Diagnostic Test Evaluation and Interpretation

Prof. Dr. med. vet. Sonja Hartnack, Dipl. ECVPH

outline

- Introduction to diagnostic tests
- Analysis of diagnostic test results: Se, Sp, PPV, NPV, likelihood ratios
- Agreement with kappa
- No gold standard: Bayesian latent class modeling
- ROC curves
- STARD reporting guidelines

learning aims

- Explain and apply sensitivity, specificity, predictive values, likelihood ratios
- Correctly interpret diagnostic test results in human and animal health
- Distinguish apparent and true prevalence
- Assess test agreement with kappa
- Explain the principles of ROC curves
- Assess if a diagnostic test accuracy study complies with STARD reporting guidelines

Indeed tests are used in virtually all problem-solving activities and therefore the understanding of the principles of test evaluation and interpretation are basic to many of our activities.

Dohoo in VER 2nd ed. 2009

Feufel *et al. BMC Medical Education* (2023) 23:75
https://doi.org/10.1186/s12909-023-04025-6

BMC Medical Education

RESEARCH Open Access

Boosting for insight and/or boosting for agency? How to maximize accurate test interpretation with natural frequencies

Markus A. Feufel[1,2,3,4*], Niklas Keller[1,3,4], Friederike Kendel[1,3,4] and Claudia D. Spies[1,2,3,4]

Background

A woman with a positive mammogram will want to know: Does that mean I have cancer? Studies have repeatedly shown that a majority of physicians do not know how to calculate the probability of a disease given a positive test result, that is, the test's *positive predictive value* or PPV [15–17]. If physicians lack these basic statistical skills, evidence-based medicine (EBM) and accurate risk communication with patients remain illusory.

Psychological Science in the Public Interest
Volume 8, Issue 2, November 2007, Pages 53-96

https://doi.org/10.1111/j.1539-6053.2008.00033.x

Original Article

Helping Doctors and Patients Make Sense of Health Statistics

Gerd Gigerenzer[1,2], Wolfgang Gaissmaier[1,2], Elke Kurz-Milcke[1,2], Lisa M. Schwartz[3], and Steven Woloshin[3]

Exercise 1

Exercise 1: answer options

app.klicker.uzh.ch/join/ouw

In %

A) 0 -15

B) 35 – 50

C) 51 – 65

D) 85 – 100

E) Not possible to calculate with information given

Exercise 1

BMJ

Health professionals' and service users' interpretation of screening test results: experimental study

Ros Bramwell, Helen West and Peter Salmon

BMJ 2006;333;284; originally published online 13 Jul 2006;
doi:10.1136/bmj.38884.663102.AE

Exercise 1

prevalence

sensitivity

specificity

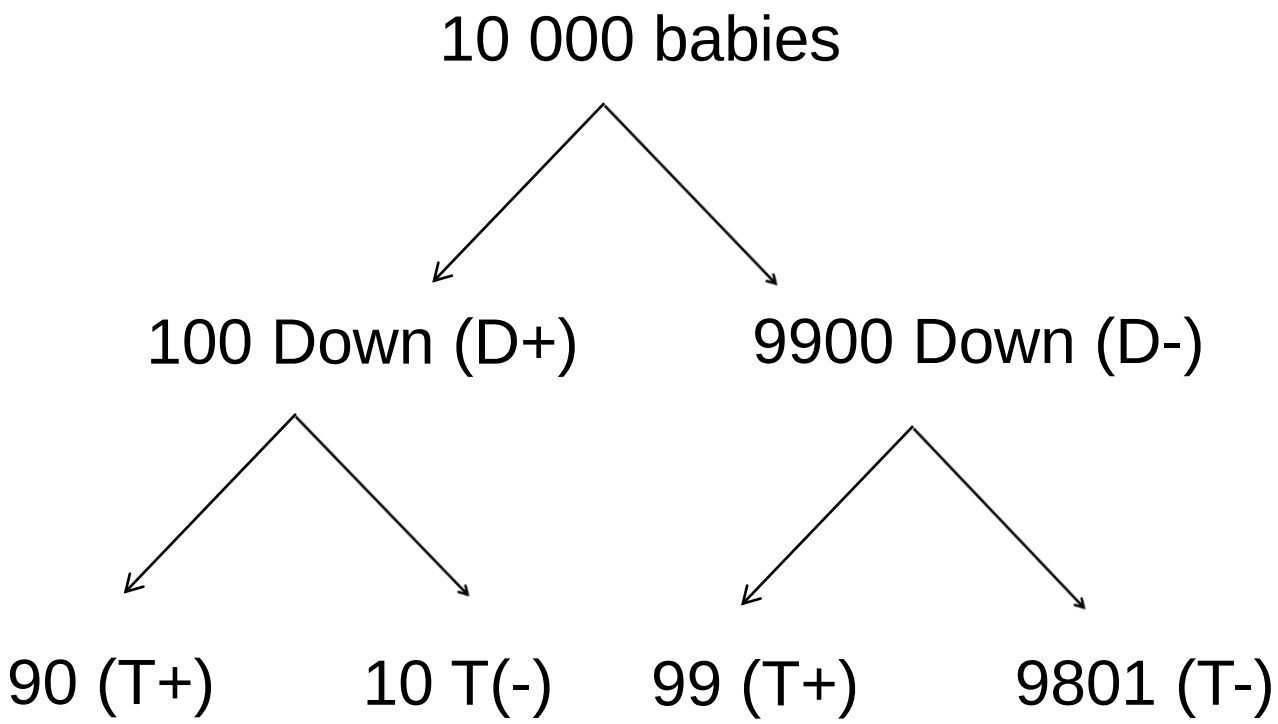
The serum test screens pregnant women for babies with Down‘s syndrome. The test is a very good one, but not perfect. Roughly 1% of babies have Down‘s syndrome. If the baby has Down‘s syndrome, there is a 90% chance that the result will be positive. If the baby is unaffected, there is still a 1% chance that the result will be positive.

A pregnant woman has been tested and the result is positive. What is the chance that her baby actually has Down‘s syndrome?

.....%?

positive predictive value

how to solve it



$$p(Down+|T+) = \frac{90}{90+99}$$

Objective To investigate the accuracy of interpretation of probabilistic screening information by different stakeholder groups and whether presentation as frequencies improves accuracy.

Participants 43 pregnant women attending their first antenatal appointment in a regional maternity service; 40 companions accompanying the women to their appointments; 42 midwives; 41 obstetricians. Participation rates were 56%, 48%, 89%, and 71% respectively.

Results Most responses (86%) were incorrect. Obstetricians gave significantly more correct answers (although still only 43%) than either midwives (0%) or pregnant women (9%). Overall, the proportion of correct answers was higher for presentation as frequencies (24%) than for presentation as percentages (6%), but further analysis showed that this difference occurred only in responses from obstetricians. Many health professionals were confident in their incorrect responses.
Conclusions Most stakeholders in pregnancy screening draw incorrect inferences from probabilistic information, and health professionals need to be aware of the difficulties that both they and their patients have with such information. Moreover, they should be aware that different people make different mistakes and that ways of conveying information that help some people will not help others.

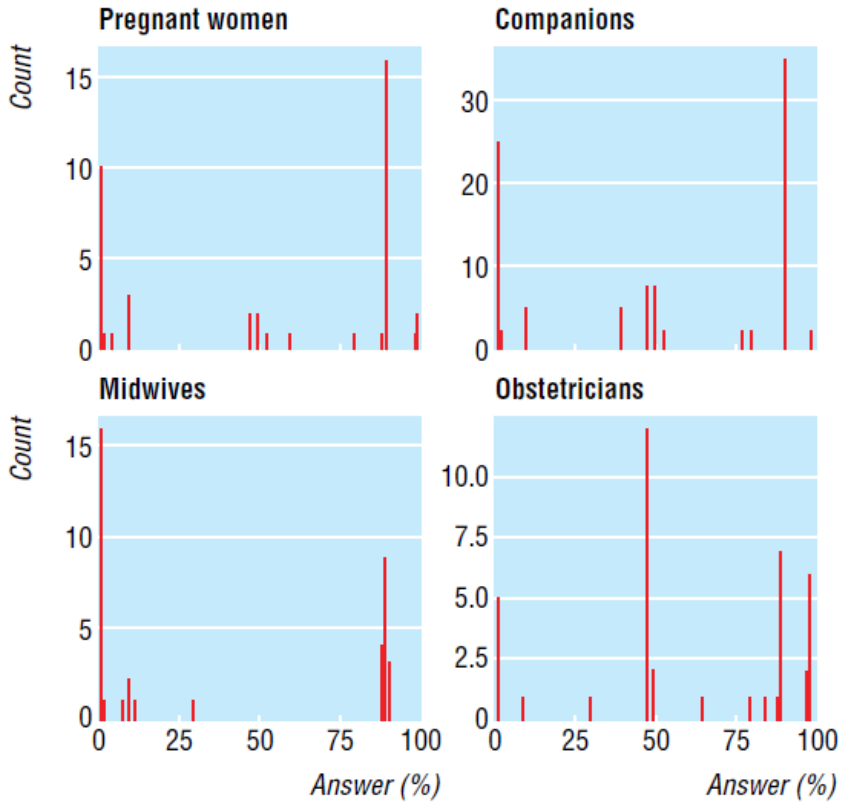


Fig 1 Distribution of responses from the four participant groups. X axis is response to scenario expressed as percentage (width interval=0.5%); y axis is number of responses

doi:10.1136/bmj.38884.663102.AE

diagnostic test

- Laboratory tests

  ELISA, PCR, Western blot, microscopy, Sorbitol fermentation, parasite egg counts…..

- Clinical investigation

  palpation, crepitation, temperature, jugular pulse, reflex, ….

- Diagnostic imaging

  radiography, ultrasound, CT,…

- Questionnaire

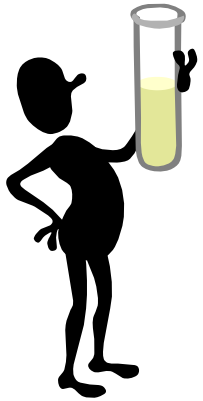
…

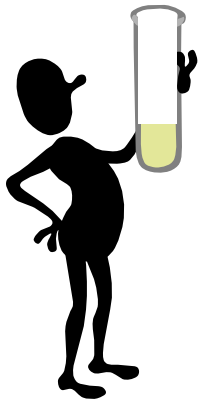
a test or testing

- is any device or process designed to detect or quantify, a sign, substance, tissue change, or body response in an individual
- test results may be categorical/qualitative (e.g. +/-) or quantitative (ordinal or continous scale)
- may lead to positive, negative or inconclusive results
- could be applied at the individual, the herd or other level of aggregation
- could be used in a single or multiple forms (serial and parallel)

analytical test performance

- analytical sensitivity





- 10 copy numbers
- 3 ppb Penicillin
- 1 larvae in 1 g muscle

„lower limit of detection“

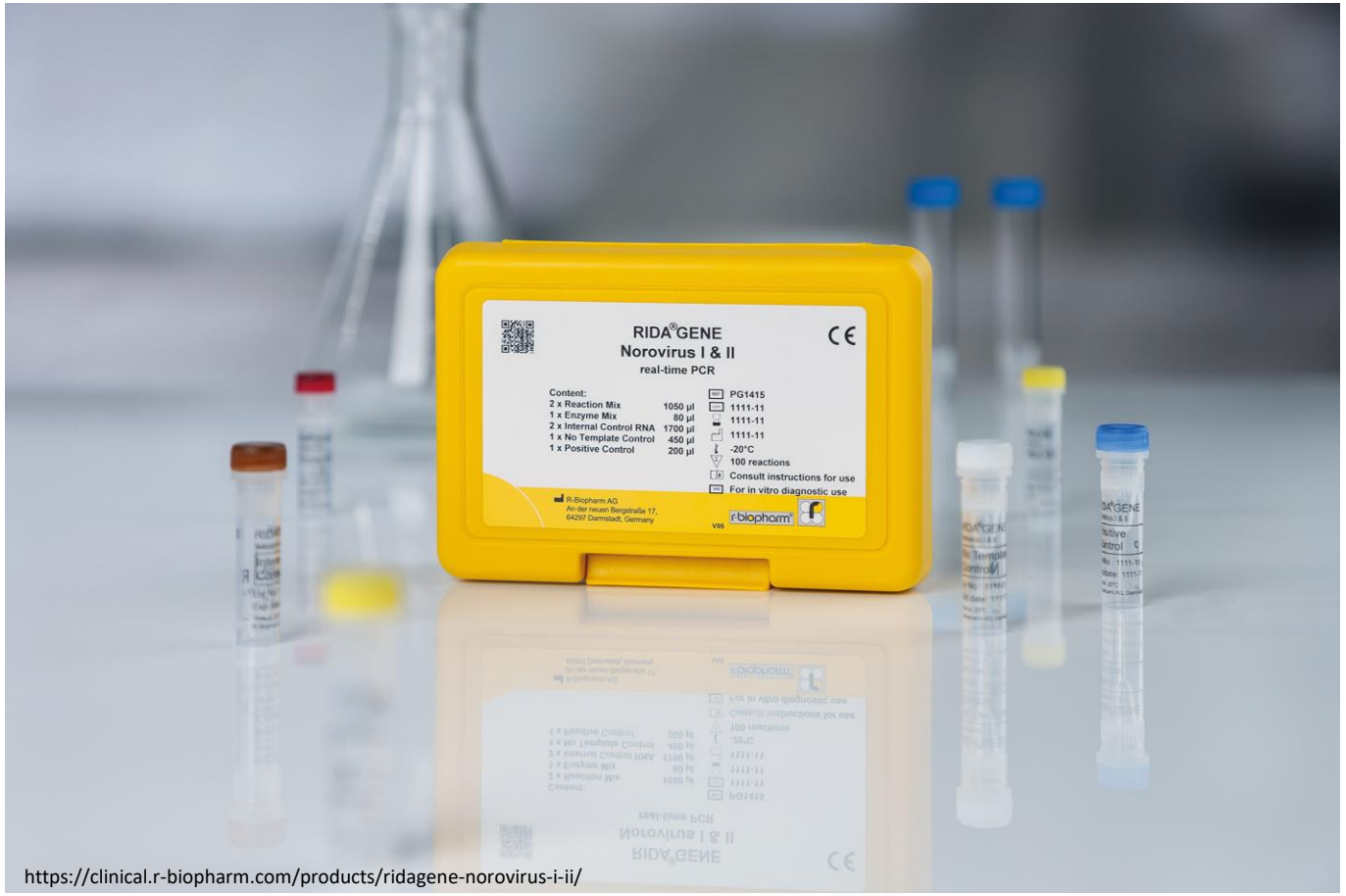
ring trials

- analytical specificity

- H7N1, not H7N7
- CSFV, no other Pestivirus
- PCV2, not PCV1
- bone AP, not liver iso-E
- *M. bovis*, not *M. avium*

the lab view…..

analytical test performance



https://clinical.r-biopharm.com/products/ridagene-norovirus-i-ii/

13.1 Analytical sensitivity

The RIDA®GENE Norovirus I & II multiplex real-time RT-PCR has a limit of detection of ≥ 50 RNA copies per reaction.

13.2 Analytical specificity

The RIDA®GENE Norovirus I & II multiplex real-time RT-PCR is specific for Norovirus of the genogroups I and II from human stool samples. No cross-reaction could be detected for the following species (see Tab. 10):

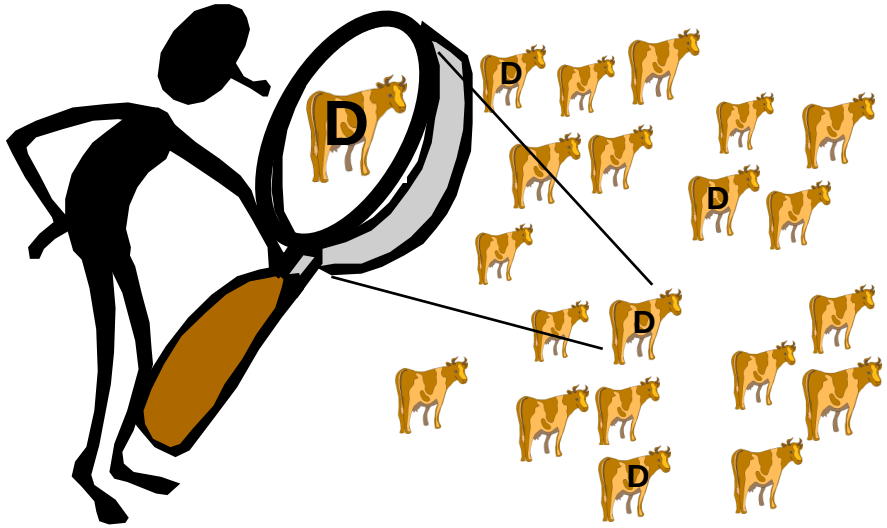
Tab. 10: Cross-reactivity testing

Adenovirus	-	*Candida albicans*	-	*Proteus vulgaris*	-
Aeromonas hydrophila	-	*Citrobacter freundii*	-	*Pseudomonas aeruginosa*	-
Arcobacter butzleri	-	*Clostridium difficile*	-	Rotavirus	-
Astrovirus	-	*Clostridium perfringens*	-	*Salmonella enteritidis*	-
Bacillus cereus	-	*Clostridium sordellii*	-	*Salmonella typhimurium*	-
Bacteroides fragilis	-	*E. coli* (O6)	-	*Serratia liquefaciens*	-
Campylobacter coli	-	*E. coli* (O26:H-)	-	*Shigella flexneri*	-
Campylobacter fetus subsp. *fetus*	-	*E. coli* (O157:H7)	-	*Staphylococcus aureus*	-
Campylobacter jejuni	-	*Enterobacter cloacae*	-	*Staphylococcus epidermidis*	-
Campylobacter lari subsp. *lari*	-	*Enterococcus faecalis*	-	*Vibrio parahaemolyticus*	-
Campylobacter upsaliensis	-	*Klebsiella oxytoca*	-	*Yersinia enterocolitica*	-

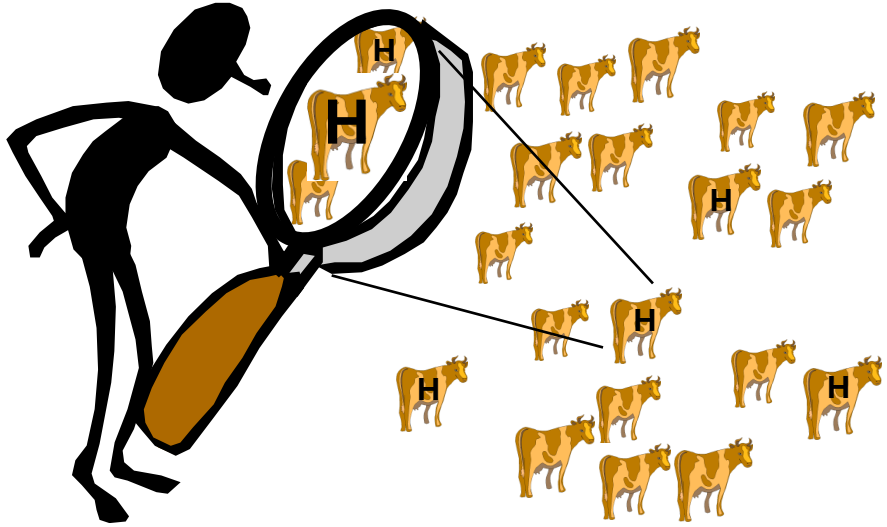
https://clinical.r-biopharm.com/wp-content/uploads/2012/06/pg1415_ridagene_norovirusiii_2021-01-28_en.pdf

diagnostic test performance

- diagnostic sensitivity



- diagnostic specificity



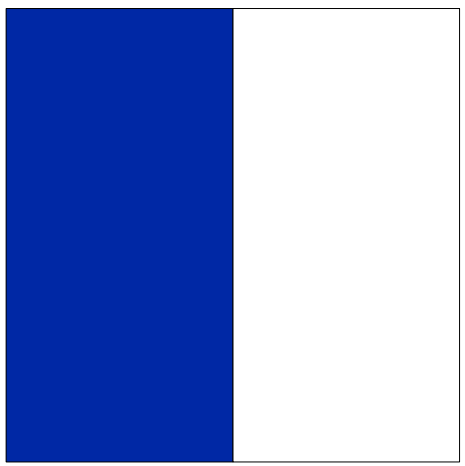
the field view…..

diagnostic test performance

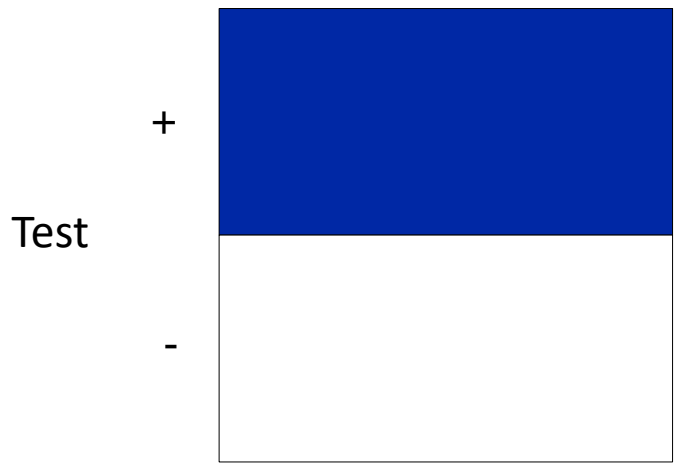
Diagnostic (epidemiologic) Se and Sp depend (in part) on analytic Se and Sp, but are distinctly different concepts.

