A Comparison between Correlated ROC Curves Using SAS and DBM MRMC
A Case Study in Image Comparison

Dachao Liu   Northwestern University   Chicago

ABSTRACT

Receiver Operating Characteristic (ROC) analysis is often used in evaluation of a diagnostic test, related to decision making in cost/benefit analysis or medical practice by assessing the performance of binary classifiers. ROC curves in the diagnosis of multi-reader multi-case setting are correlated. There are ways to compare these correlated ROC curves. In this case study, there were 3 LCD imaging machines or display configurations, and 12 observers. Each observer read 180 cases with 60 being cancer cases and 120 normal benign cases. As a case study, this paper will discuss how the data was processed in SAS and how the comparison of ROC curves was made in both SAS and DBM MRMC.

INTRODUCTION

Since its initial military use in Britain during World War II, ROC has found a wide use in a number of fields including engineering, quality control, weather forecasting, psychology and medicine. In medicine, ROC is used in evaluation of a diagnostic test that discriminates diseased from normal cases. Radiographic images are a common means of diagnostic testing. When a true status of cases is known by a prior independent means of testing (a gold standard), the subsequent evaluations by different combinations of observers and modalities result in correlated ROC analysis. This paper will deal with manipulating the raw data and subsequent comparison correlated ROC curves.

DATA STEP

Software developed in-house was used to view the images and record the probability of malignancy on a scale of 0-100 and other results. There were 12 observers who were all board certified radiologists and MQSA (the Mammography Quality Standards Act) qualified in mammography. The term "mammography" means radiography of the breast. All observers had corrected acuity of 20/20. Test patterns were displayed to five observers in a randomized order on each display. The observers were asked to identify and rate the test patterns according to how obvious those appeared to them. To minimize observer fatigue each of the 1-hour sessions had been broken down into 15-minute periods with a 5-minute break between them. A given set of 6 sessions was separated from the next set by 12 weeks so as to minimize observer recall. Each observer read 180 cases (2 views of a single breast each) in random order on each of the 3 different display configurations (3 sets of 6 sessions). To classify the case, the observer used the mouse to move a horizontal slider from 00 to 100, rating of their level of suspicion of cancer for the case (i.e., level of confidence that the case was a cancer case).

The data were sitting in three directories: 3mp, 5mp and color for 3 LCD imaging machines or display configurations. Each directory had twelve subdirectories for twelve readers. Each subdirectory had six csv files. There were 216 csv files in total. And each csv file had 28-31 columns and 30 rows (see Figure1).

Since this paper is about comparison between correlated ROC curves, I'm not going to use all the variables, but only those relevant variables, five to six of them, give or take. Each csv file had a variable to indicate the configuration (Configuration), a variable to indicate the image set (Image_Set), a variable to indicate the reader (FirstName), and a variable to indicate the id and cancer status (Image_Pair) . These features will be kept after the csv files were imported into SAS. ID was unique in each csv file.

PROC IMPORT procedure was used to get 216 csv files into SAS work directory

```
*3MP RD1;
proc import out=work.m3rd1st1
          datafile= "C:\Documents and Settings\Dachao\Desktop\resultsBackupMay22-
2007changed\results-3MP\
mammo1, reader1\imageset1.csv"
          dbms=csv replace;
datarow=2;
getnames=yes;
run;
```

