SP11 - PharmaSUG 2016

ROC Curve: Making way for correct diagnosis

Manoj Pandey, Ephicacy Lifescience Analytics Pvt. Ltd., Bangalore, India
Abhinav Jain, Ephicacy Consulting Group Inc., Iselin, NJ

ABSTRACT

The performance of a diagnostic test is based upon two factors, one how accurately it detects the disease and another how accurately it rules out the disease in healthy subjects. Scientifically, ROC (Receiver Operating Characteristics) curves help to evaluate the predictive accuracy of a diagnostic test. ROC Curves provides the ability to identify optimal cut-off points, evaluate performance comparison across multiple diagnostic tests and evaluate performance of a diagnostic test across multiple population samples. The property that makes this ROC Curves more desirable is that the indices of accuracy are least affected by arbitrarily chosen decision criteria. Calculations of Area Under Curve (AUC) and measures of accuracy determines the differentiating power of the test. This paper will focus on application of ROC Curves in clinical trial data analysis and deriving insights from ROC measures like sensitivity, specificity, AUC, Optimal Cut-off Point and Youden Index.

KEYWORDS

Sensitivity, Specificity, Receiver Operating Characteristics, Area under Curve, Optimal Cut-off Point, Youden Index

INTRODUCTION

Diagnostic Tests are health care provider's measure to confirm the presence or absence of a particular disease. These tests play a key role in determining what additional tests and interventions are needed to treat a health condition. Diagnostic tests are evaluated on criterion like accuracy and interpretability of results, ease of administering, cost effectiveness. Some commonly known diagnostics are, film screen mammography for breast screen, PSA (Prostate specific Antigen) for prostate cancer, MRI, CT-SCAN.

Diagnostic test accuracy studies are designed to study how well the test distinguishes between disease and non-disease health states. Objective of these studies is to answer, questions like: How well the test performs on diseased subjects? How well the test rules out the disease in healthy subjects? Is the performance of test comparable to existing diagnostic methods? When and in which situation test should be administered.

Results from diagnostic test can be dichotomous (binary), ordinal or continuous. For case when the outcome of the test is ordinal or continuous values, studies use ROC curves to assess the accuracy of diagnostic tests. Concepts of ROC curves is not new in the research field and over the course of time significant advancements are made to obtain conscientious result using the approach.

In the clinical research, ROC Curves can not only help in diagnostic studies but can also be employed to analyze the role of biomarkers, prognostic factors in predicting desired response.

Though the application of ROC curves is prevalent to scientific groups but it's implementation and interpretation is a unfamiliar territory in clinical programming. This paper will provide an introduction to (i) components of ROC curve, (ii) how to plot the ROC curve, (iii) advantages of ROC curves, (iv) how to find optimal cut-off point to classify diseased and non-diseased subjects, (v) measures of accuracy to determine accuracy of diagnostic test, (vi) comparing the performance of two tests, (vii) summarizing the results from ROC curve, (viii) factor that impacts the results from ROC curves.

COMPONENTS OF ROC CURVE

Ideally, results from diagnostic test should classify subjects with disease as diseased (True Positive) and health subjects as non-diseased (True Negative). But in reality tests produce two more outcomes, classifying subject with disease as healthy (False Positive) and healthy subjects as diseased (False Negative).

Here is the data from hypothetical study to analyze a serum indicator to identify subjects with renal disease. Result of serum measurement indicator will be used to diagnosis the presence or absence of disease. The larger the value of indicator, higher is the possibility that the subject has renal disease. Table 1 lists the hypothetical serum measurements of 30 subjects who have undergone the test to measure the serum indicator. Clinically, 15 subjects were found to have renal disease and 15 were found to be healthy. Serum indicator values ranged from 0.04 to 0.53 for patients with renal disease, 0.01 to 0.25 for patients with no disease.

Table 1: Hypothetical serum measurements of 15 subjects with disease and 15 subject without renal disease

Non-Disease	0.13	0.12	0.07	0.05	0.04	0.03	0.02	0.025	0.018	0.25
	0.13	0.05	0.2	0.01	0.11					
Disease	0.53	0.41	0.18	0.16	0.15	0.14	0.13	0.26	0.05	0.04
	0.32	0.18	0.22	0.19	0.1					

From this data, we randomly choose, serum indicator value of 0.1 as a decision threshold point (cut-off) to define the diagnostic test results: positive or negative. Subjects with value >= 0.1 will be test positive and with value < 0.1 will be considered test negative. Classification using this threshold point can be summarized in a 2X2 table, where the columns summarize the data with respect to true disease status and the row summarize data with respect to the test result.

Table 2: Hypothetical Serum Measurement Results in Relation to True Disease Status in 2X2 table

	Disease Status		
Diagnostic test result	Disease	Non-disease	Total
Positive	13 (True Positive)	6 (False Positive)	19 (All Test Positive)
Negative	2 (False Negative)	9 (True Negative)	11 (All Test Negative)
Total	15 (All Disease)	15 (All Non-disease)	30 (All Study Subjects)

Here are some popular indicators of inherent statistical validity of a diagnostic test. These are based on the probability of detecting correct diagnosis by the test among diseased and non-diseased subjects.

Sensitivity: The Sensitivity or true positive rate (TPR) is the proportion of subjects who are correctly diagnosed as diseased. It is the conditional probability of identifying the diseased subject.

