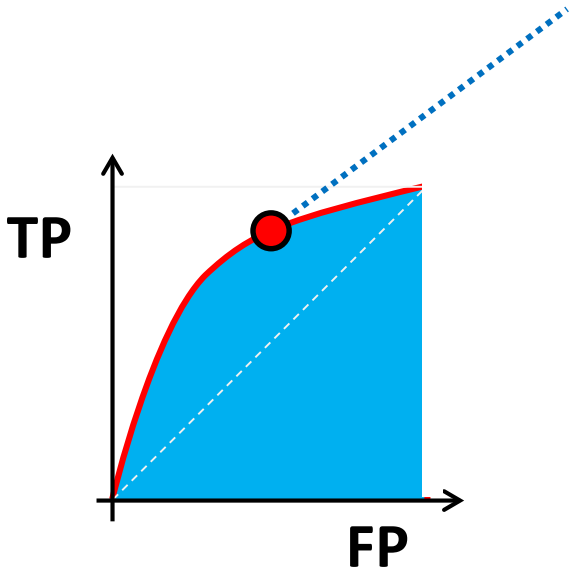
4.

Psychophysical experiments

Psychophysical (human) **experiments**
can estimate different parameters of the **ROC-curve**

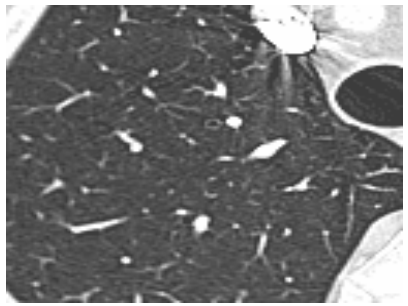


one image at a time
is the signal present?
positive
negative



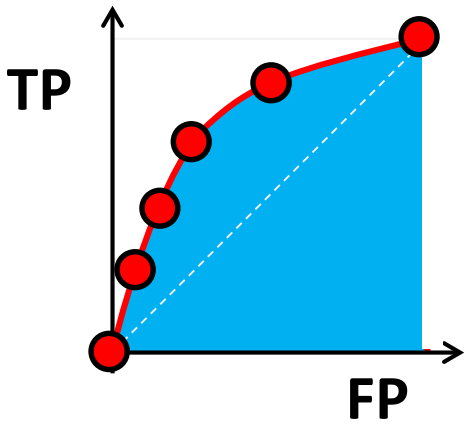
positive/negative
(YES/NO)
detection experiment
provides **one point**

Psychophysical (human) **experiments**
can estimate different parameters of the **ROC-curve**



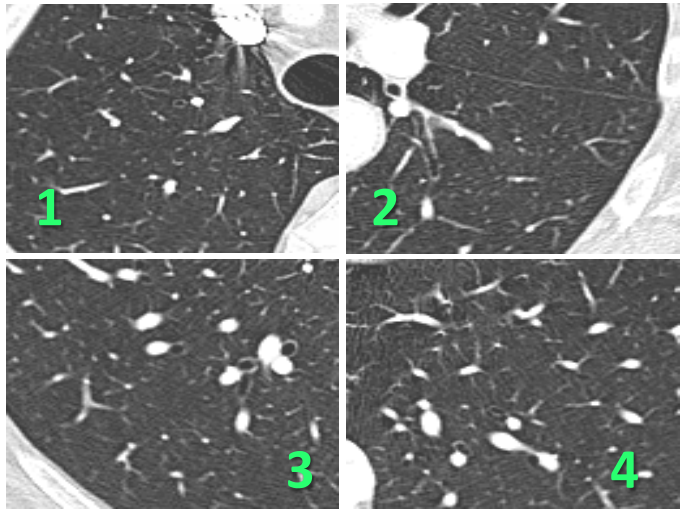
one image at a time
is the **signal present**?

No, very likely
No, probably
Maybe
Yes, probably
Yes, very likely



Rating
experiment provides the
several points
of the **curve**

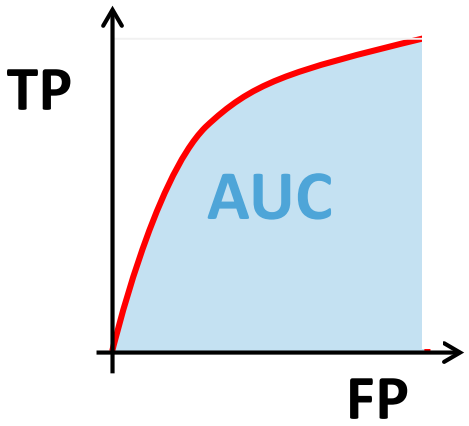
Psychophysical (human) **experiments** can estimate different parameters of the **ROC-curve**



4-AFC

Which image is the most likely to contain the signal?

1, 2, 3, 4



M-alternative forced-choice *(M-AFC)* experiment provides the **area under the curve (AUC)**

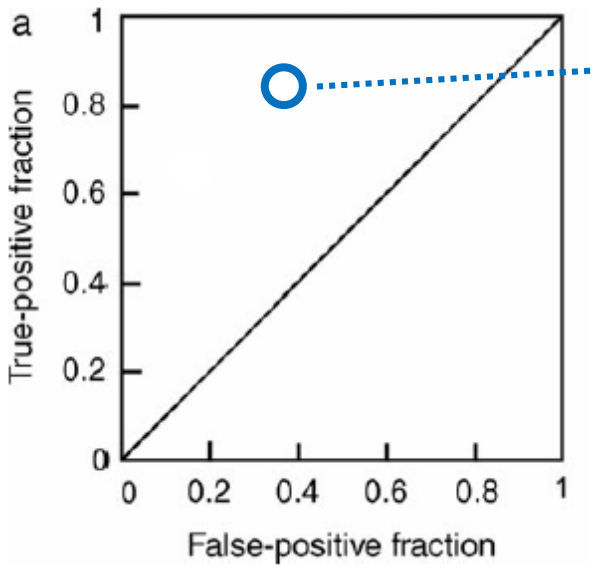
ROC theory

4.1

Detection experiments

YES/NO experiment

Result of a YES/NO experiment



one observer produces
one result

ICRU Report 79, Receiver Operating Characteristic Analysis in Medical Imaging (2008)

ROC theory

4.2

Detection experiments

Rating experiment

Rating experiment: example of scale

Table 1.5. A five-category rating scale that is frequently used for collecting data from an ROC experiment

Rating	Interpretation
1	Disease is definitely or almost definitely absent
2	Disease is probably absent
3	Disease is possibly present
4	Disease is probably present
5	Disease is definitely or almost definitely present

ICRU Report 79, Receiver Operating Characteristic Analysis in Medical Imaging (2008)

Table 1.6. Observer response data from an experiment with a test set of 40 diseased and 60 disease-free cases

	Rating of the confidence that the case is diseased					Total cases
	5 (definitely present)	4 (probably present)	3 (possibly present)	2 (probably absent)	1 (definitely absent)	
Diseased cases	20	8	6	4	2	40
Disease-free cases	3	6	6	9	36	60

Table 1.7. The cumulative frequencies of the observer response data

	Cumulative frequencies				
	5	5 + 4	5 + 4 + 3	5 + 4 + 3 + 2	5 + 4 + 3 + 2 + 1
Diseased cases	20	28	34	38	40
Disease-free cases	3	9	15	24	60

Table 1.8. The cumulative fractions of the observer response data

	Cumulative fractions				
	5	5 + 4	5 + 4 + 3	5 + 4 + 3 + 2	5 + 4 + 3 + 2 + 1
Diseased cases	0.50	0.70	0.85	0.95	1.00
Disease-free cases	0.05	0.15	0.25	0.40	1.00

ICRU Report 79, Receiver Operating Characteristic Analysis in Medical Imaging (2008)

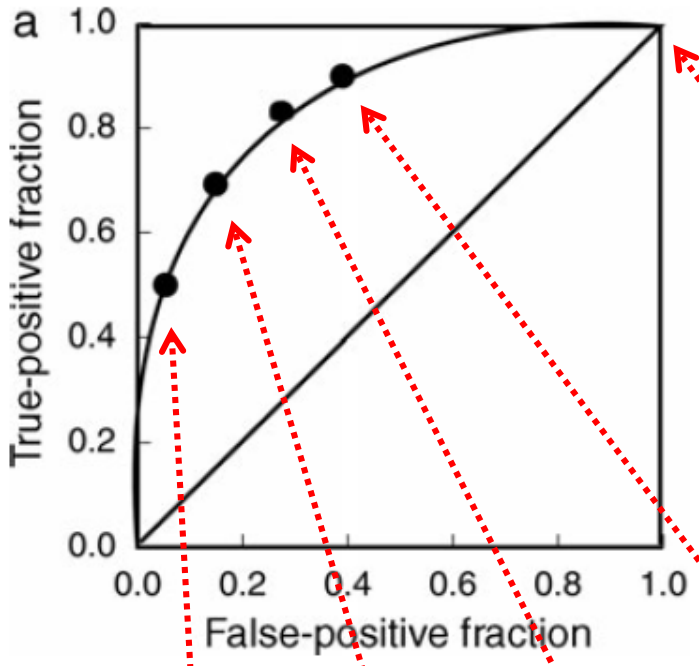
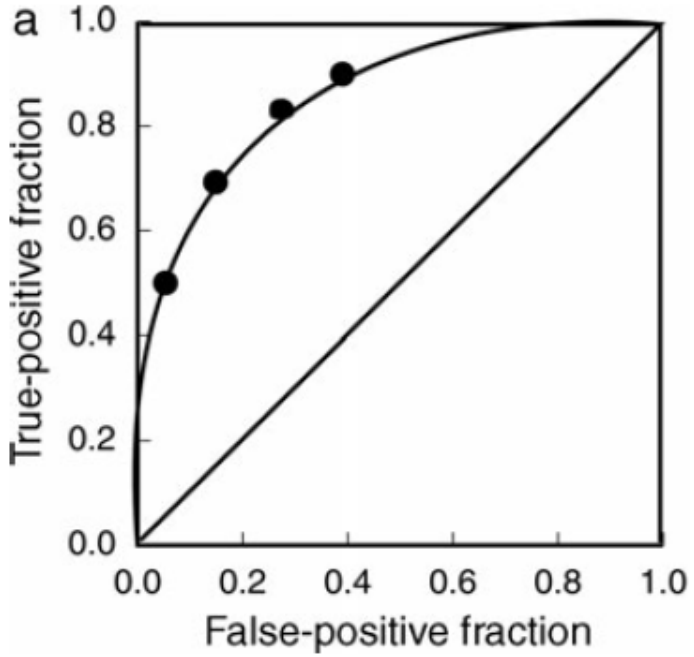