Status

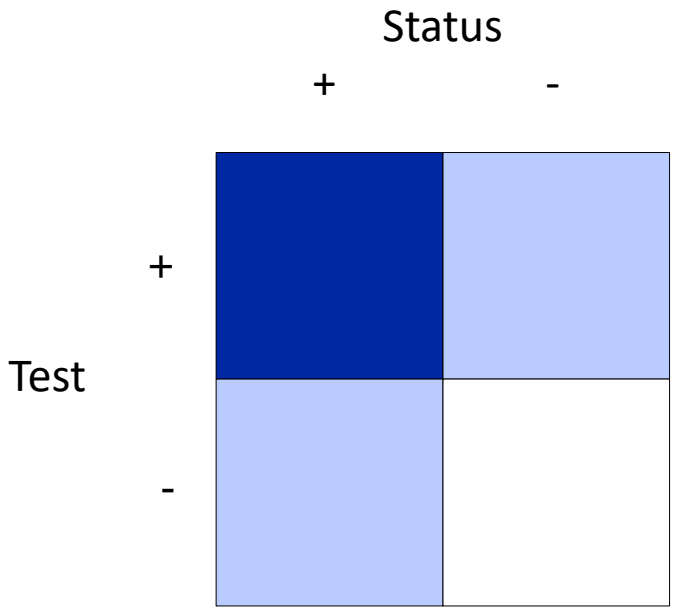
\+ -



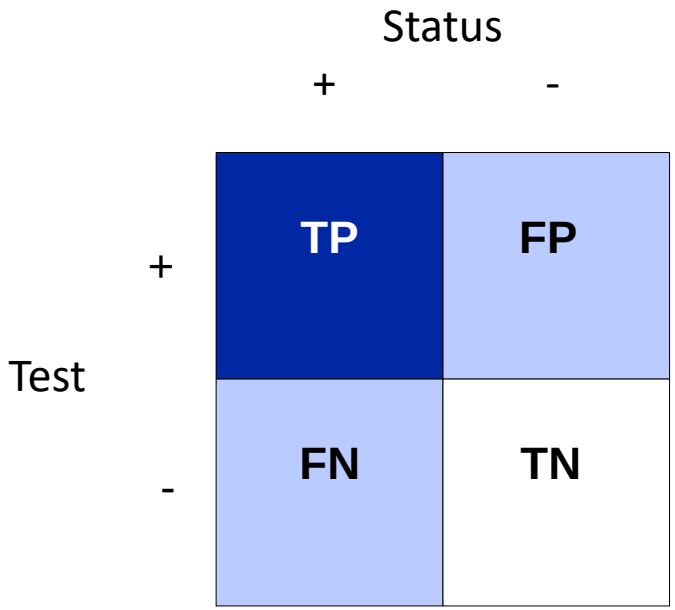
assumption, that individuals are either + or -


Test
+
-

2x2 or contingency table



2x2 or contingency table



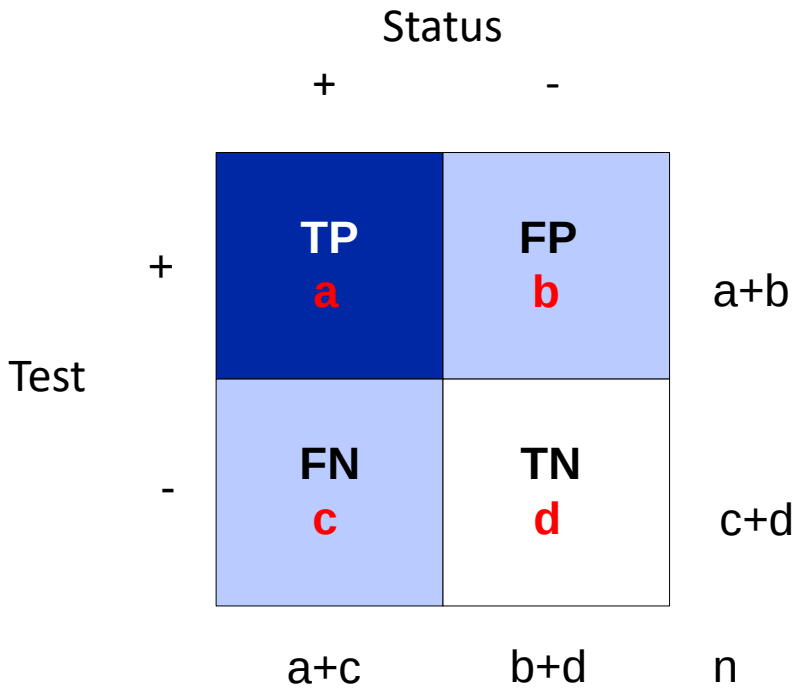
TP: true positive

FP: false positive

FN: false negative

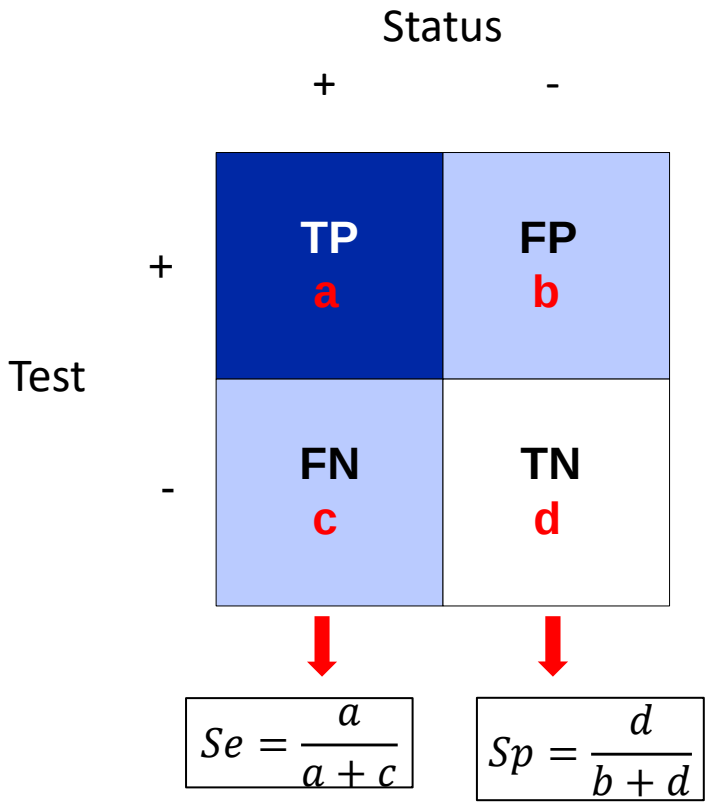
TN: true negative

2x2 or contingency table



a, b, c, d assigned by convention

sensitivity and specificity



Se: diagnostic (epidemiologic) sensitivity
Sp: diagnostic (epidemiologic) specificity

sensitivity and specificity

Se:

ability of the test to classify diseased animals correctly
Se=P(T+|D+) conditional probability

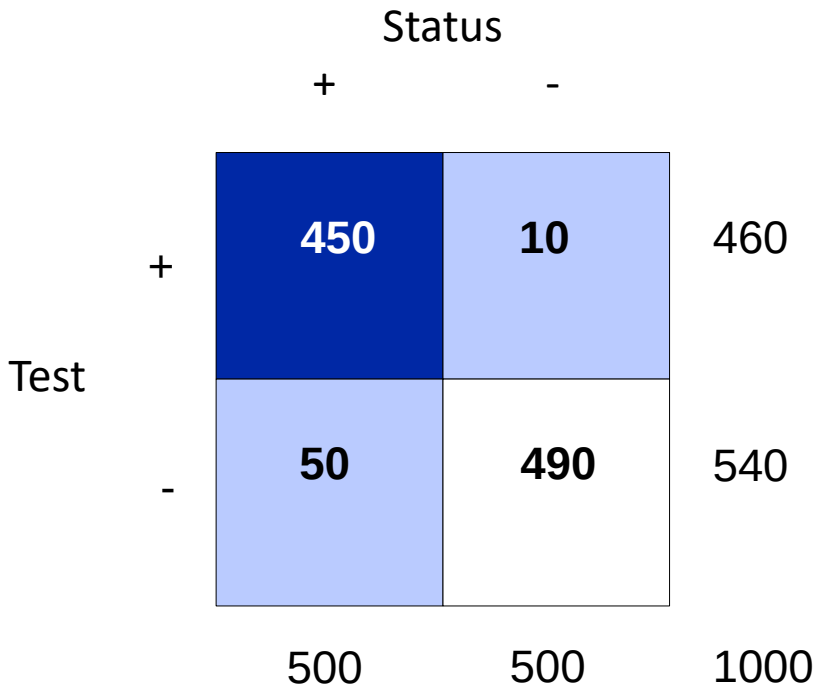
Sp:

ability of the test to classify non-diseased animals correctly
Sp=P(T-|D-) conditional probability

		Status +	Status -
Test	+	**Pr*Se**	**(1-Pr)*(1-Sp)**
	-	**Pr*(1-Se)**	**(1-Pr)*Sp**

Pr = prevalence

sensitivity and specificity

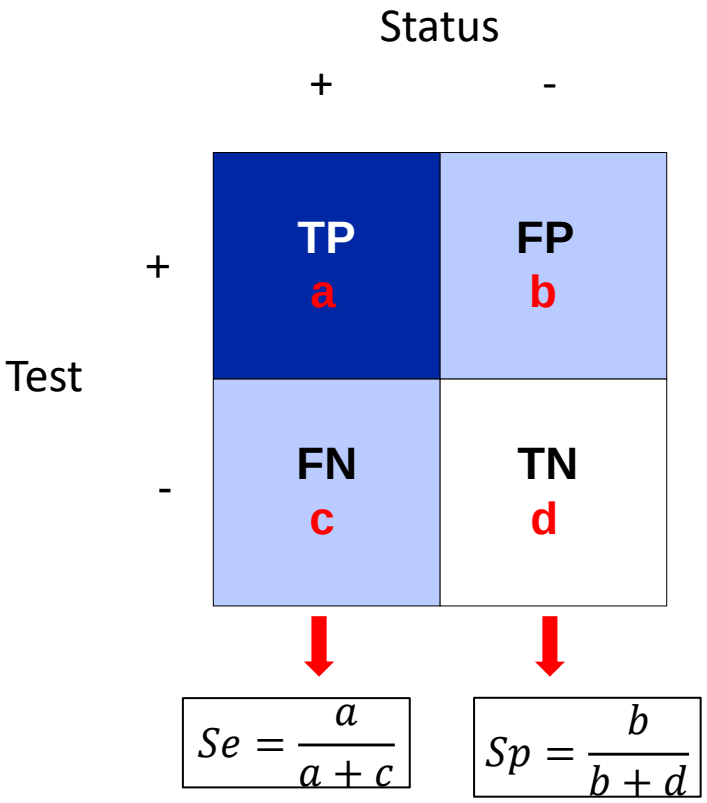


Sensitivity: 450 / 500 = 90%

Specificity: 490 / 500 = 98%

confidence intervals

$$\widehat{parameter} \pm 1.96 * SE(\widehat{parameter})$$

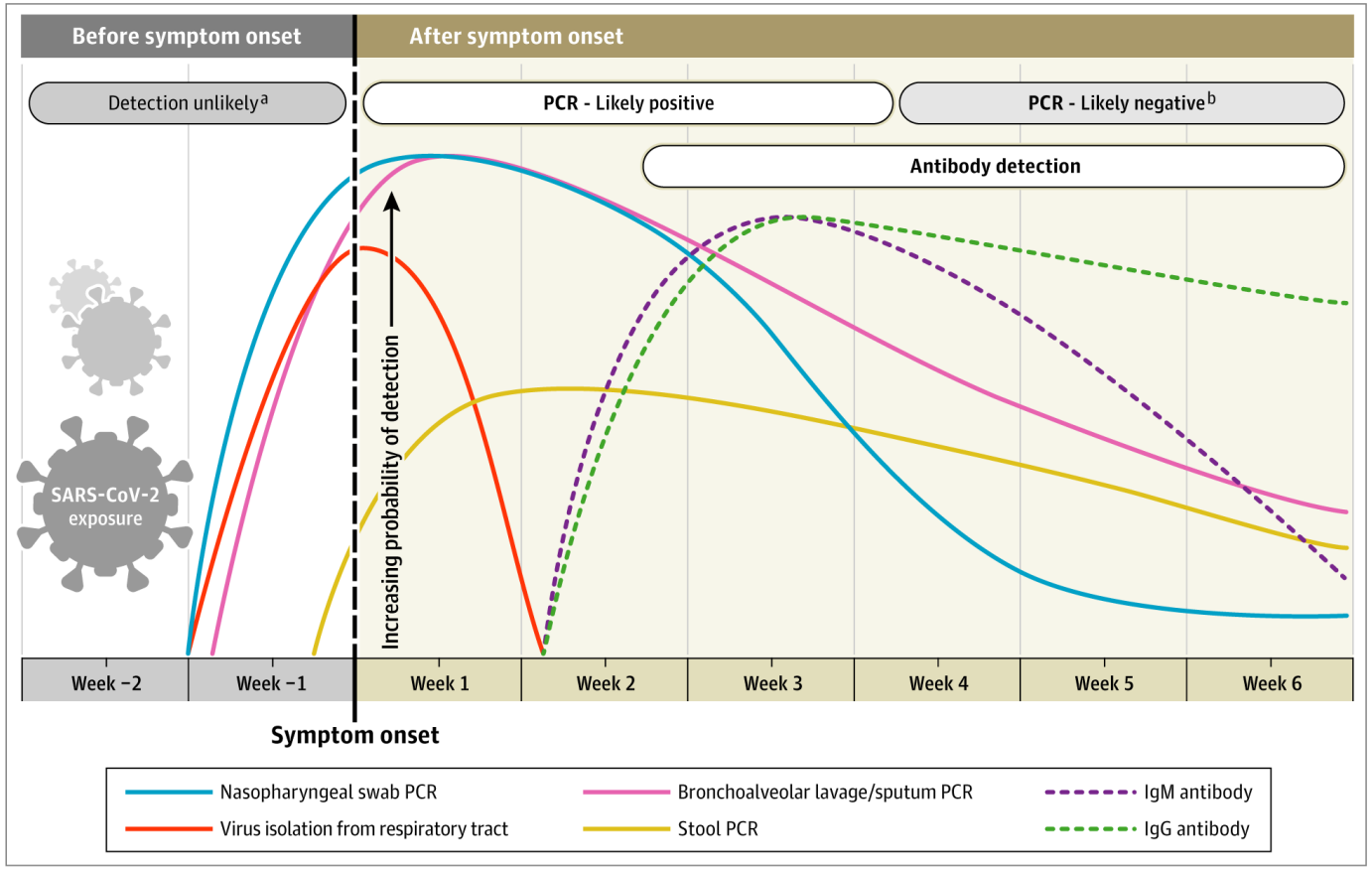



```
library(DescTools)

# For s pecificity of 98.99% (98/99)
BinomCI(98, 99, conf.level = 0.95, method = c("wilson",
                                              "wald",
                                              "jeffreys",
                                              "clopper-pearson"))

##                      est    lwr.ci    upr.ci
## wilson           0.989899 0.9449847 0.9982147
## wald             0.989899 0.9702016 1.0000000
## jeffreys         0.989899 0.9537684 0.9989080
## clopper-pearson 0.989899 0.9450032 0.9997443
```

some potential reasons for + and – serology results

Positive test results	
TP	present (or past) infection
FP	cross-reaction
FP	non-specific reactions
FP	….

Negative test results	
TN	absence of infection
FN	improper timing
FN	non-specific inhibitors
FN	immuno-tolerant (PI in BVDV)
FN	….



diagnosis of classical swine fever

individual pig													
Ab test													
virus isolation													
pathology													
clinical symptoms													
➢ 39.5°C													
d.p.i.	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	.....

factors influencing Se and Sp

- Biological factors
  - Stage of disease (Se) → prevalence-related
  - Presence of cross-reactivity (Sp)
  - Variation of covariate factors (subpopulations)
    - Estimate Se and Sp in subpopulations

- Variation in test implementation
  - Technical variation (lab, time, etc..)
  - Handling of "?" results
  - Choice of cut-off

- Biases in diagnostic studies (internal validity)

Se and Sp

Both specificity and sensitivity may change as the same diagnostic

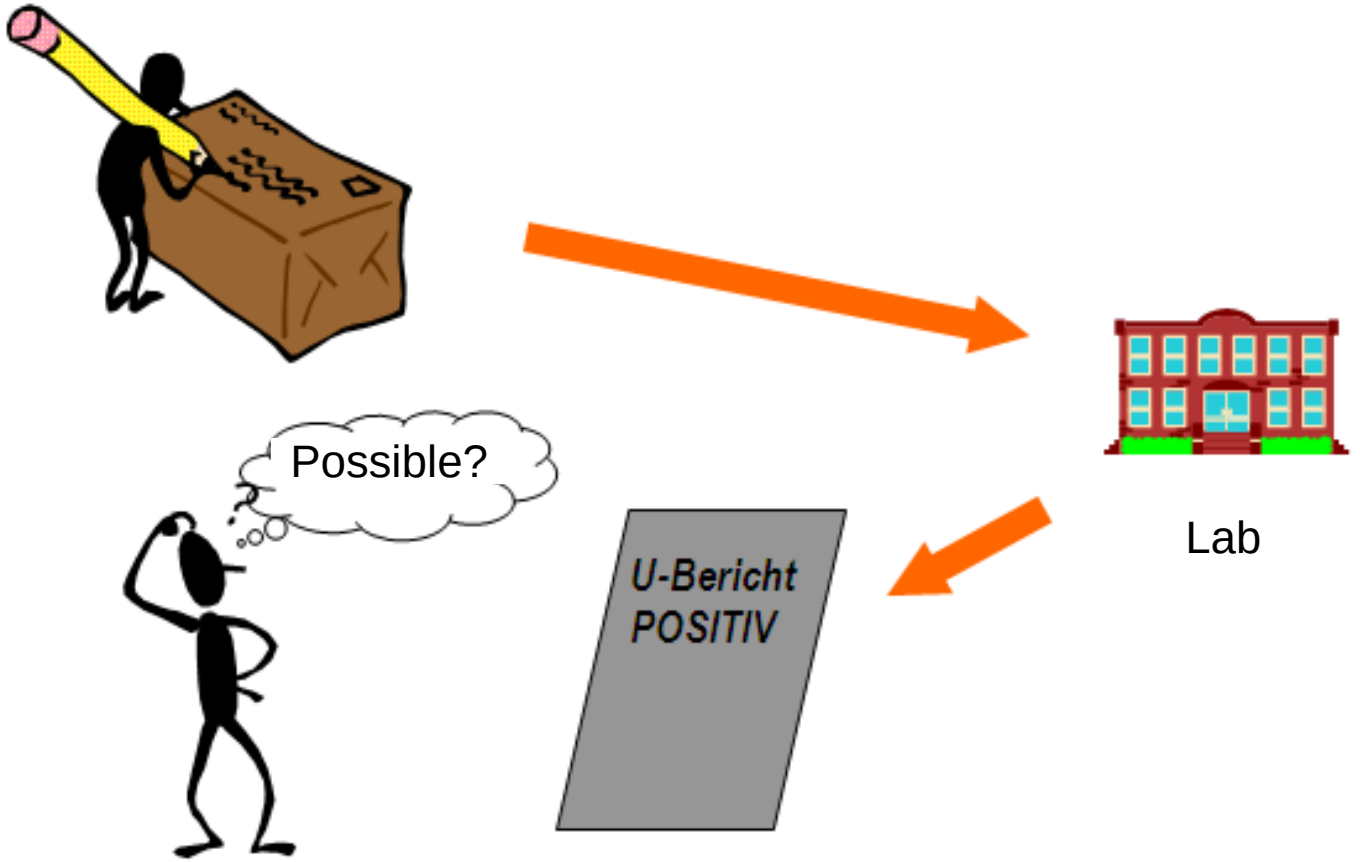
is applied in primary, secondary, and tertiary health care.

Sackett & Haynes, BMJ 2002

Some traditional textbooks on diagnostic testing still refer to the test sensitivity and specificity as values that are intrinsic to the diagnostic test. Our own experience (and that of others) indicate that both test sensitivity and specificity vary with external factors.

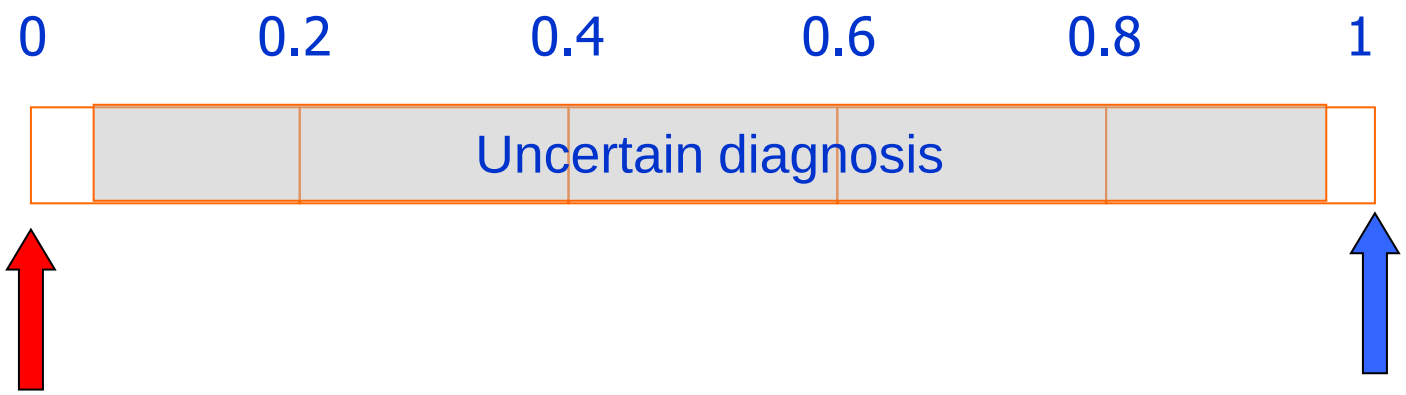
Berkvens et al., Epidemiology 2006

you are familiar with this?



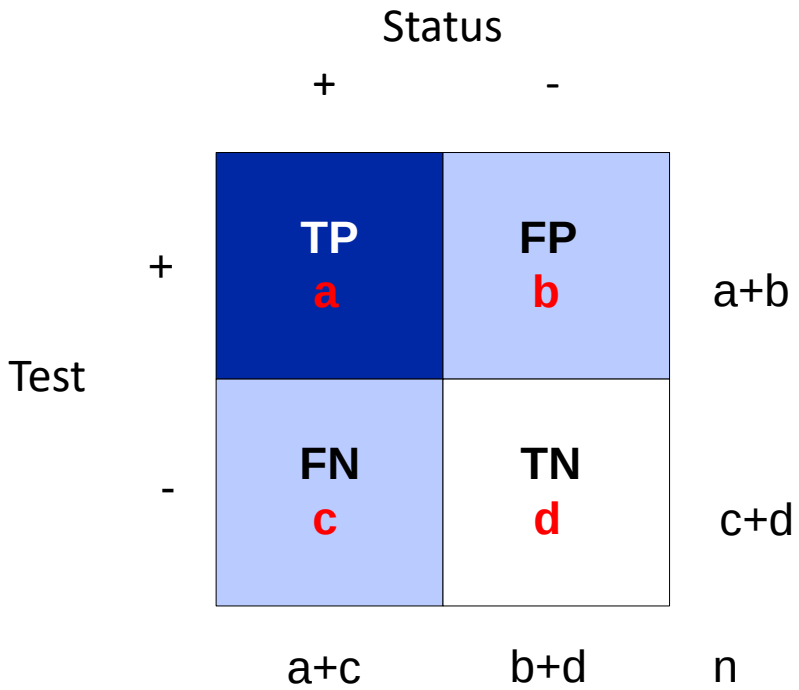
Given the known clinical information or test result about the animal ...

... what is the probability of disease ?



Confidence that the disease does NOT occur

Confidence that the disease DOES occur



If you want to know, how confident you may be in a positive or a negative test result ….

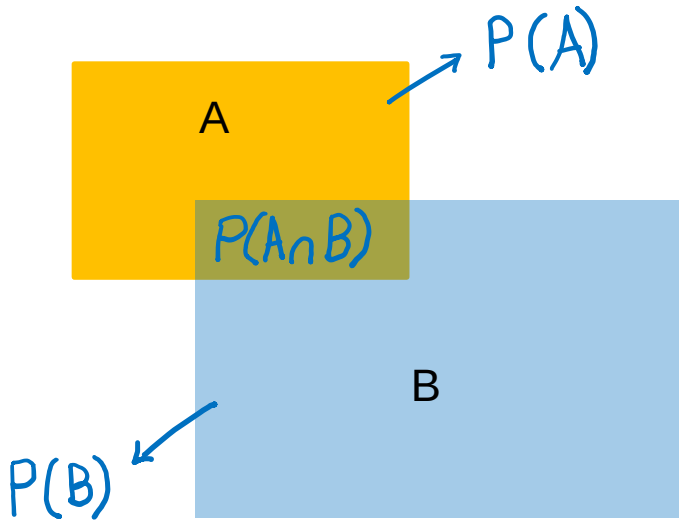
LII. *An Essay towards solving a Problem in the Doctrine of Chances. By the late Rev. Mr.* Bayes, *communicated by Mr.* Price, *in a letter to* John Canton, *M. A. and F. R. S.*

Dear Sir,

Read Dec. 23, 1763. I now send you an essay which I have found among the papers of our deceased friend Mr. Bayes, and which, in my opinion, has great merit, and well deserves to be preserved. Experimental philosophy, you will find, is nearly interested in the subject of it; and on this account there seems to be particular reason for thinking that a communication of it to the Royal Society cannot be improper.

[*Philosophical Transactions of the Royal Society of London* **53** (1763), 370–418.]