Sensitivity= (True Positive)/ (True Positive + False Negative) = 13/15 = 0.867

False Negative Rate = 1 – Sensitivity = 0.133

Interpretation: If diagnostic test is conducted on subject with disease, there is 86.7% chance that this subject will be identified as positive.

Specificity: The Specificity or true negative rate (TNR) is the proportion of subjects who are correctly diagnosed as non-diseased. It is the conditional probability of identifying the non-diseased subject.

Specificity = (True Negative)/ (True Negative + False Positive) = 9/15 = 0.60

False Positive Rate = 1 – Specificity = 0.40

Interpretation: If a diagnostic test is conducted on subject without disease, there is 60% chance that this subject will be identified as negative.

Sensitivity and specificity are the measures to assess the effectiveness of test which tells us how effectively the test identifies the diseased and non-diseased subjects. The higher the value of sensitivity and specificity, better is the diagnostic test.

Another way to look at the diagnostic test result is, if the subject has positive or negative test result, what is the likelihood of having or not having the disease? This answer will come from Positive and Negative predictive values.

Positive predictive value(PPV): The proportion of subject with actual disease in test positive subjects.

Positive predictive value = (True Positive)/ (True Positive + False Positive) = 13/19 = 0.684

Interpretation: If test result is positive, there is 68.4% chance that a subject have disease.

Negative predictive value (NPV): The proportion of true non-diseased subjects in test negative subjects.

Negative predictive value = (True Negative)/ (True Negative + False Negative) = 9/11 = 0.818

Interpretation: If test result is negative, there is 81.8% chance that a subject don't have disease.

HOW TO PLOT THE ROC CURVE?

In the previous section, we have seen the measures of predictive accuracy at a single threshold point. With a single threshold point, a diagnostic test can be analyzed as a binary predictor. But when the results from diagnostic test are continuous, there can be multiple threshold points. And the predictive accuracy with multiple threshold points can be analyzed by collectively analyzing the sensitivity and specificity at these threshold points.

ROC curve is a method of describing the overall intrinsic accuracy of the test, independent of decision thresholds. It is a graphical display of Sensitivity (TPR) on Y-axis and FPR (1 – Specificity) on X-axis for varying cut-off point (threshold values). Each data points on graph is generated by using a different cut-off point. ROC curve is generated by connecting the data points (Sensitivity, 1 – Specificity) from different cut points.

Table 3 shows the sensitivity and specificity at various cut-off points for data from hypothetical serum indicator study. One important point to note here is with the increase in the cut-off point the Sensitivity is decreasing and Specificity is increasing. This is because, with the higher value of cut-off point, number of False Negative increases which means a lot of diseased subjects are classified as non-diseased. Specificity increases because at a higher cut-off value test classifies most result as negative, contributing to higher True Negatives. This interpretation is valid when higher value of test results indicates the occurrence of disease. Figure 1 is ROC curve with data points from the Table 3 for hypothetical serum indicator data.

Table 3: Estimates of Sensitivity and Specificity at different threshold points

Threshold Cut-off point	Sensitivity	Specificity	1-Specificity (FPR)	Plot Points (X,Y)
≥0.03	1	0.266	0.733	(0.733,1)
≥0.1	0.866	0.6	0.4	(0.4,0.866)
≥0.14	0.733	0.866	0.133	(0.133,0.733)
≥0.2	0.333	0.866	0.133	(0.133,0.333)
≥0.4	0.133	1	0	(0,0.133)
≥0.5	0.066	1	0	(0,0.066)
≥0.6	0	1	0	(0,0)

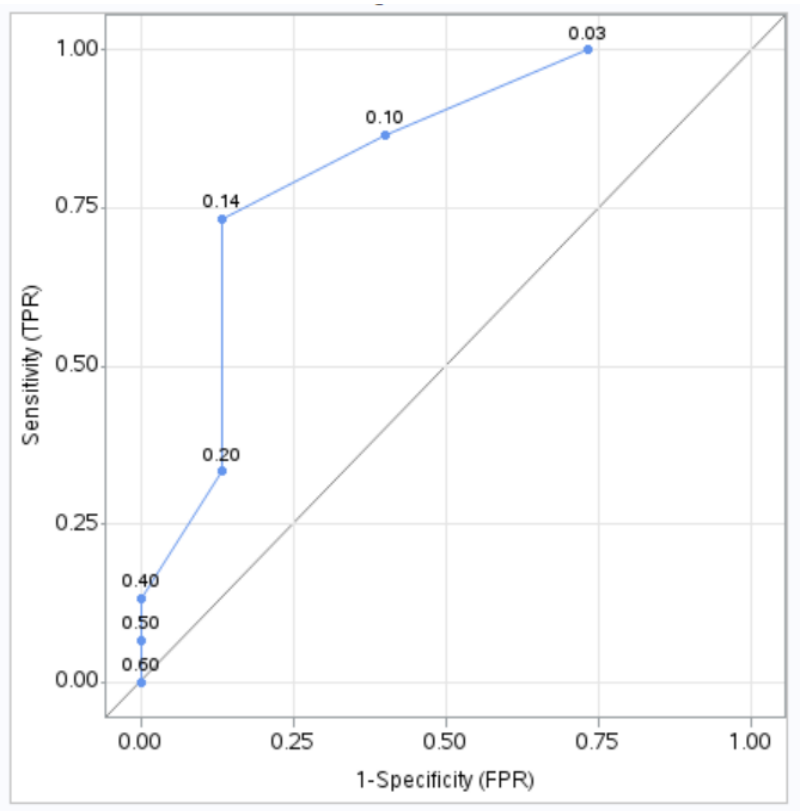

Figure 1: ROC Curve using Sensitivity and 1-Specificity values at various threshold points in Table 3.