#	LastName	FirstName	Date	Configurati	Image Set	Image Pair	Diagnosis	Total Time	Left (CC) I	Patient ID	Zoom -1 (L	Zoom Norr	Zoom +1 (Z
2	mammo10	reader10	1/31/2007	3M 0.207	imageset6	16055	95	28.451	DXm.1.2.8	16055	0	20.069	8.382
3	mammo10	reader10	1/31/2007	3M 0.207	imageset6	16054	1	47.679	DXm.1.2.8	16054	0	2.474	36.893
4	mammo10	reader10	1/31/2007	3M 0.207	imageset6	16045	95	12.398	DXm.1.2.8	16045	0	3.095	9.303
5	mammo10	reader10	1/31/2007	3M 0.207	imageset6	197	1	78.493	DXm.1.2.8	2.01E+13	0	2.464	57.692
6	mammo10	reader10	1/31/2007	3M 0.207	imageset6	16059	50	50.622	DXm.1.2.8	16059	0	4.115	22.883
7	mammo10	reader10	1/31/2007	3M 0.207	imageset6	202	0	18.046	DXm.1.2.8	2.01E+13	0	1.912	16.134
8	mammo10	reader10	1/31/2007	3M 0.207	imageset6	198	2	62.46	DXm.1.2.8	2.01E+13	0	8.753	34.95
9	mammo10	reader10	1/31/2007	3M 0.207	imageset6	210	0	23.063	DXm.1.2.8	2.01E+13	0	1.913	21.15
10	mammo10	reader10	1/31/2007	3M 0.207	imageset6	16069	100	52.656	DXm.1.2.8	16069	0	4.066	48.59
11	mammo10	reader10	1/31/2007	3M 0.207	imageset6	16066	1	18.817	DXm.1.2.8	16066	0	1.893	16.924
12	mammo10	reader10	1/31/2007	3M 0.207	imageset6	186	0	46.587	DXm.1.2.8	2.01E+13	0	2.323	44.264
13	mammo10	reader10	1/31/2007	3M 0.207	imageset6	214	0	20.91	DXm.1.2.8	2.01E+13	0	3.495	17.415
14	mammo10	reader10	1/31/2007	3M 0.207	imageset6	215	2	34.67	DXm.1.2.8	2.01E+13	0	2.354	22.322
15	mammo10	reader10	1/31/2007	3M 0.207	imageset6	190	1	47.459	DXm.1.2.8	2.01E+13	0	2.965	44.494
16	mammo10	reader10	1/31/2007	3M 0.207	imageset6	205	5	46.777	DXm.1.2.8	2.01E+13	0	3.495	34.89
17	mammo10	reader10	1/31/2007	3M 0.207	imageset6	192	1	19.879	DXm.1.2.8	2.01E+13	0	1.713	18.166
18	mammo10	reader10	1/31/2007	3M 0.207	imageset6	201	0	19.238	DXm.1.2.8	2.01E+13	0	2.334	16.904
19	mammo10	reader10	1/31/2007	3M 0.207	imageset6	188	1	20.359	DXm.1.2.8	2.01E+13	0	1.903	18.456
20	mammo10	reader10	1/31/2007	3M 0.207	imageset6	206	12	73.275	DXm.1.2.8	2.01E+13	0	2.704	53.146
21	mammo10	reader10	1/31/2007	3M 0.207	imageset6	196	1	33.759	DXm.1.2.8	2.01E+13	0	1.913	23.794
22	mammo10	reader10	1/31/2007	3M 0.207	imageset6	194	1	37.384	DXm.1.2.8	1.02E+08	0	1.873	25.526
23	mammo10	reader10	1/31/2007	3M 0.207	imageset6	213	0	19.658	DXm.1.2.8	2.01E+13	0	2.193	17.465
24	mammo10	reader10	1/31/2007	3M 0.207	imageset6	211	2	51.113	DXm.1.2.8	2.01E+13	0	6.309	30.424
25	mammo10	reader10	1/31/2007	3M 0.207	imageset6	16058	80	26.658	DXm.1.2.8	16058	0	2.393	11.087
26	mammo10	reader10	1/31/2007	3M 0.207	imageset6	212	1	44.494	DXm.1.2.8	2.01E+13	0	3.285	41.209
27	mammo10	reader10	1/31/2007	3M 0.207	imageset6	203	1	36.904	DXm.1.2.8	2.01E+13	0	1.833	23.134
28	mammo10	reader10	1/31/2007	3M 0.207	imageset6	16062	95	48.309	DXm.1.2.8	16062	0	8.261	18.637
29	mammo10	reader10	1/31/2007	3M 0.207	imageset6	16065I	100	4.927	DXm.1.2.8	16065	0	4.927	0
30	mammo10	reader10	1/31/2007	3M 0.207	imageset6	204	1	44.664	DXm.1.2.8	2.01E+13	0	3.645	22.832
31	mammo10	reader10	1/31/2007	3M 0.207	imageset6	16056	100	8.282	DXm.1.2.8	16056	0	2.664	5.618

Figure 1: One of the 216 csv files.

In SAS, each csv file got its name: beginning part for configuration, middle part for reader, and ending part for image set. For example, M3RD1ST1 indicated that the file was referring to configuration displaying machine 3m, reader 1 and image set 1.

Since each and every csv file had the same variable names in sequence, most of files having 28 variables with exception that some of the files had 29 or 30 variables. After all these csv files were imported into SAS and before these 216 SAS files were put together, the first step was to make each csv file have 28 variables.

