THE ROC CURVE METHOD FOR THE ASSESSMENT OF DIAGNOSTIC ACCURACY

William F. McCarthy, Maryland Medical Research Institute, Baltimore, MD
Nan Guo, Maryland Medical Research Institute, Baltimore, MD

ABSTRACT

This paper reviews a methodology for the assessment of diagnostic accuracy by use of the area under the receiver operating characteristic (ROC) curve. Estimation and hypothesis testing of diagnostic accuracy based on a single sample from a single diagnostic test will be discussed with examples of computation. Sample size requirements for both estimation and hypothesis testing will be discussed as well. In addition, we present a SAS® Program with a SAS® Macro which calculates the sample size for the estimation of diagnostic accuracy of a single diagnostic test.

KEY WORD

ROC Curve; Diagnostic Accuracy; Single Diagnostic Test; SAS® Program; SAS® Macro

INTRODUCTION

In order to assess the diagnostic accuracy of a new diagnostic test, we will use the area under the receiver operating characteristic (ROC) curve as the measure of diagnostic accuracy. The ROC curve is a method of describing the intrinsic accuracy of a diagnostic test apart from the decision thresholds. An ROC curve is the plot of a diagnostic test's sensitivity [plotted on the y axis] versus its false-positive rate (1-specificity) [plotted on the x axis]. An ROC analysis plots the relationship between sensitivity and specificity across all cut-points of the test and calculates the area under the curve (AUC) and its standard error. A diagnostic test with an AUC of 1 is perfectly accurate, whereas one with an AUC of 0.5 is performing no better than chance. Most diagnostic tests have AUCs between these values.

For any given diagnostic test, the relationship between sensitivity and specificity is often complex. Typically, as the diagnostic test cut-point (the value at which the diagnostic test is considered "positive") is changed, sensitivity and specificity vary inversely to each other. Without Receiver Operating Characteristic Analysis (ROC) analysis, it is difficult to summarize the performance of a diagnostic test with a manageable number of statistics or to compare the performance of different diagnostic tests.

The use of the ROC curve as the measure of accuracy allows for the following interpretations:

- The average value of sensitivity for all possible values of specificity
- The average value of specificity for all possible values of sensitivity
- The probability that a randomly selected patient with the condition of interest has a diagnostic test result indicating greater suspicion than that of a randomly chosen patient without the condition of interest.

ESTIMATION AND HYPOTHESIS TESTING IN A SINGLE SAMPLE (FOR A SINGLE DIAGNOSTIC TEST)

ESTIMATION:

Diagnostic accuracy for ordinal data (1=normal, 2=benign, 3=probably benign, 4=suspicious, 5=malignant) or continuous data (e.g., any continuous prognostic variable) can be assessed by the AUC of the ROC curve (A). This is not true for binary data (e.g., positive vs. negative). Nonparametric methods such as the Mann-Whitney and chi square statistics can be used to calculate and plot ROC curves from single samples. One can perform ROC analysis (i.e., create the ROC curve, calculate the AUC of the ROC curve, calculate the variance of the AUC of the ROC curve) on a single diagnostic test, using the method of DeLong, DeLong and Clarke-Pearson (1988), which provides an improved standard error compared to that of Hanley and McNeil (1982).
In order to perform a ROC analysis, one needs to select an appropriate gold standard that accurately determines the presence or absence of the condition of interest. A measure of accuracy of a diagnostic test (e.g., the AUC of the ROC curve of a diagnostic test) cannot be estimated without the use of an appropriate (unbiased) gold standard. The statistical formulas presented in this paper were obtained from Zhou X-H, Obuchowski NA, McClish DK (2002).

**HYPOTHESIS TESTING:**

For testing whether or not a diagnostic test has a measure of accuracy significantly different from a pre-specified value (e.g., $A=0.50$), one considers the following null and alternative hypotheses:

$$H_0 : A = 0.50$$  \[1.0\]

and

$$H_A : A \neq 0.50$$  \[1.1\]

The test statistic will be:

$$Z = \frac{\hat{A} - 0.50}{\sqrt{\text{Var}(\hat{A})}}$$  \[1.2\]

This test statistic \[1.2\] usually follows a normal distribution asymptotically (i.e., for large samples).

**SAMPLE SIZE REQUIREMENTS FOR THE ACCURACY OF A SINGLE DIAGNOSTIC TEST**

**ESTIMATION:**

A general formula for sample size estimation for constructing a 2-sided Confidence Interval (CI) for a single diagnostic test measure of accuracy is:

$$m = \frac{z_{1-\alpha/2}^2 \left[V(\hat{A})\right]}{L^2}$$  \[1.3\]

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ percentile of the standard normal distribution, $\alpha$ is the confidence level, $V(\hat{A})$ is the variance function of $\hat{A}$, and $L$ is the desired width of one-half of the CI. Often, one wishes to construct a 95% CI, in which case $\alpha = 0.05$ and $z_{1-\alpha/2} = 1.96$.

