

DIAGNOSTIC TEST ASSESSMENT: SENSITIVITY, SPECIFICITY AND PREDICTIVE VALUE

Professor René Rizzoli MD

Geneva University Hospitals & Faculty of Medicine

Geneva, Switzerland

www.iofbonehealth.org

RELIABILITY AND VALIDITY OF A DIAGNOSTIC TEST

- Introduction
- Reliability of a diagnostic test
- Validity of a diagnostic test

INTRODUCTION

- Most diagnostic tests provide incomplete information.
- Strengths of a diagnostic test: reliability and validity
- The reliability gives information on:
 - the variability of test results is linked to the test itself, its application and its interpretation.
- The validity provides information on:
 - the variability of test results is linked to the presence or absence of the disease in the tested individual.

DEFINITIONS

- **Reliability** of a diagnostic test: ability to provide the same degree in the same subjects, especially to classify subjects in the same way regardless of the circumstances of application or interpretation of the test.
- **Validity** of a diagnostic test: ability to measure what it is supposed to measure, meaning the presence or absence of disease or phenomenon to be detected or measured.

RELIABILITY OF A DIAGNOSTIC TEST

- The test result is a **categorical** variable
- The test result is a quantitative variable

THE TEST RESULT IS A CATEGORICAL VARIABLE

Kappa

Presentation of the results of a diagnostic test whose result is a dichotomous variable and is performed by two observers

Observer 2	Observer 1		Total
	Positive test	Negative test	
Positive test	a	b	$a+b$
Negative test	c	d	$c+d$
Total	$a+c$	$b+d$	$a+b+c+d = N$

THE TEST RESULT IS A CATEGORICAL VARIABLE

In the example below, the diagnoses of the two observers are the same, are concordant for 75 subjects. The proportion of concordant results is:

$$p_0 = 75/100 = 0.75 \text{ or } 75\%.$$

Comparison of diagnostic of malignant melanoma by two pathologists using a panel of 100 slides

Pathologist 2	Pathologist 1		Total
	Malignant melanoma	Other tumor	
Malignant melanoma	35	20	55
Other tumor	5	40	45
Total	40	60	100

THE TEST RESULT IS A CATEGORICAL VARIABLE

Comparison of diagnostic of malignant melanoma by two pathologists using a panel of 100 slides

Pathologist 2	Pathologist 1		Total
	Malignant melanoma	Other tumor	
Malignant melanoma	20	30	50
Other tumor	20	30	50
Total	40	60	100

The second pathologist randomly assigns his/her diagnosis: 50 “Malignant melanoma” and 50 “Other tumor”



$p_0 = 50/100 = 0.50$ or 50% ! Agreement by chance !

THE TEST RESULT IS A CATEGORICAL VARIABLE

Cohen's Kappa= Real agreement

The Kappa coefficient measures the actual agreement beyond chance related agreement, by relating it to the possible agreement beyond the agreement occurring by chance. It could be calculated by a computer.

THE TEST RESULT IS A CATEGORICAL VARIABLE

Interpretation of the value of Kappa coefficient

Quality of inter-observer agreement, measured from the Kappa coefficient.

Agreement	Kappa
Almost perfect agreement	> 0.81
Substantial agreement	$0.80 - 0.61$
Moderate agreement	$0.60 - 0.41$
Fair agreement	$0.40 - 0.21$
Slight agreement	$0.20 - 0.00$
Less than chance agreement	< 0.00

THE TEST RESULT IS A CATEGORICAL VARIABLE

Example

Evaluation of the agreement between clinicians in the diagnosis of dehydration in infants or young children.