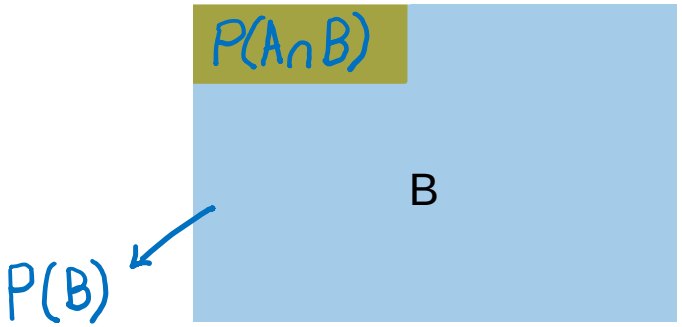
conditional probabilities



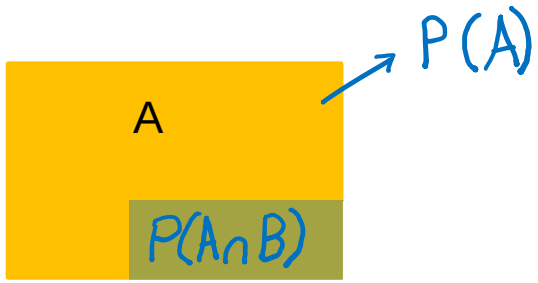
conditional probabilities

$P(A)$

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



conditional probabilities



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

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

conditional probabilities

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

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

$$P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$

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

conditional probabilities

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

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

$$P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$$

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

Bayes Theorem

Thomas Bayes
1702-1761

Bayes theorem

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

PPV — Se — Pr

$$P(D^+|T^+) = \frac{P(T^+|D^+) * P(D^+)}{P(T^+)}$$

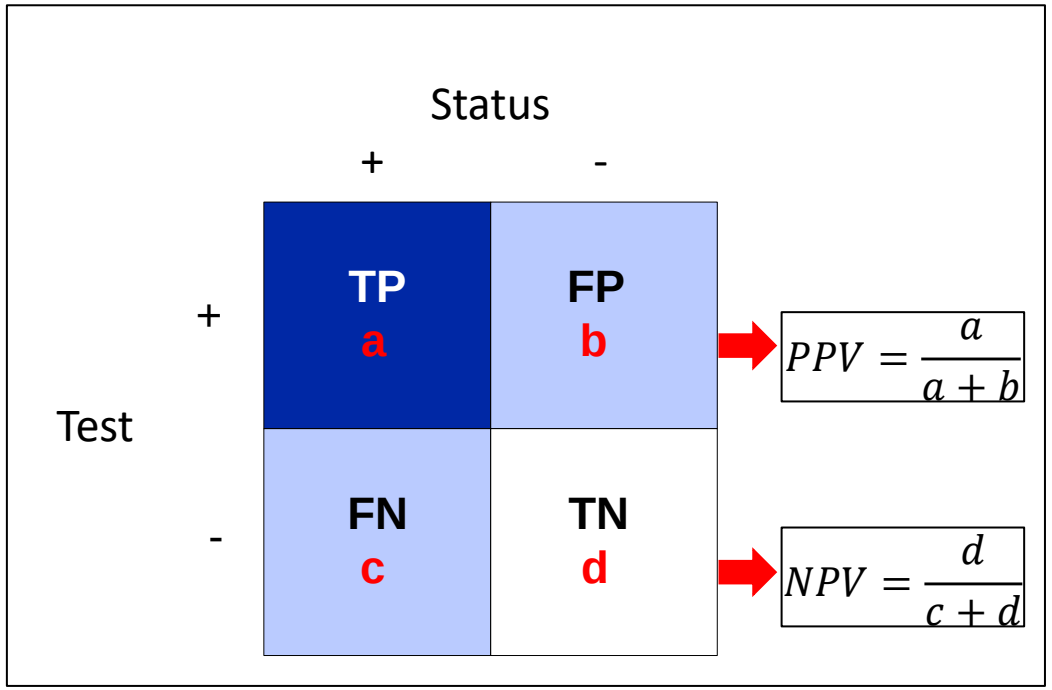
$$P(T^+|D^+) \cdot P(D^+) + P(T^+|D^-) \cdot P(D^-)$$

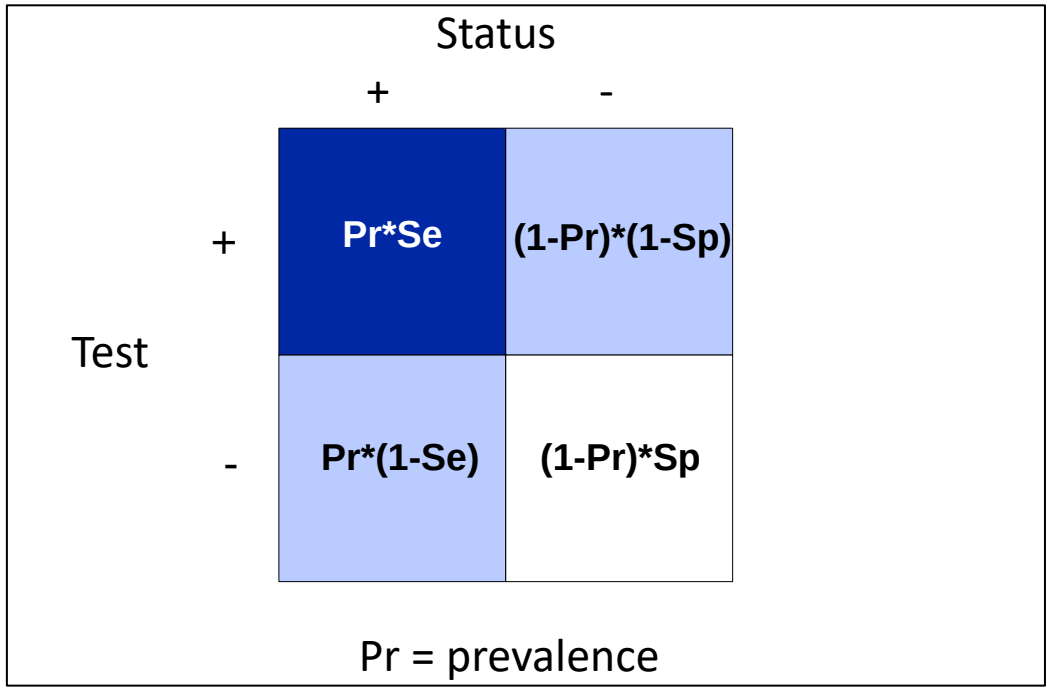
$Se \cdot Pr$ — RP

$1-Sp$ $\quad 1-Pr$ — FP

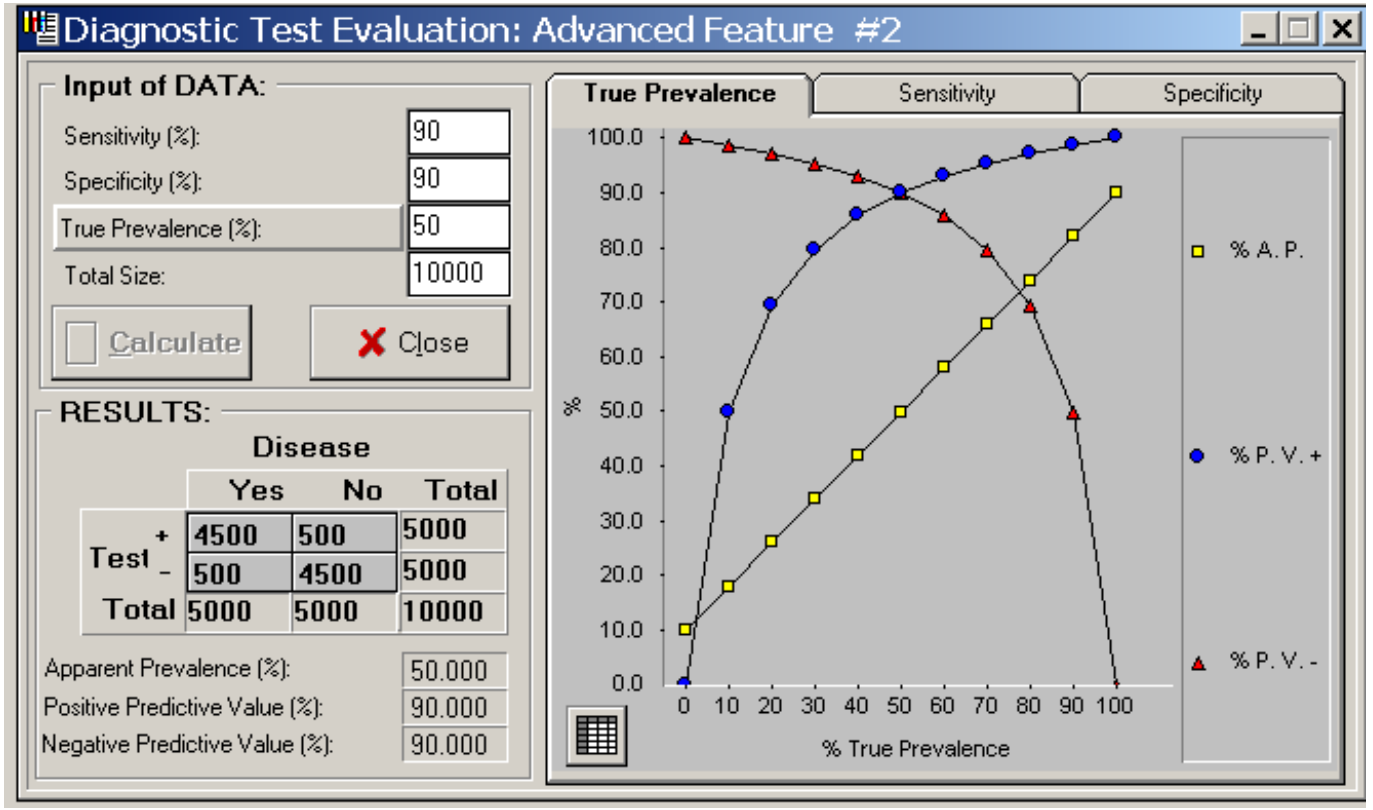
$$PPV = \frac{Pr * Se}{Pr * Se + (1 - Pr) * (1 - Sp)}$$

$$NPV = \frac{(1 - Pr) * Sp}{(1 - Pr) * Sp + Pr * (1 - Se)}$$





predictive values: influence of prevalence



- The more probable it is to find a diseased animal, the more confident you will be in a „+“ test result.
- The more probable it is to find a non-diseased healthy animal, the more confident you will be in a „-“ test result.

Why clinicians are natural bayesians

Christopher J Gill, Lora Sabin, Christopher H Schmid

Thought you didn't understand bayesian statistics? Read on and find out why doctors are expert in applying the theory, whether they realise it or not

BMJ 2005, 330: 1080-3

Every part of clinical history and examination can be viewed as a diagnostic test

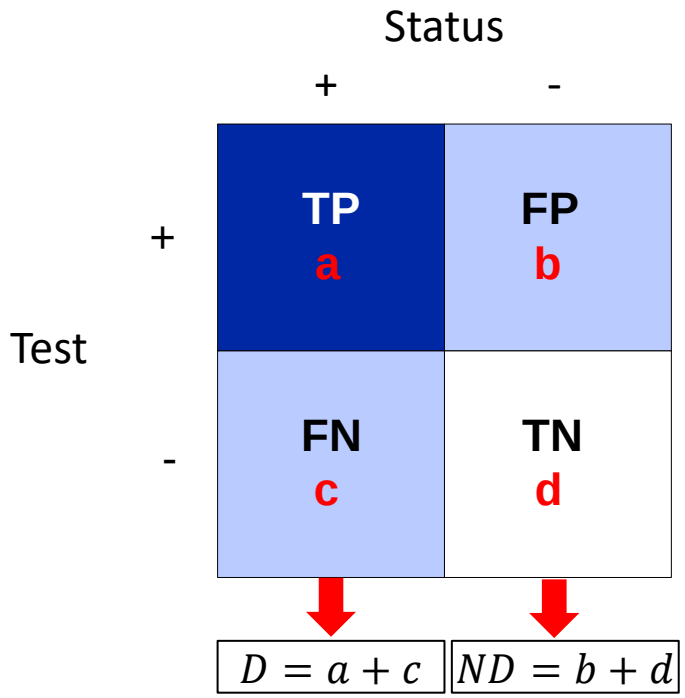
Exercise 2

likelihood ratios: LR+ and LR-

- The likelihood ratios express how much more frequent the respective result is among subjects with disease than subjects without disease.

$$LR+ = \frac{\frac{a}{D}}{\frac{b}{ND}} = \frac{Se}{1 - Sp}$$

$$LR- = \frac{\frac{c}{D}}{\frac{d}{ND}} = \frac{1 - Se}{Sp}$$



- The likelihood ratios do not depend on prevalence.
- LRs may be seen as a measure of accuracy.

diagnostic odds ratio (DOR)

$$DOR = \frac{LR+}{LR-} = \frac{a \times d}{b \times c}$$

$$SE\left(\ln DOR\right) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

- The DOR expresses how much greater the odds of having the disease are for individuals with a positive test results than for those with a negative test result.
- It is a single measure of diagnostic test performance that combines both likelihood ratios.
- Proposed for meta-analysis.
- It gives equal weight to both false positives and false negatives.

example review on diagnostic tests: tb

Ante mortem diagnosis of tuberculosis in cattle:
A review of the tuberculin tests, γ-interferon assay
and other ancillary diagnostic techniques

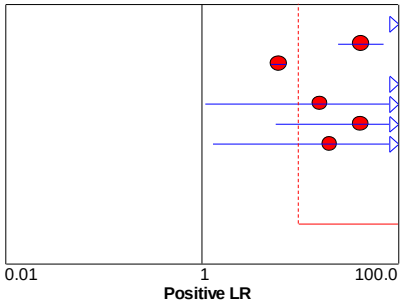
R. de la Rua-Domenech [a,*], A.T. Goodchild [b], H.M. Vordermeier [c], R.G. Hewinson [c],
K.H. Christiansen [b], R.S. Clifton-Hadley [d]

Research in Veterinary Science 81 (2006) 190–210

Table 3
Summary of the results of trials conducted to estimate the sensitivity and specificity of the single intradermal comparative cervical tuberculin (SICCT) test in cattle, in chronological order of publication

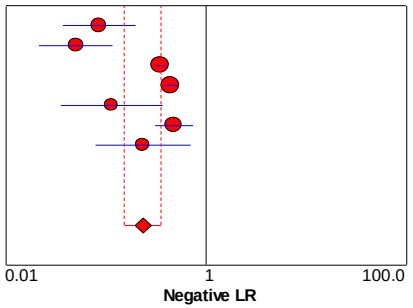
Bovine PPD	Avian PPD	Number of animals tested (of which necropsied)	Apparent proportion diseased (%)	Test interpretation	Sensitivity (%)	Specificity (%)	References (country)
				Used in series, to re-test suspects within 7 days of a caudal fold SIT			Roswurm and Konya (1973) (cited by Norby et al. (2004))
				Standard	74.4		
				Severe	88.5		
0.2 mg	0.05 mg	1110 (0) in 25 TB-free herds	0.0	Details not given	Not evaluated	100	Lesslie et al. (1975) (SE of England)
0.1 mg	0.05 mg	10305 (88)	0.56	Standard	91.4	99.9	Lesslie et al. (1975) (GB)
				Severe	100		
0.1 mg	0.05 mg	671 (316)[b] 500 (107)[c]	15.0	Standard	95.5	97.8	O'Reilly and MacClancy (1975)
Various, inc. human PPD	Various	1425 (1425)	36.9	Severe	68.6	88.8	As cited by Francis et al. (1978)[a]
0.1 mg	0.05 mg	270 (68)	25.2	Standard	75.0		O'Reilly (1986)
				Severe	94.1		
0.1 mg	0.05 mg	2799(192?)[b] 1396 (0)[c]	6.4	Standard	55.1	100	Neill et al. (1994b) (Northern Ireland)
				Severe			
0.1 mg	0.05 mg	18 (18)	83.3	Standard	93.3	Not evaluated	Doherty et al. (1995b) (Ireland)
				Severe	100		
0.1 mg	0.05 mg	2528 (2528)	14.0	Standard	90.9	Not evaluated	Costello et al. (1997) (Ireland)
2000 IU	2500 IU	30 zebu oxen (30)	73.3	Standard	90.9	100	Ameni et al. (2000) (Ethiopia)
5000 IU	2500 IU	84 (84) culture-negative cattle, reactors to a caudal fold test, from 21 herds with no history of TB	0.0	Standard	Not evaluated	94	Buddle et al. (2001) (New Zealand)
2000 IU	2500 IU	100 (100) zebu cattle ("moderate prevalence" group)	21.0	Subjective assessment (palpation of injection sites and observation of clinical signs)	52.0	99.0	Quirin et al. (2001) (Madagascar)
		22 (22) ("high prevalence" group)	45.5		80.0	100	
0.1 mg	0.1 mg	494 (494) in 7 herds		Used as a serial test, to			Norby et al. (2004) (USA)
			1.1–5.5 ("low	re-test suspects to a	75.0	Not evaluated	

example review on diagnostic tests: tb



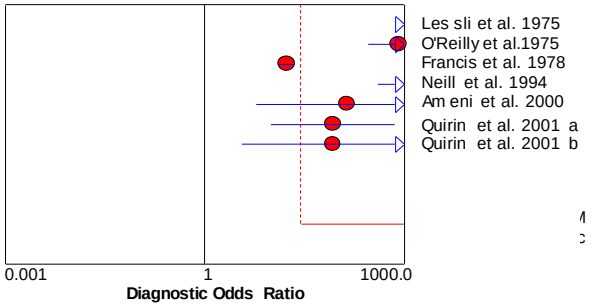
	Positive LR (95% CI)
Les sli et al. 1975	936.36 (501.44 - 1'748.51)
O'Reilly et al.1975	41.75 (24.35 - 71.58)
Francis et al. 1978	6.10 (5.03 - 7.39)
Neill et al. 1994	2'897.66 (180.73 - 46'458.94)
Ameni et al. 2000	16.04 (1.08 - 238.13)
Quirin et al. 2001 a	41.38 (5.66 - 302.65)
Quirin et al. 2001 b	20.09 (1.30 - 310.16)

Random Effects Model
Pooled Positive LR = 68.69 (9.76 to 483.55)
Cochran-Q = 271.53; df = 6 (p = 0.0000)
Inconsistency (I-square) = 97.8 %
Tau-squared = 6.0399

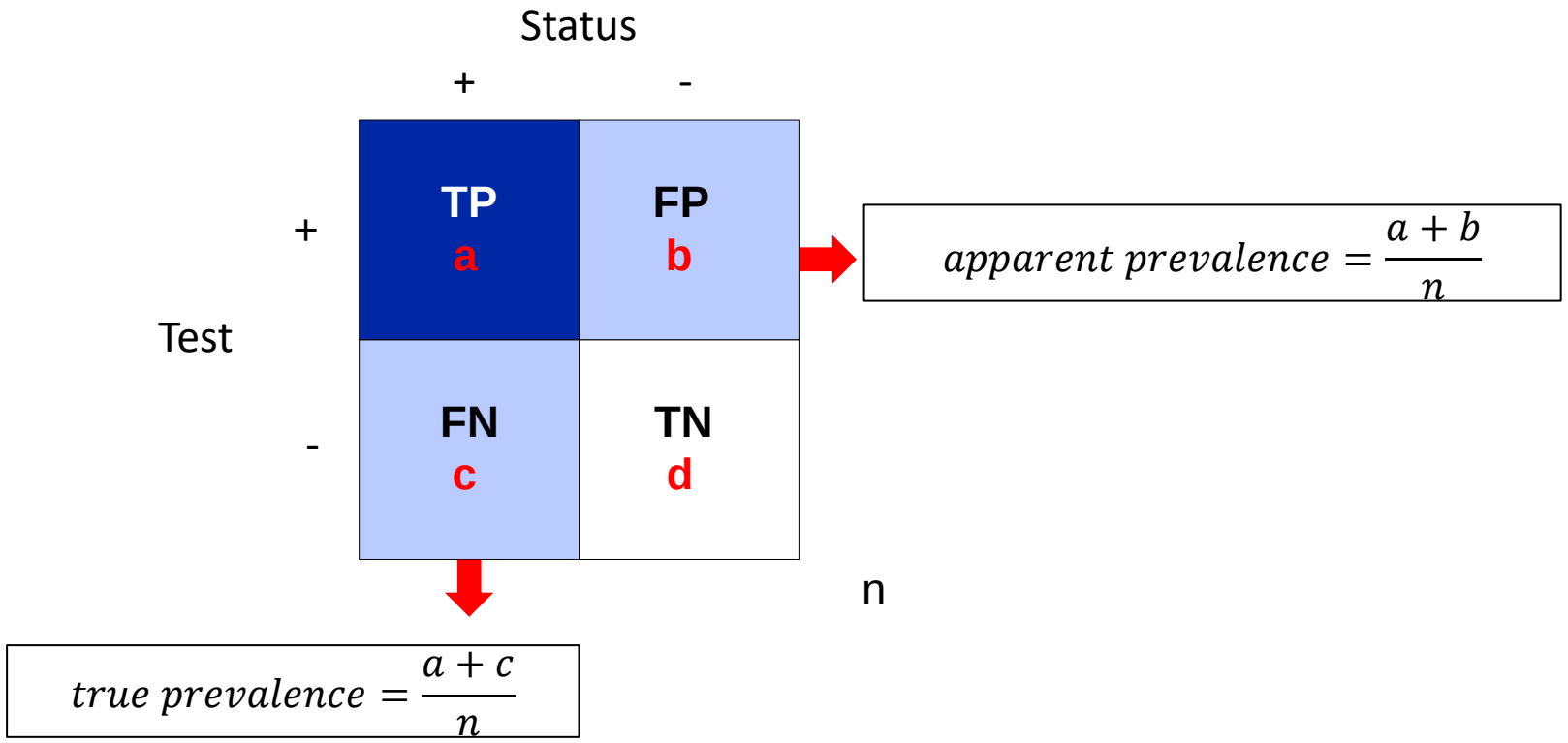


	Negative LR (95% CI)
Les sli et al. 1975	0.09 (0.04 - 0.20)
O'Reilly et al.1975	0.05 (0.02 - 0.12)
Francis et al. 1978	0.35 (0.31 - 0.40)
Neill et al. 1994	0.45 (0.38 - 0.53)
Ameni et al. 2000	0.12 (0.04 - 0.37)
Quirin et al. 2001 a	0.48 (0.31 - 0.76)
Quirin et al. 2001 b	0.24 (0.08 - 0.71)

Random Effects Model
Pooled Negative LR = 0.23 (0.15 to 0.36)
Cochran-Q = 54.68; df = 6 (p = 0.0000)
Inconsistency (I-square) = 89.0 %
Tau-squared = 0.2103



true and apparent (test) prevalence



- A test with a low sensitivity would lead to an underestimation of the prevalence
- A test with a low specificity would lead to an overestimation of the prevalence

true and apparent (test) prevalence

How to correct study results obtained with imperfect tests in order to obtain the true prevalence?

$$\hat{\mathrm{Tp}} = \frac{\mathrm{Ap} + \mathrm{Sp} - 1}{\mathrm{Se} + \mathrm{Sp} - 1}$$

(Rogan-Gladen estimator)

Tp: true prevalence

Ap: apparent or test prevalence

Prevalence **Tp**	Se,Sp 0.99 **Ap**	Se,Sp 0.95 **Ap**	Se,Sp 0.9 **Ap**
0.10	0.09	0.06	0.00
0.25	0.24	0.22	0.19
0.50	0.50	0.50	0.50
0.75	0.76	0.78	0.81
0.90	0.91	0.94	1.00

It is possible that some values for Se, Sp and Ap result in estimates of the true prevalence outside its allowed range (0-1)

→ indicates that one or both of the Se and Sp estimates used are not applicable for the population being studied

epiR

	D+	D-
T+	367	55
T-	26	678

```
library(epiR)

dat <- as.table(matrix(c(367, 55, 26, 678),
                       nrow = 2, byrow = TRUE))
colnames(dat) <- c("D+","D-")
rownames(dat) <- c("T+","T-")
```

```
dat

##     D+  D-
## T+ 367  55
## T-  26 678
```

epiR

```
epi.tests(dat, conf.level = 0.95)

##            Outcome +    Outcome -      Total
## Test +           367           55        422
## Test -            26          678        704
## Total            393          733       1126
##
## Point estimates and 95 % CIs:
## -------------------------------------------------------
## Apparent prevalence                    0.37 (0.35, 0.40)
## True prevalence                        0.35 (0.32, 0.38)
## Sensitivity                            0.93 (0.90, 0.96)
## Specificity                            0.92 (0.90, 0.94)
## Positive predictive value              0.87 (0.83, 0.90)
## Negative predictive value              0.96 (0.95, 0.98)
## Positive likelihood ratio              12.45 (9.64, 16.07)
## Negative likelihood ratio              0.07 (0.05, 0.10)
## -------------------------------------------------------
```

Exercise 3

Summary points

Diagnostic studies should match methods to diagnostic questions

Phase 1 →
- Do test results in affected patients differ from those in normal individuals?

Phase 2 →
- Are patients with certain test results more likely to have the target disorder?

Phase 3 →
- Do test results distinguish patients with and without the target disorder among those in whom it is clinically sensible to suspect the disorder?

Phase 4 →
- Do patients undergoing the diagnostic test fare better than similar untested patients?