With respect to the AUC of the ROC curve ($A$), $m$ will be the number of patients with the condition of interest. The total sample size required will be given by $m(1 + \kappa)$, where $\kappa$ denotes the ratio of the number of patients without the condition of interest to patients with the condition of interest in the study sample. One can use

$$V(\hat{A}) = \left(0.00999 e^{-a^2/2}\right) \left[5\hat{a}^2 + 8\right] + \left(a^2 + 8\right)/\kappa$$  \[1.4\]
to calculate \( V(\hat{A}) \) when the measure of accuracy is the AUC of the ROC curve.

**HYPOTHESIS TESTING:**

For testing whether or not a diagnostic test has a measure of accuracy significantly different from a pre-specified value (e.g., \( A = A_0 \)), one considers the following null and alternative hypotheses:

\[
H_0 : A = A_0 \quad \text{[1.5]}
\]

and

\[
H_A : A \neq A_0 \quad \text{[1.6]}
\]

A general formula for computing sample size for a study that tests these hypotheses is:

\[
m = \frac{\left[ z_{1-\alpha/2} \sqrt{V_0(\hat{A})} + z_{1-\beta} \sqrt{V_A(\hat{A})} \right]^2}{(A_0 - A_1)^2} \quad \text{[1.7]}
\]

where \( z_{1-\alpha} \) is the \( 1 - \alpha / 2 \) percentile of a standard normal distribution; \( \alpha \) is the type-I error rate (2-tailed test); \( z_{1-\beta} \) is the \( 1 - \beta \) percentile of a standard normal distribution; \( \beta \) is the type-II error rate (or 1-power); \( V_0(\hat{A}) \) is the variance function of the summary measure of accuracy of the diagnostic test under the null hypothesis; and \( V_A(\hat{A}) \) is the variance function of the summary measure of accuracy of the diagnostic test under the alternative hypothesis. \( A_0 \) is the pre-specified value of \( A \) under the null hypothesis and \( A_1 \) is the conjectured value of \( A \) under the alternative hypothesis.

With respect to the AUC of the ROC curve, \( m \) will be the number of patients with the condition of interest. The total sample size required will be given by \( m(1 + \kappa) \), where \( \kappa \) denotes the ratio of the number of patients without the condition of interest to patients with the condition of interest in the study sample.

An estimate of \( V_0(\hat{A}) \) and \( V_A(\hat{A}) \), which assumes an underlying binormal distribution for the diagnostic test results, is given by [1.4]:

\[
\hat{V}(\hat{A}) = \left( 0.00999e^{-a^2/2} \right) \left[ (5a^2 + 8) + (a^2 + 8)/\kappa \right]
\]

where \( a = \Phi^{-1}(\hat{A})(1.414) \) and \( \Phi^{-1} \) is the inverse of the cumulative normal distribution function. The variable \( a \) in [1.4] is parameter \( a \) from a binormal distribution.
REQUIRED SAMPLE SIZES BASED ON STUDY DESIGN PARAMETERS AND OPERATIONAL DEFINITIONS

We will consider a range of $\kappa$ based on a prevalence rate of the event of 18% to 25%.

For 18%, we have $\kappa = \frac{0.82}{0.18} = 4.56$

For 25%, we have $\kappa = \frac{0.75}{0.25} = 3.00$

FOR ESTIMATION:

A general formula for sample size estimation for constructing a 2-sided Confidence Interval (CI) for a single diagnostic test measure of accuracy is:

$$ m = \frac{z_{1-\alpha/2}^2}{L^2} \left[ V\left(\hat{A}\right) \right] $$

A range of conjectured AUC of the ROC curve (\(\hat{A}\)) will be considered: 0.85, 0.87, 0.90, 0.93, and 0.96.

$$ V\left(\hat{A}\right) \text{ is the variance function of } \hat{A}; \quad V\left(\hat{A}\right) = \left(0.0099e^{-a^2/2}\right)\left[\left(5a^2 + 8\right) + \left(a^2 + 8\right)/\kappa\right]. $$

L=0.05 (the desired width of one-half of the CI).

We will construct a 95% CI, in which case $\alpha = 0.05$ and $z_{1-\alpha/2} = 1.96$.