Table 1.8. The cumulative fractions of the observer response data

	Cumulative fractions				
	5	5 + 4	5 + 4 + 3	5 + 4 + 3 + 2	5 + 4 + 3 + 2 + 1
Diseased cases	0.50	0.70	0.85	0.95	1.00
Disease-free cases	0.05	0.15	0.25	0.40	1.00

ICRU Report 79, Receiver Operating Characteristic Analysis in Medical Imaging (2008)



The usual method consist in performing a **maximum likelihood estimation** to fit a **binormal** ROC curve

i.e. by supposing that both responses are **Gaussian**
(not necessarily with the same variances)

Fitting programs can be found on the web:

http://www.mips.ws
http://xray.bsd.uchicago.edu/krl/index.htm

ICRU Report 79, Receiver Operating Characteristic Analysis in Medical Imaging (2008)

Real experiment

You will be shown a series of 20 images.

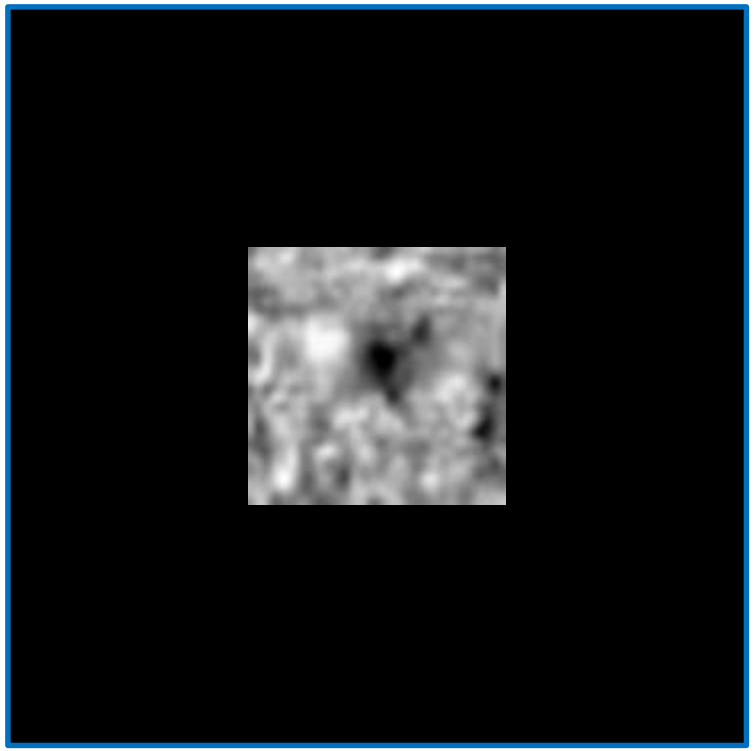
- 10 will contain a signal and a noise
- 10 will contain only noise

The images will be presented in a random order

Your task is to rate each image according to this scale

1 sure, there is no signal	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9 sure, there is a signal**

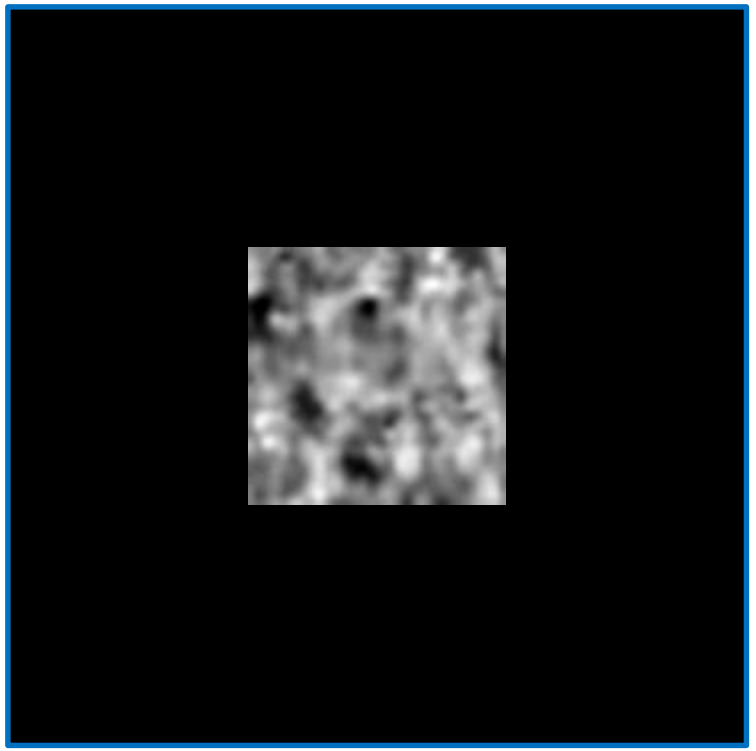
Example of **signal** present



1/2

VEO reconstruction
$CTDI_{vol}$ = 0.8 mGy
Contrast = 20 HU
Signal diameter = 6 mm