GENERATING ROC CURVE WITH SAS

In ROC curve, the Sensitivity (TPR) is plotted with 1-Specificity (FPR) on Y axis and X axis respectively for the different cut-off points. Each points on ROC curve represent the pair of (sensitivity, 1-specificity) corresponding to particular threshold point. In SAS, PROC LOGISTIC procedure is used to generate the ROC curve. Display 1 shows the input data required to generate ROC Curve using PROC LOGISITC.

	disease_stat	test_result
1	1	0.53
2	1	0.41
3	1	0.32
4	1	0.26
5	0	0.25
6	1	0.22
7	0	0.2
8	1	0.19
9	1	0.18
10	1	0.18
11	1	0.16
12	1	0.15
13	1	0.14
14	1	0.13
15	0	0.13

Display 1: Snapshot of SAS dataset with results from hypothetical serum indicator study

Here is PROC LOGISITIC code to generate ROC curve:

```
ODS GRAPHIC ON;
PROC LOGISTIC DATA = TEST_1;
     MODEL DISEASE_STAT (EVENT='1')=TEST_RESULT/OUTROC=ROCDATA;
     ROC; ROCCONTRAST;
RUN;
ODS GRAPHIC OFF;
```

The OUTROC option generates a dataset with sensitivity (_SENSIT _) and 1 - specificity (_1MSPEC_) values for each possible threshold value, which are plotted in the ROC curve.

	PROB	_POS_	_NEG_	_FALPOS_	_FALNEG_	_SENSIT_	_1MSPEC_
1	0.9972675319	1	15	0	14	0.0666666667	0
2	0.9841796729	2	15	0	13	0.1333333333	0
3	0.9428666643	3	15	0	12	0.2	0
4	0.8720142283	4	15	0	11	0.2666666667	0
5	0.8546372659	4	14	1	11	0.2666666667	0.0666666667
6	0.7906949784	5	14	1	10	0.3333333333	0.0666666667
7	0.73773634	5	13	2	10	0.3333333333	0.1333333333
8	0.7082282103	6	13	2	9	0.4	0.1333333333
9	0.6768542872	8	13	2	7	0.5333333333	0.1333333333
10	0.60932343	9	13	2	6	0.6	0.1333333333

Display 2: Snapshot of ROCDATA dataset generated with OUTROC option

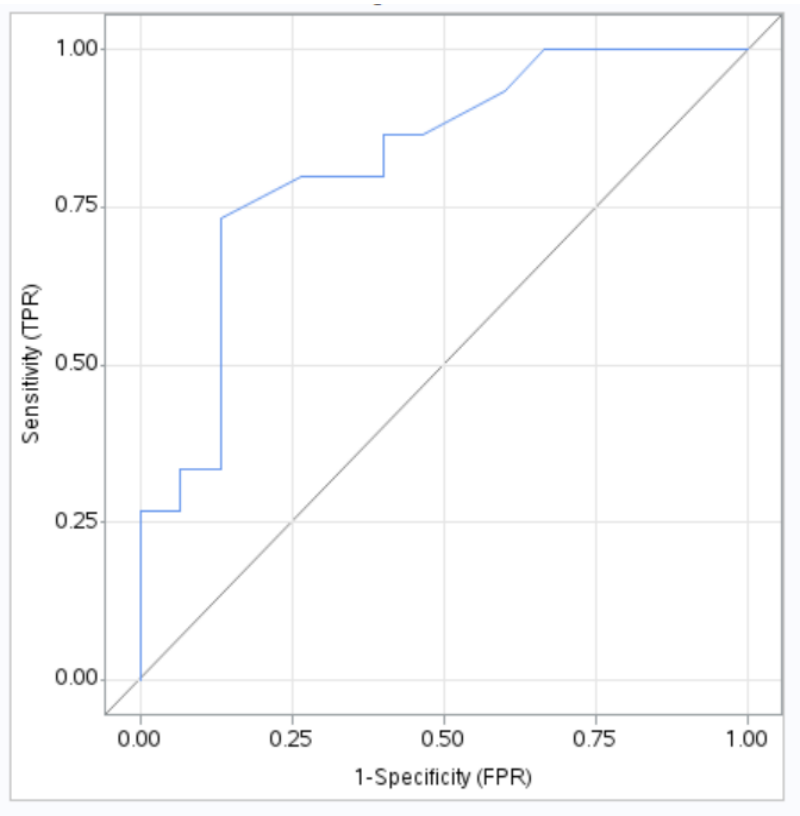

Figure 2: ROC curve generated using procedure PROC LOGISTIC

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.9146	0.8529	5.0395	0.0248
test_result	1	14.7442	6.1195	5.8051	0.0160

Display 3: Partial Logistic Regression output with estimate and slope values.

ROC CURVE INTERPRETATION

The chance diagonal is a line joining (0, 0) and (1, 1) dividing the curve into two equal parts. When ROC curve fall on this line, it indicates that results from diagnostics test are pure guess and there is random chance to distinguish subjects with versus without a disease.

Plot Point (0, 1) in upper left corner is known as Ideal Point. At this point (1 - Spec =0, Sensitivity = 1) a diagnostics test has 100% sensitivity,100% specificity and perfectly distinguishes between the diseased and non-diseased. If resulting ROC curve is closer to upper left corner, it indicates that tests has higher discriminating capabilities.