The keys to validity in diagnostic test studies are

- independent, blind comparison of test results with a reference standard among a consecutive series of patients suspected (but not known) to have the target disorder
- inclusion of missing and indeterminate results
- replication of studies in other settings

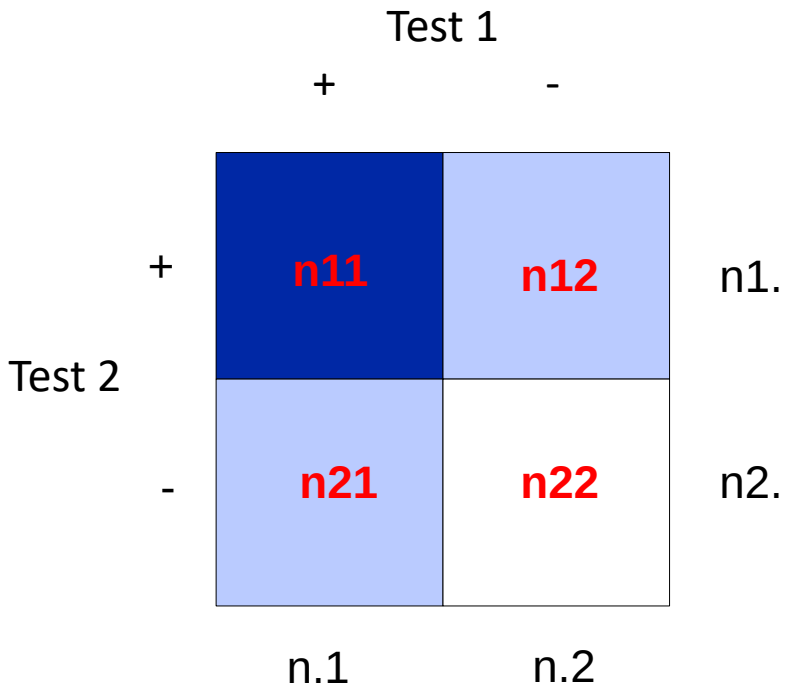
Both specificity and sensitivity may change as the same diagnostic test is applied in primary, secondary, and tertiary care

kappa statistic

Assessment of agreement (beyond chance), without assuming that one test is the best

→ Measures the level of agreement between two (or more) sets of test results having in mind that there will always be some agreement due to chance.

kappa



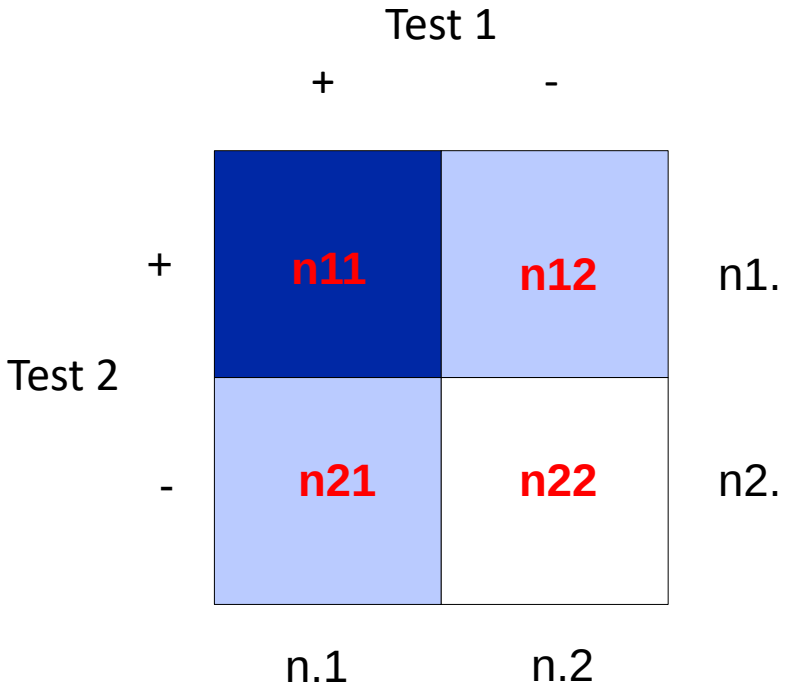
n11 (+,+) → concordant

n12 (-,+) → disconcordant

n21 (+,-) → disconcordant

n22 (-,-) → concordant

kappa



$$\text{observed agreement} = (n11 + n22)/n$$

$$\text{expected agreement (chance)} = \left[\frac{(n1.* n.1)}{n * n} + \frac{(n2.* n.2)}{n * n}\right]$$

$$\text{actual agreement beyond chance} = \text{observed} - \text{expected}$$

$$\text{potential agreement beyond chance} = (1 - \text{expected})$$

$$\kappa = \frac{\text{actual agreement beyond chance}}{\text{potential agreement beyond chance}} = 2\frac{(n11n22 - n12n21)}{n1.n2. + n.2n.1}$$

Cohen's kappa

- Kappa may be extended to situations where there are more than two tests (or raters).

- For tests on an ordinal scale (e.g. 5-point scale), a weighted kappa might be applied which is sensitive to the number of categories and the choice of weights.

Guidelines proposed by	Kappa coefficient scale								
	<0	0.00	0.20	0.40	0.50	0.60	0.75	0.80	1.00
Landis and Koch, 1977	Poor	Slight	Fair	Moderate		Substantial		Excellent or "Almost perfect"	Perfect
Altman, 1999	Poor		Fair	Moderate		Good		Very good	
Fleiss, 1981	Poor			Fair to Good				Excellent	

Beware that these scales are somewhat arbitrary !

Cohen’s kappa

- Interpretation with caution
  - low kappa value, because
    - only one test is good or
    - both tests are bad or
    - both tests are good, but negatively correlated (some Ag and Ab tests)
  - high kappa value, because
    - both tests make the same errors
- Kappa depends on the prevalence of the attribute of concern
- A bias (e.g. tendency of one test or rater to assign more positive test results than the other) may affect Kappa

Mc Nemar‘s X^2

Before quantifying the level of agreement, one needs to determine if the two tests or raters are classifying approximately the same proportion of individuals as positive.

$$Mc\ Nemar's\ X^2 = \frac{(n12 - n21)^2}{n12 + n21}$$

Or use an exact binomial test for correlated proportions

kappa in R

```
# kappa from a data set
library(irr)
data(anxiety)
head(anxiety)

##   rater1 rater2 rater3
## 1      3      3      2
## 2      3      6      1
## 3      3      4      4
## 4      4      6      4
## 5      5      2      3
## 6      5      4      2

kappa2(anxiety[,1:2])

##  Cohen's Kappa for 2 Raters (Weights: unweighted)
##
##  Subjects = 20
##    Raters = 2
##     Kappa = 0.119
##
##         z = 1.16
##   p-value = 0.245
```

kappa in R

```
library(psych)
# kappa from two vectors
rater1 = c(1,2,3,4,5,6,7,8,9) # rater one's ratings
rater2 = c(1,3,1,6,1,5,5,6,7) # rater one's ratings
cohen.kappa(x=cbind(rater1,rater2))

## Call: cohen.kappa1(x = x, w = w, n.obs = n.obs, alpha = alpha, levels = levels)
##
## Cohen Kappa and Weighted Kappa correlation coefficients and confidence boundaries
##                  lower estimate upper
## unweighted kappa -0.18     0.00  0.18
## weighted kappa    0.43     0.68  0.93
##
##  Number of subjects = 9
```

```
# kappa from a table
mydat <- matrix(c(34,2,5,13),
                ncol=2,byrow=TRUE)
cohen.kappa(mydat)

## Call: cohen.kappa1(x = x, w = w, n.obs = n.obs, alpha = alpha, levels = levels)
##
## Cohen Kappa and Weighted Kappa correlation coefficients and confidence boundaries
##                  lower estimate upper
## unweighted kappa  0.49      0.7   0.9
## weighted kappa    0.49      0.7   0.9
##
##  Number of subjects = 54
```

Exercise 4

status?

How to obtain a „status“ for the evaluation of diagnostic tests?

- use of gold standard populations
- use of a gold standard test
- use of a pseudo-gold standard test

  (or combination of tests)
- reference test with known Se and Sp
- evaluation when there is no „gold standard“

reference status

- absolute reference status
  - presence of pathogen (e.g. trichinella)
  - characteristic histopathological lesions
- relative reference status
  - comparable indirect detection methods (other serological tests)
- additional information reference status
  - experimentally infected animals and/or animals from populations free of the disease
  - information gained later in time

gold standard

A test or procedure that is absolutely accurate:

⟶ diagnoses all of the specific diseases that exist and misdiagnoses none or correctly classifies all samples from individuals being affected or not.

In reality there are very few true gold standards, because

- imperfections of the test itself
- a good portion of the error is due to biological variability

www.theguardian.com

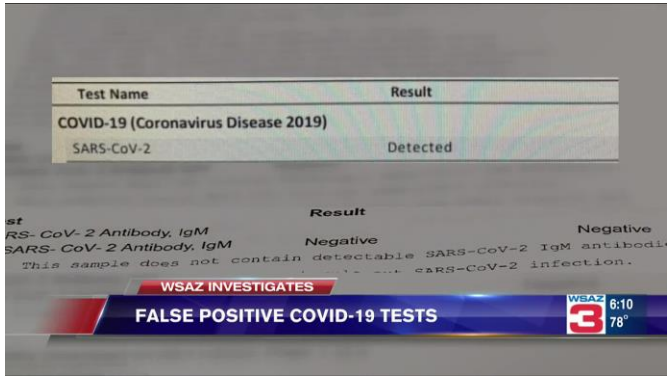
www.swissinfo.ch

Health officials say false negative COVID-19 test results happen 10% of the time

www.wtvy.com

www.youtube.com/

www.nbcsports.com



www.wsaz.com

JOURNAL OF
MEDICAL VIROLOGY

LETTER TO THE EDITOR | Free Access

Should RT-PCR be considered a gold standard in the diagnosis of COVID-19?

Moustapha Dramé MD, PhD, Maturin Tabue Teguo MD, PhD, Emeline Proye MD, Fanny Hequet MD, Maxime Hentzien MD, PhD, Lukshe Kanagaratnam MD, PhD, Lidvine Godaert MD, PhD

First published: 08 May 2020 | https://doi.org/10.1002/jmv.25996 | Citations: 7

Dramé et al., 2020
«… indeed, when an existing test is considered as a reference, this suggests that the test in question is always correct, and that all misclassifications (false negatives, false positives) are due to the new test…»

*«Consequently, the new test will **never** be able to achieve sensitivity of 100%, since it is considered responsible for all misclassifications.»*

Sensitivity & Specificity

- Sensitivity is the ability of a diagnostic test, to correctly classify infected individuals
- Specificity is the ability of a diagnostic test, to correctly classify healthy individuals

	Infected	Healthy	
Test (+)	80	5	85
Test (-)	20	95	115
	100	100	200

JOURNAL OF
MEDICAL VIROLOGY

LETTER TO THE EDITOR | Free Access

Performance of VivaDiag COVID-19 IgM/IgG Rapid Test is inadequate for diagnosis of COVID-19 in acute patients referring to emergency room department

Irene Cassaniti, Federica Novazzi, Federica Giardina, Francesco Salinaro, Michele Sachs, Stefano Perlini, Raffaele Bruno, Francesco Mojoli, Fausto Baldanti ... See all authors

First published: 30 March 2020 | https://doi.org/10.1002/jmv.25800 | Citations: 61

JOURNAL OF
MEDICAL VIROLOGY

LETTER TO THE EDITOR | Free Access

Should RT-PCR be considered a gold standard in the diagnosis of COVID-19?

Moustapha Dramé MD, PhD, Maturin Tabue Teguo MD, PhD, Emeline Proye MD, Fanny Hequet MD, Maxime Hentzien MD, PhD, Lukshe Kanagaratnam MD, PhD, Lidvine Godaert MD, PhD

First published: 08 May 2020 | https://doi.org/10.1002/jmv.25996 | Citations: 7

Dramé et al., 2020
«… indeed, when an existing test is considered as a reference, this suggests that the test in question is always correct, and that all misclassifications (false negatives, false positives) are due to the new test…»

*«Consequently, the new test will **never** be able to achieve sensitivity of 100%, since it is considered responsible for all misclassifications.»*

Sensitivity & Specificity

- Sensitivity is the ability of a diagnostic test, to correctly classify infected individuals
- Specificity is the ability of a diagnostic test, to correctly classify healthy individuals

	RT PCR (+)	RT PCR (-)	
Test (+)	80	5	85
Test (-)	20	95	115
	100	100	200

JOURNAL OF
MEDICAL VIROLOGY

LETTER TO THE EDITOR | Free Access

Performance of VivaDiag COVID-19 IgM/IgG Rapid Test is inadequate for diagnosis of COVID-19 in acute patients referring to emergency room department

Irene Cassaniti, Federica Novazzi, Federica Giardina, Francesco Salinaro, Michele Sachs, Stefano Perlini, Raffaele Bruno, Francesco Mojoli, Fausto Baldanti ... See all authors

First published: 30 March 2020 | https://doi.org/10.1002/jmv.25800 | Citations: 61

JOURNAL OF
MEDICAL VIROLOGY

LETTER TO THE EDITOR | Open Access

Bayesian latent class models to estimate diagnostic test accuracies of COVID-19 tests

Sonja Hartnack ✉, Paolo Eusebi, Polychronis Kostoulas

First published: 08 August 2020 | https://doi.org/10.1002/jmv.26405

University of Zurich

HARMONY

Novel tools for test evaluation and disease prevalence estimation

https://harmony-net.eu/

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Bayesian Latent Class Models (BLCM)

BLCMs can be applied:

- To estimate sensitivity and specificity in the absence of a gold standard.
- Here, "latent" means that the true disease/infection status is not observed but can be estimated from the data.

historical sketch LCM

– 1980 Hui-Walter paradigm

BIOMETRICS 36, 167–171
March, 1980

Estimating the Error Rates of Diagnostic Tests

S. L. Hui[1] and S. D. Walter

two tests, one population

Population 1

		T2+	T2-
D+	T1+	P1*Se1*Se2	P1*Se1*(1-Se2)
	T1-	P1*(1-Se1)*Se2	P1*(1-Se1)*(1-Se2)

		T2+	T2-
D-	T1+	(1-P1)*(1-Sp1)*(1-Sp2)	(1-P1)*(1-Sp1)*Sp2
	T1-	(1-P1)*Sp1*(1-Sp2)	(1-P1)*Sp1*Sp2

two tests, one population

Population 1

		T2+	T2-
D+	T1+	P1*Se1*Se2	P1*Se1*(1-Se2)
	T1-	P1*(1-Se1)*Se2	P1*(1-Se1)*(1-Se2)

		T2+	T2-
D-	T1+	(1-P1)*(1-Sp1)*(1-Sp2)	(1-P1)*(1-Sp1)*Sp2
	T1-	(1-P1)*Sp1*(1-Sp2)	(1-P1)*Sp1*Sp2

two tests, one population

Population 1

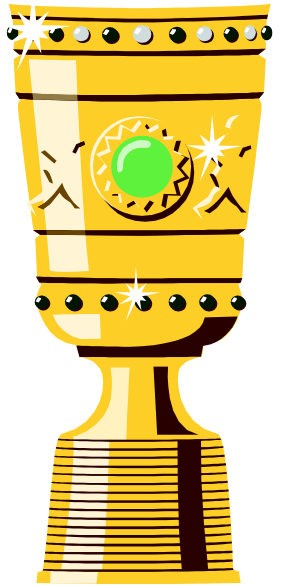
T1+T2+:	P1*Se1*Se2+(1-P1)*(1-Sp1)*(1-Sp2)
T1+T2-:	P1*Se1*(1-Se2)+(1-P1)*(1-Sp1)*Sp2
T1-T2+:	P1*(1-Se1)*Se2+(1-P1)*Sp1*(1-Sp2)
T1-T2-:	P1*(1-Se1)*(1-Se2)+(1-P1)*Sp1*Sp2

- 5 parameter and 3 degrees of freedom
- Non identifiable model

two tests, two populations

Population 1	
T1+T2+:	P1*Se1*Se2+(1-P1)*(1-Sp1)*(1-Sp2)
T1+T2-:	P1*Se1*(1-Se2)+(1-P1)*(1-Sp1)*Sp2
T1-T2+:	P1*(1-Se1)*Se2+(1-P1)*Sp1*(1-Sp2)
T1-T2-:	P1*(1-Se1)*(1-Se2)+(1-P1)*Sp1*Sp2

Population 2	
T1+T2+:	P2*Se1*Se2+(1-P2)*(1-Sp1)*(1-Sp2)
T1+T2-:	P2*Se1*(1-Se2)+(1-P2)*(1-Sp1)*Sp2
T1-T2+:	P2*(1-Se1)*Se2+(1-P2)*Sp1*(1-Sp2)
T1-T2-:	P2*(1-Se1)*(1-Se2)+(1-P2)*Sp1*Sp2



Identifiable model!

6=6

Hui-Walter Paradigm (1980)

$$S \geq \frac{R}{(2^{R-1} - 1)}$$

S: Populations, R: Tests

Assumptions

1. The population is divided into two or more populations in which two or more tests are evaluated,
2. sensitivity and specificity are the same in all populations.
3. The tests are conditionally independent given the disease status.

$$P(T_1^+ | T_2^+) = P(T_1^+ | T_2^-)$$

historical sketch LCM

BIOMETRICS 36, 167–171
March, 1980

Estimating the Error Rates of Diagnostic Tests

S. L. Hui[1] and S. D. Walter

- 1980 **Hui-Walter paradigma**
- 1985 Vacek **The effect of conditional dependence on the evaluation of diagnostic tests**

TABLE 2. Maximum Number of Estimable Parameters and Number of Parameters to Be Estimated in the Absence of Conditional Independence and Under Conditional Independence as a Function of the Number of Tests per Subject

Number of Tests	Maximum Number of Estimable Parameters	Parameters to be Estimated Under Conditional Dependence	Parameters to Be Estimated Under Conditional Independence
1	1	3	3
2	3	7	5
3	7	15	7
4	15	31	9
5	31	63	11
h	$2^h - 1$	$2^{h+1} - 1$	$2h + 1$

Berkvens D et al. (2006) Estimating Disease Prevalence in a Bayesian Framework Using Probabilistic Constraints. doi: 10.1097/01.ede.0000198422.64801.8d

historical sketch BLCM

BIOMETRICS 36, 167–171
March, 1980

Estimating the Error Rates of Diagnostic Tests

S. L. Hui[1] and S. D. Walter

- 1980 **Hui-Walter paradigm**
- 1985 Vacek **The effect of conditional dependence on the evaluation of diagnostic tests**
- 1995 Joseph et al. **Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard**

historical sketch BLCM

BIOMETRICS 36, 167–171
March, 1980

Estimating the Error Rates of Diagnostic Tests

S. L. Hui[1] and S. D. Walter

- 1980 **Hui-Walter paradigm**
- 1985 Vacek **The effect of conditional dependence on the evaluation of diagnostic tests**
- 1995 Joseph et al. **Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard**

prevalence

$\pi = P(D)$

sensitivity

$\eta_i = \mathrm{P}(+|\mathrm{D}, T_i)$

specificity

$\theta_i = P(-|\bar{D}, T_i)$

prior beta distributions

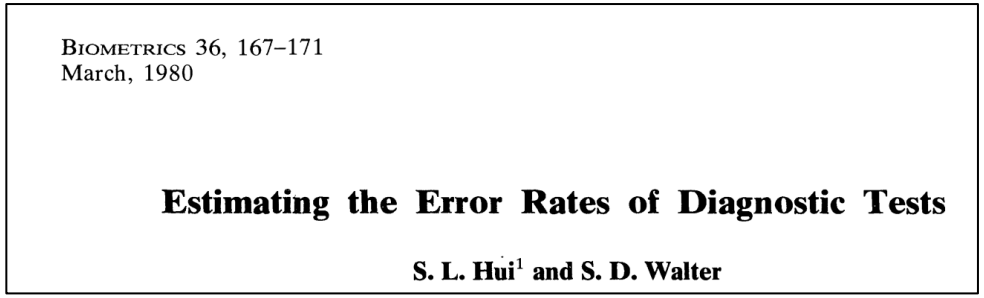
$\pi \sim Beta(a_\pi, b_\pi)$

$\eta_i \sim Beta(a_{\eta_i}, b_{\eta_i})$

$\theta_i \sim Beta(a_{\theta_i}, b_{\theta_i})$

$Posterior \propto Likelihood * Prior$

historical sketch BLCM



BIOMETRICS 36, 167–171
March, 1980

Estimating the Error Rates of Diagnostic Tests

S. L. Hui[1] and S. D. Walter

- 1980 **Hui-Walter paradigm**
- 1985 Vacek **The effect of conditional dependence on the evaluation of diagnostic tests**
- 1995 Joseph et al. **Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard**
- 1997 Bayesian Using Gibbs Sampling (BUGS)
- 2000 Enoe et al. **Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease is unknown**
- 2005 **OpenBUGS**
- 2007 Plummer **Just another Gibbs sampler (JAGS)**
- 2007 Toft et al. **Assessing the convergence of Markov Chain Monte Carlo methods: an example from evaluation of diagnostic tests in absence of a gold standard**
- 2017 Kostoulas et al. **STARD-BLCM: Standards for the Reporting of Diagnostic accuracy studies that use Bayesian Latent Class Models**

OIE Manual of Diagnostic Tests and Vaccines for Terrestrial Animals **endorsed BLCM** *(2016)*

World Organisation for Animal Health
Founded as OIE

literature search in PubMed

((bayesian analysis[MeSH Terms]) AND (sensitivity[MeSH Terms])) AND (covid-19[MeSH Terms])

1 Cite Share
Validation of RT-qPCR test for SARS-CoV-2 in saliva specimens.
Ávila LMS, Galvis MLD, Campos MAJ, Lozano-Parra A, Villamizar LAR, Arenas MO, Martínez-Vega RA, Cala LMV, Bautista LE.
J Infect Public Health. 2022 Dec;15(12):1403-1408. doi: 10.1016/j.jiph.2022.10.028. Epub 2022 Nov 2.
PMID: 36371937 Free PMC article.

2 Cite Share
Evaluating diagnostic accuracies of Panbio™ test and RT-PCR for the detection of SARS-CoV-2 in Addis Ababa, Ethiopia using Bayesian Latent-Class Models (BLCM).
Sisay A, Hartnack S, Tiruneh A, Desalegn Y, Tesfaye A, Desta AF.
PLoS One. 2022 Oct 19;17(10):e0268160. doi: 10.1371/journal.pone.0268160. eCollection 2022.
PMID: 36260547 Free PMC article.
Item in Clipboard

3 Cite Share
[Covid-19: Thomas Bayes warns and we should listen. False negatives and the probability of encountering them.].
Recchia M, Serra G.
Recenti Prog Med. 2022 May;113(5):317-323. doi: 10.1701/3803.37893.
PMID: 35587553 Italian.

4 Cite Share
Bayes Lines Tool (BLT): a SQL-script for analyzing diagnostic test results with an application to SARS-CoV-2-testing.
Aukema W, Malhotra BR, Goddek S, Kämmerer U, Borger P, McKernan K, Klement RJ.
F1000Res. 2021 May 10;10:369. doi: 10.12688/f1000research.51061.3. eCollection 2021.
PMID: 35284065 Free PMC article.

5 Cite Share
Diagnostic accuracy of three commercially available one step RT-PCR assays for the detection of SARS-CoV-2 in resource limited settings.
Sisay A, Abera A, Dufera B, Endrias T, Tasew G, Tesfaye A, Hartnack S, Beyene D, Desta AF.
PLoS One. 2022 Jan 20;17(1):e0262178. doi: 10.1371/journal.pone.0262178. eCollection 2022.
PMID: 35051204 Free PMC article.

6 Cite Share
Performance of Three Tests for SARS-CoV-2 on a University Campus Estimated Jointly with Bayesian Latent Class Modeling.
Perkins TA, Stephens M, Alvarez Barrios W, Cavany S, Rulli L, Pfrender ME.
Microbiol Spectr. 2022 Feb 23;10(1):e0122021. doi: 10.1128/spectrum.01220-21. Epub 2022 Jan 19.
PMID: 35044220 Free PMC article.

7 Cite Share
SPECT myocardial perfusion imaging identifies myocardial ischemia in patients with a history of COVID-19 without coronary artery disease.
Çap M, Bilge Ö, Gündoğan C, Tatlı İ, Öztürk C, Taştan E, Kepenek F, Işık F, Okşul M, Oktay M, Akyüz A, Erdoğan E, Burak C, Süleymanoğlu M, Karagöz A, Tanboğa IH.
Int J Cardiovasc Imaging. 2022 Feb;38(2):447-456. doi: 10.1007/s10554-021-02477-9. Epub 2021 Nov 22.
PMID: 34811596 Free PMC article.

8 Cite Share
Adjusting COVID-19 Seroprevalence Survey Results to Account for Test Sensitivity and Specificity.
Meyer MJ, Yan S, Schlageter S, Kraemer JD, Rosenberg ES, Stoto MA.
Am J Epidemiol. 2022 Mar 24;191(4):681-688. doi: 10.1093/aje/kwab273.
PMID: 34791024

9 Cite Share
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) seroprevalence: Navigating the absence of a gold standard.
Saeed S, O'Brien SF, Abe K, Yi QL, Rathod B, Wang J, Fazel-Zarandi M, Tuite A, Fisman D, Wood H, Colwill K, Gingras AC, Drews SJ.
PLoS One. 2021 Sep 23;16(9):e0257743. doi: 10.1371/journal.pone.0257743. eCollection 2021.
PMID: 34555095 Free PMC article.