Evaluated clinical signs	Range of Kappa values
Depressed fontanelle	0.10 to 0.27
Dry mucous membranes	0.28 to 0.59
Sunken eyes	0.06 to 0.59
Lack of tears	0.12 to 0.75
Anomaly of respiratory rhythm	-0.04 to 0.40
Perfusion abnormality of extremities	0.23 to 0.66

RELIABILITY OF A DIAGNOSTIC TEST

- The test result is a categorical variable
- The test result is a **quantitative** variable

THE TEST RESULT IS A QUANTITATIVE VARIABLE

The intraclass correlation coefficient (ICC) is a measure of the inter-rater reliability on quantitative data

- ICC values between 1 and -1.
- The closer the ICC is to 1, the better the agreement between the two observers, or for any source of variation of the measure, the more reliable the measurement.
- The closer ICC is to 0, the worse the agreement
- A value of -1 means a perfect disagreement.

PARAMETERS MEASURING THE DIAGNOSTIC VALIDITY

- The test result is a **categorical** variable
- The test result is a quantitative variable

THE TEST RESULT IS A CATEGORICAL VARIABLE

Presentation of the results of a validity study of a diagnostic test whose result is a dichotomous variable

Test evaluated	Reference test		
	Diseased	Non diseased	Total
Positive (diseased)	TP	FP	TP + FP
Negative (non diseased)	FN	TN	TN + FN
Total	TP + FN	FP + TN	TP+FN+FP+TN

TP: true positive, TN: true negative, FP: false positive, FN: false negative

THE TEST RESULT IS A CATEGORICAL VARIABLE

- The sensitivity $Se = \frac{TP}{TP + FN}$
- The specificity $Sp = \frac{TN}{TN + FP}$
- The positive predictive value $PPV = \frac{TP}{TP + FP}$
- The negative predictive value $NPV = \frac{TN}{TN + FN}$

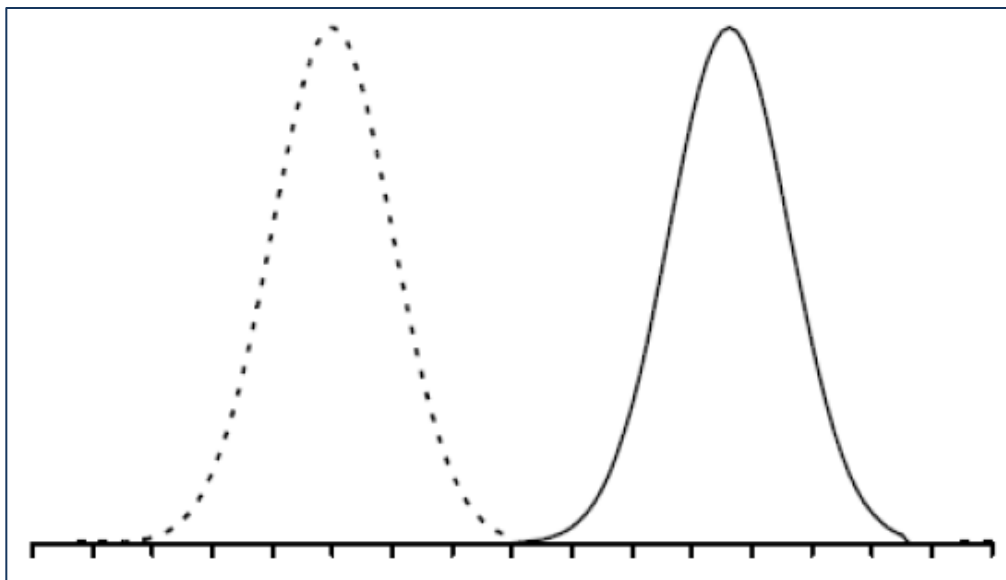
TP: true positive, TN: true negative, FP: false positive, FN: false negative

PARAMETERS MEASURING THE DIAGNOSTIC VALIDITY

- The test result is a categorical variable
- The test result is a **quantitative** variable