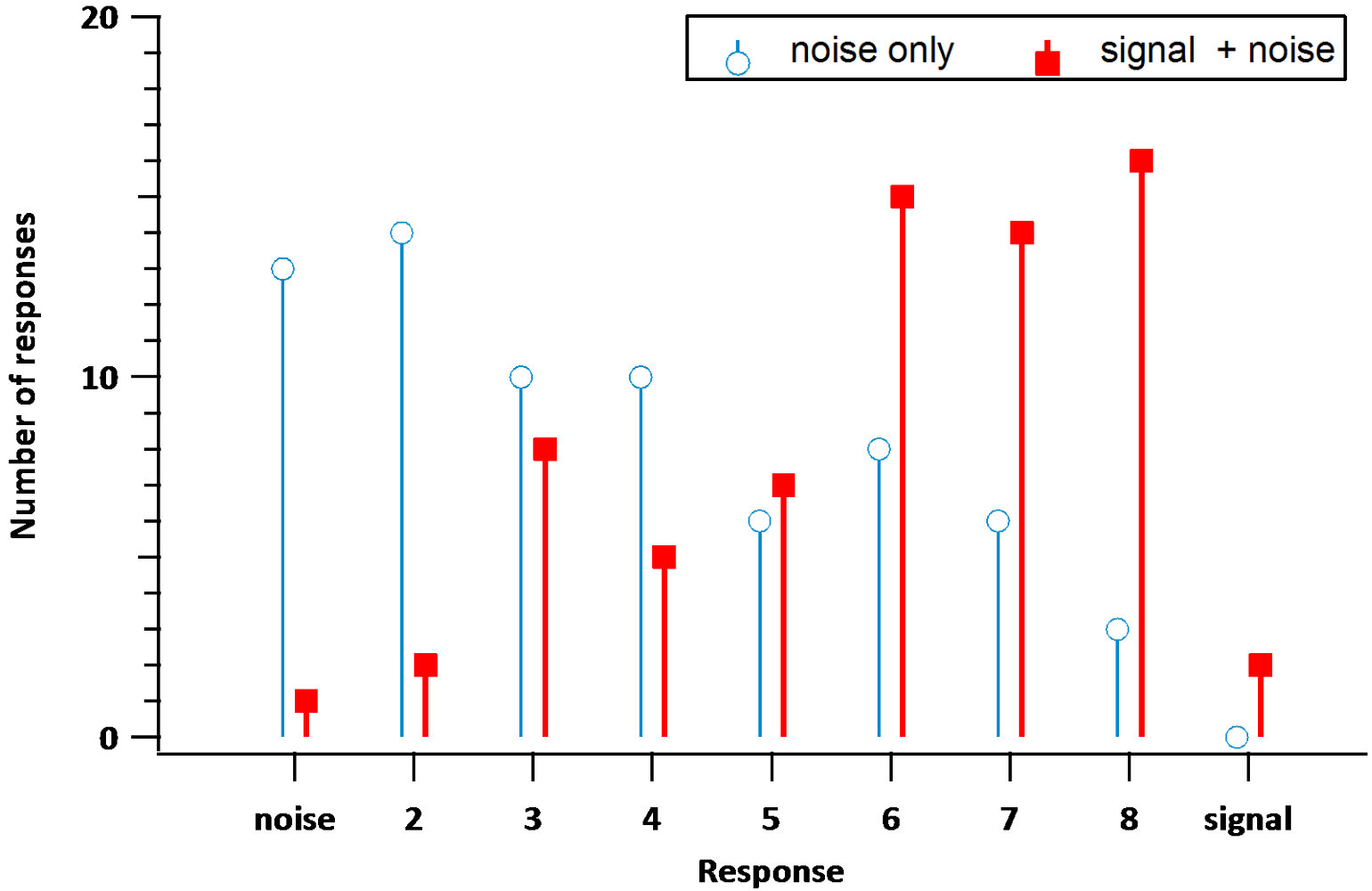
Example of **noise** only



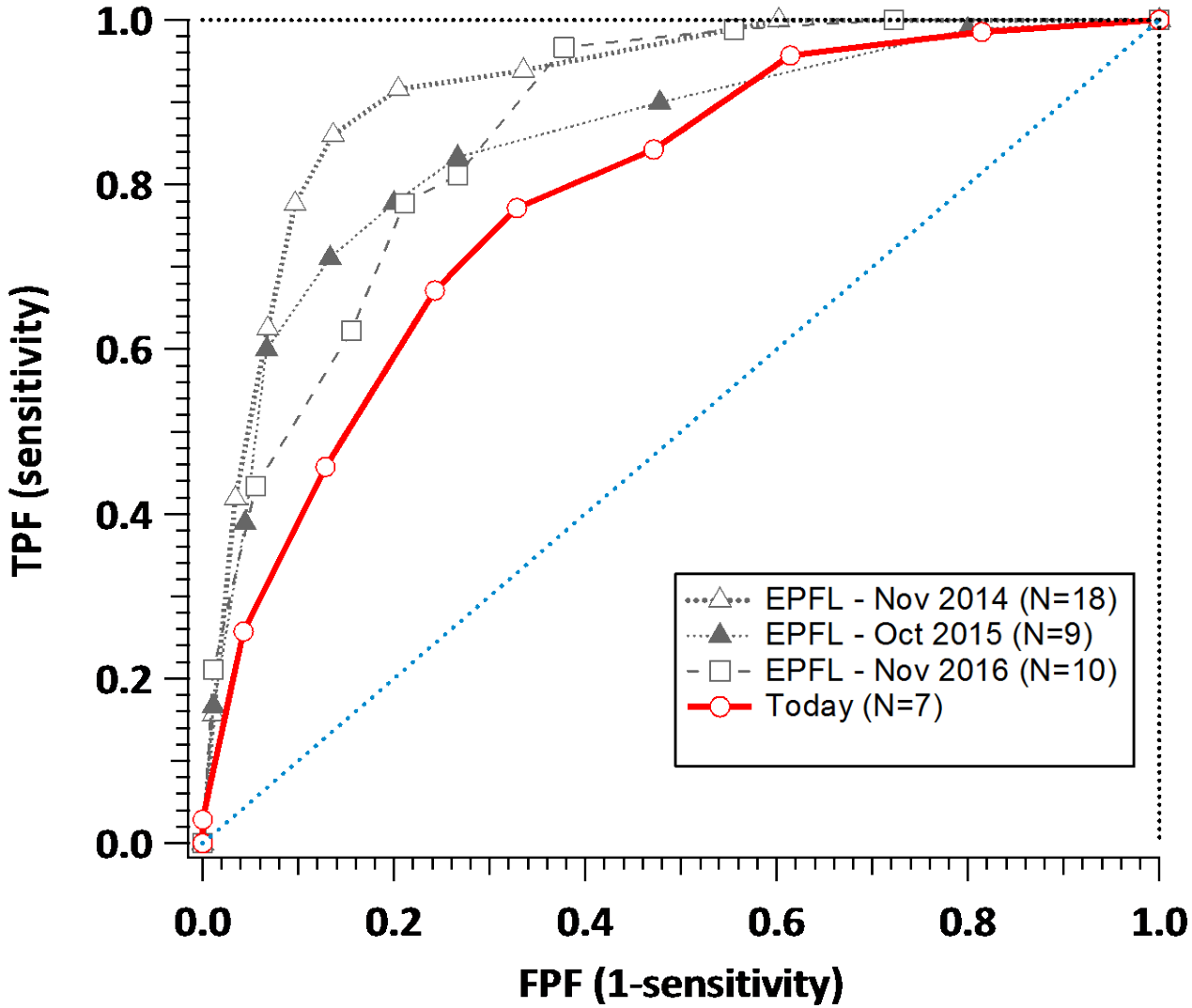
VEO reconstruction
$CTDI_{vol}$ = 0.8 mGy

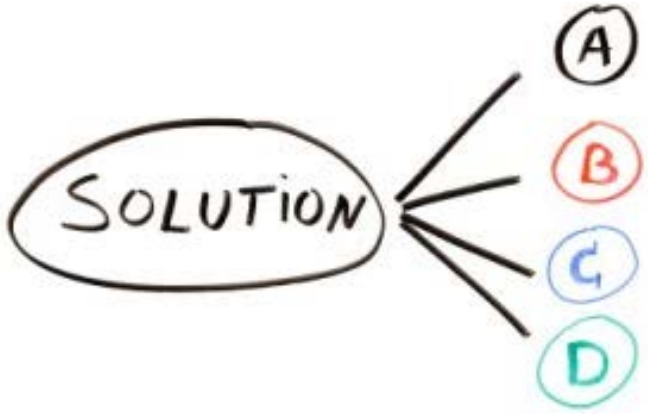
1/2

Obtained results (2018)



Obtained results (2018)





