University of York Department of Health Sciences

Measurement in Health and Disease

Interpretation of Diagnostic Tests

Martin Bland

http://martinbland.co.uk/

Some artificial test and diagnosis data

Test 1	Disease diagnosis positive	Disease diagnosis negative	Total	Agreement kappa	J a/(a+b+c)
positive	4	5	9		
negative	1	90	91		
Total	5	95	100	**0.54**	**0.40**

Test 2	Disease diagnosis positive	Disease diagnosis negative	Total		
positive	0	0	0		
negative	5	95	100		
Total	5	95	100	**0.00**	**0.00**

Test 3	Disease diagnosis positive	Disease diagnosis negative	Total		
positive	2	0	2		
negative	3	95	98		
Total	5	95	100	**0.56**	**0.40**

Some artificial test and diagnosis data

Test 1	Disease diagnosis positive	Disease diagnosis negative	Total
positive	**4**	5	9
negative	1	90	91
Total	5	95	100

More true positives

Test 2	Disease diagnosis positive	Disease diagnosis negative	Total
positive	0	0	0
negative	5	95	100
Total	5	95	100

Test 3	Disease diagnosis positive	Disease diagnosis negative	Total
positive	2	**0**	2
negative	3	95	98
Total	5	95	100

Fewer false positives

Sensitivity and Specificity

There is no one simple index which enables us to compare different tests in all the ways we would like.

Two things we need to measure:

- how good the test is at finding disease positives,
- how good the test is at excluding disease negatives.

$$\text{sensitivity} = \frac{\text{disease} + \text{ve who are also test} + \text{ve}}{\text{disease} + \text{ve}}$$

$$\text{specificity} = \frac{\text{disease } - \text{ve who are also test } - \text{ve}}{\text{disease } - \text{ve}}$$

Some artificial test and diagnosis data

Test 1	Disease diagnosis positive	Disease diagnosis negative	Total	**Sensitivity**	**Specificity**
positive	4	5	9		
negative	1	90	91		
Total	5	95	100	**0.80**	**0.95**

Test 2	Disease diagnosis positive	Disease diagnosis negative	Total		
positive	0	0	0		
negative	5	95	100		
Total	5	95	100	**0.00**	**1.00**

Test 3	Disease diagnosis positive	Disease diagnosis negative	Total		
positive	2	0	2		
negative	3	95	98		
Total	5	95	100	**0.40**	**1.00**

Example: many alcoholics have evidence at X-ray of past rib fractures.

Would this be of any value in the detection of alcoholism in patients?

74 patients with alcoholic liver disease, 20 had evidence of at least one past fracture on chest X-ray.

Sensitivity 20/74 = 0.27.

In a control group of 181 patients with non-alcoholic liver disease or gastro-intestinal disorders, 6 had evidence of at least one fracture.

Specificity (181-6)/181 = 0.97.

Example: alcoholism and past rib fractures at X-ray.

74 patients with alcoholic liver disease, 20 had evidence of at least one past fracture on chest X-ray.

Sensitivity 20/74 = 0.27.

181 controls, 6 had evidence of at least one fracture.

Specificity (181-6)/181 = 0.97.

11 alcoholics had evidence of bilateral or multiple fractures.

Sensitivity 11/74 = 0.15.

Two controls had evidence of bilateral or multiple fractures

Specificity (181-2)/181 = 0.99.

More stringent test was more specific and less sensitive.

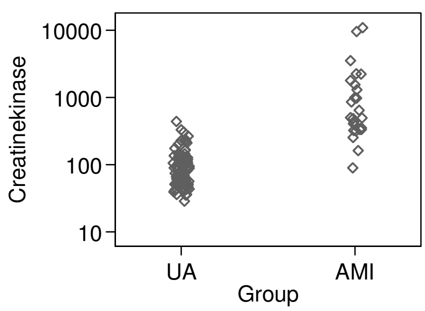
ROC curves

Sometimes a test is based on a continuous variable.

Creatinekinase in patients with unstable angina and acute myocardial infarction (AMI) (data of Frances Boa)

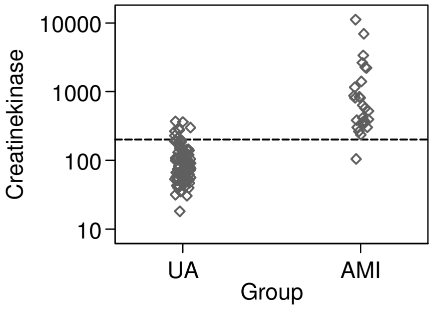
Unstable angina							AMI	
23	48	62	83	104	130	307	90	648
33	49	63	84	105	139	351	196	894
36	52	63	85	105	150	360	302	962
37	52	65	86	107	155		311	1015
37	52	65	88	108	157		325	1143
41	53	66	88	109	162		335	1458
41	54	67	88	111	176		347	1955
41	57	71	89	114	180		349	2139
42	57	72	91	116	188		363	2200
42	58	72	94	118	198		377	3044
43	58	73	94	121	226		390	7590
45	58	73	95	121	232		398	11138
47	60	75	97	122	257		545	
48	60	80	100	126	257		577	
48	60	80	103	130	297		629	

Creatinekinase in patients with unstable angina and acute myocardial infarction (AMI)



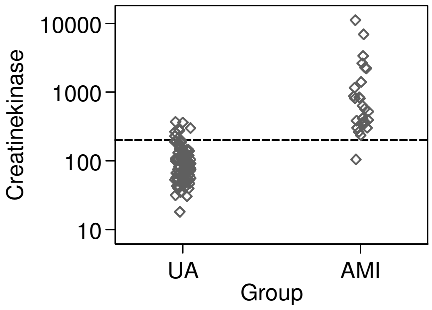
Need a cutoff to make a diagnosis.