Figure 2 is ROC curve for data from study of hypothetical serum indicator as diagnostic test for renal failure. ROC curve is well above the chance diagonal. This indicates that the hypothetical serum indicator is effectively discriminating disease vs. non-disease subjects to diagnose renal disease.

Display 3 is partial logistic regression output, regression estimate and chi-square p-value. Significant p-value indicates that the test results can effectively predict the disease status. The model's intercept (-1.9146) and slope (14.7442) are needed to compute the cutoff scores.

ADVANTAGES OF ROC CURVE

When we use line segment to connect each point from all possible decision thresholds, and test results data does not require to follow any distribution pattern, the resulting curve is called *Empirical* ROC curve. This is the easiest version of ROC curve, both easy to create and interpret.

A smooth fitted curve can be obtained, when the results from diagnostic test follow a binomial model (i.e. test result data from both diseased and non-diseased subject follow normal distribution).

ROC curve has several advantages over isolated measurement of Sensitivity and Specificity. Here are some advantages of ROC curve, which makes it preferred tool for diagnostics studies.

1. ROC curve is visual representation of accuracy data. Plotting of ROC Curve does not required selection of a particular cut-off point since it displays all possible cut-off points. Also one can obtain an optimal cut-off point for correctly identifying diseased or non-diseased subjects.

2.The ROC curve is independent of scale of test result, that is it is invariant to the transformation (linear, logarithm or square root) of the result values. It's a rank based measure, that depends on the rank of the magnitude value.

3. Associated measures of accuracy like AUC and partial area under the curve can be obtained describing the discriminating capability of diagnostic tests.

4. It provides visual comparison of two or more diagnostic tests on a common scale, facilitating easy comparison.

5. ROC curve helps in finding the required value of sensitivity at fixed value of specificity.

In following sections, we have described some of these advantages in detail and methods for obtaining these results.

AUC CALCULATION

ROC curve is summary of accuracy information of continuous predictor. Area Under Curve (AUC) is the most commonly used quantitative summary measures of the ROC curve. AUC facilitates comparison and improves interpretation. It shows and measures that how good a diagnostic test can classify between diseased and non-diseased subjects. AUC values range from 0 to 1, a perfect diagnostic test will have an AUC value 1, whereas worthless diagnostic test will have an AUC <=0.5.

AUC can be calculated with different approaches.

1.In an empirical curve, AUC is estimated by trapezoidal rule that is forming trapezoids using observed points, computing their areas and adding them together. But when the results have many threshold values it can be a tedious task to compute AUC using this approach.

2. C-index for calculating area. AUC estimates are reported under Association of Predicted Probabilities and Observed Responses section of PROC LOGISTIC output using this approach.

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	81.3	Somers' D	0.649
Percent Discordant	16.4	Gamma	0.664
Percent Tied	2.2	Tau-a	0.336
Pairs	225	c	0.824

Display 4: AUC results with c-index approach for ROC curve from hypothetical serum indicator study.

Calculations in display 4 are based on subject pairs from the study data. These pairs can be classified as:
Concordant Pair - If in a pair, subject with higher test result is diseased and other subject with lower test result is non-diseased.
Discordant Pair - If in a pair, subject with higher test result is non-diseased and other subject with lower test result is diseased.

Tied Pair – If in a pair, both subjects in pair have same test result value
Non-informative Pair – If in a pair, both subjects are diseased or both are non-diseased.
Informative Pair – All possible subject pairs excluding the non-informative pair. (shown as "Pairs" in Display 4.)

AUC is probability of concordance which is defined as the number of concordant pairs plus one-half the number of tied pairs divided by the number of all informative pairs i.e. sum of percent concordant and 0.5 * (percent tied).

From logistic output AUC can be obtained from value marked with “c”.

For hypothetical serum indicator study the AUC is 0.824 (or 82.4%). This means that if we randomly select two subjects, one with and other with non-disease, the probability is 0.824 that the subject with disease will have more suspicious test result.

3. Using Wilcoxon Mann-Whitney test using 2-sample rank-sum statistics. AUC estimates calculated with this approach are reported under ROC Association Statistics section of PROC LOGISTIC output.

ROC Association Statistics

ROC Model	Mann-Whitney Area	Standard Error	95% Wald Confidence Limits		Somers' D (Gini)	Gamma	Tau-a
Model	0.8244	0.0781	0.6713	0.9776	0.6489	0.6636	0.3356
ROC1	0.5000	0	0.5000	0.5000	0	.	0

ROC Contrast Test Results

Contrast	DF	Chi-Square	Pr > ChiSq
Reference = Model	1	17.2420	<.0001

Display 5: AUC results with Wilcoxon Mann-Whitney test for ROC curve from hypothetical serum indicator study.

The area under curve (AUC) is 0.824 as compared to that of the diagonal line which is always 0.500 (half the graph). The AUC of 0.824 means that serum indicator correctly classifies randomly drawn pairs of diseased/non-diseased subjects 82.4% of the time. Display 5 shows significance tests for AUC. Since the 95% confidence interval (0.6713, 0.9776) does not contain 0.500, we can conclude that our AUC is significantly better than chance. A chi-square test is also in the output and provides a p-value ($p < .0001$) associated with the null hypothesis that our AUC equals 0.500. P-value also confirms that the test is significant.