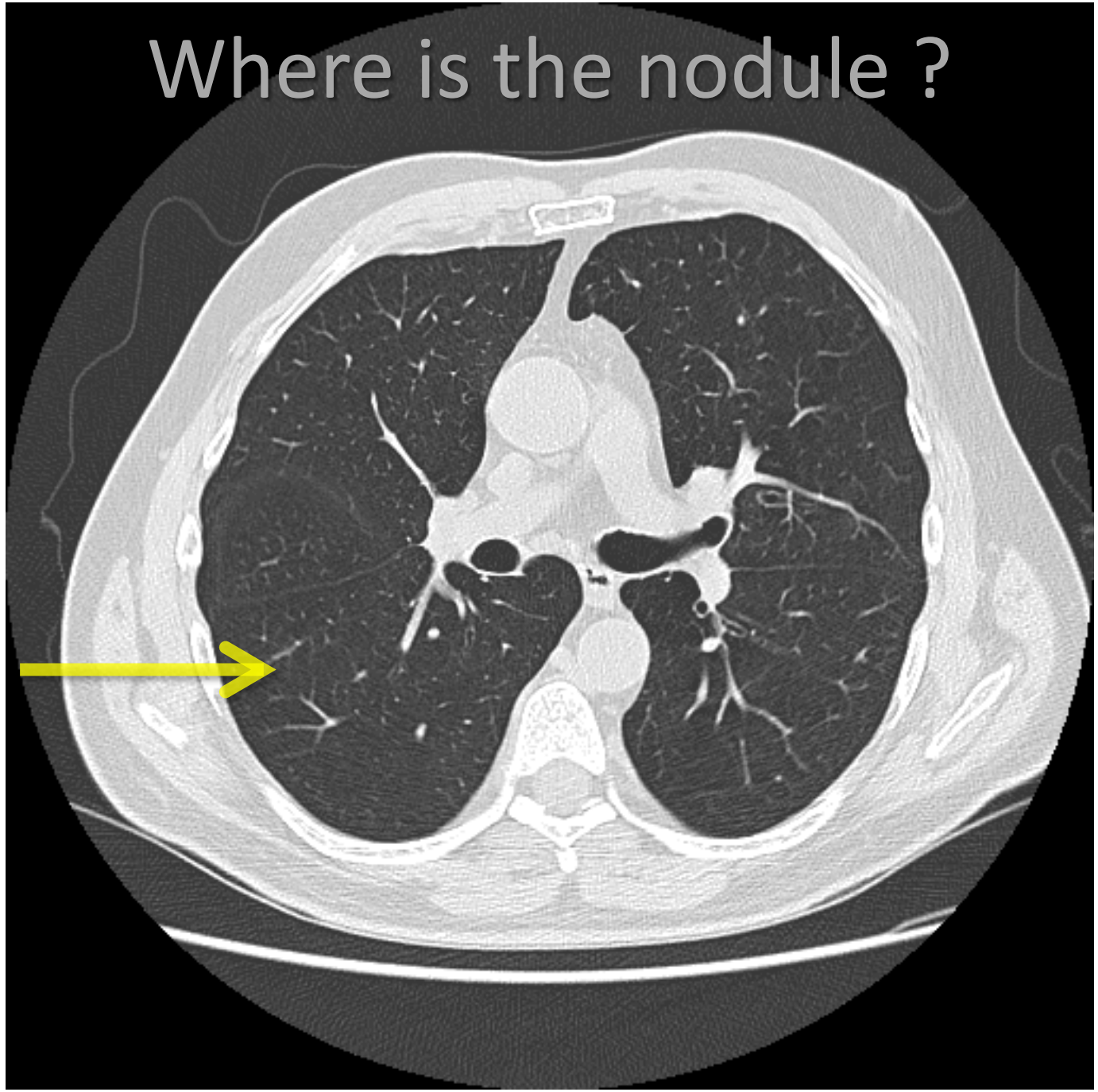
ROC theory

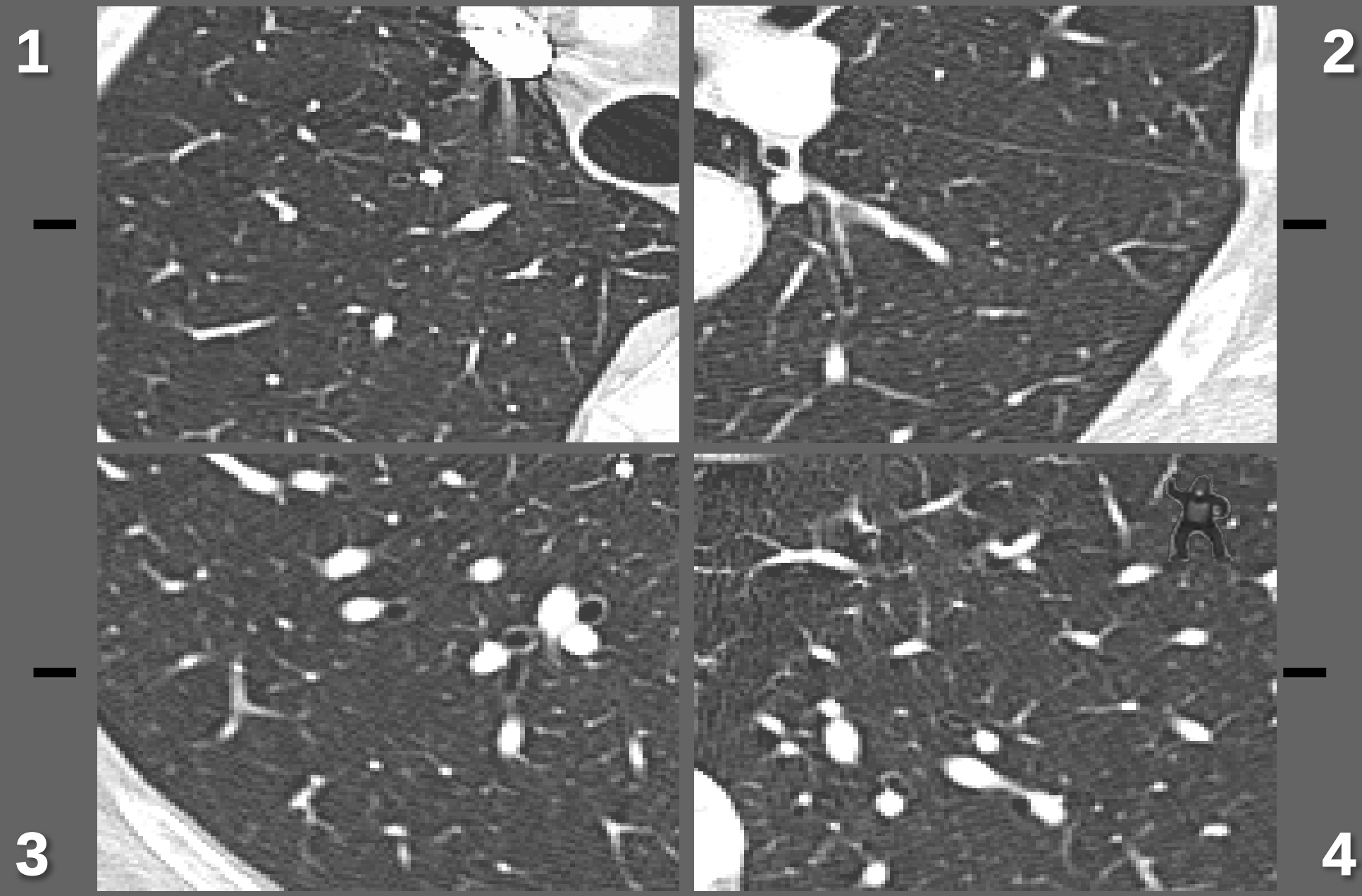
4.3

Detection experiments

Multi-alternative forced-choice experiment (M-AFC)

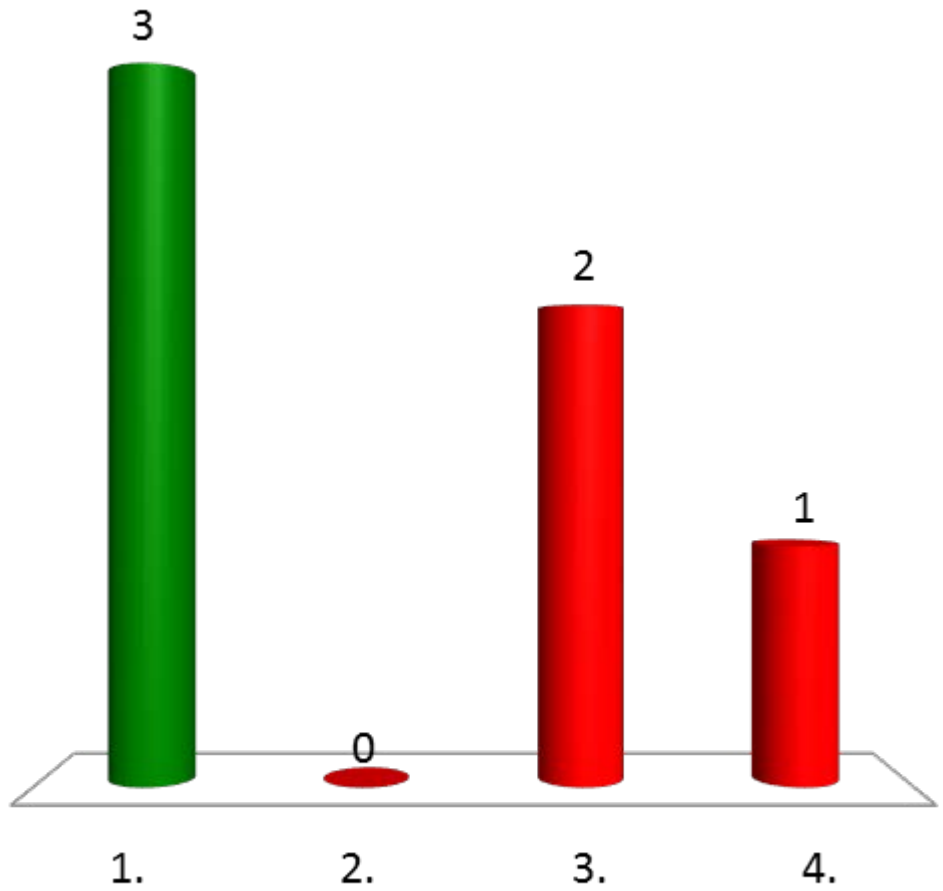
Where is the nodule ?




1
2
3
4

Where is the nodule ?