Also, particular mention should be made on variable Image_Pair. It has dual functions: one being the indicator, the other being ID. Each patient had a value assigned by Image_Pair. In this sense, variable Image_Pair was an id variable. As an indicator, it indicated which patient had cancer and which had not. Patients whose Image_Pair values larger than 16000, equal to set… or starting with 200 were all cancer patients. Otherwise, they were normal patients. Variable Image_Pair should be a character variable, but in some files, it was a numeric variable. PUT function was used to convert it from numeric into character variable.

```
%macro dat(dsn);
data &dsn;
set &dsn;
Image_P=put(Image_Pair, $7.);
drop Image_Pair var29 var30;
rename Image_P=Image_Pair;
run;
%mend dat;
```

In each csv file, each column had its variable name. For some unknown reason, after being imported into SAS, some column variable names would not show their original names, they showed VAR23, VAR25, etc. instead, They should be renamed back to their original names.

```
data m3p (rename=(var23=Zoom1Right var25=Harder var26=Normal var27=Soft));
set
m3one
m3two
m3three
m3four
m3five
m3six
m3seven
m3eight
m3nine
m3ten
m3eleven
m3twelve
;
run;
```

After these files were set together, another problem was detected. Some values of variable Image_Pair were missing or truncated values, although they existed in the corresponding csv files. And this problem must be solved. I had to find out which one was missing or truncated and replaced it with its value. Here were some of them.

```
if Configuration in ('3M 0.207') then Configuration='3m';
if firstname='reader10' and Image_Set='imageset6' and Image_Pair='        .' then
Image_Pair='16065l';
if firstname='reader11' and Image_Set='imageset4' and Image_Pair='         .' then
Image_Pair='16063l';
if firstname='reader11' and Image_Set='imageset6' and Image_Pair='         .' then
Image_Pair='16065l';
if firstname='reader12' and Image_Set='imageset4' and Image_Pair='         .' then
Image_Pair='16063l';
if firstname='reader12' and Image_Set='imageset6' and Image_Pair='         .' then
Image_Pair='16065l';
if firstname='reader2' and Image_Set='imageset4' and Image_Pair='        .' then
Image_Pair='16063l';
if firstname='reader2' and Image_Set='imageset6' and Image_Pair='         .' then
Image_Pair='16065l';
if firstname='reader5' and Image_Set='imageset6' and Image_Pair='        .' then
Image_Pair='16065l';
if firstname='reader6' and Image_Set='imageset4' and Image_Pair='        .' then
Image_Pair='16063l';
if firstname='reader6' and Image_Set='imageset6' and Image_Pair='        .' then
Image_Pair='16065l';
if firstname='reader8' and Image_Set='imageset4' and Image_Pair='        .' then
Image_Pair='16063l';
if firstname='reader9' and Image_Set='imageset4' and Image_Pair='       .' then
Image_Pair='16063l';

if firstname='reader7' and Image_Set='imageset1' and Image_Pair='1600' then
Image_Pair='16001';

if firstname='reader7' and Image_Set='imageset1' AND Diagnosis=98 and
Image_Pair='set1' then Image_Pair='set10';
```

I had to check the data in other two configurations in the same way. It's very tedious, isn't? Good news. Beginning in SAS 9.1 a new statement has been added to PROC IMPORT which allows you to tell SAS how many rows you want SAS to scan to determine variable attributes. The new statement is GUESSINGROWS=N. The N can range from 1 to 32767. It saves a lot of time and effort.

Now it's time to put the data in three configurations together by using the SET statement.

```
data m35c;
 set m3p m5p cop;
 run;
```

With this setup, it's easy to subdivide the data. For example, if I want a subset data for reader 1 and configuration 3mp, I can use

```
data r1m3;
set m35c;
where substr(Configuration,1,2)='3m' and substr(FirstName,7,2)='1';
run;
```

If I want to a subset data for reader 12 regardless of configuration, I can use

```
data r12;
set m35c;
where substr(FirstName,7,2)='12';
run;
```

This subdividing was crucial for the comparison of ROCs across readers or configurations. There were 12 readers and 3 display configurations, therefore altogether I could get 36 ROC curves with their corresponding AUC values (see Figure2).