THE TEST RESULT IS A QUANTITATIVE VARIABLE

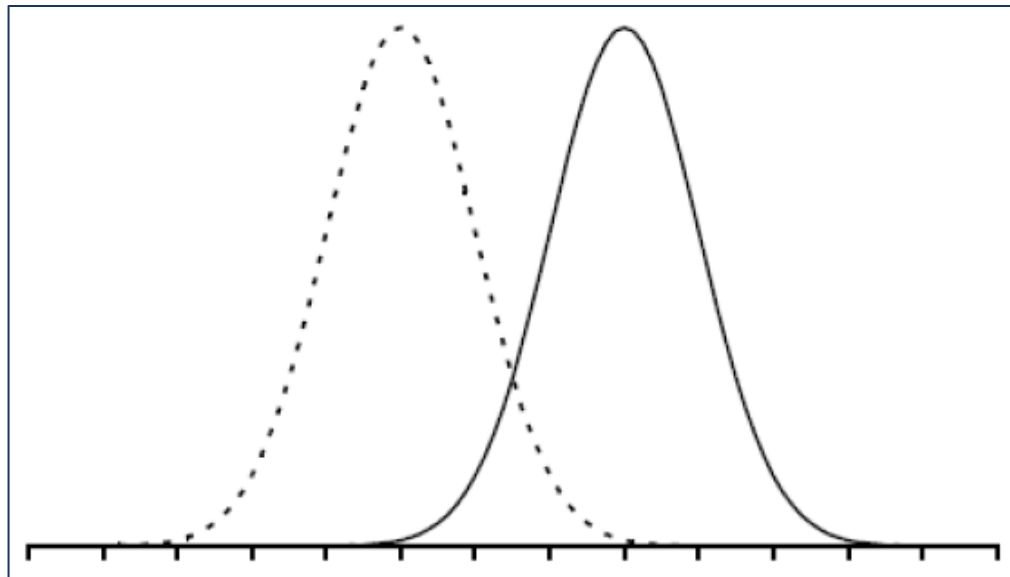
Distribution of the quantitative results of a diagnostic test for diseased patients (solid lines) and non-diseased (dashed): ideal situation for the choice of a threshold.



Threshold of the test's positivity

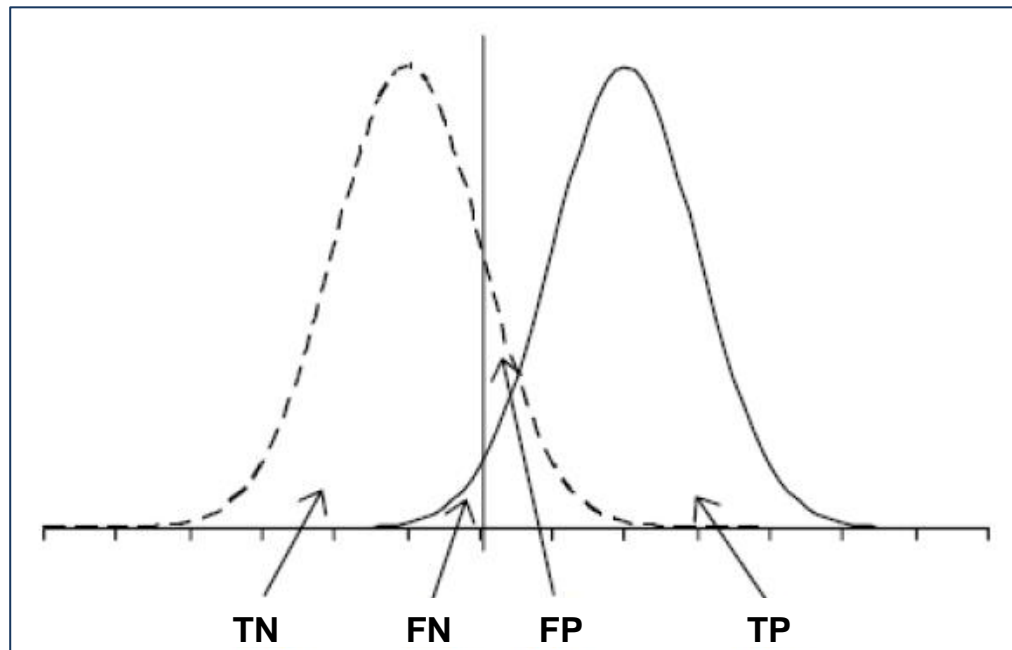
THE TEST RESULT IS A QUANTITATIVE VARIABLE

Distribution of the quantitative results of a diagnostic test for diseased patients (solid lines) and non-diseased (dotted lines): likely situation.