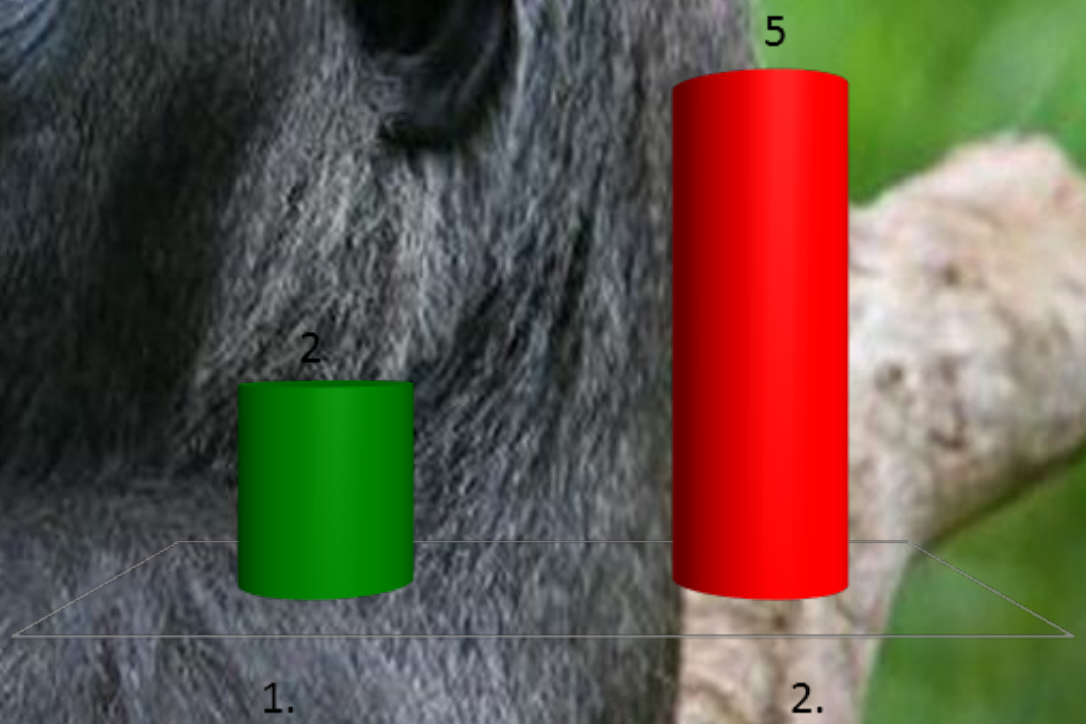
1	2
3	4



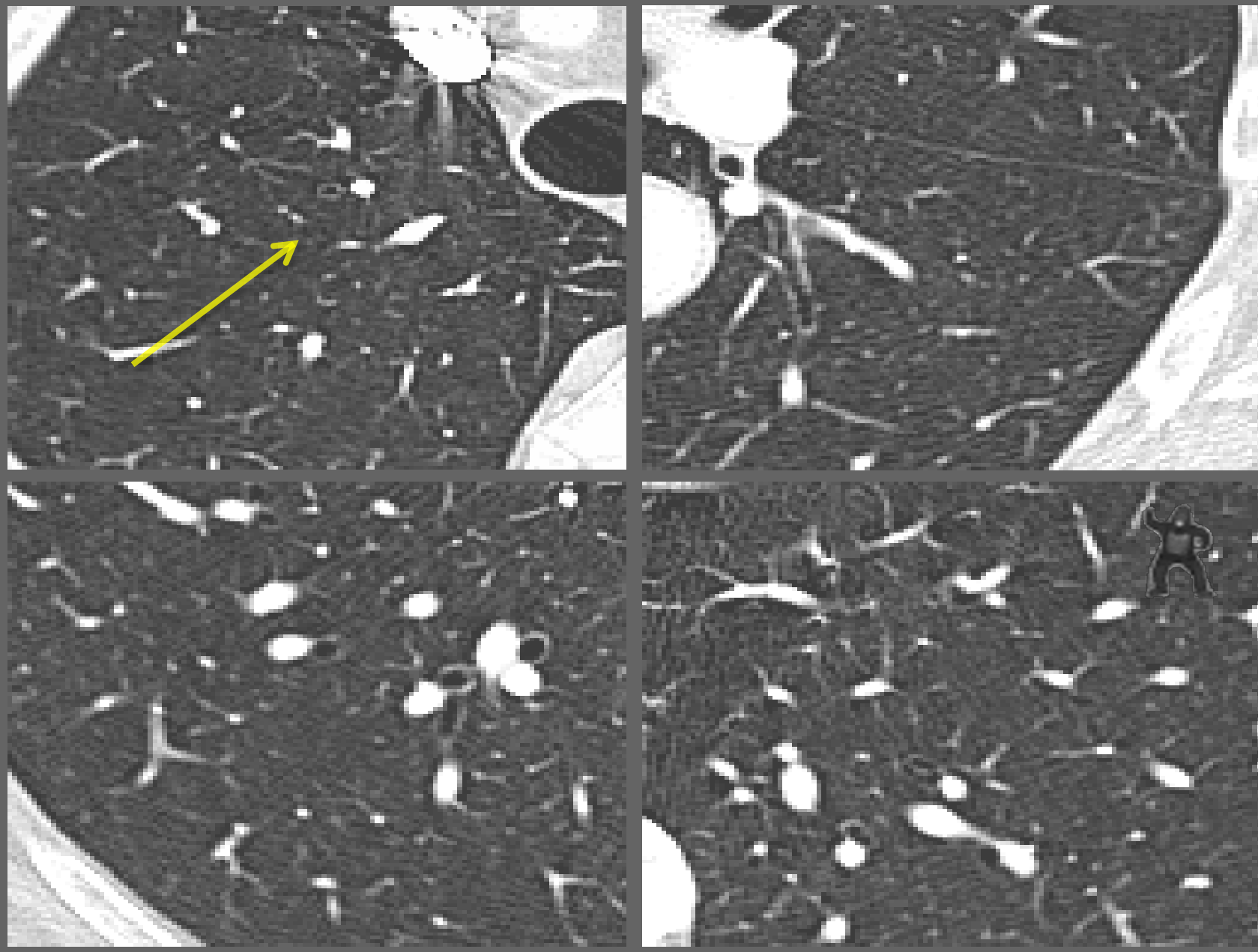
6 sur 7

Did you see a gorilla?

1. yes
2. no



7 sur 7

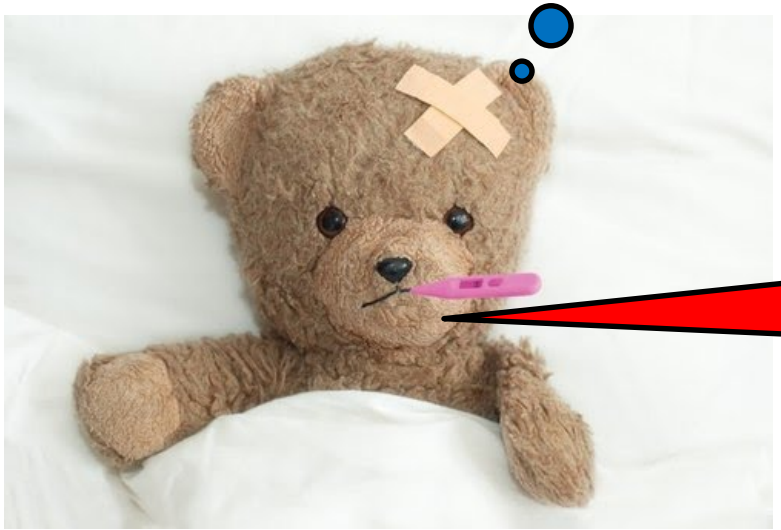

1
2
3
4

ROC theory

5.

How to communicate about the performance of a test

What should the patient think if the result of the **diagnostic is positive**?



I am sick

After
a **positive test**,
the interesting
question is:

Doctor, what is the probability that I am actually sick?

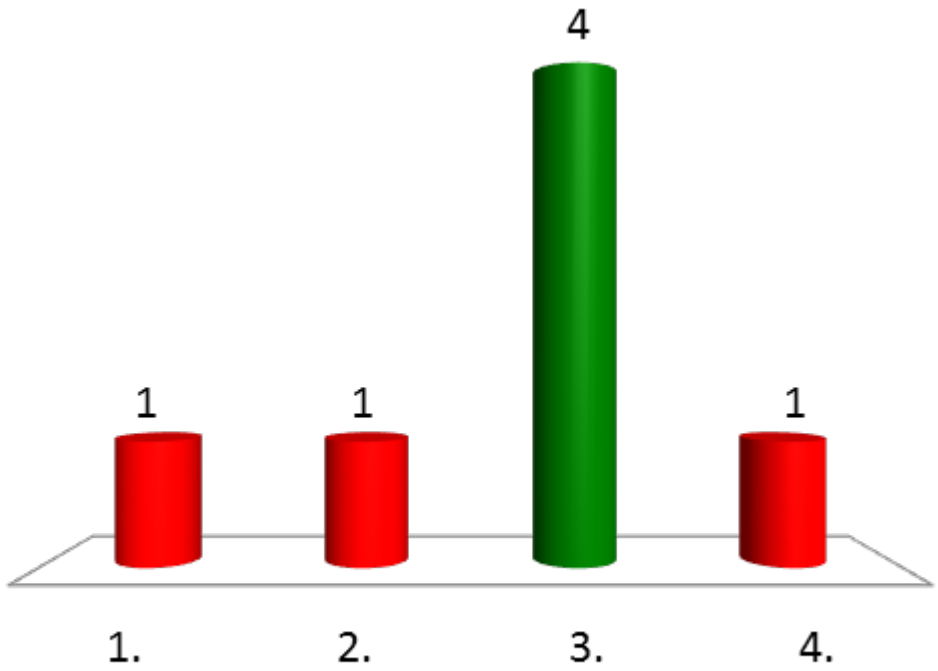
Your patient is 40 years old and participate in a
mammography screening

Information

- about 1% of 40 years-old women have breast cancer
- sensitivity 80%
- specificity 90%

If the diagnostic is positive, what is the probability that your patient actually has a breast cancer?