10 Cite Share
SARS-CoV-2 rapid antigen testing in the healthcare sector: A clinical prediction model for identifying false negative results.
Leiner J, Pellissier V, Nitsche A, König S, Hohenstein S, Nachtigall I, Hindricks G, Kutschker C, Rolinski B, Gebauer J, Prantz A, Schubert J, Patzschke J, Bollmann A, Wolz M.
Int J Infect Dis. 2021 Nov;112:117-123. doi: 10.1016/j.ijid.2021.09.008. Epub 2021 Sep 10.
PMID: 34517045 Free PMC article.

11 Cite Share
Test, trace, isolate: evidence for declining SARS-CoV-2 PCR sensitivity in a clinical cohort.
Bergmans BJM, Reusken CBEM, van Oudheusden AJG, Godeke GJ, Bonačić Marinović AA, de Vries E, Kluiters-de Hingh YCM, Vingerhoets R, Berrevoets MAH, Verweij JJ, Nieman AE, Reimerink J, Murk JL, Swart A.
Diagn Microbiol Infect Dis. 2021 Oct;101(2):115392. doi: 10.1016/j.diagmicrobio.2021.115392. Epub

12 Cite Share
Efficient and effective single-step screening of individual samples for SARS-CoV-2 RNA using multi-dimensional pooling and Bayesian inference.
Sobczyk J, Pyne MT, Barker A, Mayer J, Hanson KE, Samore MH, Noriega R.
J R Soc Interface. 2021 Jun;18(179):20210155. doi: 10.1098/rsif.2021.0155. Epub 2021 Jun 16.
PMID: 34129787 Free PMC article.

13 Cite Share
The detection dogs test is more sensitive than real-time PCR in screening for SARS-CoV-2.
Hag-Ali M, AlShamsi AS, Boeijen L, Mahmmod Y, Manzoor R, Rutten H, Mweu MM, El-Tholoth M, AlShamsi AA.
Commun Biol. 2021 Jun 3;4(1):686. doi: 10.1038/s42003-021-02232-9.
PMID: 34083749 Free PMC article.

14 Cite Share
Prediction of COVID-19 with Computed Tomography Images using Hybrid Learning Techniques.
Perumal V, Narayanan V, Rajasekar SJS.
Dis Markers. 2021 Apr 22;2021:5522729. doi: 10.1155/2021/5522729. eCollection 2021.
PMID: 33968281 Free PMC article.

15 Cite Share
Seroprevalence of severe acute respiratory syndrome coronavirus 2 in Slovenia: results of two rounds of a nationwide population study on a probability-based sample, challenges and lessons learned.
Poljak M, Oštrbenk Valenčak A, Štrumbelj E, Maver Vodičar P, Vehovar V, Resman Rus K, Korva M, Knap N, Seme K, Petrovec M, Zupan B, Demšar J, Kurdija S, Avšič Županc T.
Clin Microbiol Infect. 2021 Jul;27(7):1039.e1-1039.e7. doi: 10.1016/j.cmi.2021.03.009. Epub 2021 Apr 7.
PMID: 33838303 Free PMC article.

16 Cite Share
Diagnostic Accuracy Estimates for COVID-19 Real-Time Polymerase Chain Reaction and Lateral Flow Immunoassay Tests With Bayesian Latent-Class Models.
Kostoulas P, Eusebi P, Hartnack S.
Am J Epidemiol. 2021 Aug 1;190(8):1689-1695. doi: 10.1093/aje/kwab093.
PMID: 33823529 Free PMC article. Review.

17 Cite Share
Occupancy modeling and resampling overcomes low test sensitivity to produce accurate SARS-CoV-2 prevalence estimates.
Sanderlin JS, Golding JD, Wilcox T, Mason DH, McKelvey KS, Pearson DE, Schwartz MK.
BMC Public Health. 2021 Mar 23;21(1):577. doi: 10.1186/s12889-021-10609-y.
PMID: 33757468 Free PMC article.

18 Cite Share
Evaluating short-term forecasting of COVID-19 cases among different epidemiological models under a Bayesian framework.
Li Q, Bedi T, Lehmann CU, Xiao G, Xie Y.
Gigascience. 2021 Feb 19;10(2):giab009. doi: 10.1093/gigascience/giab009.
PMID: 33604654 Free PMC article.

19 Cite Share
Multiplex assays for the identification of serological signatures of SARS-CoV-2 infection: an antibody-based diagnostic and machine learning study.
Rosado J, Pelleau S, Cockram C, Merkling SH, Nekkab N, Demeret C, Meola A, Kerneis S, Terrier B, Fafi-Kremer S, de Seze J, Bruel T, Dejardin F, Petres S, Longley R, Fontanet A, Backovic M, Mueller I, White MT.
Lancet Microbe. 2021 Feb;2(2):e60-e69. doi: 10.1016/S2666-5247(20)30197-X. Epub 2020 Dec 21.
PMID: 33521709 Free PMC article.

20 Cite Share
A Bayesian approach to improving spatial estimates of prevalence of COVID-19 after accounting for misclassification bias in surveillance data in Philadelphia, PA.
Goldstein ND, Wheeler DC, Gustafson P, Burstyn I.
Spat Spatiotemporal Epidemiol. 2021 Feb;36:100401. doi: 10.1016/j.sste.2021.100401. Epub 2021 Jan 8.
PMID: 33509436 Free PMC article.

21 Cite Share
Temporal variations in the diagnostic performance of chest CT for Covid-19 depending on disease prevalence: Experience from North-Eastern France.
Ohana M, Muller J, Severac F, Bilbault P, Behr M, Oberlin M, Leyendecker P, Roy C.
Eur J Radiol. 2021 Jan;134:109425. doi: 10.1016/j.ejrad.2020.109425. Epub 2020 Nov 20.
PMID: 33254065 Free PMC article.

22 Cite Share
Bayesian latent class models to estimate diagnostic test accuracies of COVID-19 tests.
Hartnack S, Eusebi P, Kostoulas P.
J Med Virol. 2021 Feb;93(2):639-640. doi: 10.1002/jmv.26405. Epub 2020 Aug 25.
PMID: 32770741 Free PMC article. No abstract available.

23 Cite Share
Towards reduction in bias in epidemic curves due to outcome misclassification through Bayesian analysis of time-series of laboratory test results: case study of COVID-19 in Alberta, Canada and Philadelphia, USA.
Burstyn I, Goldstein ND, Gustafson P.
BMC Med Res Methodol. 2020 Jun 6;20(1):146. doi: 10.1186/s12874-020-01037-4.
PMID: 32505172 Free PMC article.

24 Cite Share
Interpreting COVID-19 Test Results: a Bayesian Approach.
Good CB, Hernandez I, Smith K.
J Gen Intern Med. 2020 Aug;35(8):2490-2491. doi: 10.1007/s11606-020-05918-8. Epub 2020 Jun 3.
PMID: 32495086 Free PMC article. No abstract available.

25 Cite Share
Constructing co-occurrence network embeddings to assist association extraction for COVID-19 and other coronavirus infectious diseases.
Oniani D, Jiang G, Liu H, Shen F.
J Am Med Inform Assoc. 2020 Aug 1;27(8):1259-1267. doi: 10.1093/jamia/ocaa117.
PMID: 32458963 Free PMC article.

JOURNAL OF
MEDICAL VIROLOGY

LETTER TO THE EDITOR | Open Access

Bayesian latent class models to estimate diagnostic test accuracies of COVID-19 tests

Sonja Hartnack ✉, Paolo Eusebi, Polychronis Kostoulas

First published: 08 August 2020 | https://doi.org/10.1002/jmv.26405

University of Zurich

HARMONY

Novel tools for test evaluation and disease prevalence estimation

https://harmony-net.eu/

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

BLCM Application

Patients in Cassaniti et al., 2020

- 50 patients emergency room

30 COVID-19 positive patients

30 healthy volunteers

Keys to validity in diagnostic test studies:

«… independent, blind comparison of test results with a reference standard among a consecutive series of patients suspected (but not known) to have the target disorder».

Sackett and Haynes (2002),BMJ 2002;324:539

Code and technical details are here:

https://github.com/shartn/BLCM-COVID19

JOURNAL OF MEDICAL VIROLOGY

LETTER TO THE EDITOR | Free Access

Performance of VivaDiag COVID-19 IgM/IgG Rapid Test is inadequate for diagnosis of COVID-19 in acute patients referring to emergency room department

Irene Cassaniti, Federica Novazzi, Federica Giardina, Francesco Salinaro, Michele Sachs, Stefano Perlini, Raffaele Bruno, Francesco Mojoli, Fausto Baldanti … See all authors

First published: 30 March 2020 | https://doi.org/10.1002/jmv.25800 | Citations: 61

PCR	IgG	IgM	N
+	+	+	28
+	+	-	1
+	-	+	3
+	-	-	36
-	+	+	0
-	+	-	0
-	-	+	1
-	-	-	41

code BLCM https://github.com/shartn/BLCM-COVID19

```
#########################################################################################
## Bayesian latent-class model code for three diagnostic tests
#########################################################################################

#########################################################
##Definition of the variables in the model
#########################################################

var p[N], q[N,8], pr[N], L[N],checks[N,16];

# N    <- observations (N = 110 patients)
# p    <- individual samples
# q    <- different combinations of test results
# prc  <- prevalence
# s    <- test sensitivities
# c    <- test specificities
# covs <- conditional dependency between tests sensitivities
# covc <- conditional dependency between tests specificities
# m.ca <- data set name
```

code BLCM https://github.com/shartn/BLCM-COVID19

```
############################################################
## Modelling the different probabilities of combinations of tests results
############################################################

model {

  for(i in 1:N){

  q[i,1]<-prc*(s1*s2*s3+covs12+covs13+covs23)+(1-prc)*((1-c1)*(1-c2)*(1-c3)+covc12+covc13+covc23);
  q[i,2]<-prc*(s1*s2*(1-s3)+covs12-covs13-covs23)+(1-prc)*((1-c1)*(1-c2)*c3+covc12-covc13-covc23);
  q[i,3]<-prc*(s1*(1-s2)*s3-covs12+covs13-covs23)+(1-prc)*((1-c1)*c2*(1-c3)-covc12+covc13-covc23);
  q[i,4]<-prc*(s1*(1-s2)*(1-s3)-covs12-covs13+covs23)+(1-prc)*((1-c1)*c2*c3-covc12-covc13+covc23);
  q[i,5]<-prc*((1-s1)*s2*s3-covs12-covs13+covs23)+(1-prc)*(c1*(1-c2)*(1-c3)-covc12-covc13+covc23);
  q[i,6]<-prc*((1-s1)*s2*(1-s3)-covs12+covs13-covs23)+(1-prc)*(c1*(1-c2)*c3-covc12+covc13-covc23);
  q[i,7]<-prc*((1-s1)*(1-s2)*s3+covs12-covs13-covs23)+(1-prc)*(c1*c2*(1-c3)+covc12-covc13-covc23);
  q[i,8]<-prc*((1-s1)*(1-s2)*(1-s3)+covs12+covs13+covs23)+(1-prc)*(c1*c2*c3+covc12+covc13+covc23);
```

code BLCM https://github.com/shartn/BLCM-COVID19

```
##########################################################
## Check and correct potential errors of probabilities exceeding (0,1) bounds
##########################################################

  checks[i,1]<-   s1*s2*s3+covs12+covs13+covs23;
  checks[i,2]<-   (1-c1)*(1-c2)*(1-c3)+covc12+covc13+covc23;
  checks[i,3]<-  s1*s2*(1-s3)+covs12-covs13-covs23;
  checks[i,4]<-  (1-c1)*(1-c2)*c3+covc12-covc13-covc23;
  checks[i,5]<-  s1*(1-s2)*s3-covs12+covs13-covs23;
  checks[i,6]<-  (1-c1)*c2*(1-c3)-covc12+covc13-covc23;
  checks[i,7]<-  s1*(1-s2)*(1-s3)-covs12-covs13+covs23;
  checks[i,8]<-  (1-c1)*c2*c3-covc12-covc13+covc23;
  checks[i,9]<-  (1-s1)*s2*s3-covs12-covs13+covs23;
  checks[i,10]<- c1*(1-c2)*(1-c3)-covc12-covc13+covc23;
  checks[i,11]<- (1-s1)*s2*(1-s3)-covs12+covs13-covs23;
  checks[i,12]<- c1*(1-c2)*c3-covc12+covc13-covc23;
  checks[i,13]<- (1-s1)*(1-s2)*s3+covs12-covs13-covs23;
  checks[i,14]<- c1*c2*(1-c3)+covc12-covc13-covc23;
  checks[i,15]<- (1-s1)*(1-s2)*(1-s3)+covs12+covs13+covs23;
  checks[i,16]<- c1*c2*c3+covc12+covc13+covc23;

  valid[i]<- step(1-q[i,1])*step(q[i,1])*
             step(1-q[i,2])*step(q[i,2])*
             step(1-q[i,3])*step(q[i,3])*
```

code BLCM https://github.com/shartn/BLCM-COVID19

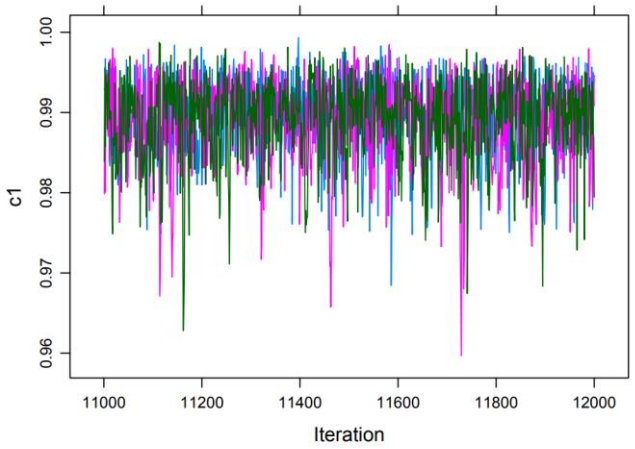
```
#########################################################
## Contribution to the likelihood for each observation
#########################################################

  L[i]<- equals(valid[i],1)*(
           equals(m.ca [i,1],1)*equals(m.ca[i,2],1)*equals(m.ca [i,3],1)*q[i,1]
         + equals(m.ca [i,1],1)*equals(m.ca[i,2],1)*equals(m.ca [i,3],0)*q[i,2]
         + equals(m.ca [i,1],1)*equals(m.ca[i,2],0)*equals(m.ca [i,3],1)*q[i,3]
         + equals(m.ca [i,1],1)*equals(m.ca[i,2],0)*equals(m.ca [i,3],0)*q[i,4]
         + equals(m.ca [i,1],0)*equals(m.ca[i,2],1)*equals(m.ca [i,3],1)*q[i,5]
         + equals(m.ca [i,1],0)*equals(m.ca[i,2],1)*equals(m.ca [i,3],0)*q[i,6]
         + equals(m.ca [i,1],0)*equals(m.ca[i,2],0)*equals(m.ca [i,3],1)*q[i,7]
         + equals(m.ca [i,1],0)*equals(m.ca[i,2],0)*equals(m.ca [i,3],0)*q[i,8]
        ) +(1-equals(valid[i],1)) *(1e-14);
```

code BLCM https://github.com/shartn/BLCM-COVID19

```
############################################################
## Definition of model priors which should be modified for sensitivity analysis.
## Assuming a uniform distribution for the covariance terms
## Assuming a beta distribution for prevalence and test accuracy estimates.
## Parameters of the beta prior distribution obtained by Betabuster
## https://betabuster.software.informer.com/1.0/
############################################################

  covs12 <- 0;
  covs13 <- 0;
  covs23 ~ dunif(-1,1);
  covc12 <- 0;
  covc13 <- 0;
  covc23 <- 0;

  prc ~ dbeta(1,1);
  c1  ~ dbeta(426.36,4.64) ;                    # SP RT-PCR #median=0.99, 5th percentile=0.98
  c2  ~ dbeta(108.19,2.53) I(0.5,);             # SP IGG #median=0.98, 5th percentile=0.95
  c3  ~ dbeta(108.19,2.53) I(0.5,);             # SP IGM #median=0.98, 5th percentile=0.95
  s1  ~ dbeta(1,1) I(0.1,);
  s2  ~ dbeta(1,1) I(0.1,);
  s3  ~ dbeta(1,1) I(0.1,);
  logL<-sum(log(p[1:N]));

    }
```

BLCM results

Specificity RT PCR



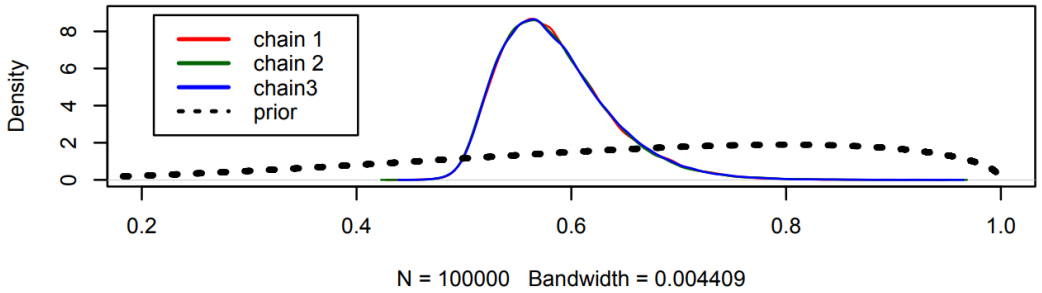
Gelman- Rubin diagnostics

JAGS model summary statistics from 300000 samples (chains = 3; adapt+burnin = 11000):

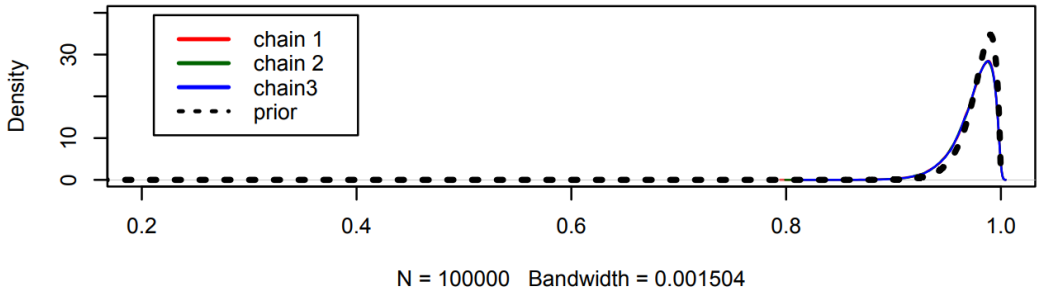
	Lower95	Median	Upper95	Mean	SD	Mode	MCerr	MC%ofSD	SSeff	AC.10	psrf
prc	0.74661	0.87936	1	0.87418	0.071603	--	0.00058172	0.8	15151	0.19255	0.99996
c1	0.97907	0.9899	0.99758	0.98915	0.0050545	--	0.000029382	0.6	29592	0.00040012	0.99998
c2	0.95439	0.98217	0.99859	0.97957	0.012772	--	0.000073739	0.6	30000	0.00078812	1
c3	0.95095	0.98051	0.99831	0.97779	0.013694	--	0.000079065	0.6	30000	0.0042698	1.0001
s1	0.49006	0.5523	0.64279	0.55868	0.041174	--	0.00040996	1	10087	0.44678	1.0007
s2	0.21678	0.31285	0.41958	0.31511	0.051762	--	0.00066389	1.3	6079	0.65377	1.0002
s3	0.2351	0.33248	0.43675	0.33436	0.051669	--	0.00062846	1.2	6759	0.54848	1.0002
covs23	0.064806	0.085744	0.1048	0.085196	0.010377	--	0.00012592	1.2	6792	0.59999	1.0003

sensitivity analysis

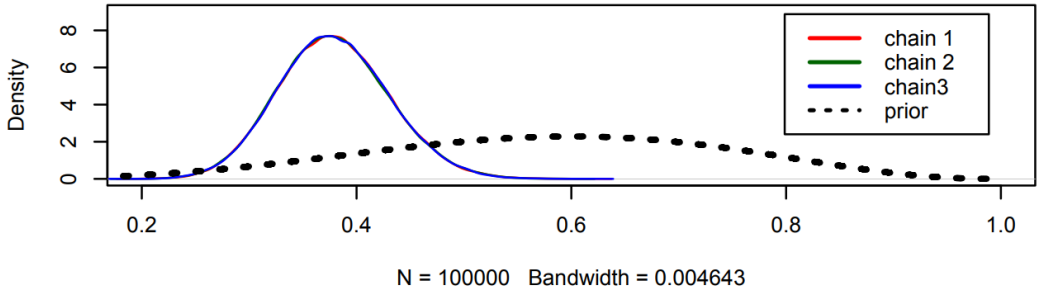
se RT-PCR: prior 95% sure >0.3, mode 0.8



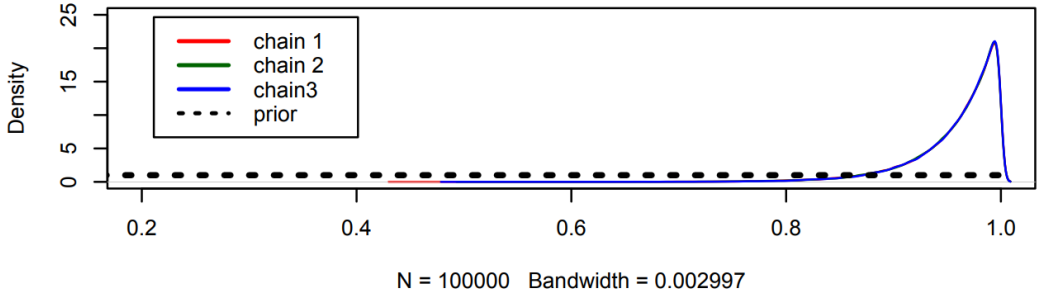
sp RT-PCR: prior 95%; >0.95, mode 0.99



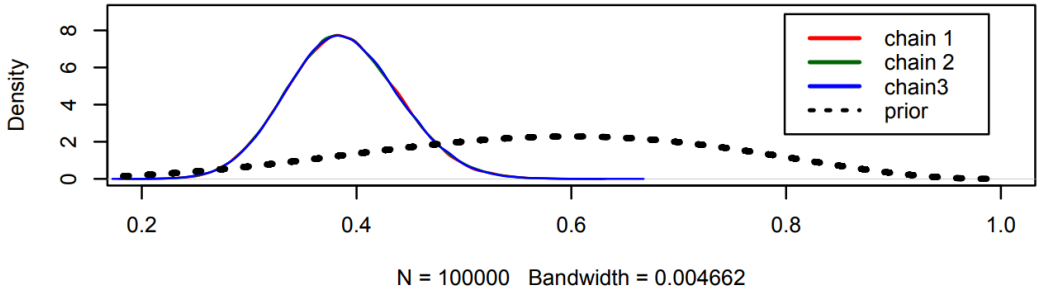
se IGG: prior 95% sure >0.3, mode 0.6



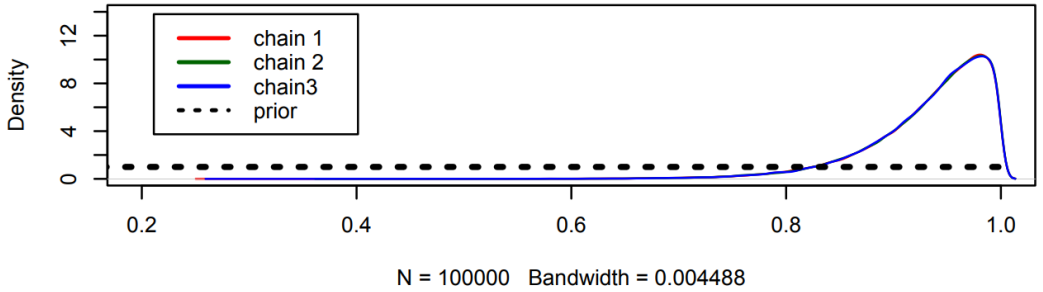
sp IGG: uninformative prior



se IGM: prior 95% sure >0.3, mode 0.6



sp IGM: uninformative prior



BLCM results

It is not possible to generalise the BLCM results beyond the study population.