OPTIMAL CUT-OFF POINT TO CLASSIFY DISEASED AND NON-DISEASED SUBJECTS

An ideal diagnostic test will have 100% sensitivity and 100% specificity, where test would correctly identify the subjects with disease and non-disease. In reality test are not 100% perfect and the intent is to maximize the ability. Optimal cutoff point or decision threshold is the point that gives maximum correct classification.

Each observation in the data, generates a binary response classification matrix in form of predicted probability of the positive or negative result. Predicted probability is a continuous value between 0 and 1. But it is desired to have a binary prediction whether test result is positive or negative. This leads to choose a cut-off point on probability scale where if the predicted probability exceeds the chosen cut-off point, then the result is test positive. Otherwise, the predicted response is test negative.

There are different methods to obtain this cut-off point.

1.Minimize Cost criterion: This method considers the financial cost, health impact, discomfort to patient and further investigative cost (downstream cost) for correct and false diagnosis. These also considers factors like disease prevalence and prior probability.

Other methods to obtain cut-off point are based upon sensitivity and specificity. They give equal weightage to sensitivity and specificity.

2. Youden Index: Maximum (Sensitivity + Specificity – 1)

Youden index (J) is maximum potential of effectiveness of diagnostic test. This maximizes the distance between chance diagonal to the point [x, y] on the curve, a point farthest from random. It ranges between 0 to 1. Where J=1 indicates that there is no false negative rate (FNR) or false positive rate(FPR), i.e. effectiveness of diagnostic test is perfect, and J=0 indicates that diagnostic test gives the same proportion of positive results for the group with and without disease, i.e. test is useless.

With SAS, as of now there is no direct way to calculate Youden index. But it can be calculated using the dataset generated by OUTROC option. Firstly, calculate J value for each sensitivity and specificity and out of all J values select maximum J value as Youden index. Corresponding diagnostics test result value will be cut-off threshold point for the diagnostic test/marker.

3. Distance to (0, 1): SQRT ((1-Sensitivity)^2+(1-Specificity)^2)

Distance to (0,1), D is point closest to ideal point. This minimizes the distance from the "perfect" point at the upper-left corner of the ROC plot where 1-Specificity=0 and Sensitivity=1. The point with shortest distance is D, test result value corresponding to this point is threshold cut-off point.

4. Sensitivity, Specificity equality: ABS(Sensitivity - Specificity)
This method of identifying cut-off point minimizes the difference between sensitivity and specificity values at all possible cut-off points. The point between lowest absolute difference between sensitivity and specificity is the equality point. And the test result value corresponding to this is threshold cut-off point.

SAS IMPLEMENTATION OF OPTIMAL CUT-OFF POINT CALCULATION

Here is SAS code to generate optimal cut-off points from 3 methods giving equal weightage to sensitivity and specificity.

```
DATA CUTOFF;
  SET ROCDATA;
  _SPECIF_ = (1 - _1MSPEC_);
   LOGIT=LOG(_PROB_/(1-_PROB_));
  CUT_POINT=(LOGIT + 1.914)/14.744;
  J = _SENSIT_ + _SPECIF_ - 1;
  D = SQRT ((1-_SENSIT_)**2 + (1-_SPECIF_)**2);
  DIFF= ABS (_SENSIT_ - _SPECIF_);
RUN;
```

	PROB	_SENSIT_	_SPECIF_	_1MSPEC_	CUT_POINT	J	D	DIFF
1	0.9972675319	0.0666666667	1	0	0.53	0.07	0.9333333333	0.9333333333
2	0.9841796729	0.1333333333	1	0	0.41	0.13	0.8666666667	0.8666666667
3	0.9428666643	0.2	1	0	0.32	0.2	0.8	0.8
4	0.8720142283	0.2666666667	1	0	0.26	0.27	0.7333333333	0.7333333333
5	0.8546372659	0.2666666667	0.9333333333	0.0666666667	0.25	0.2	0.7363574011	0.6666666667
6	0.7906949784	0.3333333333	0.9333333333	0.0666666667	0.22	0.27	0.6699917081	0.6
7	0.73773634	0.3333333333	0.8666666667	0.1333333333	0.2	0.2	0.6798692685	0.5333333333
8	0.7082282103	0.4	0.8666666667	0.1333333333	0.19	0.27	0.6146362972	0.4666666667
9	0.6768542872	0.5333333333	0.8666666667	0.1333333333	0.18	0.4	0.4853406593	0.3333333333
10	0.60932343	0.6	0.8666666667	0.1333333333	0.16	0.47	0.4216370214	0.2666666667
11	0.5737156879	0.6666666667	0.8666666667	0.1333333333	0.15	0.53	0.3590109871	0.2
12	0.537326846	0.7333333333	0.8666666667	0.1333333333	0.14	0.6	0.298142397	0.1333333333
13	0.500535985	0.8	0.7333333333	0.2666666667	0.13	0.53	0.3333333333	0.0666666667
14	0.4637393192	0.8	0.6666666667	0.3333333333	0.12	0.47	0.3887301263	0.1333333333
15	0.4273333122	0.8	0.6	0.4	0.11	0.4	0.4472135955	0.2
16	0.3916977705	0.8666666667	0.6	0.4	0.1	0.47	0.4216370214	0.2666666667
17	0.2926586324	0.8666666667	0.5333333333	0.4666666667	0.07	0.4	0.4853406593	0.3333333333
18	0.2355218005	0.9333333333	0.4	0.6	0.05	0.33	0.6036923425	0.5333333333
19	0.2100155353	1	0.3333333333	0.6666666667	0.04	0.33	0.6666666667	0.6666666667
20	0.1865972959	1	0.2666666667	0.7333333333	0.03	0.27	0.7333333333	0.7333333333