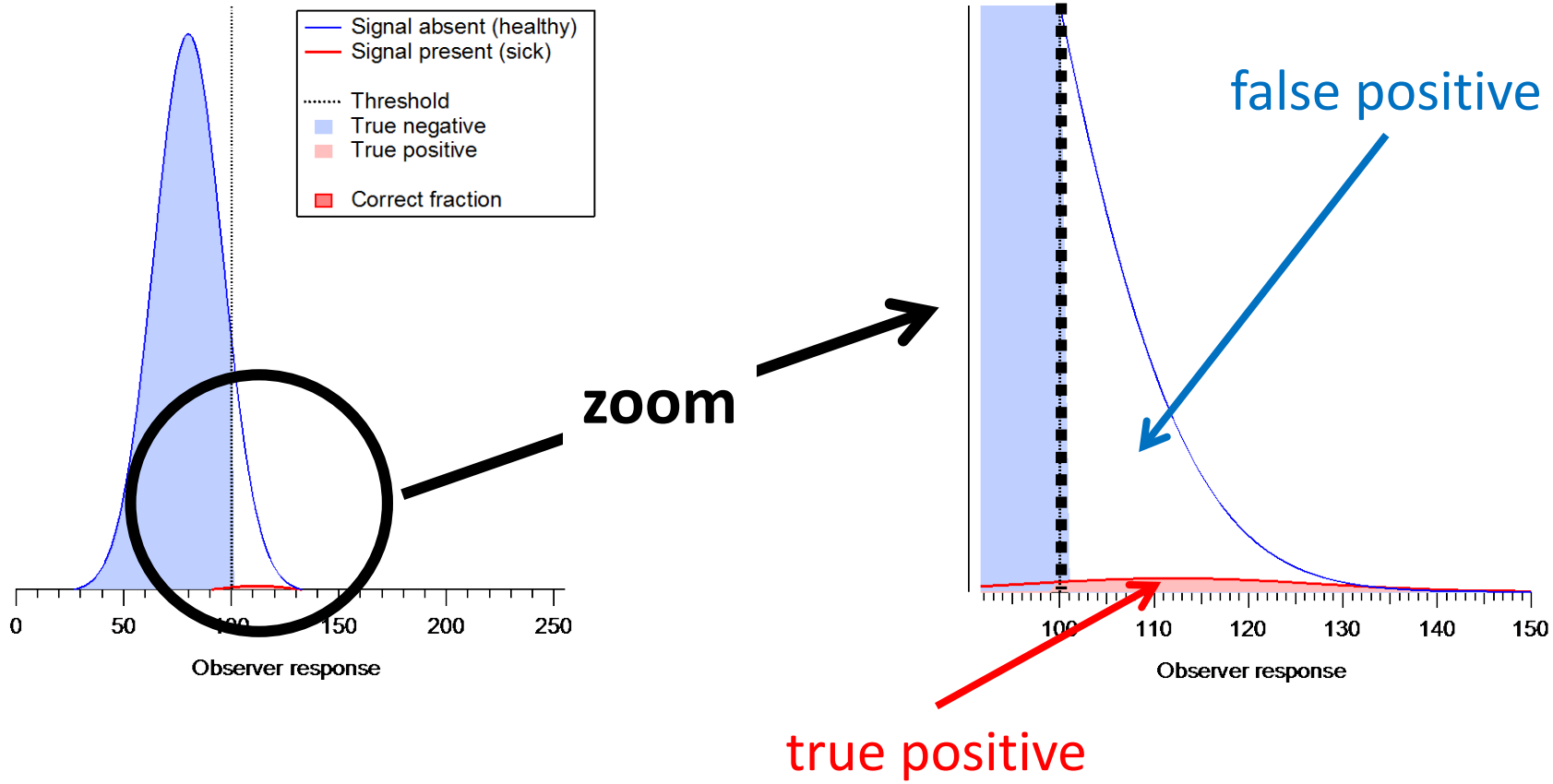
1. 0.075%
2. 0.75%
3. 7.5%
4. 75%



7 sur 7

Mammography screening: positive predicting value

prevalence: 0.01
sensitivity: 0.80
specificity: 0.90



Mammography screening: positive predicting value

prevalence: **0.01**
sensitivity: **0.80**
specificity: **0.90**

Bayes' theorem

$$p(\text{sick}|+) = \frac{P(+|\text{sick}) \times P(\text{sick})}{P(+)}$$

$$= \frac{P(+|\text{sick}) \times P(\text{sick})}{p(+|\text{sick}) \times P(\text{sick}) + p(+|\text{healthy})(1 - P(\text{sick}))}$$

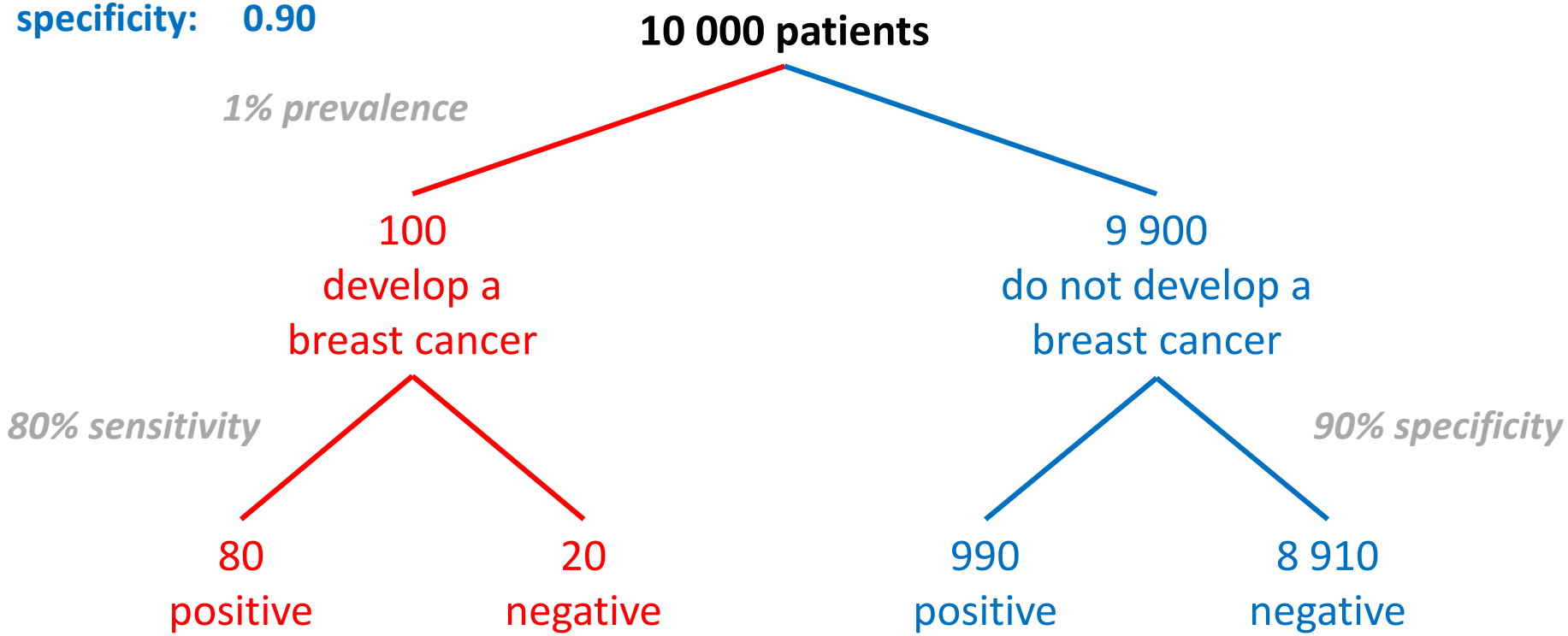
$$= \frac{\text{sensitivity} \times \text{prevalence}}{\text{sensitivity} \times \text{prevalence} + (1 - \text{sensitivity})(1 - \text{prevalence})}$$

$$= 0.075$$

Mammography screening: positive predicting value

(alternative description)

prevalence: 0.01
sensitivity: 0.80
specificity: 0.90



$$\text{PPV} = \frac{80}{80 + 990} = 0.075$$

ROC theory

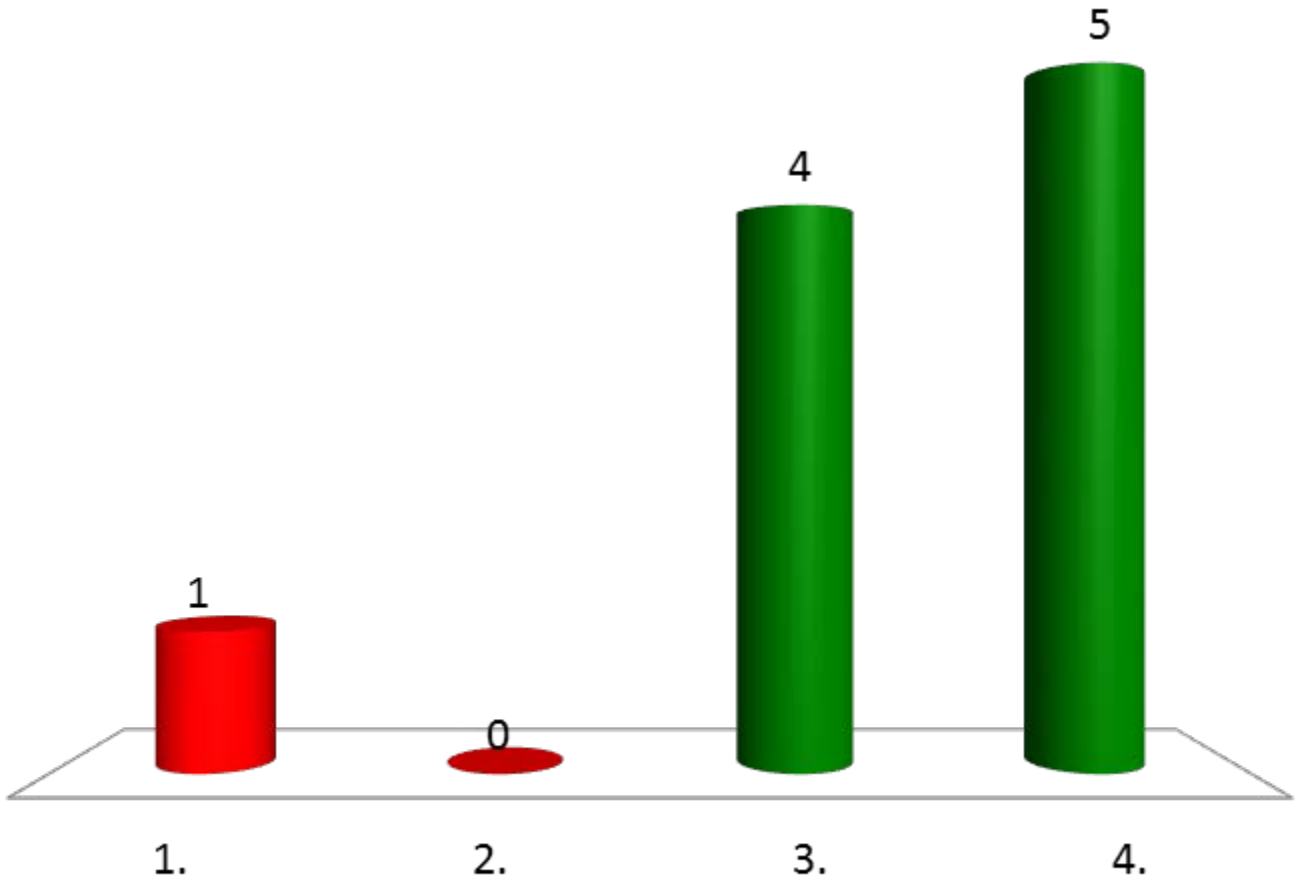
6.