For configuration 3mp, the data should be like the following (printing partially)

```
                              DIAGR12 (m3r12templet)     13:25 Thursday, July 3, 2008   1

 Obs   IP    TRUTH    diagr1m3    diagr2m3    diagr6m3    diagr7m3    diagr11m3    diagr12m3

   1  100      0          2           3           1           1           0            1
   2  103      0          2           1           1           3           0            3
   3  104      0          2           1           0           1           0            2
   4  105      0          3           1           5           4           1            5
   5  106      0          2          50           1           1           2            3
   . . . . . . . . . . . .
  41  16010    1          2           3           2           2           2            3
  42  16013    1         99          99         100         100         100           97
  43  16014    1         50           3          80          99          90           20
  44  16015    1         65          90          50          87          85            4
  45  16016    1         50          10          95         100          50           15
   . . . . . . . . . . . .
  59  16042l   1          1           2           2           3           1            3
  60  16042r   1         98          25          95         100          98           80
  61  16045    1         75          99          95          99          99           80
  62  16047    1         60          90          50          89          95           50
  63  16049    1         99          99          95          98          85           75
   . . . . . . . . . . . .
 169  set16l   1          1           1           1          97           0            2
 170  set16r   1         75          95          70         100          80           52
 171  set17    1         10           2          10          99           2           41
 172  set18    1          1           1           1           8           1            3
 173  set19    1         97          90          95          99          99           92
   . . . . . . . . . . . .
 176  set5     1         50          90          95          99         100           97
 177  set6     1         97          20          80          92          98           90
```

4

178 set7	1	98	98	99	99	100	99
179 set8	1	85	70	30	99	95	20
180 set9	1	95	96	90	100	99	99

It was true of configurations 5mp and color. Similarly, the data could also be shaped by 12 readers on a certain configuration.

After the data was in shape, I could do ROC curves comparison now.

ROC CURVES COMPARISON

A receiver operating characteristic curve is a statistical tool to assess the performance of binary classifiers and further the accuracy of predictions through the concepts of the sensitivity and specificity. ROC curves provide a comprehensive way to summarize the accuracy of predictions by using the area under the ROC curve (AUC) which is a popular indicator of the performance of a binary classifier. A ROC curve is a plot of sensitivity versus 1-specificity. Sensitivity is defined as the probability that a test result is positive given the subject is a truly positive case and the specificity is defined as the probability that a test result is negative given the subject is a truly negative case. The area under the ROC curve (AUC) is commonly used as a summary measure of diagnostic accuracy. It can take values from 0 to 1. The closer to 1, the better the measure of diagnostic accuracy.

When two or more different diagnostic tests are performed on the same individual, correlation arises definitely and must be taken into account. When comparing the accuracy (AUC) stemming from correlated data, we must take into account the fact that the 2 AUCs are correlated because they are based on the same individual. Using the theory on generalized U-statistics, DeLong developed the methodology to do ROC analysis of correlated data (1988). A macro was created based on Delong's methodology in SAS. Here is the code how to input data into the macro.

```
title "ROC plot for r1m3";
proc logistic data=DIAGR12;
   model TRUTH(event='1') = diagr1m3 / outroc=or roceps=0;
   output out=out p=p;
   ods output association=assoc;
   run;

data _null_;
   set assoc;
   if label2='c' then call symput("area",cvalue2);
   run;

title2 "Approximate area under curve = &area";
%rocplot(out=out, outroc=or, p=p, id=diagr1m3)
title2;

data joint;
   set _rocplot;
   length index $  13;
   Index='r1m3';
   run;
```

For reader N, where N had values of 1,2,…12 in this paper, I used

```
data diagrNm3;
   set _rocplot;
   length index $ 13;
   Index='rNm3';
   run;

data joint;
   set joint diagrNm3;
   run;
```

Data joint was used to plot joint ROC curves for configuration 3m.

For the configurations 5mp and color, the same thing was done. And three p-values were obtained for three configurations across 12 readers (see Figure2).