Display 6: Dataset CUTOFF with estimate of J, D and DIFF for different threshold points

Youden Index: Display 6 shows that among all the values (Sensitivity+ Specificity-1) maximum value is 0.6 which is Youden index (J) value and corresponding to this value sensitivity is 0.73 and specificity is 0.86. Optimal cut-off value corresponding to Youden index is 0.14. With the help of Youden index it can be inferred that a subject with the test result value >= 0.14 when categorized as test positive, there is 73% probability that the subject will be classified in diseased group and a subject with the test result value < 0.14 when categorized as test negative, there is 86% probability that the subject will be classified into non-diseased group.

Distance to (0, 1): Display 6 shows the D = SQRT ((1-Sensitivity)^2+(1-Specificity)^2) for each possible cut-off point, minimum value 0.298 is D value. Cut-off value corresponding with value is 0.14, which is same cut-off value as shown by Youden index.

Sensitivity, Specificity equality: Display 6 shows the diff = abs(sensitivity – specificity) for all points, diff value = 0.066 is equality point. Threshold cut-off point 0.13 is optimal cut-off point as per this method.

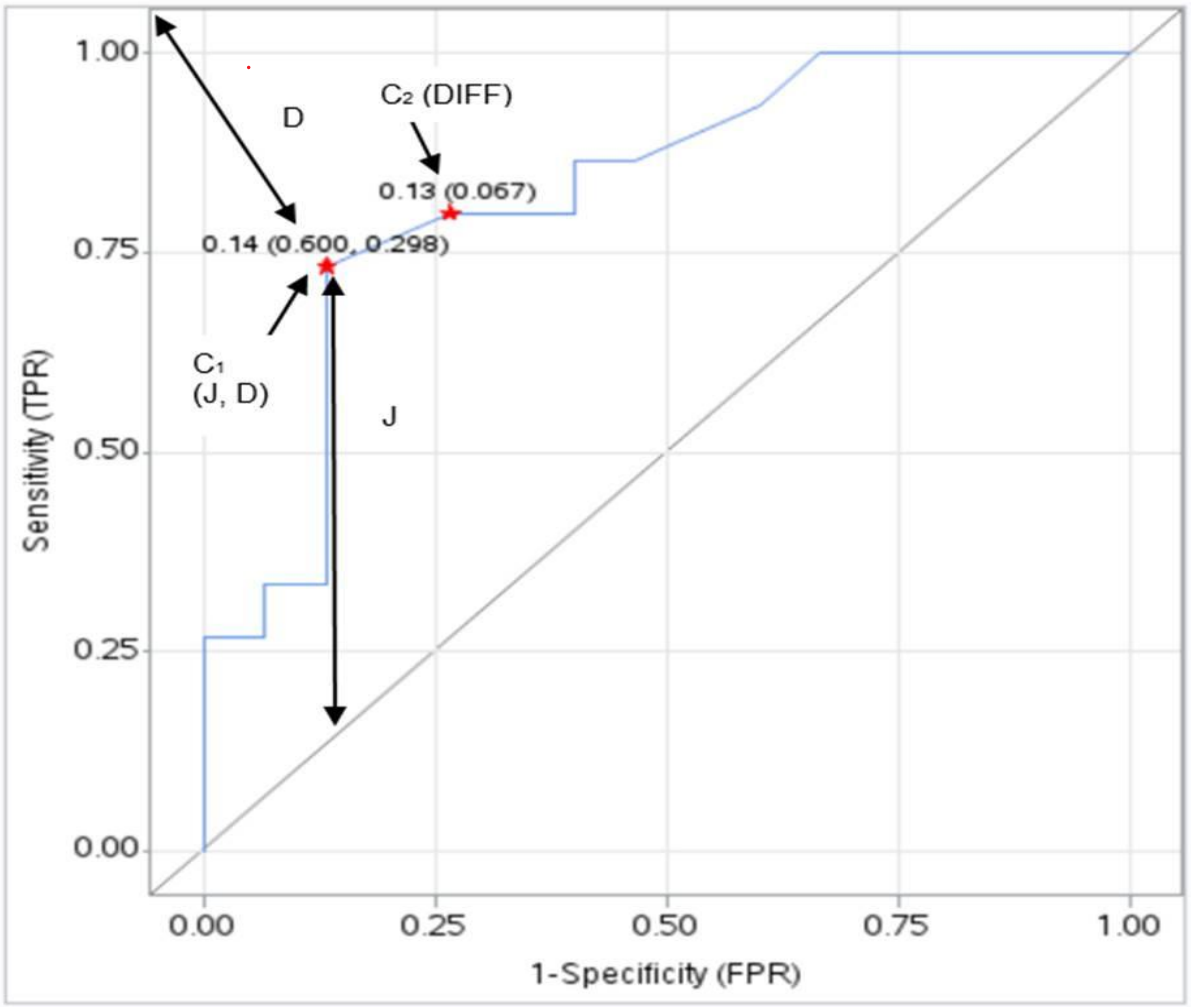

Figure 3: ROC curve showing optimal cut-point (C_1 = 0.14) with Youden Index (J = 0.6) and Distance to 0,1 (D = 0.298). Optimal cut-point (C_2 = 0.13) with sensitivity, specificity equality (DIFF = 0.067)