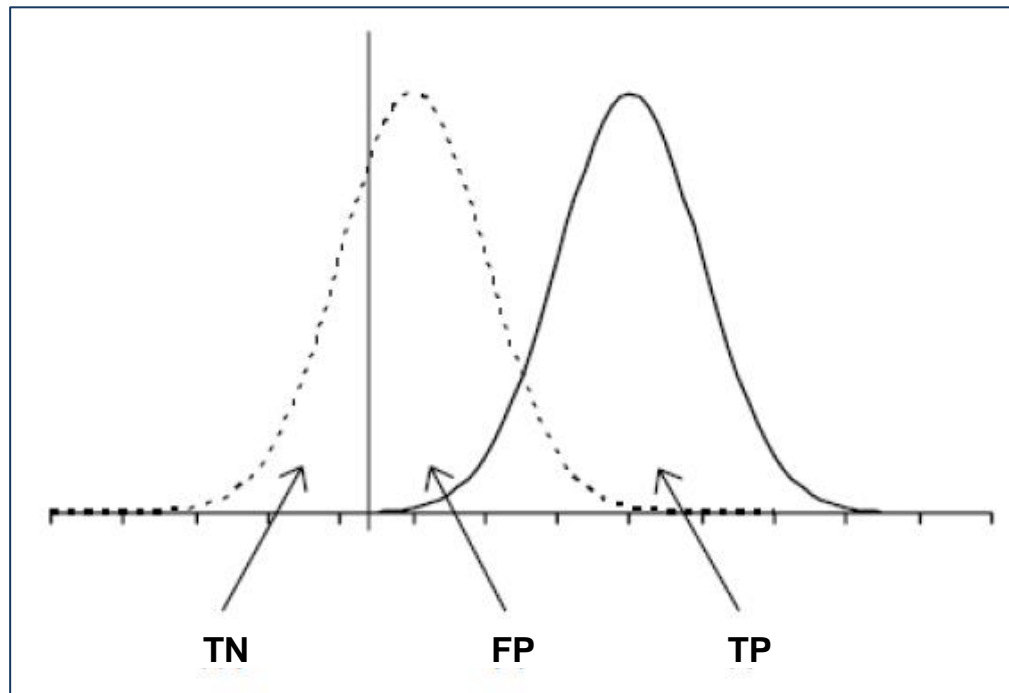
THE TEST RESULT IS A QUANTITATIVE VARIABLE

Distribution of the quantitative results of a diagnostic test for diseased patients (solid lines) and non-diseased (dashed): consequences of applying a threshold on any diagnosis.



THE TEST RESULT IS A QUANTITATIVE VARIABLE

Distribution of the quantitative results of a diagnostic test for diseased patients (solid lines) and non-diseased (dashed): implications for diagnosis when applying a low threshold eliminating false negatives.



THE TEST RESULT IS A QUANTITATIVE VARIABLE

Example

We want to evaluate the discriminating power of two prostate cancer screening tests: free PSA and the ratio free/bound PSA .

For this purpose, we measure these two values in subjects, already knowing if they have (1) or not (0) prostate cancer (biopsy).

THE TEST RESULT IS A QUANTITATIVE VARIABLE

Example

Subject #	Real status	Total PSA (micro g/l)	Free PSA (%)
1	0	1,63	22,38
2	0	2,03	20,00
3	0	4,06	4,74
4	0	2,04	17,57
5	0	4,72	14,40
6	0	4,91	7,38
7	0	5,36	8,97
8	0	8,68	28,00
9	0	6,26	10,39
10	0	5,21	5,06
11	1	10,58	20,00
12	1	15,10	34,76
13	1	15,21	48,84
14	1	15,80	33,38
15	1	13,60	47,79
16	1	3,89	41,73
17	1	1,78	28,27
18	1	7,31	6,74
19	1	8,45	49,79
20	1	4,88	36,71

