

Plotting and Interpreting an ROC Curve

[Previous Section](#) | [Main Menu](#) | [Next Section](#) |

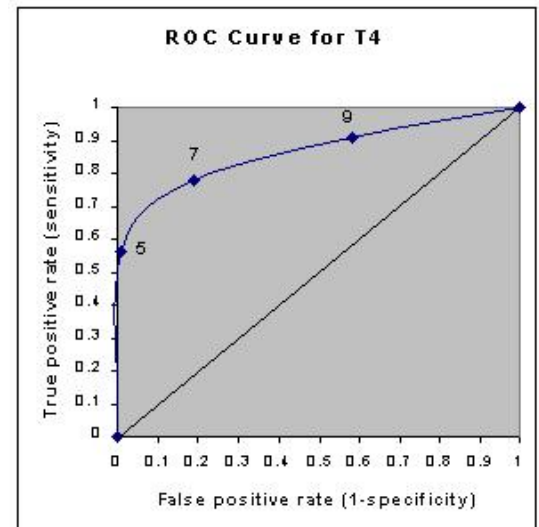
This section continues the hypothyroidism example started in the the previous section. We showed that the table at left can be summarized by the operating characteristics at right:

T4 value	Hypothyroid	Euthyroid
5 or less	18	1
5.1 - 7	7	17
7.1 - 9	4	36
9 or more	3	39
Totals:	32	93

Cutpoint	Sensitivity	Specificity
5	0.56	0.99
7	0.78	0.81
9	0.91	0.42

The operating characteristics (above right) can be reformulated slightly and then presented graphically as shown below to the right:

Cutpoint	True Positives	False Positives
5	0.56	0.01
7	0.78	0.19
9	0.91	0.58



This type of graph is called a **Receiver Operating Characteristic curve** (or ROC curve.) It is a plot of the true positive rate against the false positive rate for the different possible cutpoints of a diagnostic test.

An ROC curve demonstrates several things:

1. It shows the tradeoff between sensitivity and specificity (any increase in sensitivity will be accompanied by a decrease in specificity).
2. The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test.
3. The closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the test.
4. The slope of the tangent line at a cutpoint gives the likelihood ratio (LR) for that value of the test. You can check this out on the graph above. Recall that the LR for $T4 < 5$ is 52. This corresponds to the far left, steep portion of the curve. The LR for $T4 > 9$ is 0.2. This corresponds to the far right, nearly horizontal portion of the curve.
5. The area under the curve is a measure of test accuracy. This is discussed further in the [next section](#).

[Previous Section](#) | [Main Menu](#) | [Next Section](#) |

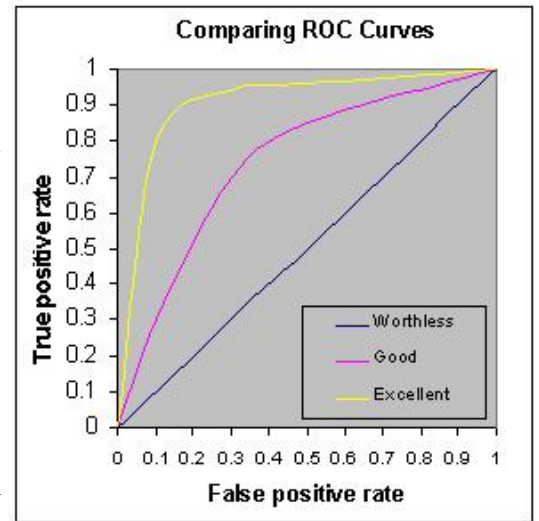
The Area Under an ROC Curve

[Previous Section](#) | [Main Menu](#) | [Next Section](#) |

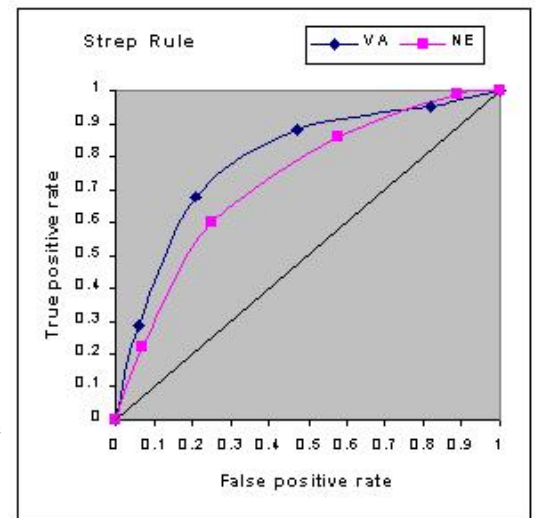
The graph at right shows three ROC curves representing excellent, good, and worthless tests plotted on the same graph. The accuracy of the test depends on how well the test separates the group being tested into those with and without the disease in question. Accuracy is measured by the area under the ROC curve. An area of 1 represents a perfect test; an area of .5 represents a worthless test. A rough guide for classifying the accuracy of a diagnostic test is the traditional academic point system:

- .90-1 = excellent (A)
- .80-.90 = good (B)
- .70-.80 = fair (C)
- .60-.70 = poor (D)
- .50-.60 = fail (F)

Recall the T4 data from the [previous section](#). The area under the T4 ROC curve is .86. The T4 would be considered to be "good" at separating hypothyroid from euthyroid patients.



ROC curves can also be constructed from clinical prediction rules. The graphs at right come from a study of how clinical findings predict strep throat (Wigton RS, Connor JL, Centor RM. Transportability of a decision rule for the diagnosis of streptococcal pharyngitis. Arch Intern Med. 1986;146:81-83.) In that study, the presence of tonsillar exudate, fever, adenopathy and the absence of cough all predicted strep. The curves were constructed by computing the sensitivity and specificity of increasing numbers of clinical findings (from 0 to 4) in predicting strep. The study compared patients in Virginia and Nebraska and found that the rule performed more accurately in Virginia (area under the curve = .78) compared to Nebraska (area under the curve = .73). These differences turn out not to be statistically different, however.



At this point, you may be wondering what this area number really means and how it is computed. The area measures **discrimination**, that is, the ability of the test to correctly classify those with and without the disease. Consider the situation in which patients are already correctly classified into two groups. You randomly pick one from the disease group and one from the no-disease group and do the test on both. The patient with the more abnormal test result should be the one from the disease group. The area under the curve is the percentage of randomly drawn pairs for which this is true (that is, the test correctly classifies the two patients in the random pair).

Computing the area is more difficult to explain and beyond the scope of this introductory material. Two methods are commonly used: a non-parametric method based on constructing trapezoids under the curve as an approximation of area and a parametric method using a maximum likelihood estimator to fit a smooth curve to the data points. Both methods are available as computer programs and give an estimate of area and standard error that can be used to compare different tests or the same test in different patient populations. For more on quantitative ROC analysis, see Metz CE. Basic principles of ROC analysis. Sem Nuc Med. 1978;8:283-298.

A final note of historical interest

You may be wondering where the name "Receiver Operating Characteristic" came from. ROC analysis is part of a field called "Signal Detection Theory" developed during World War II for the analysis of radar images. Radar operators had to decide whether a blip on the screen represented an enemy target, a friendly ship, or just noise. Signal detection theory measures the ability of radar receiver operators to make these important distinctions. Their ability to do so was called the Receiver Operating Characteristics. It was not until the 1970's that signal detection theory was recognized as useful for interpreting medical test results.

[Previous Section](#) | [Main Menu](#) | [Next Section](#) |