Optimal cut-off point can also be calculated by using the macro %ROCPLOT from SAS® Institute. Here is the sample call to the macro.

```
PROC LOGISTIC DATA= TEST_1;
     MODEL DISEASE_STAT (EVENT ='1') = TEST_RESULT/OUTROC=ROC1;
     OUTPUT OUT=OUT P=PHAT;
RUN;

%ROCPLOT( INROC = ROC1, INPRED = OUT, P = PHAT,ID = TEST_RESULT _CUTPT_  _OPTDIST_
         _OPTY_ _OPTSESP_,OPTCRIT =  DIST YOUDEN SESPDIFF, OPTSYMBOLSTYLE = SIZE=0,
          OPTBYX = PANELALL, X = TEST_RESULT);
```

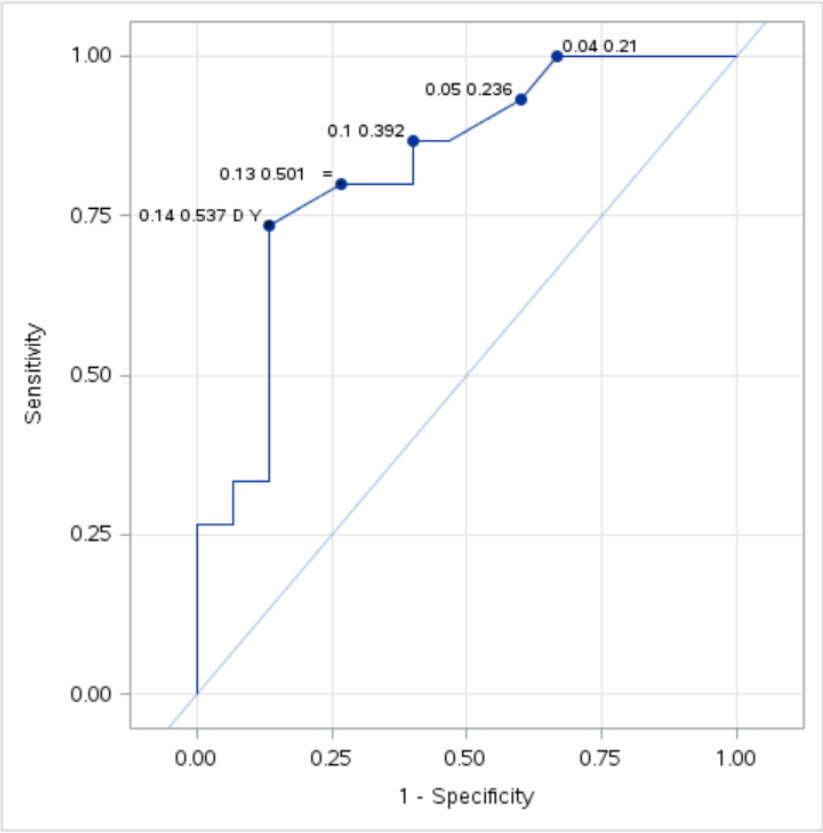

Figure 4: ROC curve with optimal cut-points with Youden Index (Symbol: Y) and Distance to 0,1 (Symbol: D), Sensitivity, specificity equality (Symbol: =) generated using the ROCPLOT macro.

Optimal Cutpoints

Criterion	Symbol	Cutpoint	Label	Value
Dist To 0,1	D	0.53733	0.14 0.537 D Y	0.29814
Sens-Spec	=	0.50054	0.13 0.501 =	0.06667
Youden	Y	0.53733	0.14 0.537 D Y	0.60000

Display 7: Cutoff values, Youden Index, Distance to (0,1), Sensitivity, specificity equality from the ROCPLOT macro.

Comparing values from Display 6 and Display 7 shows the optimal cut-off points calculated in DATA steps are same as the results produced by macro output.

COMPARING THE PERFORMANCE OF TWO TEST

Another feature of ROC curve is that it facilitates the comparison of two or more diagnostics tests. Let's consider two diagnostic tests, Test 1 and Test 2, where results are collected from both test on same population along with the gold standard result. Here we want to compare, which test is most effective in classification of subjects in two categories (Disease and Non-disease).

Below is the SAS code for comparing the performance of two diagnostic tests. The ODS SELECT is used to display only three selected output, plot of two ROC curves (ROCOVERLAY), the estimated AUCs and their 95% confidence intervals (ROCASSOCIATION) and the test of difference of the two AUCs (ROCCONTRASTESTIMATE).

```
ODS GRAPHICS ON;
ODS SELECT ROCOVERLAY ROCASSOCIATION ROCCONTRASTESTIMATE;
PROC LOGISTIC DATA=MULTI DESCENDING PLOTS=ROC;
     MODEL DISEASE_STAT = TEST_RESULT1 TEST_RESULT2 / NOFIT;
     ROC 'DIAGNOSTIC TEST 1' TEST_RESULT1;
     ROC 'DIAGNOSTIC TEST 2' TEST_RESULT2;
     ROCCONTRAST REFERENCE ('DIAGNOSTIC TEST 2') / ESTIMATE;
RUN;
ODS GRAPHICS OFF;
```