Above = AMI, below = UA.

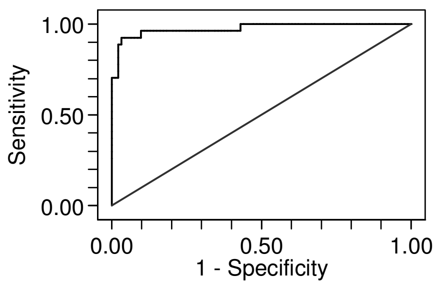


Ck=100: sensitivity = 0.96 and specificity 0.62

Ck=200: sensitivity = 0.93 and specificity 0.91



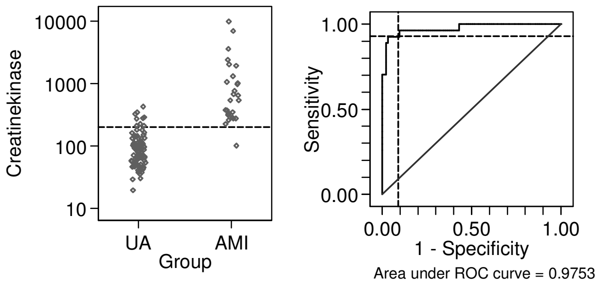
Plot sensitivity against specificity (usually 1 – specificity) to give the Receiver Operating Characteristic (ROC) curve.



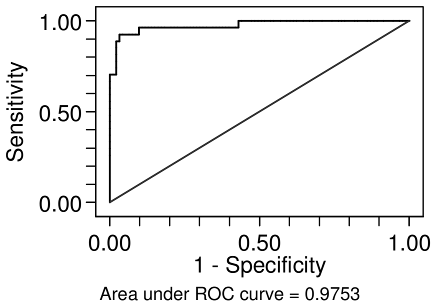
Area under ROC curve = 0.9753

Each point on the curve correspond to a cut-off.

Ck=200: sensitivity = 0.93 and specificity 0.91.



Area under the ROC curve estimates the probability that an observation from a member of one population (disease positive) chosen at random will exceed a member of the other population (disease negative).



Positive and Negative Predictive Value

Positive predictive value or **PPV** = probability that a subject who is test positive will also be a disease positive.

Depends on the prevalence of the condition.

If test and true diagnosis data are from a simple random sample of the population in which we are interested, we can estimate these as simple proportions.

If this is not the case, the usual situation, we can calculate the PPV for any population prevalence.

PPV for any population prevalence.

Sensitivity = p_{sens}, specificity = p_{spec}, prevalence = p_{prev}.

Probability (disease positive and test positive) = $p_{\text{prev}} \times p_{\text{sens}}$.

Probability (disease negative and test positive) =

$(1 - p_{\text{prev}}) \times (1 - p_{\text{spec}})$.

Total probability (test positive) = $p_{\text{prev}} \times p_{\text{sens}} + (1 - p_{\text{prev}}) \times (1 - p_{\text{spec}})$.

Positive predictive value is the proportion of test positives who are disease positives:

$$\text{PPV} = \frac{p_{\text{prev}} p_{\text{sens}}}{p_{\text{prev}} p_{\text{sens}} + (1 - p_{\text{prev}})(1 - p_{\text{spec}})}$$

$$\text{PPV} = \frac{p_{\text{prev}} p_{\text{sens}}}{p_{\text{prev}} p_{\text{sens}} + (1 - p_{\text{prev}})(1 - p_{\text{spec}})}$$

In screening situations the prevalence is almost always small and the PPV is low. Suppose we have a test which is both sensitive and specific, p_{sens} = 0.95 and p_{spec} = 0.95, and the disease has prevalence p_{prev} = 0.01 (1%). Then

$$\text{PPV} = \frac{0.01 \times 0.95}{0.01 \times 0.95 + (1 - 0.01) \times (1 - 0.95)} = 0.16$$

so only 16% of test positives would be disease positives.

The probability that a subject who is test negative will not have the disease is the **negative predictive value** or **NPV**.

$$\text{NPV} = \frac{(1 - p_{\text{prev}}) p_{\text{spec}}}{p_{\text{prev}}(1 - p_{\text{sens}}) + (1 - p_{\text{prev}}) p_{\text{spec}}}$$

NPV is usually high.

PPV and NPV are what we really want to know to interpret a test result, but they are properties of the test in a particular population, not just of the test.

There are other statistics quoted for tests, such as the odds ratio and the likelihood ratio, but they are beyond the scope of this course.