With respect to the AUC of the ROC curve, $m$ will be the number of patients with the event. The total sample size required will be given by $m(1 + \kappa)$, where $\kappa$ denotes the ratio of the number of patients without the event to patients with the event in the study sample.

Table 1. Required Sample Size for Estimation

<table>
<thead>
<tr>
<th>$L=0.05$</th>
<th></th>
<th>$L=0.05$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa = 4.56$</td>
<td>$\kappa = 3.00$</td>
<td>$\kappa = 3.00$</td>
<td>$\kappa = 3.00$</td>
</tr>
<tr>
<td>m</td>
<td>Total N</td>
<td>m</td>
<td>Total N</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>0.85</td>
<td>109</td>
<td>606</td>
<td>0.85</td>
</tr>
<tr>
<td>0.90</td>
<td>79</td>
<td>439</td>
<td>0.90</td>
</tr>
<tr>
<td>0.93</td>
<td>66</td>
<td>367</td>
<td>0.93</td>
</tr>
<tr>
<td>0.96*</td>
<td>37</td>
<td>206</td>
<td>0.96*</td>
</tr>
</tbody>
</table>

* One-sided 95% CI
FOR HYPOTHESIS TESTING:

For testing whether or not a diagnostic test has a measure of accuracy significantly different from a pre-specified value (e.g., $A = A_0$), one considers the following null and alternative hypotheses:

$$H_0 : A = A_0$$

where $A_0 = 0.85$ (assumed minimum acceptable AUC of the ROC curve)

and

$$H_A : A \neq A_0$$

A general formula for computing sample size for a study that tests these hypotheses is:

$$m = \frac{z_{1-\alpha/2}^2 \sqrt{V_0}(A) + z_{1-\beta}^2 \sqrt{V_A}(A)}{(A_0 - A_1)^2}$$

where

- $z_{1-\alpha} = 1.96$ (the $1 - \alpha / 2$ percentile of a standard normal distribution) when $\alpha = 0.05$ (the type-I error rate; 2-tailed test).
- $z_{1-\beta} = 0.84$ (the $1 - \beta$ percentile of a standard normal distribution) when $\beta = 0.20$ (the type-II error rate).

$V_0(A)$ is the variance function of the summary measure of accuracy of the diagnostic test under the null hypothesis and $V_A(A)$ is the variance function of the summary measure of accuracy of the diagnostic test under the alternative hypothesis. $A_0$ is the pre-specified value of $A$ under the null hypothesis and $A_1$ is the conjectured value of $A$ under the alternative hypothesis ($0.87, 0.90, \text{and } 0.93$).

With respect to the AUC of the ROC curve, $m$ will be the number of patients with the event. The total sample size required will be given by $m(1 + \kappa)$, where $\kappa$ denotes the ratio of the number of patients without the event to patients with the event in the study sample ($\kappa = 4.56, \kappa = 3.00$).

An estimate of $V_0(A)$ and $V_A(A)$, which assumes an underlying binormal distribution for the diagnostic test results, is given by:

$$\hat{V}(A) = \left(0.00999e^{-a^2/2}\right)\left[5a^2 + 8\right] + \left(a^2 + 8\right)/\kappa$$
where \( a = \Phi^{-1}(A)(1.414) \) and \( \Phi^{-1} \) is the inverse of the cumulative normal distribution function. The variable \( a \) in is parameter \( a \) from a binormal distribution.

Table 2. Required Sample Size for Hypothesis Testing

<table>
<thead>
<tr>
<th>( \kappa = 4.56 \quad A_0 = 0.85 )</th>
<th>( \kappa = 3.00 \quad A_0 = 0.85 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m )</td>
<td>Total N</td>
</tr>
<tr>
<td>( A_1 )</td>
<td>( A_1 )</td>
</tr>
<tr>
<td>0.87</td>
<td>1065</td>
</tr>
<tr>
<td>0.90</td>
<td>204</td>
</tr>
<tr>
<td>0.93</td>
<td>67</td>
</tr>
<tr>
<td>0.95</td>
<td>43</td>
</tr>
</tbody>
</table>

REFERENCES


CONTACT INFORMATION

William F. McCarthy
Principal Statistician
Director of Clinical Trial Statistics and SAS Programming
Maryland Medical Research Institute
600 Wyndhurst Avenue
Baltimore, Maryland 21210-2425
(410) 435-4200
www.mmri.org
wmccarthy@mmri.org

Nan Guo
Senior Lead SAS Programmer
SAS Certified Advanced Programmer for SAS9
Maryland Medical Research Institute
600 Wyndhurst Avenue
Baltimore, Maryland 21210-2425
(410) 435-4200
www.mmri.org
nguo@mmri.org