Summary questions

What is presented by the ROC curve?

(several possible answers)

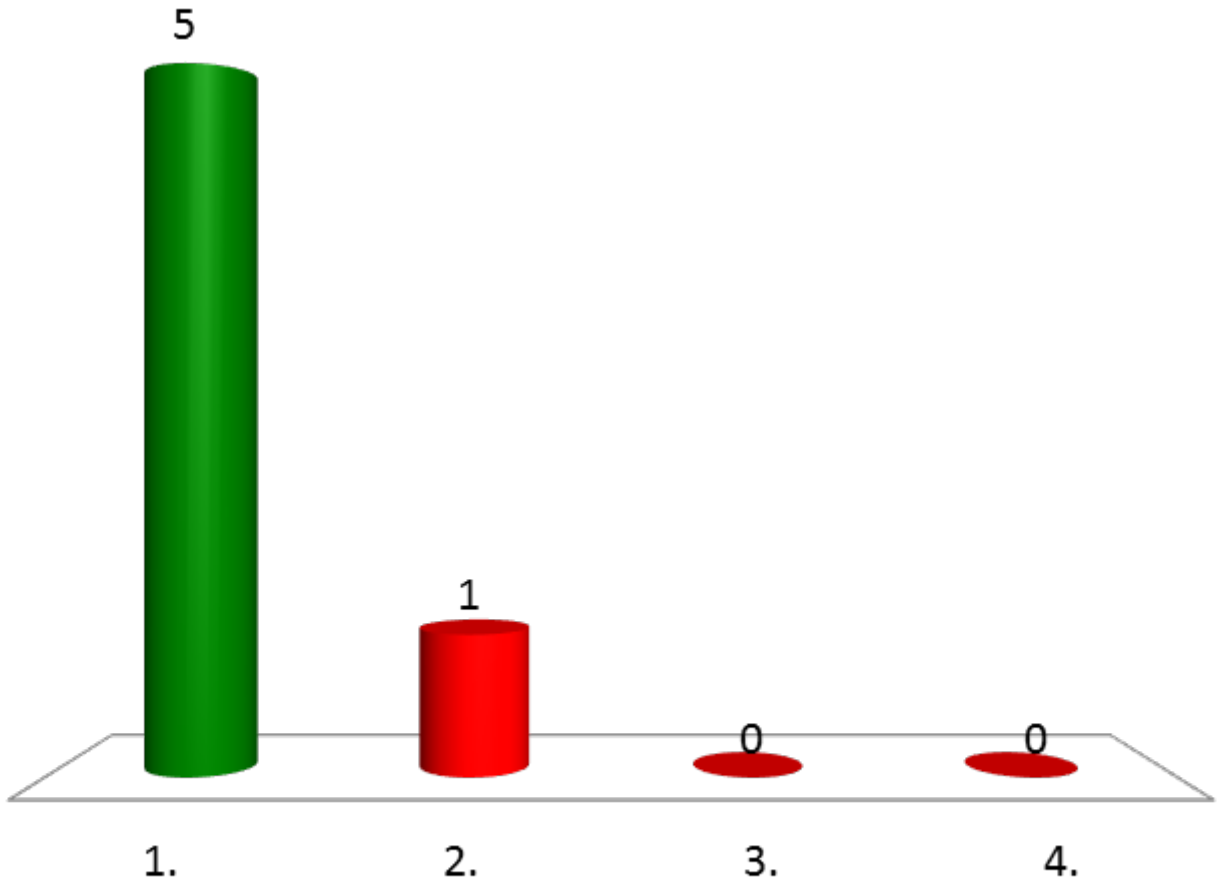
1. **sensitivity**
   versus **specificity**
2. **specificity**
   versus **sensitivity**
3. **sensitivity**
   versus **1-specificity**
4. **true positive** fraction
   versus
   false positive fraction



7 sur 7

What is the sensitivity?

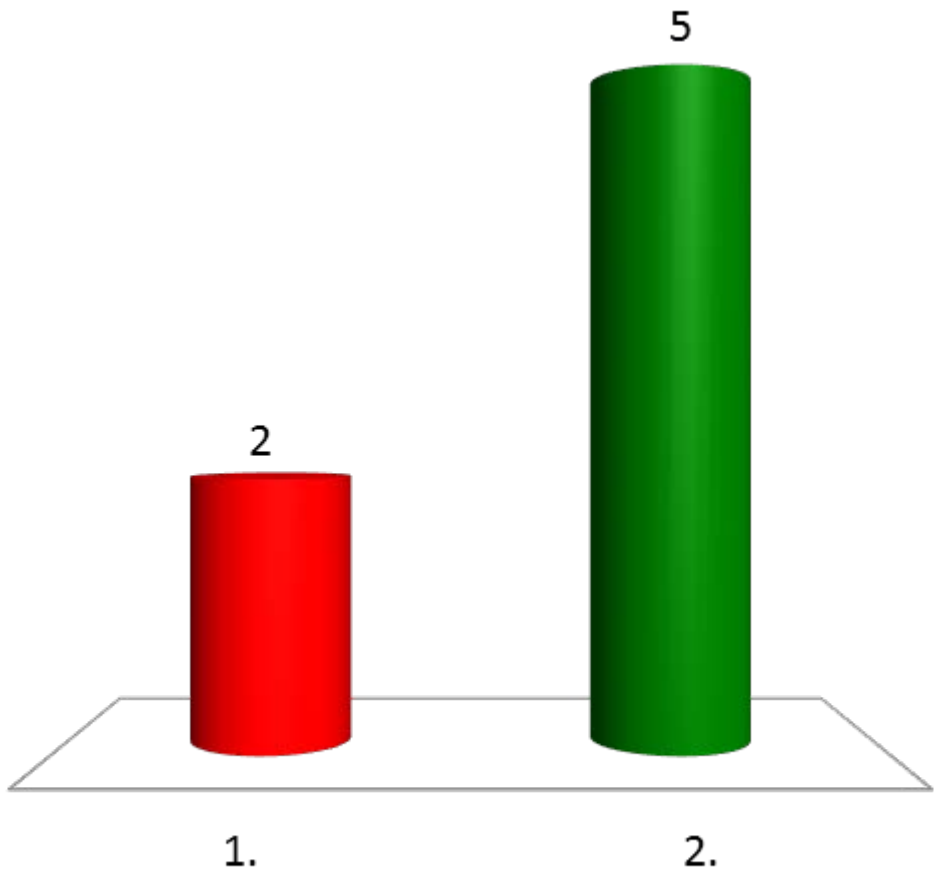
1. Probability
   of **correct responses**
   for a **sick** patient
2. Probability
   of **correct responses**
   for a **healthy** patient
3. Probability
   to be **sick**
   with a **positive diagnostic**
4. Probability
   to be **healthy**
   with a **negative diagnostic**



6 sur 7

Does **sensitivity** depend on the **prevalence**?

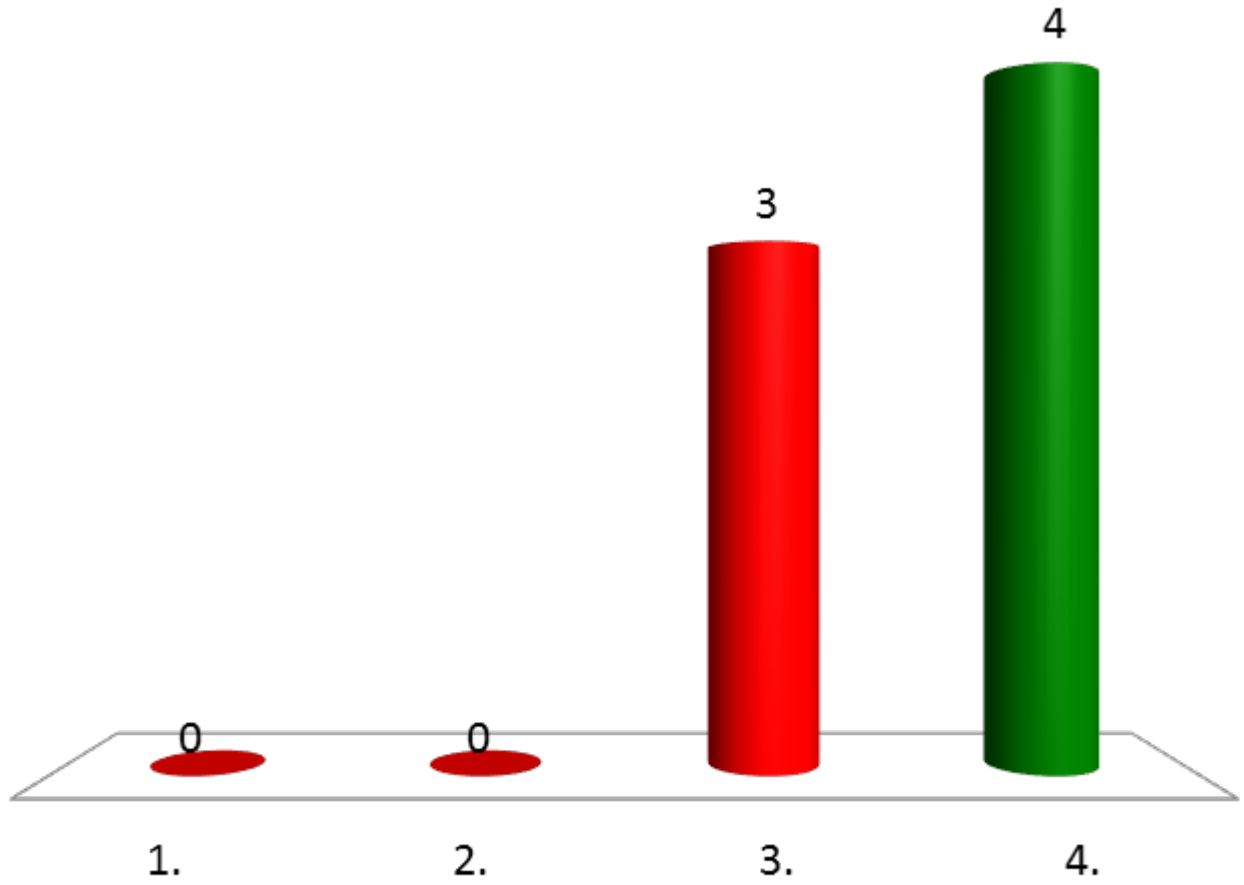
1. YES
2. NO



7 sur 7

What is the **negative predicting value**?