THE TEST RESULT IS A QUANTITATIVE VARIABLE

Example

Based on the positive
threshold value > 6 microg/l
for total PSA

Se=

Sp=

VPP=

VPN=

Subject #	Real status	Total PSA (micro g/l)
1	0	1,63
2	0	2,03
3	0	4,06
4	0	2,04
5	0	4,72
6	0	4,91
7	0	5,36
8	0	8,68
9	0	6,26
10	0	5,21
11	1	10,58
12	1	15,10
13	1	15,21
14	1	15,80
15	1	13,60
16	1	3,89
17	1	1,78
18	1	7,31
19	1	8,45
20	1	4,88

THE TEST RESULT IS A QUANTITATIVE VARIABLE

Example

Based on the positive threshold value > 6 microg/l for total PSA

$$Se = 7 / (7 + 3) = 70\%$$

$$Sp = 8 / (8 + 2) = 80\%$$

$$PPV = 7 / (7 + 2) = 77\%$$

$$NPV = 8 / (8 + 3) = 73\%$$

Subject #	Real status	Total PSA (micro g/l)	TEST
1	0	1,63	TN
2	0	2,03	TN
3	0	4,06	TN
4	0	2,04	TN
5	0	4,72	TN
6	0	4,91	TN
7	0	5,36	TN
8	0	8,68	FP
9	0	6,26	FP
10	0	5,21	TN
11	1	10,58	TP
12	1	15,10	TP
13	1	15,21	TP
14	1	15,80	TP
15	1	13,60	TP
16	1	3,89	FN
17	1	1,78	FN
18	1	7,31	TP
19	1	8,45	TP
20	1	4,88	FN

THE TEST RESULT IS A QUANTITATIVE VARIABLE

Example

Based on the positive threshold value $\geq 20\%$ for free PSA:

$$Se = 9 / (9 + 1) = 90\%$$

$$Sp = 8 / (8 + 2) = 80\%$$

$$PPV = 9 / (9 + 2) = 82\%$$

$$NPV = 7 / (7 + 1) = 88\%$$

Subject #	Real status	Free PSA (micro g/l)	TEST
1	0	22,38	FP
2	0	20,00	FP
3	0	4,74	TN
4	0	17,57	TN
5	0	14,40	TN
6	0	7,38	TN
7	0	8,97	TN
8	0	28,00	FP
9	0	10,39	TN
10	0	5,06	TN
11	1	20,00	TP
12	1	34,76	TP
13	1	48,84	TP
14	1	33,38	TP
15	1	47,79	TP
16	1	41,73	TP
17	1	28,27	TP
18	1	6,74	FN
19	1	49,79	TP
20	1	36,71	TP

THE TEST RESULT IS A QUANTITATIVE VARIABLE

Example

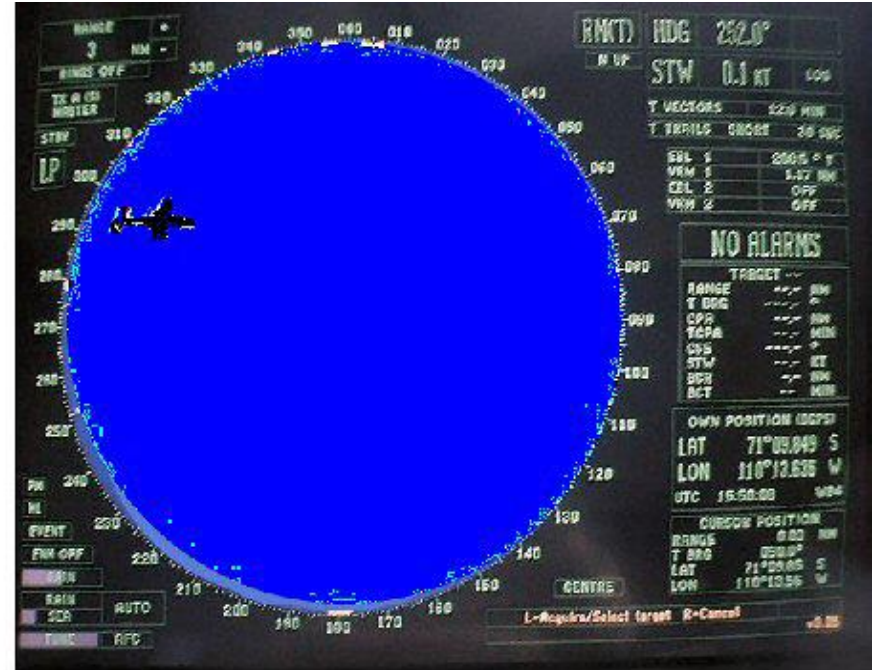
How to compare these values with each other?