Table 1. AUC by rater and machine					
		Machine			
		3M	*5M*	*Color*	*Pvalue*
	1	0.915	0.915	0.900	0.617
	2	0.860	0.879	0.875	0.748
	3	0.892	0.895	0.883	0.937
	4	0.866	0.849	0.883	0.480
	5	0.908	0.832	0.890	0.160
Reader	6	0.931	0.923	0.924	0.892
	7	0.884	0.857	0.879	0.534
	8	0.899	0.886	0.874	0.606
	9	0.843	0.846	0.861	0.821
	10	0.916	0.915	0.923	0.905
	11	0.896	0.917	0.903	0.241
	12	0.906	0.928	0.891	0.210
Pvalue		0.052	0.004	0.275	

Figure 2: AUC by reader and display configuration and p-values

The following is the code to print 12 ROC curves together. Two graphics are showed after the code.

```
  %roc(data=DIAGR12,
       var=diagr1m3 diagr2m3 diagr3m3 diagr4m3 diagr5m3 diagr6m3 diagr7m3 diagr8m3
       diagr9m3 diagr10m3 diagr11m3 diagr12m3,
       response=TRUTH)
/* Plot all indices and p-value of overall test r3m3mparing areas */
  symbol1 i=join v=circle c=yellow line=1;
  symbol2 i=join v=dot c=green line=2;
  symbol3 i=join v=triangle c=red line=3;
  symbol4 i=join v=circle c=blue line=4;
  symbol5 i=join v=dot c=green line=5;
  symbol6 i=join v=triangle c=red line=6;
  symbol7 i=join v=circle c=blue line=7;
  symbol8 i=join v=dot c=green line=8;
  symbol9 i=join v=triangle c=red line=9;
    symbol10 i=join v=circle c=blue line=10;
      symbol11 i=join v=dot c=black line=11;
  symbol12 i=join v=triangle c=pink line=12;

  proc gplot data=joint;
    title  "Comparison of Twelve ROC Curves for 3m";
    title2 "Test of H0: equal areas under curves -- p=&pvalue";
    footnote "Comparison of Breast Imaging Techniques";
    label index="Index";
    plot _sensit_ * _1mspec_ = Index /
      vaxis=0 to 1 by .1 haxis=0 to 1 by .1 cframe=ligr;
    run;
     quit;
```
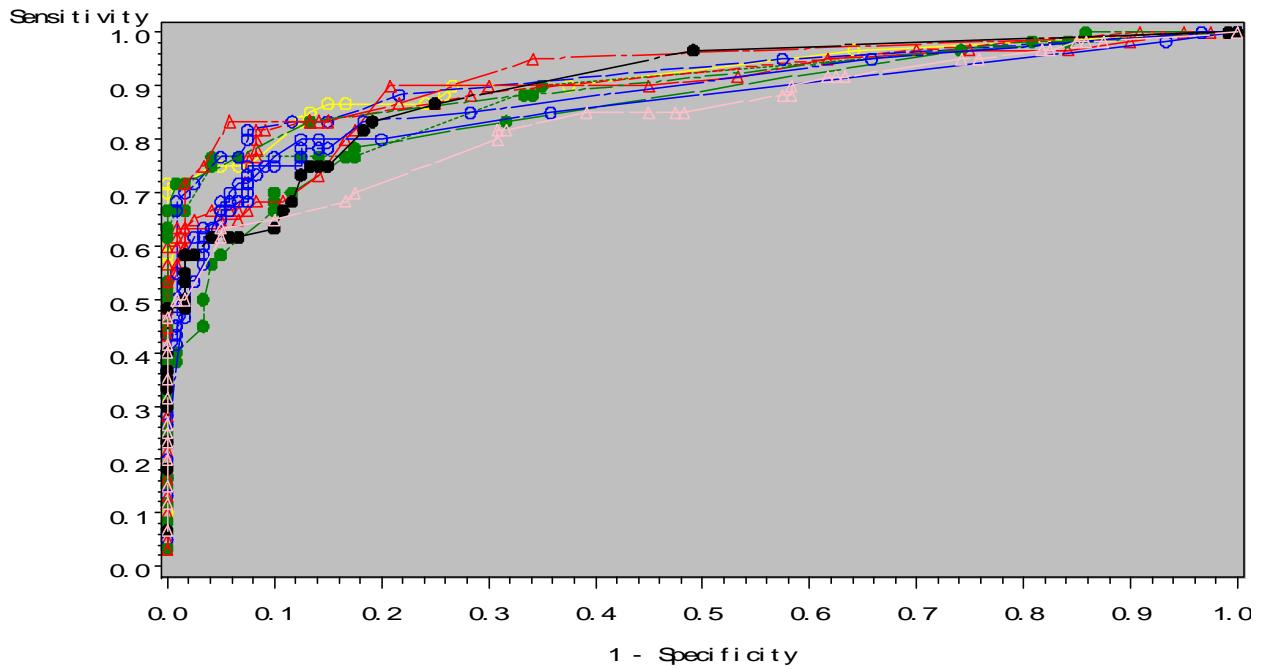
6