Parameter	Median (95% Credibility interval)
Prevalence	87.9 (72.2;99.0)
RT-PCR Sensitivity	55.2 (49.7;65.6)
RT-PCR Specificity	99.0 (97.7;99.7)
IgG Sensitivity	31.3 (21.9;42.4)
IgG Specificity	98.2 (94.8;99.6)
IgM Sensitivity	33.2 (23.8;44.1)
IgM Specificity	98.1 (94.4;99.6)
Covs IgG/IgM	8.6 (6.3;10.4)

discussion

- «All models are wrong, but some are useful» (Box, 1976)

- Despite enormous worldwide COVID-19 activities, BLCMs are hardly ever not applied,

  and suitable data sets are virtually absent.

- Reasons?

- For *pandemic preparedness*,
  - Plan to include BLCM studies to assess test accuracies.
  - Test accuracy monitoring: Systematically, a certain percentage of all samples are examined with several diagnostic tests in parallel to estimate sensitivity and specificity with the help of BLCM.

acknowledgements

Paolo Eusebi, Regional Authority of Umbria, Perugia, Italy

Polychronis Kostoulas, School of Health Sciences, Thessaly, Greece

HARMONY
Novel tools for test evaluation and disease prevalence estimation

https://harmony-net.eu/

Eusebi *et al.*
BMC Medical Research Methodology (2023) 23:55
https://doi.org/10.1186/s12874-023-01853-4

BMC Medical Research Methodology

RESEARCH Open Access

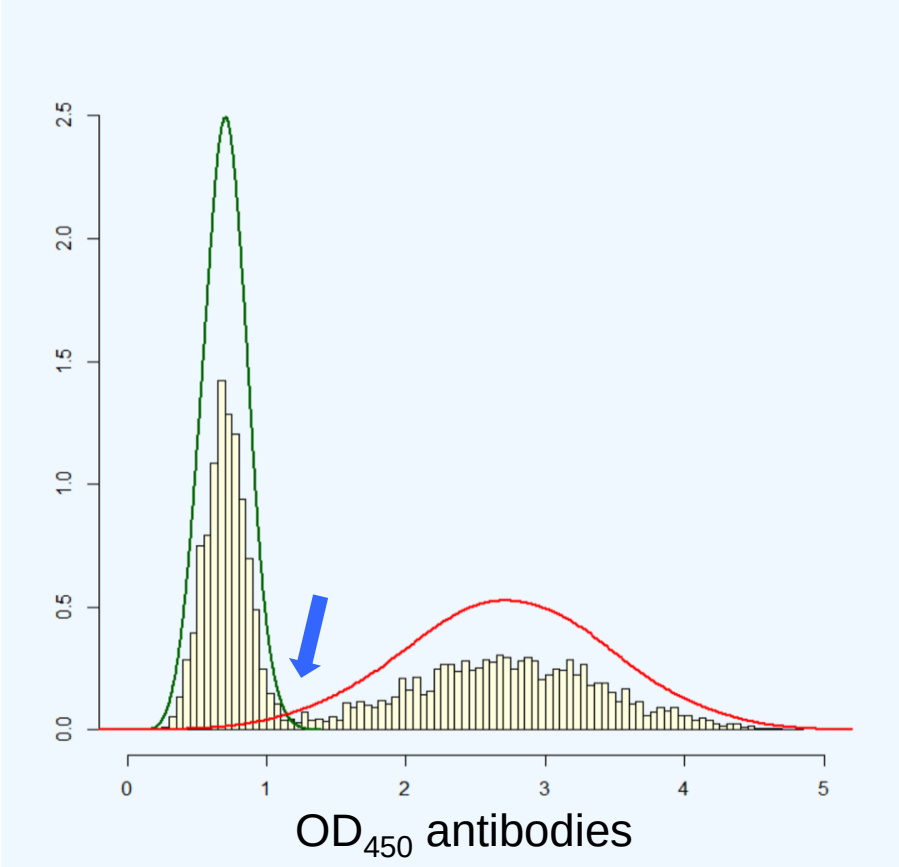
Check for updates

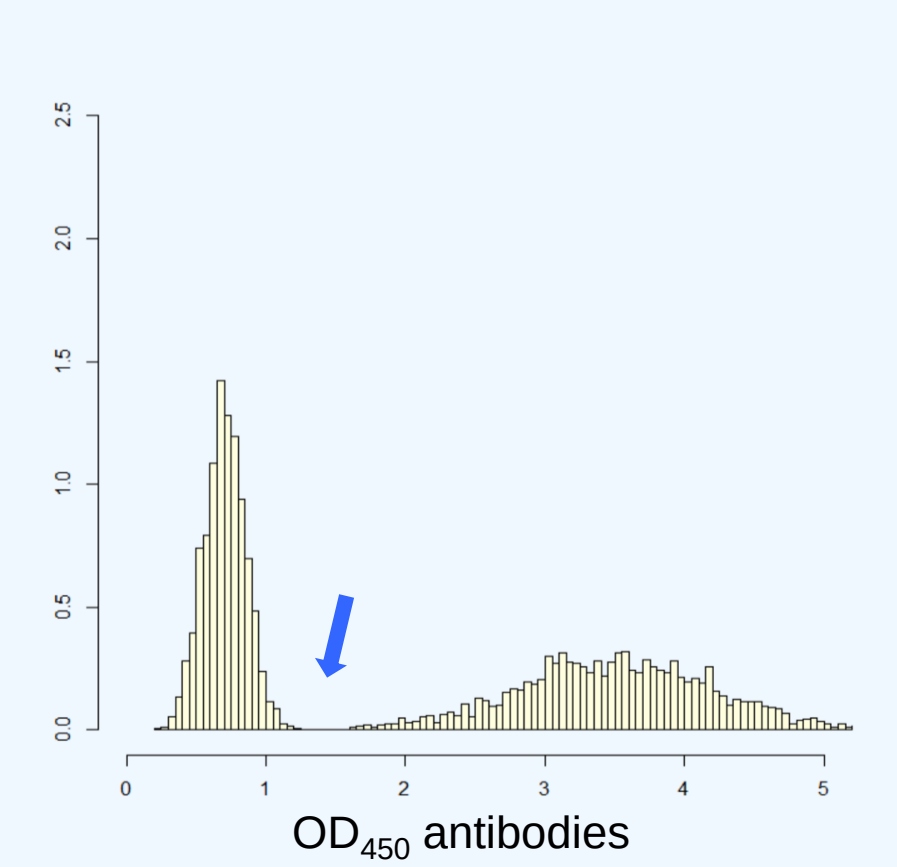
Addressing misclassification bias in vaccine effectiveness studies with an application to Covid-19

Paolo Eusebi[1,2*], Niko Speybroeck[3], Sonja Hartnack[4], Jacob Stærk-Østergaard[5], Matthew J. Denwood[5] and Polychronis Kostoulas[6]

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

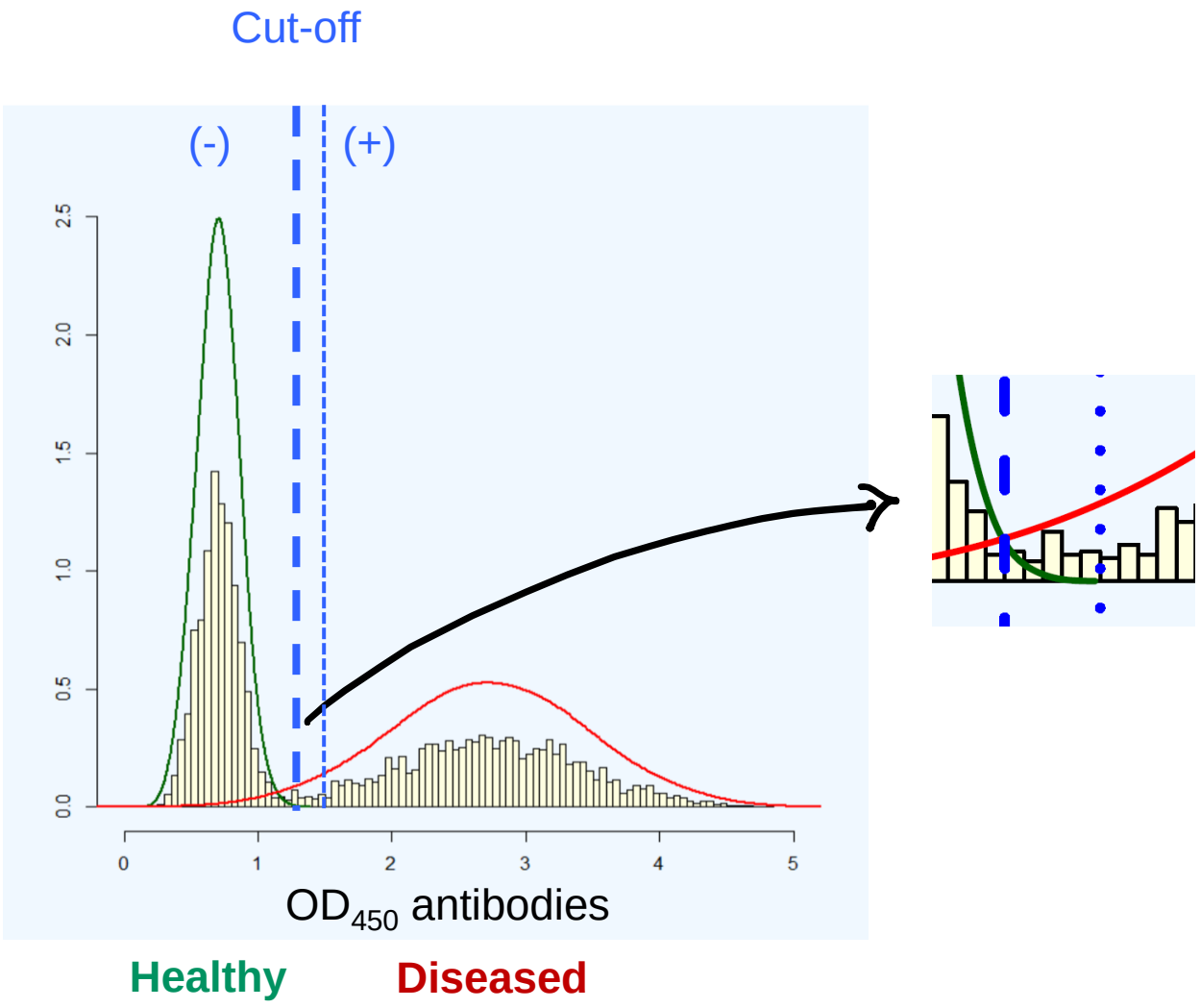
ROC curves



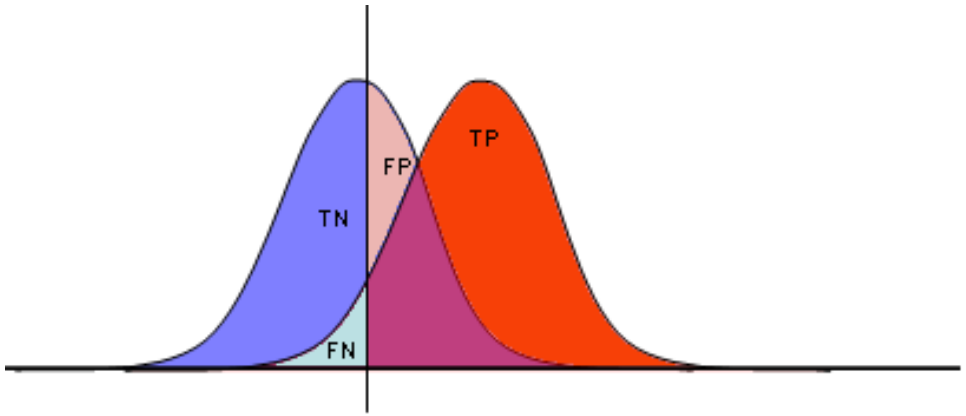


Healthy **Diseased**

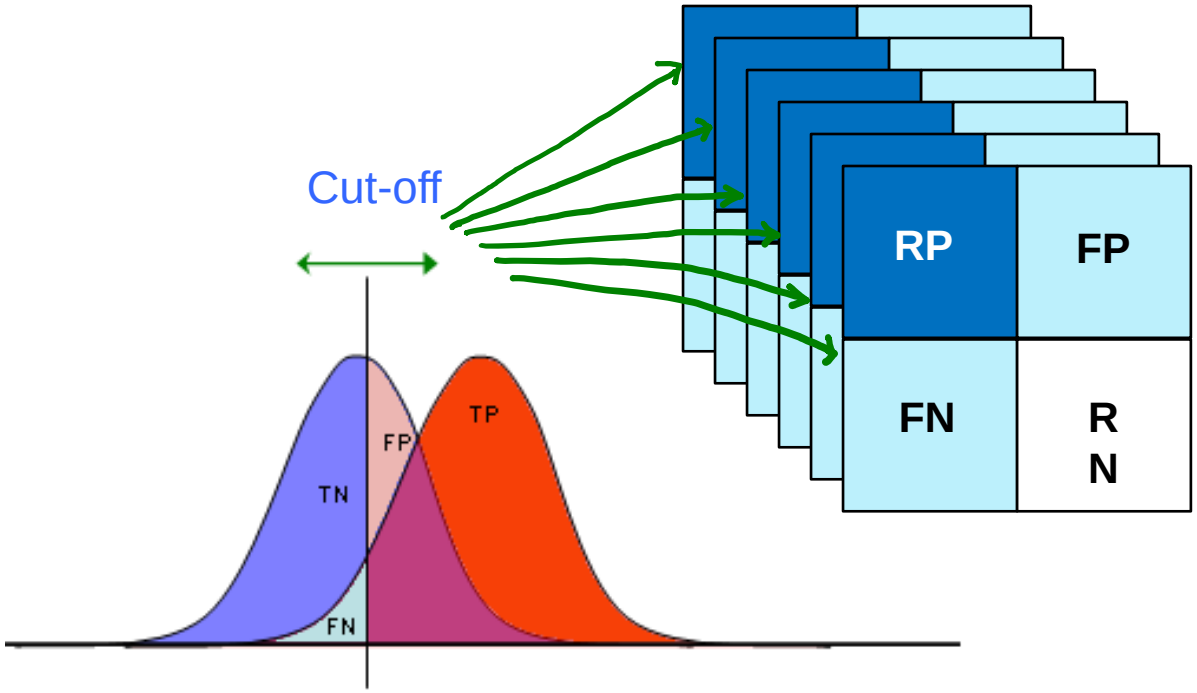
ROC curves


Cut-off
(-)
(+)
0.0
0.5
1.0
1.5
2.0
2.5
0
1
2
3
4
5
OD450 antibodies
Healthy
Diseased