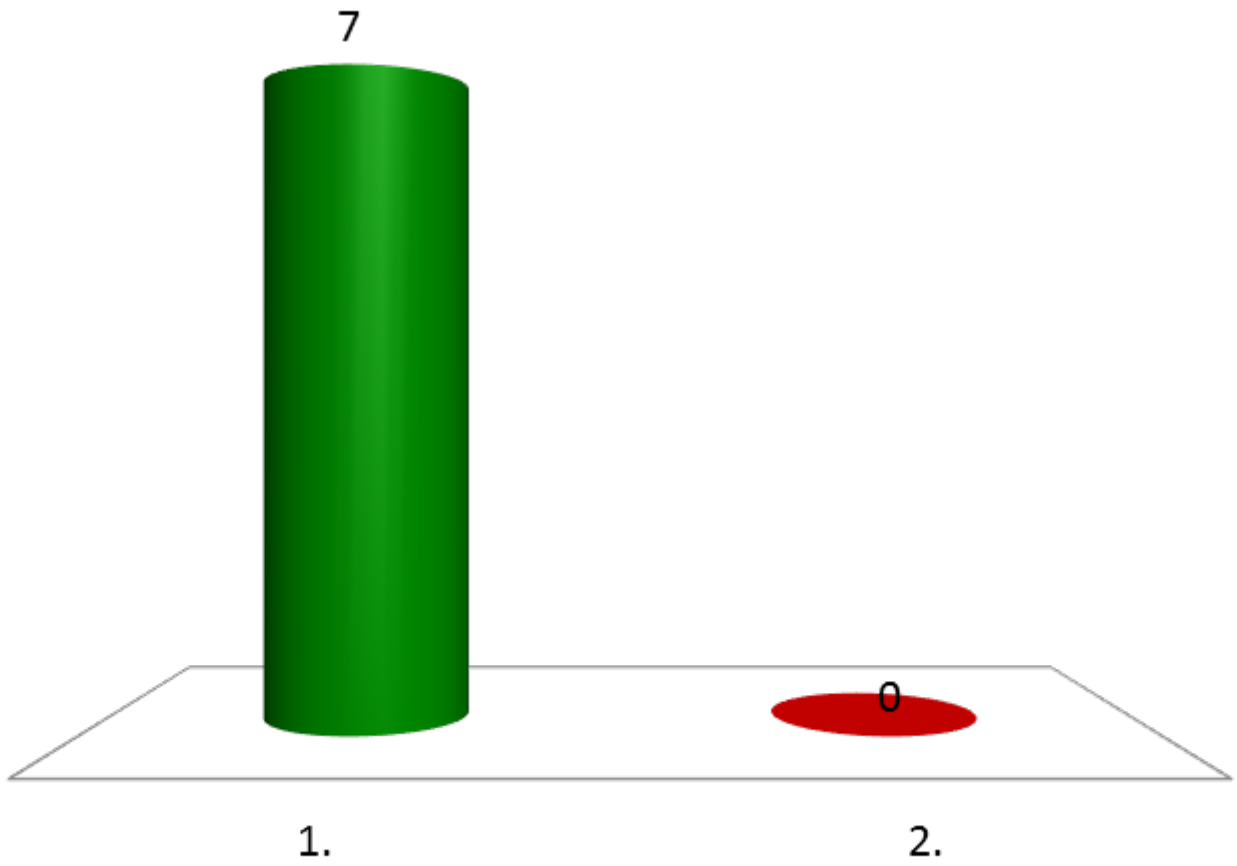
1. Probability
   of **correct responses**
   for a **sick** patient
2. Probability
   of **correct responses**
   for a **healthy** patient
3. Probability
   to be **sick**
   with a **positive diagnostic**
4. Probability
   to be **healthy**
   with a **negative diagnostic**



7 sur 7

Does the **predictive negative value** depend on the **prevalence**?

1. YES
2. NO



7 sur 7

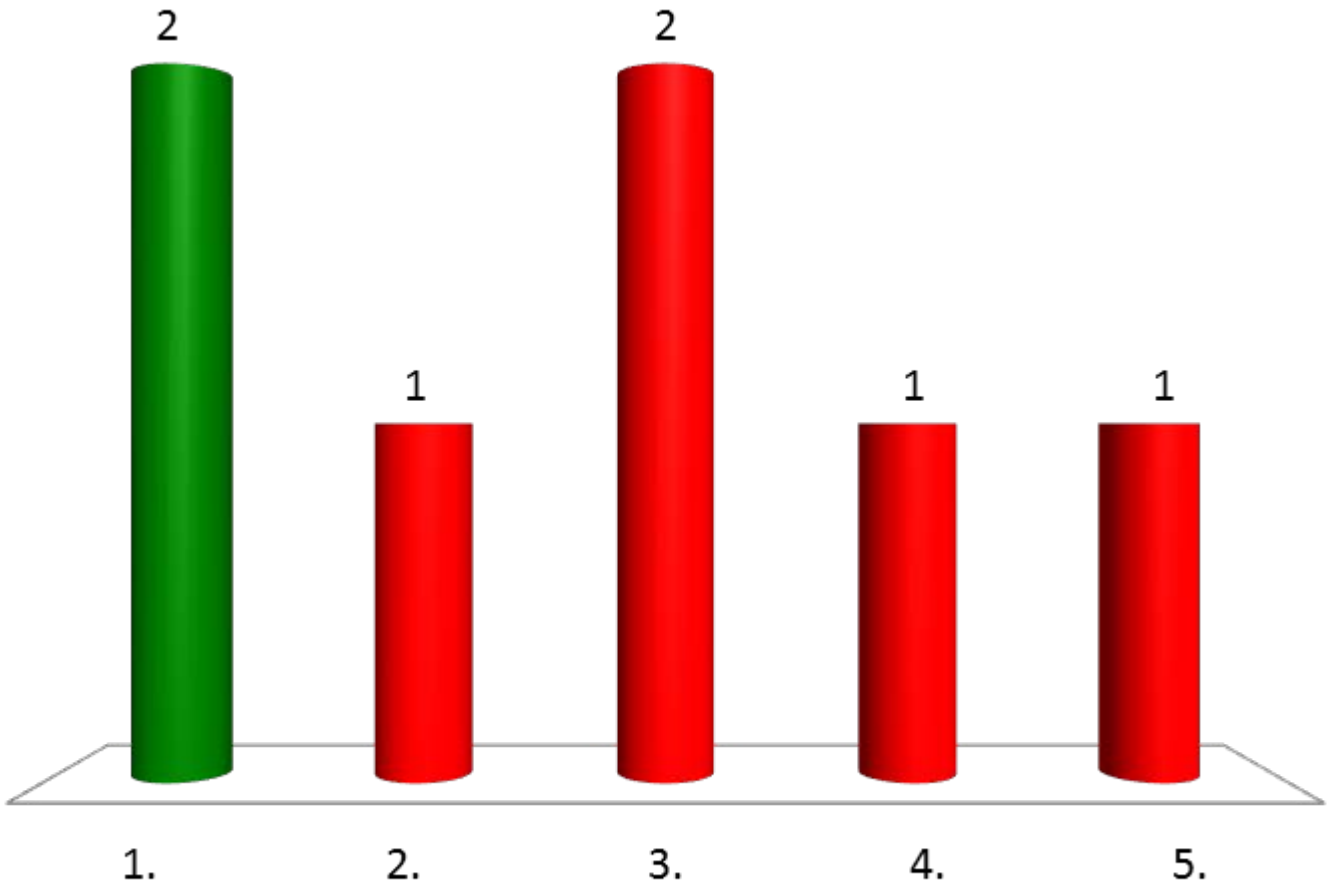
Low prevalence pathology

- A sickness has a **prevalence** of **1/10'000**
- A test has the following characteristics
  - **specificity 0.95**
  - **sensitivity 1.00**
- The test gives a **positive result**
- What is the **probability to be sick**?

Probability to be sick?

(sensitivity= 100% ; specificity = 95% ; prevalence = 0.01%)

1. less than 1%
2. 1 to 3%
3. 3 to 10%
4. 10 to 30%
5. more than 30%



7 sur 7