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

SAS CODE FOR USER INTERFACE:

```sas
options symbolgen mlogic mprint mcompilenote=all nocenter nonumber nodate
ORIENTATION=LANDSCAPE ps=60 ls=120 ;

%let path=C:\Documents and Settings\wmccarthy\My Documents\sugi 2007 paper 2\;

libname mylib "&path\data";

%inc "&path\macros\Cal_m_Esti_Accu_single_Dx.sas";

filename errinp "&path\csv file\Error_levl.csv";

* ------------------------------------------------------------------ *
* Creates a Look-Up Table *
* Z value for N(0,1) distribution for *
* selected error levels *
* ------------------------------------------------------------------ *

data dslkup;
infile errinp dso firstobs=2 missover;
```
length err tail1 tail2 5.3;
input err taill tail2;
run;

* --- Prints out look-up table --- *
proc print data= dslkup;
run cancel ;

* -------------------------------------------------------------- *
* ---- Creates the FIRST PAGE OF THE USER INTERFACE windows ---- *
* -------------------------------------------------------------- *
%window Greeting color=green
    #10 @10 "Sample Size Requirement for Estimation of Accuracy of a Single Diagnostic Test"
    #13 @30 "When using AUC of ROC curves "
    #16 @38 "Written by"
    #18 @40 "Nan Guo"
    #20 @34 "William F. McCarthy"
    #24 @27 "Maryland Medical Research Institute"
    #40 @2 "Press ENTER to continue.";

* displays window greeting;
%display Greeting;

* -------------------------------------------------------------- *
* --- Creates user input interface --- *
* -------------------------------------------------------------- *
%macro getnu_sn;
%global a L tail kappa alpha chyn;
%geta:   %let a     = ;
    %let L     = ;
    %let tail  = ;
    %let kappa = ;
    %let alpha = ;
    %let chyn  = ;

%window Get_nusn
    #12 @12 "Please Enter conjectured AUC of ROC curve for the technology"
    #78 a 4 attr=underline color=pink
    #14 @12 "Please enter width of one-half of the CI, L= "
    #78 L 4 attr=underline color=pink
    #15 @12 "one-tailed OR two-tailed (1=one-tailed 2=two-tailed)?"
    #78 tail 1 attr=underline color=red
    #16 @12 "kappa = " @78 kappa 5 attr=underline color=red
    #17 @12 "alpha = " @78 alpha 5 attr=underline color=red
    #50 @12 "Press ENTER to continue.";

%display Get_nusn;

    %let a     = %sysevalf(&a*1);
    %let L     = %sysevalf(&L*1);
    %let tail  = &tail ;
    %let kappa = %sysevalf(&kappa*1);
    %let alpha = %sysevalf(&alpha*1);
    %let chyn  = &chyn ;
    %if %index(%upcase(&tail),1) eq 0 and
        %index(%upcase(&tail),2) eq 0 %then %goto geta;

%window repeat
    #11 @12 "YOU HAVE ENTERED: "
    #12 @12 "---------------------------------------------------------------"
#13 @12 "conjectured AUC of ROC curve for the technology = &a"
#14 @12 "width of one-half of the CI = &L"
#15 @12 "One-tailed OR two-tailed (1=one-tailed 2=two-tailed) = &tail"
#16 @12 "kappa = &kappa"
#17 @12 "alpha = &alpha"
#18 @12 "------------------------------------------------------------------"
#25 @12 "Do you want to change any of these values (y=YES or n=NO)? "
    chyn chyn attr=underline color=red
#78 chyn 1 attr=underline color=red
#50 @12 "Press ENTER to continue."

%display repeat;
%if %upcase(&chyn) = Y %then %goto geta;
%else %if %upcase(&chyn) = N %then %Cre_M_Single_Dx;

%mend getnu_sn;

* --- Calls macro getnu_sn --- *
%getnu_sn

SAS Macro:

%* --- C:\From_Bill\TOSHIBA\Protocol Sample Size
%* --- Calculates Values Aa --- *
%macro cal(var);
  select (&var);
    when(0.75) A&var=0.95;
    when(0.80) A&var=1.19;
    when(0.85) A&var=1.46;
    when(0.90) A&var=1.82;
  otherwise A&var=round(-3.39700+5.76000*&var,0.01);
end;
%mend;

* ---------------------------------------------------------------- *
* --- calculates sample size m for Test of Estimation Accuracy --- *
* ---------------------------------------------------------------- *
%macro Cre_M_Single_Dx;
data abc(drop=err tail:);
  set dskup end=EOF;
  %* ------ Z value ------ *
  if round(err,.001) = &alpha then call symput('Za',trim(left(put(tail&tail,5.3))));
    a=symget('a');
    if EOF then output;
  run;

data Cre_cal_M;
  set abc;
  %* --- calculates Aa --- *
  %cal(a)