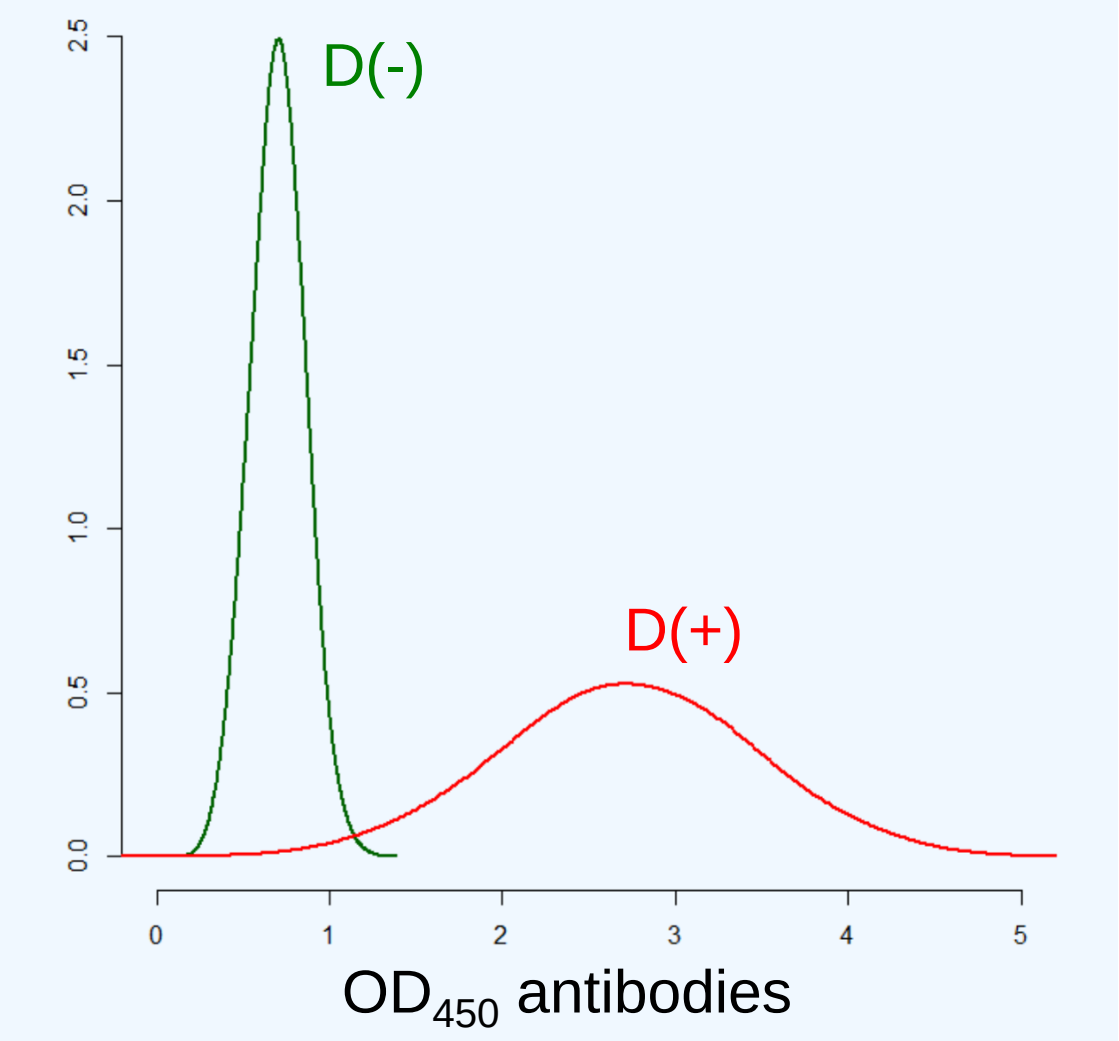
ROC curves



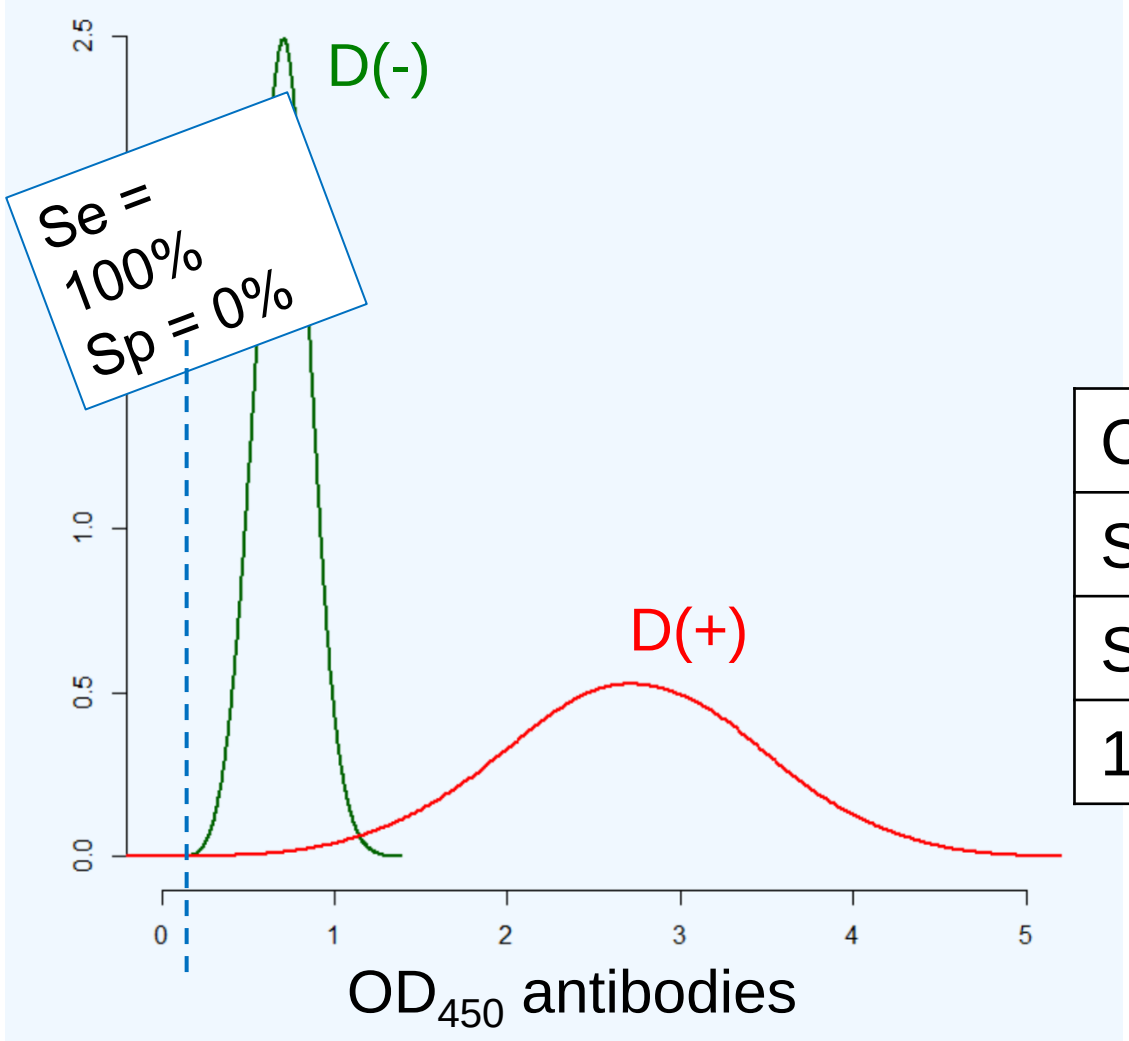
ROC curves



Receiver
Operating
Characteristic
curve analysis

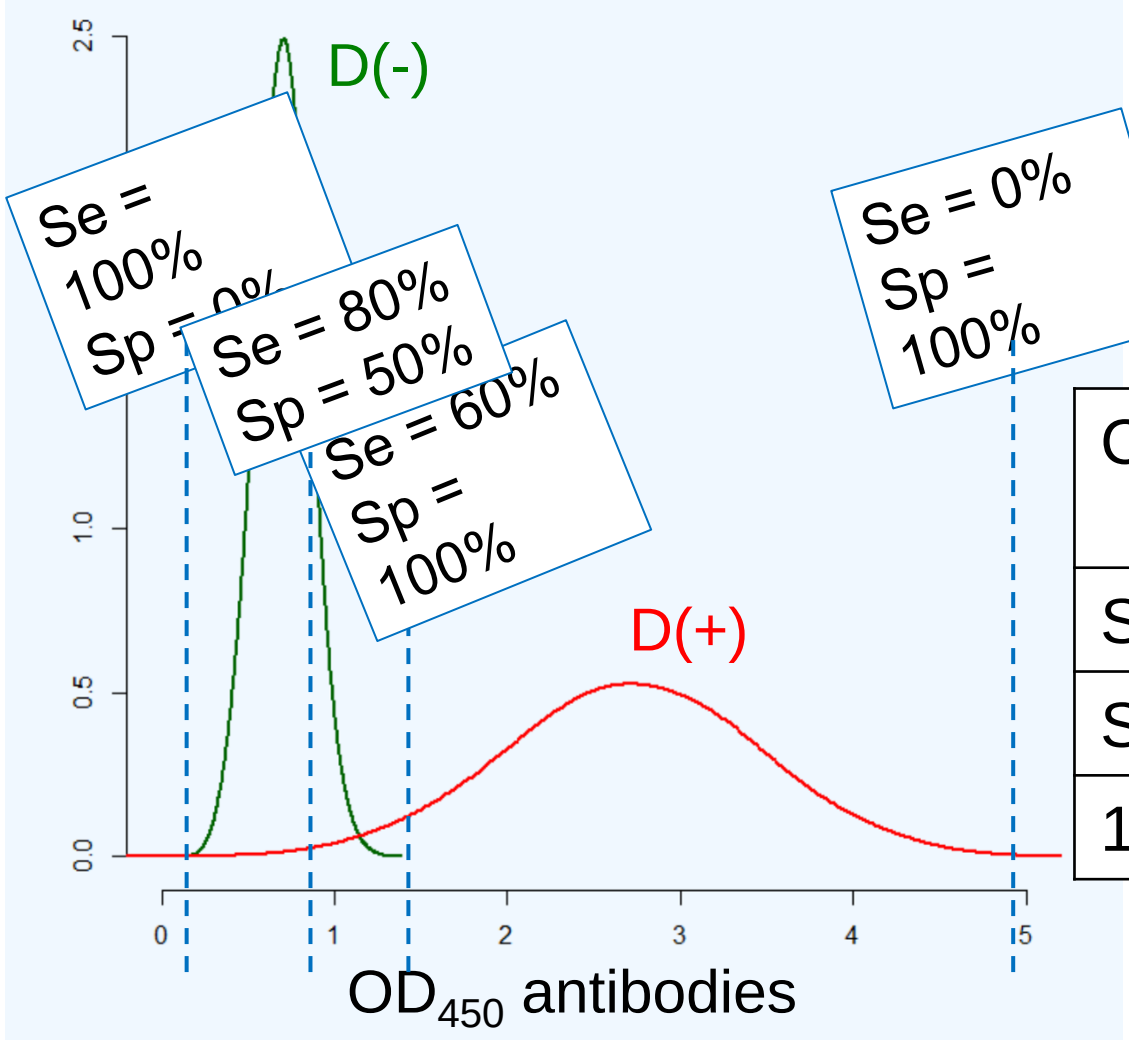


Receiver Operating Characteristic curve analysis



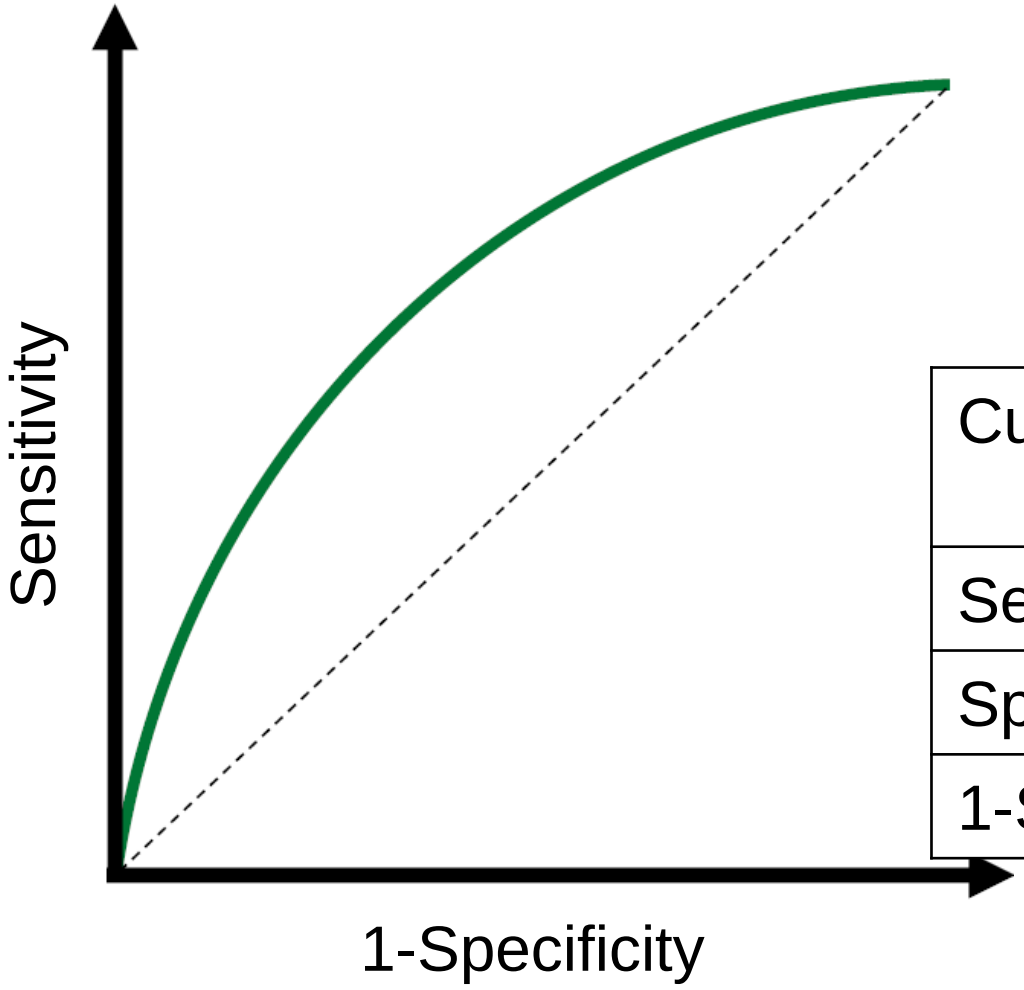
Cut-off	0.2	…	…	…	…	…
Se	100	…	…	…	…	…
Sp	0	…	…	…	…	…
1-Sp	100	…	…	…	…	…

Receiver Operating Characteristic curve analysis

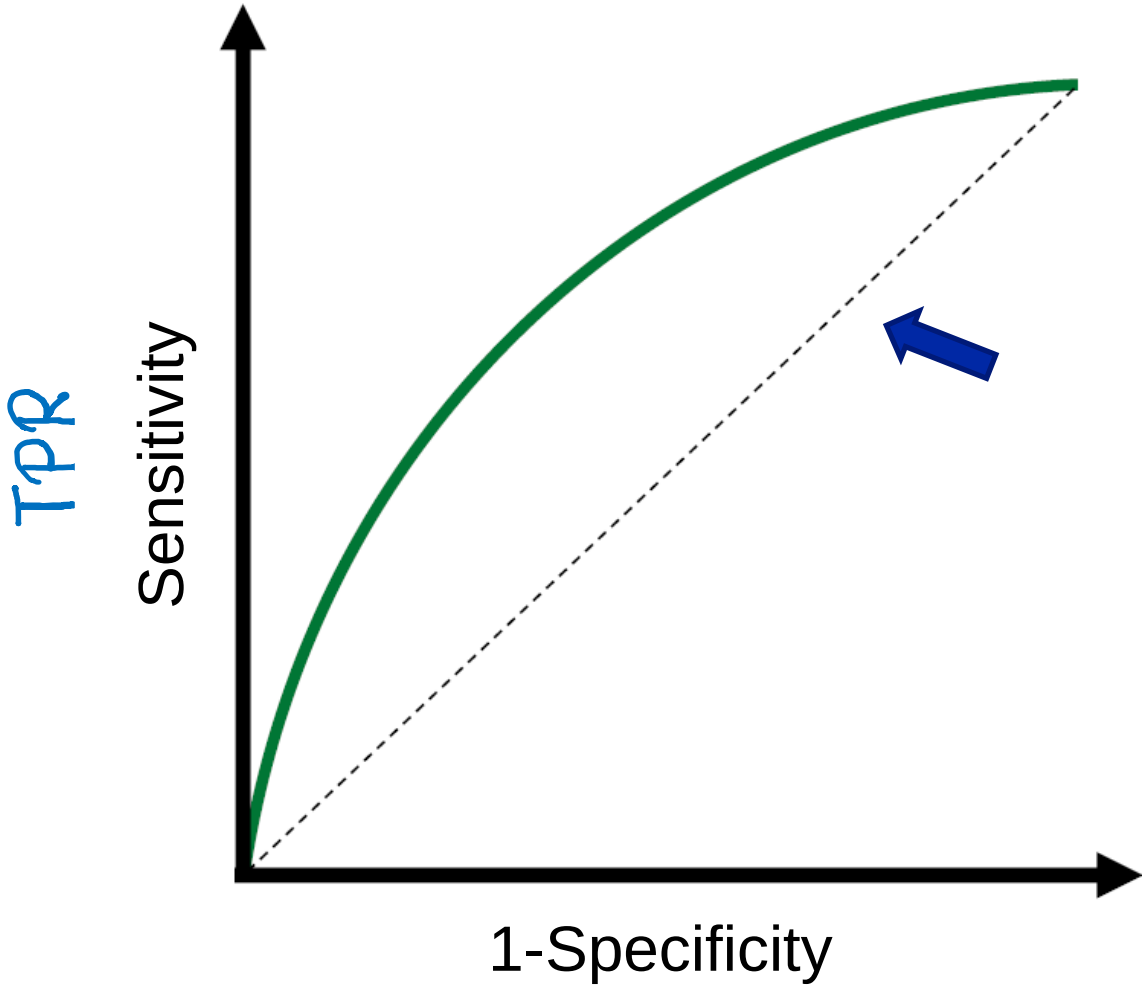


Cut-off	0.2	0.21	0.85	…	1.45	4.9
Se	100	100	80	…	60	0
Sp	0	…	50	…	100	100
1-Sp	100	…	50	…	0	0

Receiver Operating Characteristic curve analysis

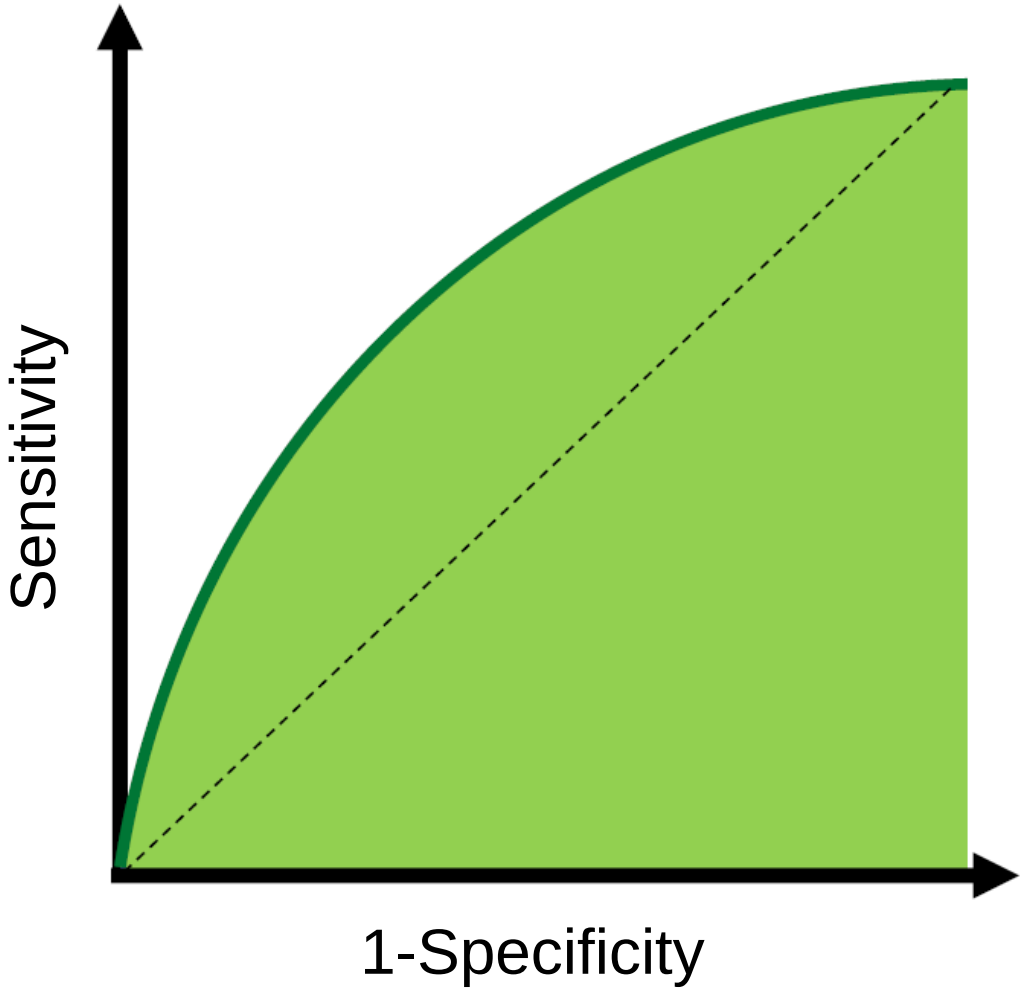


Cut-off	0.2	0.21	0.85	…	1.45	4.9
Se	100	100	80	…	60	0
Sp	0	…	50	…	100	100
1-Sp	100	…	50	…	0	0



True positive rate
False positive rate

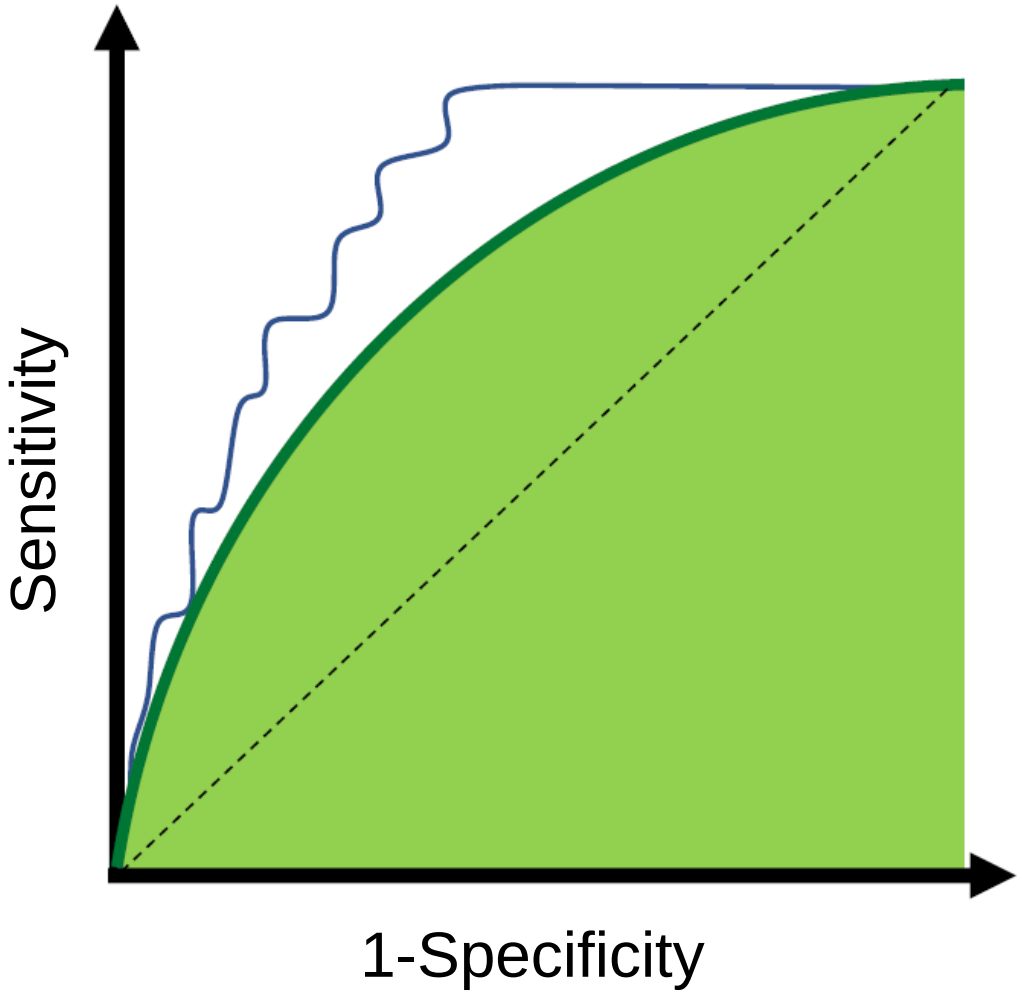
- For a range of cut-off values the relationship between pairs of true positives rates (Se) and false positives rates (1-Sp) are displayed.
- The perfect test (Se and Sp=1) would be situated at the upper left corner.
- The 45° equality line represents a test with no discriminatory ability (chance alone).
- Use of the ROC curve has the advantage over a „one cut-point“ value for determining Se and Sp in that it describes the overall ability of the test to discriminate diseased from non-diseased animals over a range of cut-points.



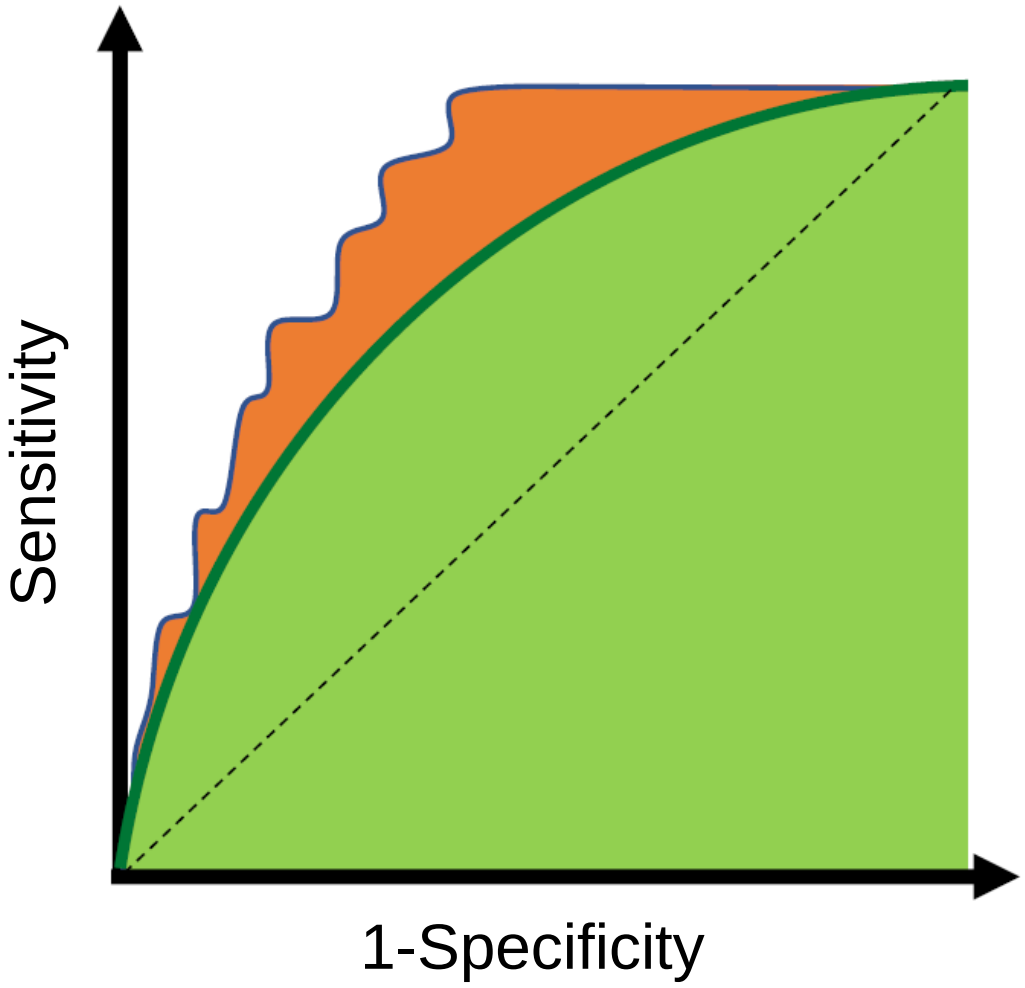
«A global assessment of a test's performance»
AUC (area under the curve)
AUC 0.63

AUC:

- the probability that a randomly-chosen positive example is ranked more highly than a randomly-chosen negative example.
- «a global assessment of a test's performance»
- a perfect test would yield an AUC of 1, an uninformative test an AUC of 0.5.
- calculation of the AUC is related to the Mann-Whitney statistic and is based on every comparison of individuals in the known-positive and known-negative groups.



AUC:
area under the curve
AUC 0.63



AUC:

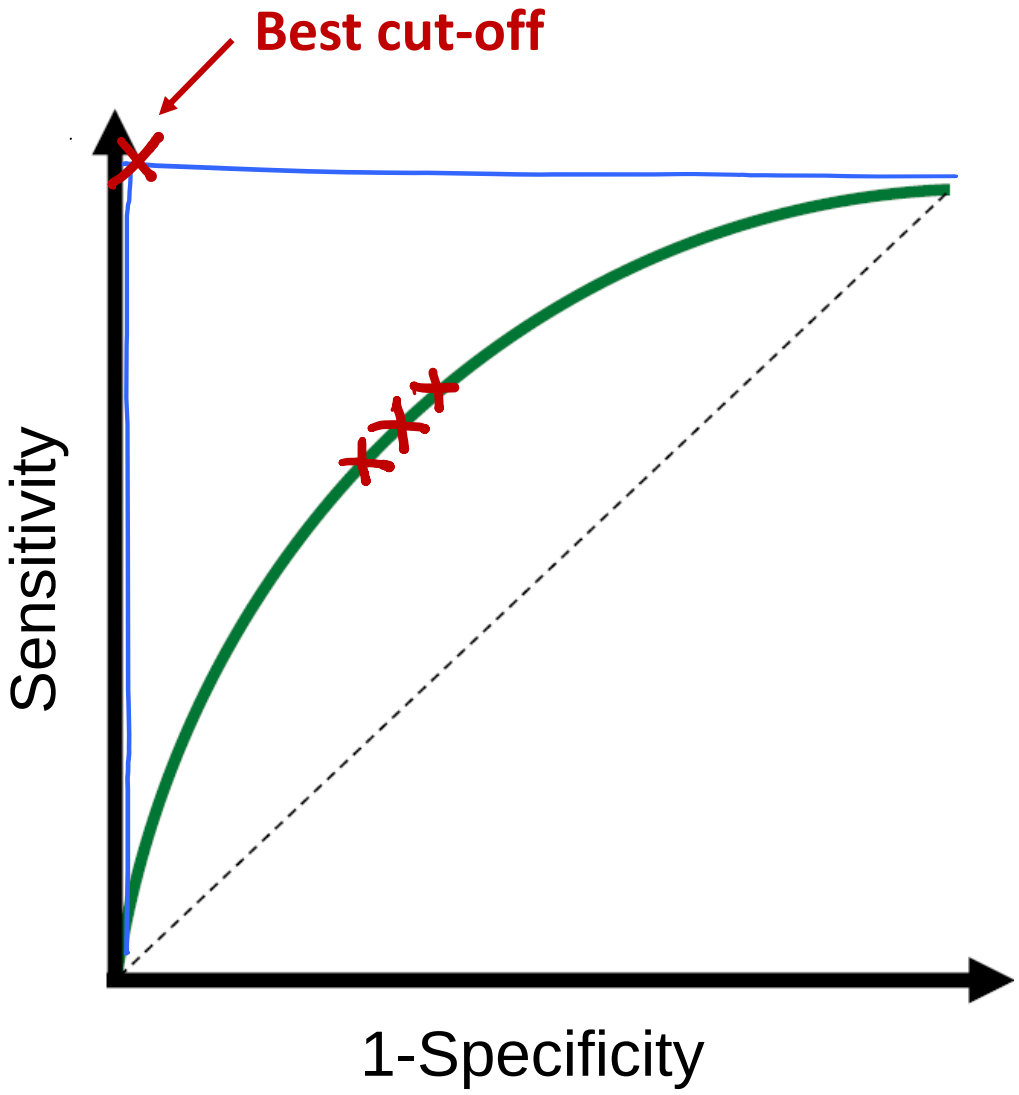
area under the curve

AUC 0.63 [95% CI: 0.56;0.70]

AUC 0.78 [95% CI: 0.72;0.84]

ROC, AUC and pAUC

- You might want to minimize FP (or FN) if one is causing more serious consequences than the other and obtain mainly test results without FP (or FN).
- Then there is the option to obtain mainly test results in one particular region of the ROC curve such as an area that constraints Se (or Sp) within defined limits (partial AUC).



Youden's index: Sensitivity + Specificity -1

- This index assumes that Se and Sp are of equal importance, which may not be the case.

Se ↗ : screening test
Sp ↗ : confirmatory test

pROC

- build ROC curves
- estimate AUC, pAUC
- compute CIs
- statistical tests to compare AUC, pAUC
- smoothing ROC curves
- determine cut-off

https://www.expasy.org/resources/proc

Package 'pROC'

May 13, 2023

Type Package

Title Display and Analyze ROC Curves

Version 1.18.2

Date 2023-05-13

Encoding UTF-8

Depends R (>= 2.14)

Imports methods, plyr, Rcpp (>= 0.11.1)

Suggests microbenchmark, tcltk, MASS, logcondens, doParallel, testthat, vdiffr, ggplot2

LinkingTo Rcpp

Description Tools for visualizing, smoothing and comparing receiver operating characteristic (ROC curves). (Partial) area under the curve (AUC) can be compared with statistical tests based on U-statistics or bootstrap. Confidence intervals can be computed for (p)AUC or ROC curves.