⇒ Se, Sp, VPP, VPN vary depending on the threshold value used for both tests

ROC CURVES (RECEIVER OPERATING CHARACTERISTIC)



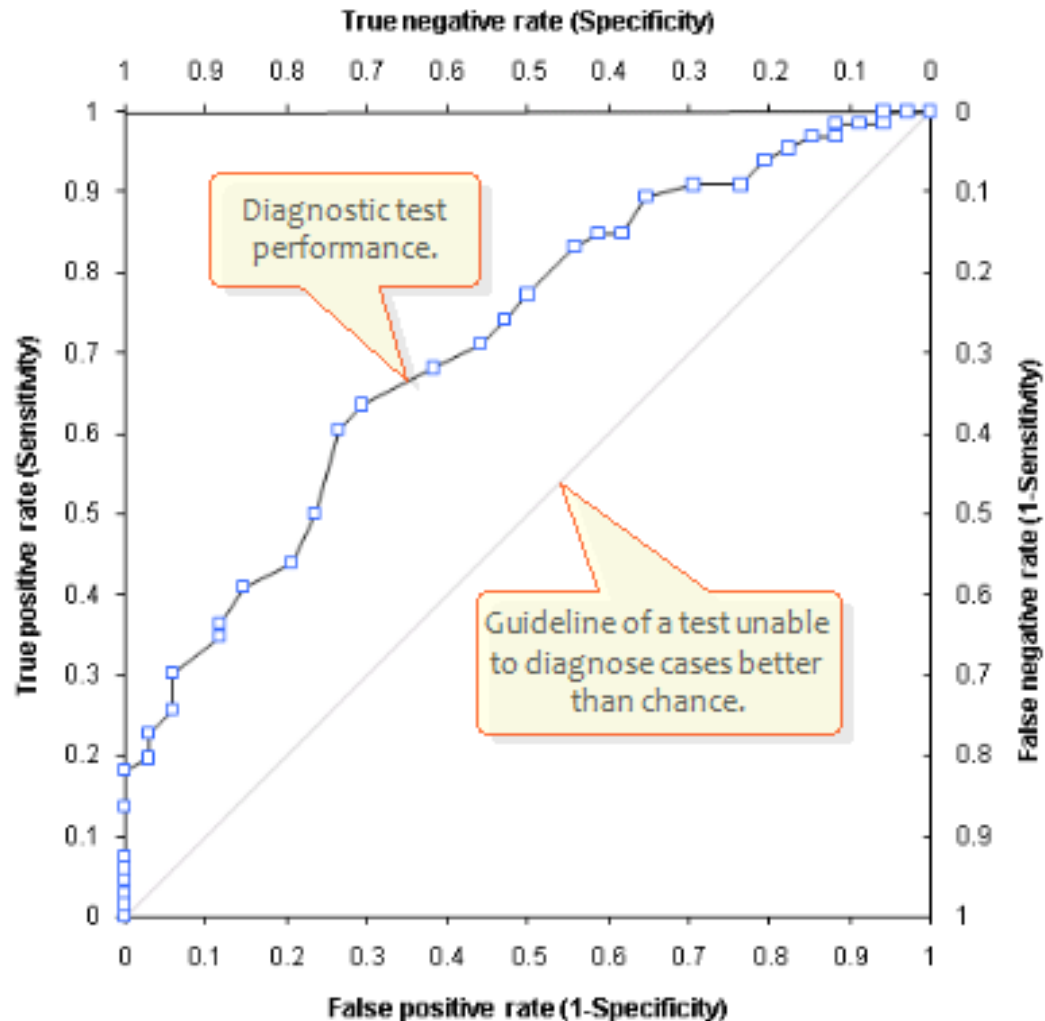
Too sensitive



Too specific

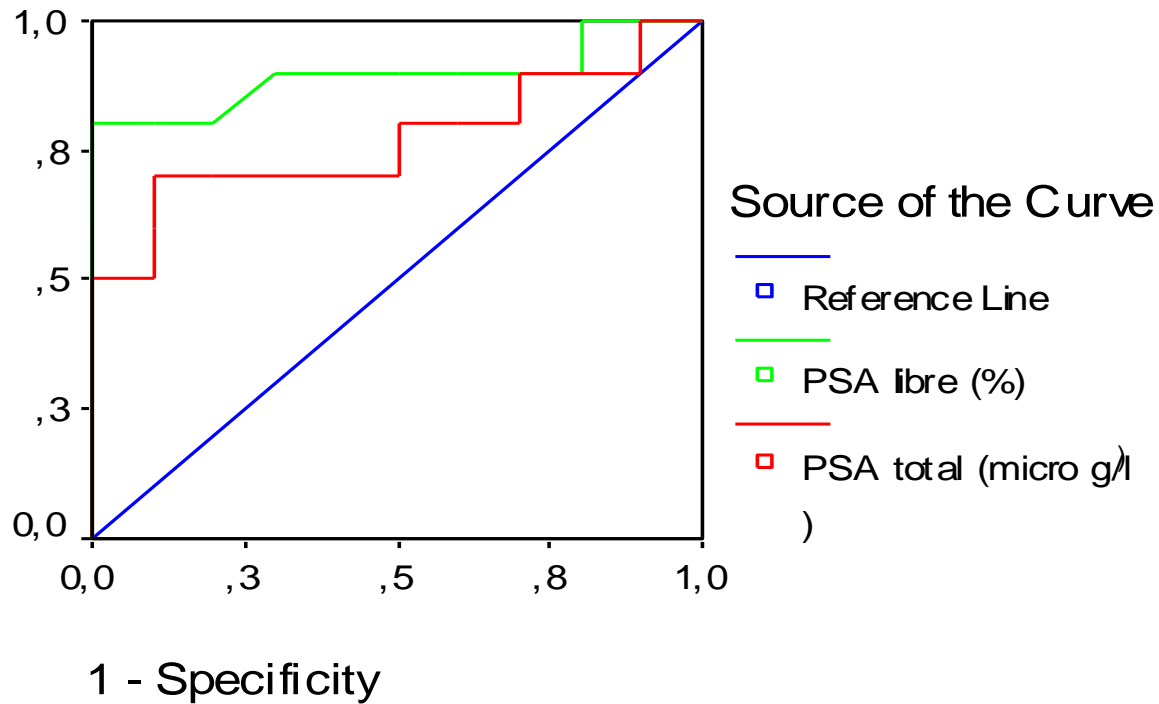
Compromise ?

ROC CURVES (RECEIVER OPERATING CHARACTERISTIC)



APPLICATION EXAMPLE

ROC Curve



Conclusion:
PSA libre (%)
has the highest
area under the
curve: so it is
better as a
diagnostic test

DXA AND FRACTURE RISK PREDICTION

- Osteoporosis diagnosis is based on areal BMD
- areal BMD is a good predictor of bone strength and of fracture risk
- 50% of hip fractures are occurring in patients without areal BMD-determined osteoporosis diagnosis
- -> areal BMD may not be a sensitive tool to screen population for osteoporosis

CONCLUSION

Reliability:

- Test-retest
- Interobserver agreement
 - Kappa
 - ICC
- Validity :
 - Diseased / non-diseased
 - Sensitivity – Specificity
 - PPV –NPV
 - ROC