ROC plot for r1m3

Approximate area under curve = 0.915
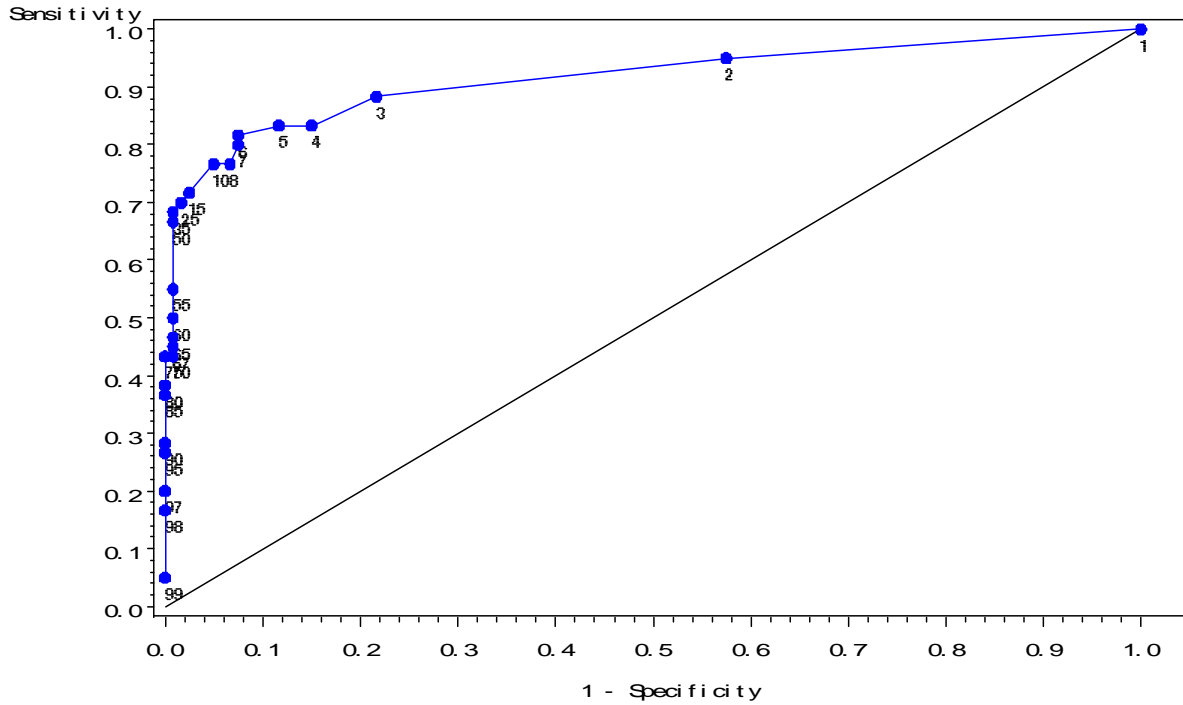
Point labels are values of diagr1m3
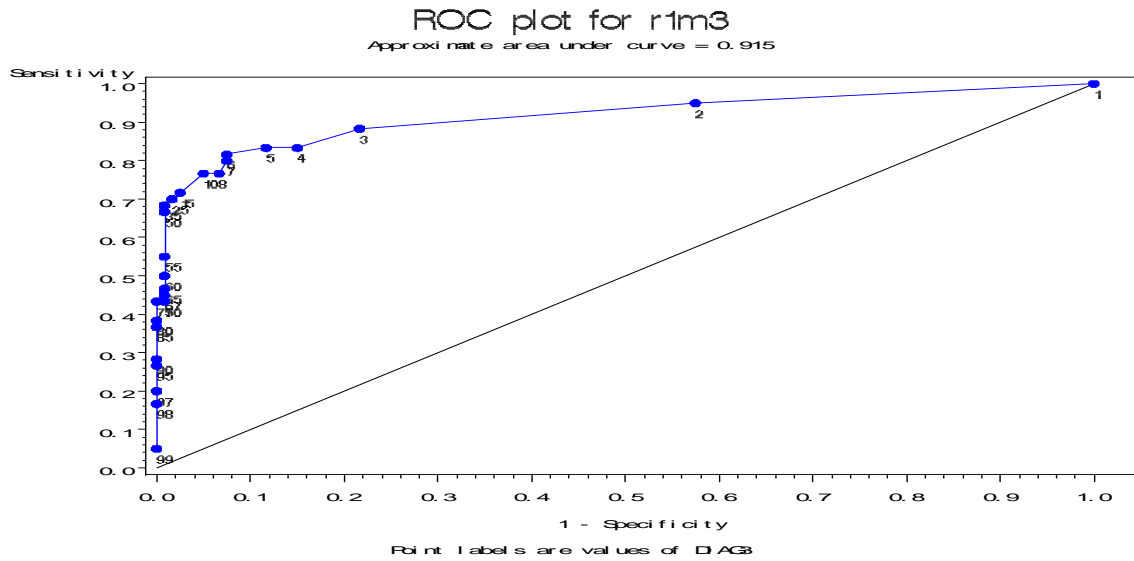
Comparison of Twelve ROC Curves for 3m