License GPL (>= 3)

URL http://expasy.org/tools/pROC/

A multiparameter panel method for outcome prediction following aneurysmal subarachnoid hemorrhage

Turck et al.

Intensive Care Med (2010) 36:107–115
DOI 10.1007/s00134-009-1641-y

pROC: aSAH data set

Several biomarkers measured in the blood of 141 patients at hospital admission after aneurysmal subarachnoid haemorrhage (aSAH) to predict the six-month outcome.

- Response: outcome (good, poor)
- Biomarkers: s100b, ndka
  - S100b
  - ndka (Nucleoside diphosphate kinase A)
- Clinical variables:
  - wfns: Word Federation of Neurological Surgeons (scale)
  - gos6: standard Glasgow outcome scale
- Demographic data: gender, age

pROC: aSAH data set

```
library(pROC)
data(aSAH)
head(aSAH)
```

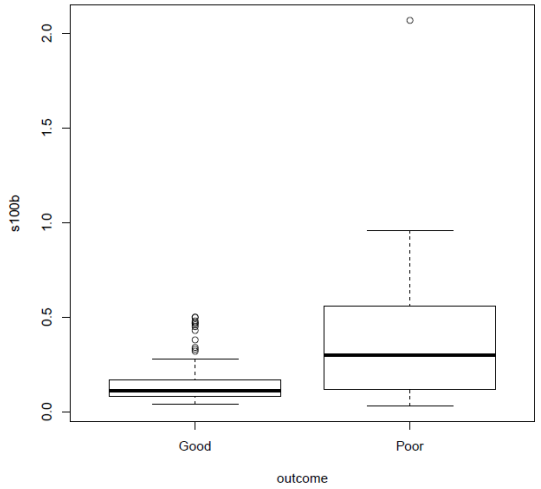
```
##    gos6 outcome gender age wfns s100b  ndka
## 29    5    Good Female  42    1  0.13  3.01
## 30    5    Good Female  37    1  0.14  8.54
## 31    5    Good Female  42    1  0.10  8.09
## 32    5    Good Female  27    1  0.04 10.42
## 33    1    Poor Female  42    3  0.13 17.40
## 34    1    Poor   Male  48    2  0.10 12.75
```

```
summary(aSAH)
```

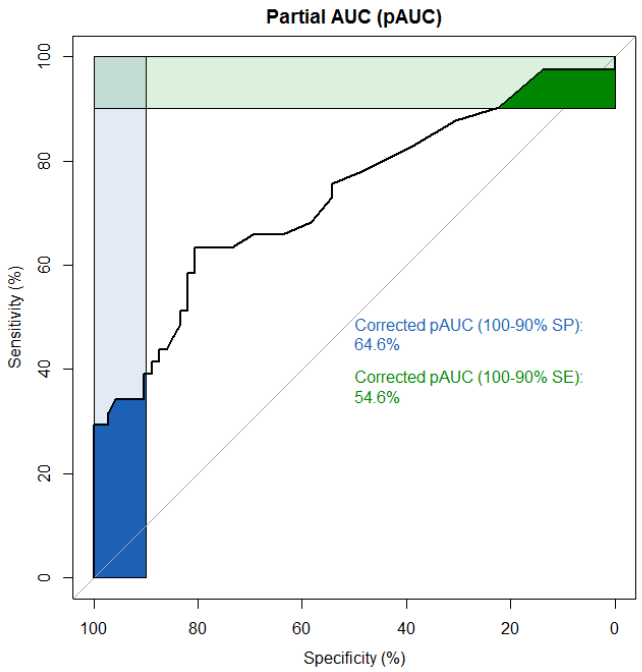
```
##  gos6   outcome      gender        age        wfns       s100b
##  1:28   Good:72   Male  :42   Min.   :18.0   1:39   Min.   :0.030
##  2: 0   Poor:41   Female:71   1st Qu.:42.0   2:32   1st Qu.:0.090
##  3:13                         Median :51.0   3: 4   Median :0.140
##  4: 6                         Mean   :51.1   4:16   Mean   :0.247
##  5:66                         3rd Qu.:61.0   5:22   3rd Qu.:0.330
##                               Max.   :81.0          Max.   :2.070
##       ndka
##  Min.   :  3.01
##  1st Qu.:  9.01
##  Median : 12.22
##  Mean   : 19.66
##  3rd Qu.: 17.30
##  Max.   :419.19
```

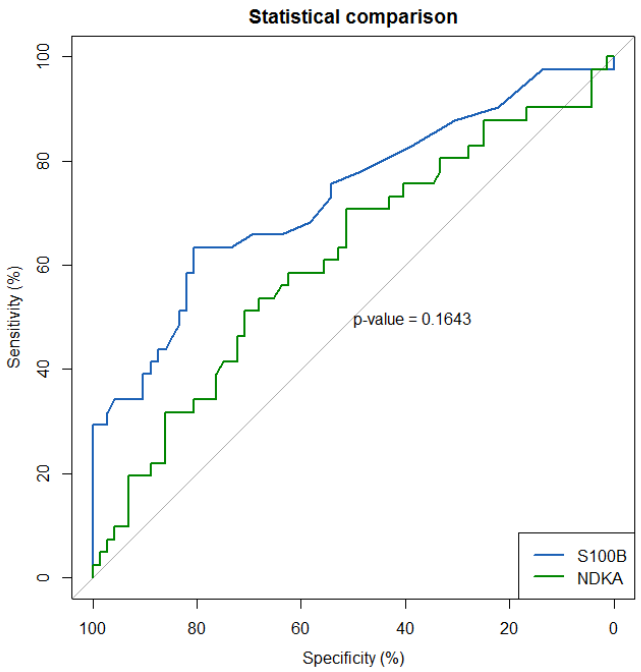
```
str(aSAH)
```

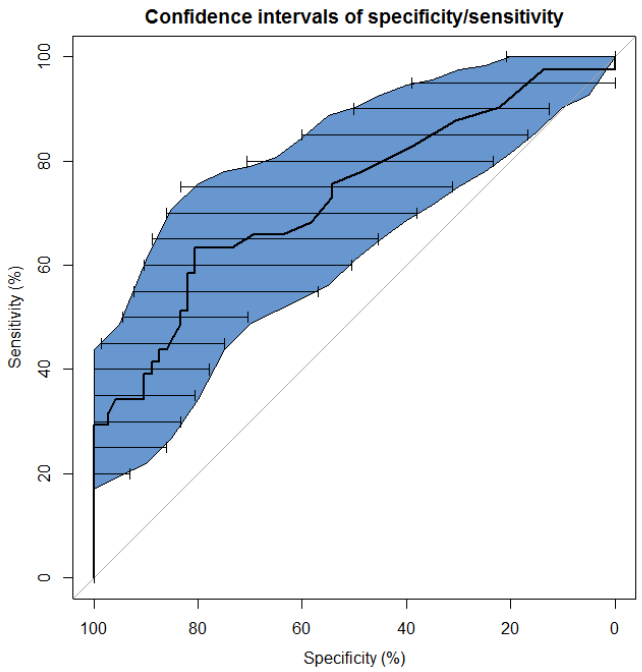
```
## 'data.frame': 113 obs. of  7 variables:
##  $ gos6   : Ord.factor w/ 5 levels "1"<"2"<"3"<"4"<..: 5 5 5 5 1 1 4 1 5 4 ...
##  $ outcome: Factor w/ 2 levels "Good","Poor": 1 1 1 1 2 2 1 2 1 1 ...
##  $ gender : Factor w/ 2 levels "Male","Female": 2 2 2 2 2 1 1 1 2 2 ...
##  $ age    : int  42 37 42 27 42 48 57 41 49 75 ...
##  $ wfns   : Ord.factor w/ 5 levels "1"<"2"<"3"<"4"<..: 1 1 1 1 3 2 5 4 1 2 ...
##  $ s100b  : num  0.13 0.14 0.1 0.04 0.13 0.1 0.47 0.16 0.18 0.1 ...
##  $ ndka   : num  3.01 8.54 8.09 10.42 17.4 ...
```



pROC: ROC analysis of aSAH data set







Exercise 5

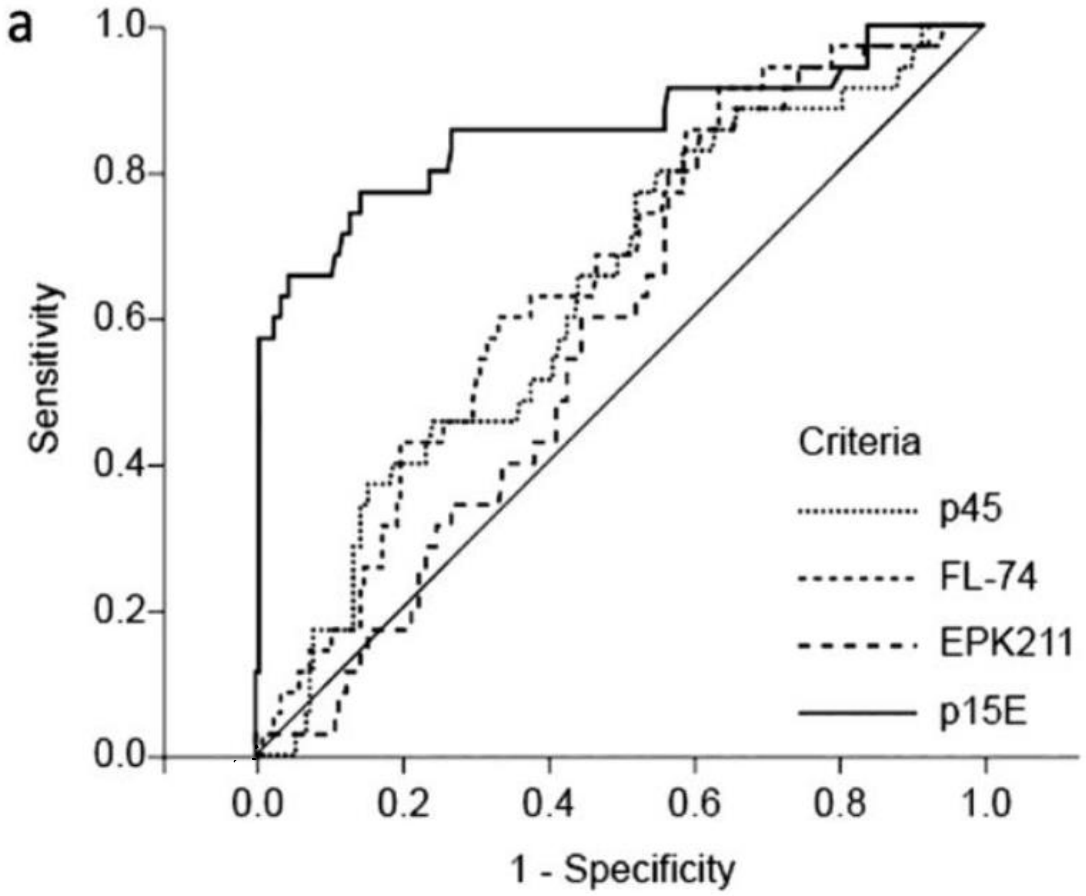
American Society for Microbiology
Journal of Clinical Microbiology
Volume 52, Issue 6, June 2014, Pages 2046-2052
https://doi.org/10.1128/JCM.02584-13

Clinical Veterinary Microbiology

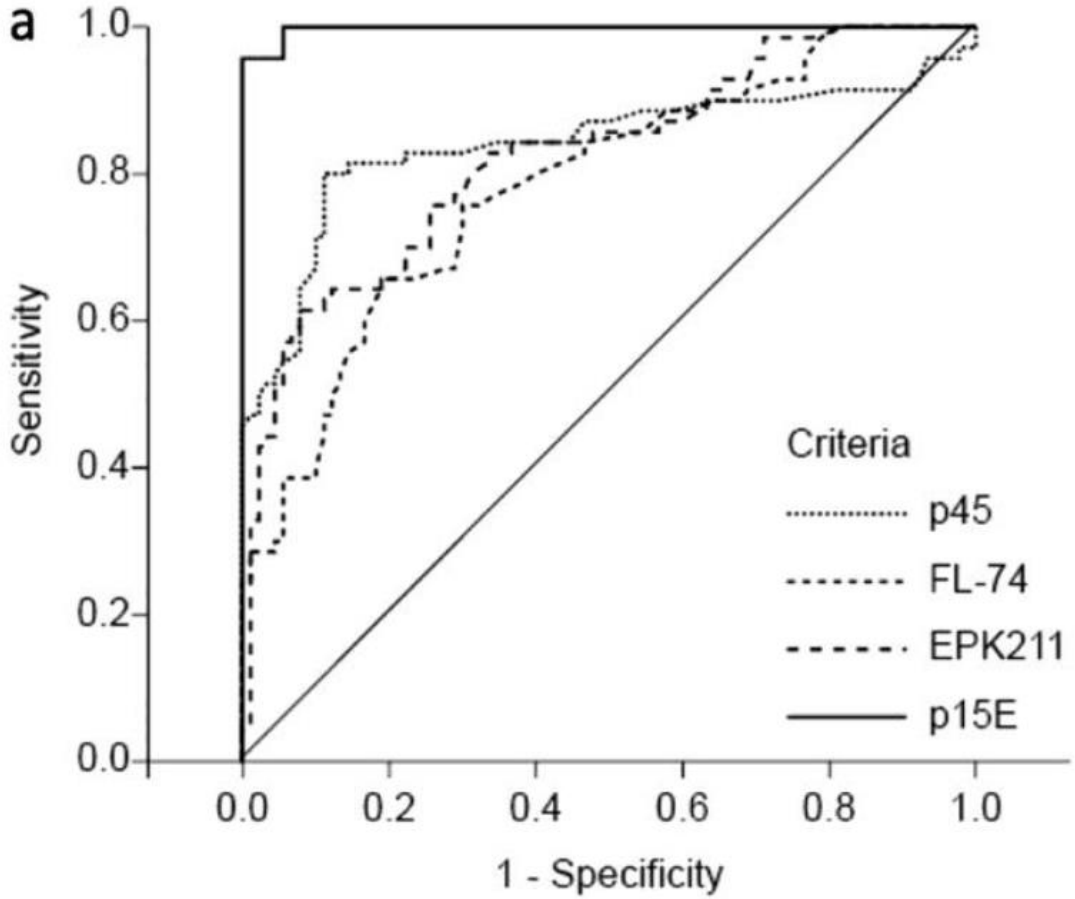
Detection of Antibodies to the Feline Leukemia Virus (FeLV) Transmembrane Protein p15E: an Alternative Approach for Serological FeLV Detection Based on Antibodies to p15E

Eva Boenzli[a,*], Maik Hadorn[b], Sonja Hartnack[c], Jon Huder[d], Regina Hofmann-Lehmann[a], and Hans Lutz[a]

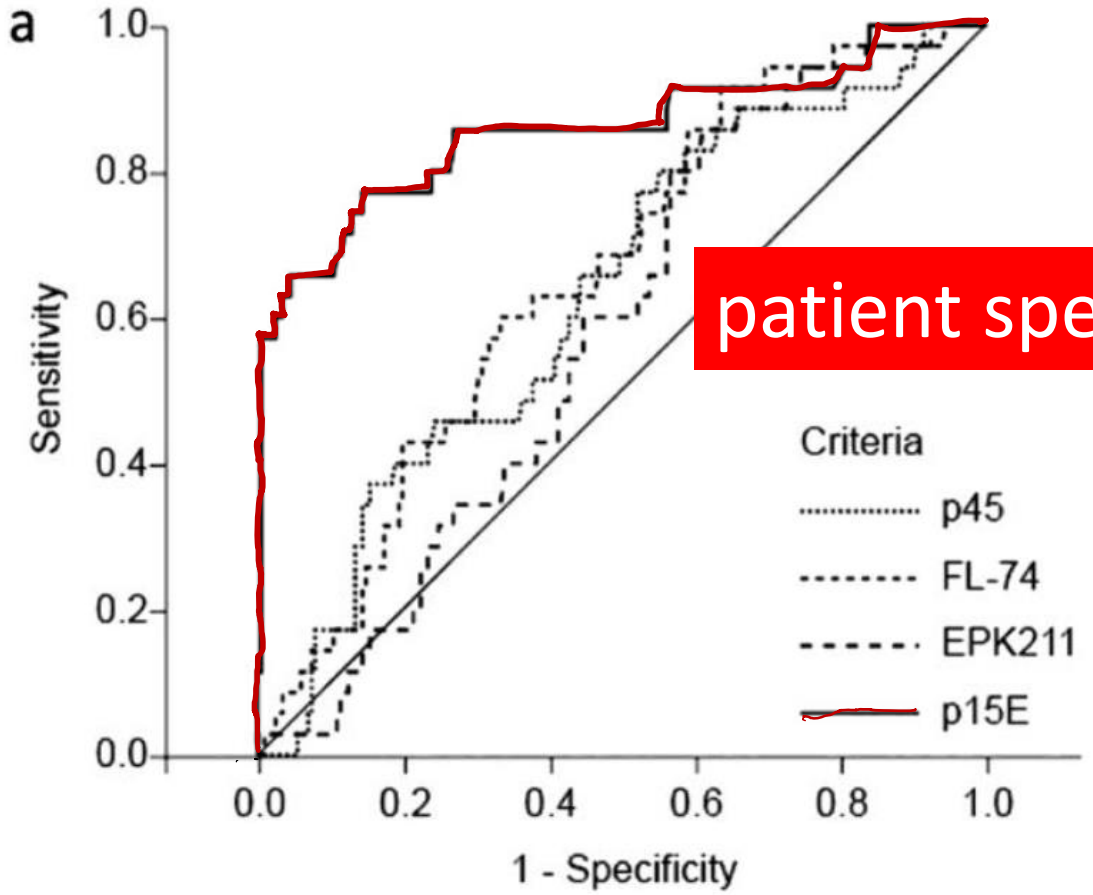
naturally infected



experimentally infected

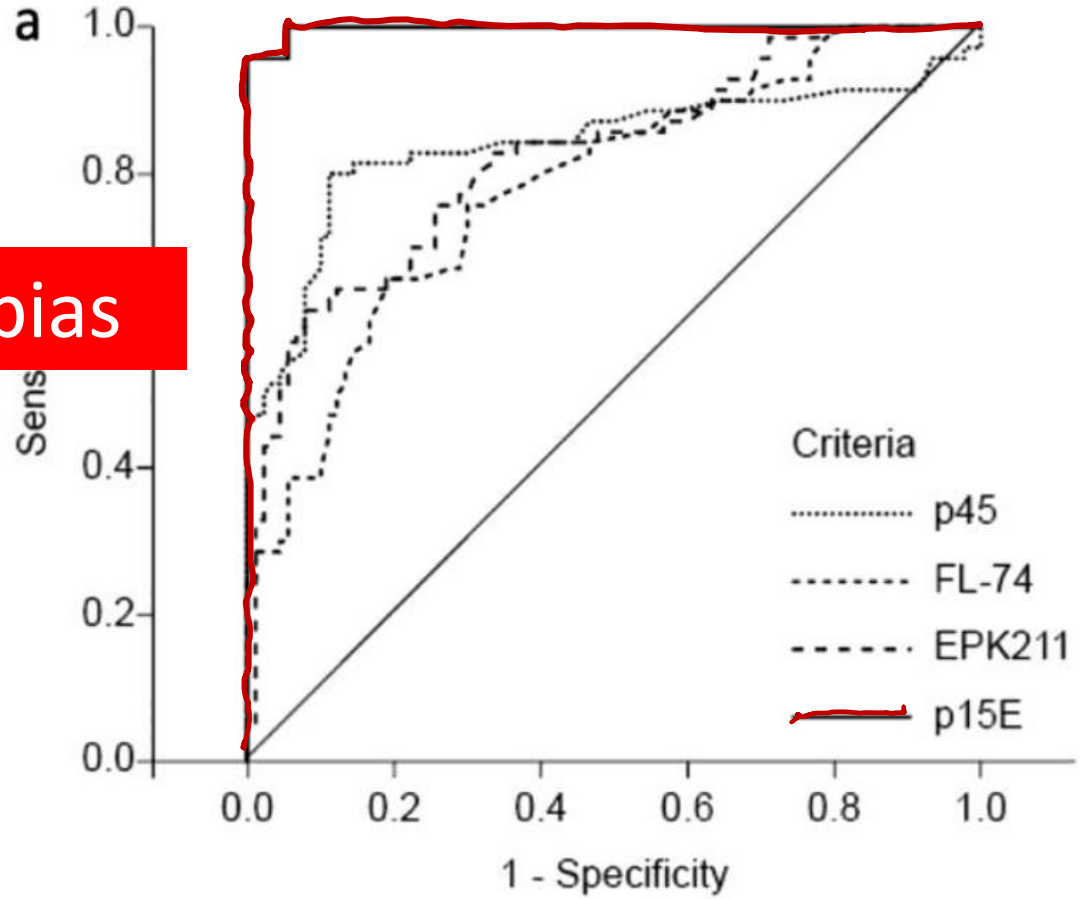


naturally infected



patient spectrum bias

experimentally infected



Keys to validity in diagnostic test studies:

«… independent, blind comparison of test results with a reference standard among a consecutive series of patients suspected (but not known) to have the target disorder».

Sackett and Haynes (2002),BMJ 2002;324:539

reporting guidelines

Search for reporting guidelines

Use your browser's Back button to return to your search results

STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies

Reporting guideline provided for? (i.e. exactly what the authors state in the paper)

Studies of diagnostic accuracy.

STARD 2015 checklist (PDF) STARD 2015 flow diagram (PDF)

STARD 2015 checklist (Word)

Full bibliographic reference

Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, LijmerJG Moher D, Rennie D, de Vet HCW, Kressel HY, Rifai N, Golub RM, Altman DG, Hooft L, Korevaar DA, Cohen JF, For the STARD Group. STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies.

This guideline was published simultaneously in 3 journals. You can read the guideline in any of these journals using the links below.

BMJ. 2015;351:h5527. PMID: 26511519
Radiology. 2015:151516. PMID: 26509226
Clinical Chemistry. 2015. pii: clinchem.2015.246280. PMID: 26510957

Reporting guidelines for main study types

Randomised trials	CONSORT	Extensions
Observational studies	STROBE	Extensions
Systematic reviews	PRISMA	Extensions
Study protocols	SPIRIT	PRISMA-P
Diagnostic/prognostic studies	STARD	TRIPOD
Case reports	CARE	Extensions
Clinical practice guidelines	AGREE	RIGHT
Qualitative research	SRQR	COREQ
Animal pre-clinical studies	ARRIVE	
Quality improvement studies	SQUIRE	Extensions
Economic evaluations	CHEERS	

Translations

Some reporting guidelines are also available in languages other than English. Find out more in our Translations section.

Exercise 6

Exercise 6: STARD guidelines

Which STARD items have been considered?
Read the paper and assess the adherence to the STARD guidelines (group work)

RESEARCH METHODS & REPORTING

BMJ2015;351:h5527 doi: 10.1136/bmj.h5527

STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies

Incomplete reporting has been identified as one of the sources of avoidable waste in biomedical research.[1] Since STARD was initiated, several other initiatives have been undertaken to

Section & Topic	No	Item
TITLE OR ABSTRACT		
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)
ABSTRACT		
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)
INTRODUCTION		
	3	Scientific and clinical background, including the intended use and clinical role of the index test
	4	Study objectives and hypotheses
METHODS		
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)
Participants	**6**	Eligibility criteria
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)
	8	Where and when potentially eligible participants were identified (setting, location and dates)
	9	Whether participants formed a consecutive, random or convenience series
Test methods	**10a**	Index test, in sufficient detail to allow replication
	10b	Reference standard, in sufficient detail to allow replication
	11	Rationale for choosing the reference standard (if alternatives exist)
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test
	13b	Whether clinical information and index test results were available to the assessors of the reference standard
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy
	15	How indeterminate index test or reference standard results were handled
	16	How missing data on the index test and reference standard were handled
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory
	18	Intended sample size and how it was determined

STARD

RESULTS		
Participants	**19**	Flow of participants, using a diagram
	20	Baseline demographic and clinical characteristics of participants
	21a	Distribution of severity of disease in those with the target condition
	21b	Distribution of alternative diagnoses in those without the target condition
	22	Time interval and any clinical interventions between index test and reference standard
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)
	25	Any adverse events from performing the index test or the reference standard
DISCUSSION		
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability
	27	Implications for practice, including the intended use and clinical role of the index test
OTHER INFORMATION		
	28	Registration number and name of registry
	29	Where the full study protocol can be accessed
	30	Sources of funding and other support; role of funders

Thank you.