  %* --- calculates variance of Aa --- *
  Va=(0.0099*exp((-Aa**2)/2))*((5*(Aa**2)+8)+(Aa**2+8)/&kappa);

  %* --- Zalpha critical value --- *
  Za = input(symget('Za'),best12.);
%* --- Sample Size of Test --- *
m   = round((Za**2) * Va/(&L**2));

%* --- Total sample size for single Dx;
tot = round(m *(*1+&kappa));

label m = "Sample Size for Estimation Accuracy Single Diagnostic Test"
tot = "Total Sample Size for Estimation Accuracy Single Diagnostic Test";

run;
%* --- prints out macro Za & Zb values in the .LOG --- *
$put macro Za= &Za L=&L a=&a Kappa=&kappa;

%* --- Prints data set out --- *
proc print data=Cre_cal_M label;
run;

* ---- ODS begins ---- *
proc template;
define style Newstyle;
  style cellcontents / background=white
                  font_face="arial, helvetica"
                  font_style=roman
                  font_weight=bold
                  font_size=3;
  style header /
    background=white
    borderwidth = 0
    BORDERCOLOR = white
    font_face="times"
    font_weight=bold
    font_style=roman
    cellwidth=5 in
    font_size=3;
  style systemtitle /
    background=white
    borderwidth = 0
    BORDERCOLOR = white
    font_face="arial, helvetica"
    font_weight=bold
    font_style=italic
    font_size=3;
  style SysTitleAndFooterContainer from systemtitle /
    font_size=3;
  style systemfooter from systemtitle /
    font_size=3;
  style footer from systemtitle /
    font_size=3;
  style table /
    borderwidth = 1
    background = white
    font_style = Roman
    font_weight = Bold
    font_face = "arial, helvetica"
    cellspacing = 1;
end;
run;

proc template;
  define table Cellstyle;
    mvar sysdate9;
    dynamic colhd;
    define column all;
      generic=on;
      header=colhd;
      just=center;
      justify=on;
      style=cellcontents;
    end;
    define footer table_footer;
      text 'Run on ' sysdate9;
    end;
  end;
run;

ods listing close;

ods rtf file="&path\rtf File\Sample_Size_Esti_Accu_Single_Dx.rtf"
style=newstyle;
title1 j=C "Sample Size Requirement for Estimation of Accuracy of a Single Diagnostic Test";
title2 j=C "When using AUC of ROC curves ";
title3 " ";
title4 h=3 "Given values are:" ;
title5 "one-tailed OR two-tailed (1=one-tailed 2=two-tailed) = &tail";
title6 "conjectured AUC of ROC curve for the test = &a   width of one-half of the CI = &L   kappa = &kappa   alpha = &alpha ";

footnotel italic bold h=3 j=l "Pgm:\\ Estimation_accuracy_Test_Single_Dx.sas" j=R "Run:&sysdate ";

data _null_;  
  set Cre_cal_M;
  file print ods=(template='cellstyle'
    columns=(all= m(generic=on dynamic=(colhd='Sample Size for Estimation Accuracy Single Diagnostic Test'))
             all=tot(generic=on dynamic=(colhd='Total Sample Size for Estimation Accuracy Single Diagnostic Test')))  
  );
  put _ods_;
run;

ods rtf close;
ods listing;

%mend Cre_M_Single_Dx;

%Cre_M_Single_Dx
Look-Up Table

<table>
<thead>
<tr>
<th>Error Level</th>
<th>One-tailed</th>
<th>Two-tailed</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.500</td>
<td>0.674</td>
<td></td>
</tr>
<tr>
<td>0.400</td>
<td>0.253, 0.842</td>
<td></td>
</tr>
<tr>
<td>0.300</td>
<td>0.524, 1.036</td>
<td></td>
</tr>
<tr>
<td>0.200</td>
<td>0.842, 1.282</td>
<td></td>
</tr>
<tr>
<td>0.100</td>
<td>1.282, 1.645</td>
<td></td>
</tr>
<tr>
<td>0.050</td>
<td>1.645, 1.96</td>
<td></td>
</tr>
<tr>
<td>0.025</td>
<td>1.96, 2.248</td>
<td></td>
</tr>
<tr>
<td>0.010</td>
<td>2.326, 2.576</td>
<td></td>
</tr>
<tr>
<td>0.005</td>
<td>2.576, 2.813</td>
<td></td>
</tr>
</tbody>
</table>