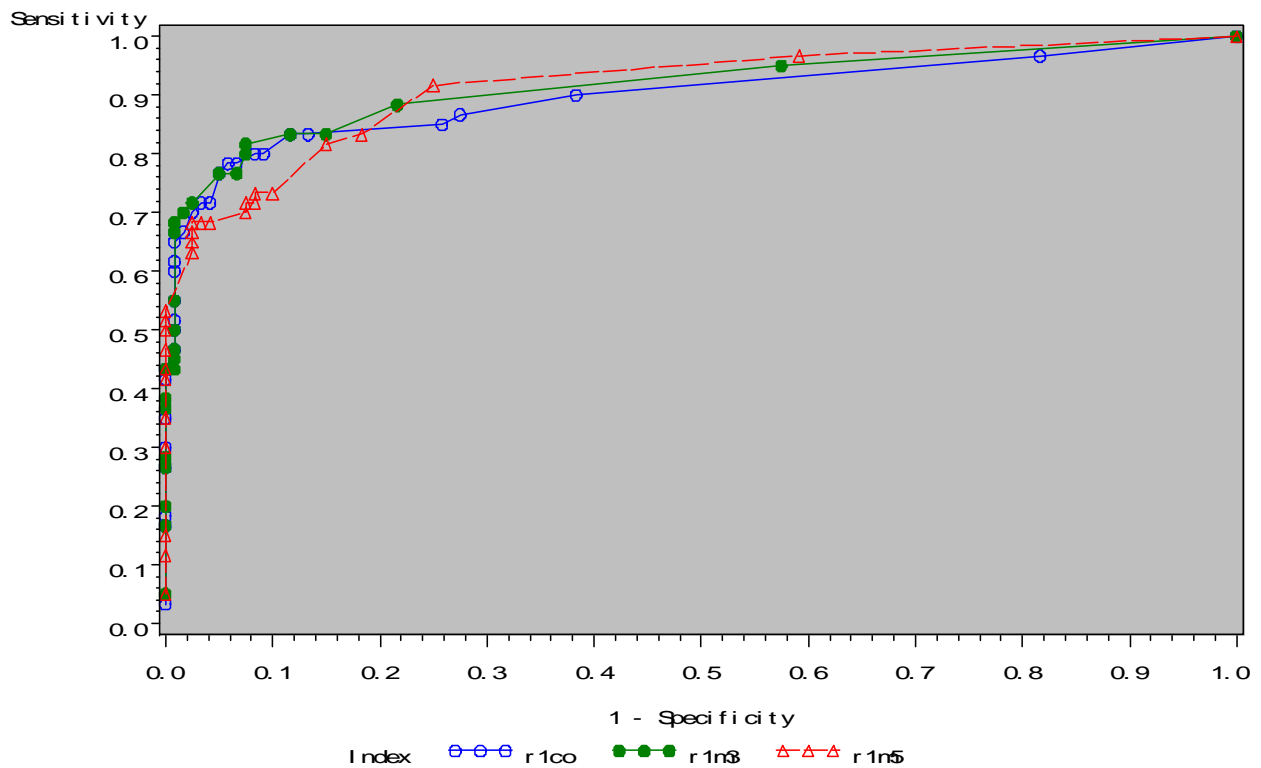Test of H0: equal areas under curves -- p=0.0521

Comparison of Breast Imaging Techniques

7

Similarly, I did the same thing. And 12 p-values were obtained for 12 readers across three configurations (see Figure2) and graphics below.