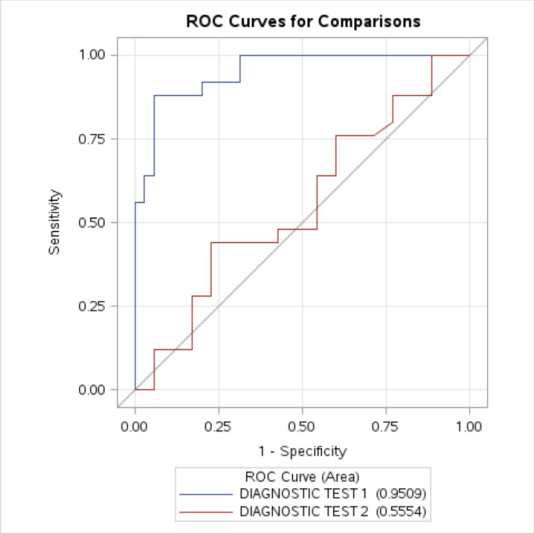

Figure 5: ROC curve comparing performance of Diagnostic Test 1 (Blue) and Diagnostic Test 2 (Red)

ROC Association Statistics

ROC Model	Mann-Whitney Area	Mann-Whitney Standard Error	Mann-Whitney 95% Wald Confidence Limits		Somers' D (Gini)	Gamma	Tau-a
DIAGNOSTIC TEST 1	0.9509	0.0248	0.9023	0.9994	0.9017	0.9017	0.4458
DIAGNOSTIC TEST 2	0.5554	0.0765	0.4054	0.7054	0.1109	0.1111	0.0548

ROC Contrast Estimation and Testing Results by Row

Contrast	Estimate	Standard Error	95% Wald Confidence Limits		Chi-Square	Pr > ChiSq
DIAGNOSTIC TEST 1 - DIAGNOSTIC TEST 2	0.3954	0.0827	0.2333	0.5576	22.8471	<.0001

Display 8: PROC LOGISTIC output to compare AUC from two tests.

In Figure 5, Blue, Red colored curves are ROC curves for Test 1, Test 2 respectively. Test 1 have AUC of 0.950, closer to ideal point (0, 1) whereas Test 2 have AUC value 0.555, closer to the chance of line (AUC=0.5). Based on AUC values of ROC curve, Diagnostic Test 1 have superior diagnostic capability than Test 2. Results from ROC contrast estimation in display 8, confirms that the AUC (0.950) of Test 1 is significantly greater than the AUC (0.555) of Test 2 (p-value<0.0001).

CONSIDERATIONS THAT CAN LIMIT THE APPLICATION OF ROC CURVES

AUC is a measure of the overall performance of a diagnostic test. The higher AUC is indicator of better performance of the diagnostic test. But in some cases when comparing the ROC curves of two tests, AUC values can be equal. But equal AUC values does mean that the two tests yield the same overall diagnostic performance. Shape of the ROC curves with equal AUC might not be identical. One test can be better in high FPR range (or high sensitivity range), whereas the other test can be better in the low FPR range (or low sensitivity range). So right tests should be chosen based upon the clinical requirement.

Depending upon disease severity and associated risk profile, diagnostic test should be chosen with appropriate balance in sensitivity and specificity values. For example, in high risk group (cancer screening) test should have high sensitivity even if FPR is high, because a false negative test result may have serious downstream consequences. On the contrary in low risk group (low prevalence, risky subsequent proceedings) test should have high specificity. Other False Positive subjects have to undergo unnecessary follow-up procedures. In cases like these AUC cannot be used as a measure of the overall diagnostic performance. Two techniques that are used for these type of studies are:

1.Use ROC curve to estimate the sensitivity at a particular FPR, and to compare the sensitivities of different ROC curves at a particular FPR

2. Use partial area under the ROC curve. Partial ROC area is the area between two FPRs or between two sensitivities.

A diagnostic test should be judged in the context of clinical situation and health outcome involved. Over the years of research, significant advancement has been made on how to use the ROC curve and associated measures to judge the diagnostic capabilities of the test under study.

CONCLUSION

ROC analysis is a powerful and effective means to assess the accuracy of a diagnostic test when the outcome of test are continuous values. For a similar outcome it compares the performance of two or more diagnostic tests. Sensitivity, specificity and AUC are basic measure of ROC analysis which provide the ability to select the right diagnostic tests to diagnose a disease condition. With the help of various cut-off criterion, it can provide the optimal cut-off points that can maximize the discriminating power of diagnostic tests.

REFERENCES

Zhou, Xiao-hua, Obuchowski, Nancy a. and Mcclish, Donna: copyright 2002.Statistical Methods in Diagnostic Medicine. 17, 28,139.New York, USA: John Wiley & Sons, Inc.

Mithat Gönen: 2007. Analyzing Receiver Operating Characteristic Curves with SAS®. Cary, NC, USA: SAS Institute.

Kumar, Rajeev, and Indrayan, Abhaya.17 april,2007. “Receiver Operating Characteristic (ROC) Curve for Medical Researchers”. Volume 48 New Delhi, India: Indian Pediatrics

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Manoj Pandey
Enterprise: Ephicacy Lifescience Analytics Pvt. Ltd., India
Address: No.6, 2nd Floor, 2nd Main Rd, Arekere, Off. Bannerghatta Road,
City, State ZIP: Bangalore, Karnataka 560076
E-mail: manoj.pandey@ephicacy.in
Web: www.ephicacy.com

Name: Abhinav Jain
Enterprise: Ephicacy Consulting Group
Address: 33 Wood Avenue South, Suite 600
City, State ZIP: Iselin, NJ 08830
E-mail:abhinav.jain@ephicacy.com
Web: www.ephicacy.com