ROC plot for r1m3

Approximate area under curve = 0.915

Point labels are values of DIAG3

Comparison of Three ROC curves

Test of H0: equal areas under curves -- p=0.6166

Index r1co r1m3 r1m5

From DeLong, et. al. (1988, Biometrics 44)

For the 3mp and 5mp display configurations there were statistically significant differences across the readers (p=0.052 and 0.004 respectively) but such a difference across readers was not significant for the color display configuration (p=0.275).

Using De-Long's method, I could get a p-value within a configuration across readers. If generalization to populations of both of readers and configuration is required, the programs DBM MRMC 2.1, should be used. It is a freely available computer program for Windows, though still in its beta version. It can be downloaded from http://krl.bsd.uchicago.edu/KRL_ROC/software_index.htm . If we want to obtain a p-value across configurations and readers, we have to use DBM MRMC software (Dorfman-Berbaum- Metz multireader multicase). Then we can get p-value either across configurations or readers. MRMC uses the Dorfman-Berbaum-Metz algorithm to compare multiple treatments (e.g., imagining modalities) by using data from multiple readers and multiple cases. This program employs jackknifing and ANOVA techniques. This program allows the conclusions drawn from a study to be generalized to both a population of readers and a population of cases, which has bridged a gap SAS hasn't filled in.

DBM MRMC has a special way of inputting data. For the sake of space, I will not describe its process here. I used the data already shaped in SAS to do further manipulations and ran 12 readers across three configurations to get a p-value for the comparison of the ROCs for the three configurations (see Figure3).

```
3 treatments, 12 readers, 180 cases (120 normal, 60 abnormal)

 Curve fitting methodology is PROPROC
 Dependent variable is AREA
 Normalized pseudovalues used in analyses.


 ==========================================================================
 *****                            Estimates                          *****
 ==========================================================================


                    TREATMENT
          ----------------------------------------
 READER        1           2           3
 ------    ----------  ----------  ----------
    1      0.92001702  0.91674003  0.90558526
    2      0.86582276  0.88713858  0.88731591
    3      0.87713770  0.87753592  0.88783978
    4      0.87353354  0.85858549  0.88882430
    5      0.91242541  0.86521648  0.91543934
    6      0.93662130  0.93583796  0.93062478
    7      0.88829355  0.87946967  0.90324372
    8      0.90376982  0.89216973  0.88187289
    9      0.84065714  0.82985945  0.86718542
   10      0.91355382  0.90254114  0.91101602
   11      0.89815777  0.92048217  0.90400442
   12      0.90852719  0.93559496  0.88762410



 TREATMENT MEANS (averaged across readers)
 ----------------------------------------
      1        0.89487642
      2        0.89176430
      3        0.89754800



a) Test for H0: Treatments have the same AUC
```

```
 Source           DF    Mean Square      F value  Pr > F
 ----------   ------   ---------------   -------  -------
 Treatment        2       0.01809855      0.34    0.7170
 Error        36.50       0.05391076
 Error term: MS(TR) + max[MS(TC)-MS(TRC),0]


 Conclusion: The treatment AUCs are not significantly different, F(2,36) = 0.34, p = 0.7170.
```

Figure 3: output from DMB MRMC

From output of DMB MRMC analysis, we can see that AUCs differ slightly from those in Figure1, whose values are a little bit consistently smaller than its counterpart in DMB MRMC analysis. Similarly I did DBM MRMC analysis for three configurations across 12 readers. There was no significant difference in accuracy across the 3 configurations (p=0.72). There was significant variation in accuracy across the 12 readers (p<0.0001). Printout is not provided for this p-value for the sake of space.

CONCLUSION
In multi-reader and multi-case setting, ROC curves are correlated. This paper has discussed the whole process of data step and correlated ROC comparison analysis in detail from a practical approach, by which it can save you time and effort in similar situation. Both SAS and DBM MRMC are useful tools in correlated ROC analysis.

REFERENCES
DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. Biometrics 1988;44:837-845

http://support.sas.com/kb/25/017.html for SAS macro roc.sas

http://support.sas.com/kb/25/018.html for SAS macro rocplot.sas

http://krl.bsd.uchicago.edu/KRL_ROC/software_index.htm   for DBM MRMC 2.1

CONTACT INFORMATION
Your comments and questions are valued and encouraged. Contact the author at:
     Dachao Liu
     Northwestern University
     Suite 1102
     680 N Lake Shore Dr.
     Chicago, IL 60611
     Phone (312)503-2809
     Email: dachao-liu@northwestern.edu