

Chapter 537

Binary Diagnostic Tests – Two Independent Samples

Introduction

An important task in diagnostic medicine is to measure the accuracy of two diagnostic tests. This can be done by comparing summary measures of diagnostic accuracy such as *sensitivity* or *specificity* using a statistical test. Often, you want to show that a new test is similar to another test, in which case you use an equivalence test. Or, you may wish to show that a new diagnostic test is not inferior to the existing test, so you use a non-inferiority test. All of these hypothesis tests are available in this procedure for the important case when the diagnostic tests provide a binary (yes or no) result.

The results of such studies can be displayed in two 2-by-2 tables in which the true condition is shown as the rows and the diagnostic test results are shown as the columns.

True Condition	Diagnostic Test 1 Result			Diagnostic Test 2 Result		
	Positive	Negative	Total	Positive	Negative	Total
Present (True)	T_{11}	T_{10}	n_{11}	T_{21}	T_{20}	n_{21}
Absent (False)	F_{11}	F_{10}	n_{10}	F_{21}	F_{20}	n_{20}
Total	m_{11}	m_{10}	N_1	m_{21}	m_{20}	N_2

Data such as this can be analyzed using standard techniques for comparing two proportions which are presented in the chapter on Two Proportions. However, specialized techniques have been developed for dealing specifically with the questions that arise from such a study. These techniques are presented in chapter 5 of the book by Zhou, Obuchowski, and McClish (2002).

Test Accuracy

Several measures of a diagnostic test's accuracy are available. Probably the most popular measures are the test's *sensitivity* and the *specificity*. Sensitivity is the proportion of those that have the condition for which the diagnostic test is positive. Specificity is the proportion of those that do not have the condition for which the diagnostic test is negative. Other accuracy measures that have been proposed are the likelihood ratio and the odds ratio. Study designs anticipate that the sensitivity and specificity of the two tests will be compared.

Comparing Sensitivity and Specificity

Suppose you arrange the results of two diagnostic tests into two 2-by-2 tables as follows:

True Condition	Standard or Reference Diagnostic Test 1 Result			Treatment or Experimental Diagnostic Test 2 Result		
	Positive	Negative	Total	Positive	Negative	Total
Present (True)	T_{11}	T_{10}	n_{11}	T_{21}	T_{20}	n_{21}
Absent (False)	F_{11}	F_{10}	n_{10}	F_{21}	F_{20}	n_{20}
Total	m_{11}	m_{10}	N_1	m_{21}	m_{20}	N_2

Hence, the study design includes $N = n_1 + n_2$ patients.

The hypotheses of interest when comparing the sensitivities (Se) of two diagnostic tests are either the difference hypotheses

$$H_0: Se_1 - Se_2 = 0 \text{ versus } H_A: Se_1 - Se_2 \neq 0$$

or the ratio hypothesis

$$H_0: Se_1 / Se_2 = 1 \text{ versus } H_A: Se_1 / Se_2 \neq 1$$

Similar sets of hypotheses may be defined for the difference or ratio of the specificities (Sp) as

$$H_0: Sp_1 - Sp_2 = 0 \text{ versus } H_A: Sp_1 - Sp_2 \neq 0$$

and

$$H_0: Sp_1 / Sp_2 = 1 \text{ versus } H_A: Sp_1 / Sp_2 \neq 1$$

Note that the difference hypotheses usually require a smaller sample size for comparable statistical power, but the ratio hypotheses may be more convenient.

The sensitivities are estimated as

$$\hat{Se}_1 = \frac{T_{11}}{n_{11}} \text{ and } \hat{Se}_2 = \frac{T_{21}}{n_{21}}$$

and the specificities are estimated as

$$\hat{Sp}_1 = \frac{F_{10}}{n_{10}} \text{ and } \hat{Sp}_2 = \frac{F_{20}}{n_{20}}$$

The sensitivities and specificities of the two diagnostic tests may be compared using either their difference or their ratio.

As can be seen from the above, comparison of the sensitivity or the specificity reduces to the problem of comparing two independent binomial proportions. Hence the formulas used for hypothesis testing and confidence intervals are the same as presented in the chapter on testing two independent proportions. We refer you to that chapter for further details.

Data Structure

This procedure does not use data from a dataset. Instead, you enter the values directly into the panel. The data are entered into two tables. The table on the left represents the existing (standard) test. The table on the right contains the data for the new diagnostic test.

Zero Cells

Although zeroes are valid values, they make direct calculation of ratios difficult. One popular technique for dealing with the difficulties of zero values is to enter a small 'delta' value such as 0.50 or 0.25 in the zero cells so that division by zero does not occur. Such methods are controversial, but they are commonly used. Probably the safest method is to use the hypotheses in terms of the differences rather than ratios when zeroes occur, since these may be calculated without adding a delta.

Example 1 – Binary Diagnostic Test of Two Independent Samples

This section presents an example of how to enter data and run an analysis. In this example, samples of 50 individuals known to have a certain disease and 50 individuals without the disease were divided into two, equal-sized groups. Half of each group was given diagnostic test 1 and the other half was given diagnostic test 2. The results are summarized into the following tables:

True Condition	Standard or Reference Diagnostic Test 1 Result			Treatment or Experimental Diagnostic Test 2 Result		
	Positive	Negative	Total	Positive	Negative	Total
Present (True)	20	5	25	21	4	25
Absent (False)	7	18	25	5	20	25
Total	27	23	50	26	24	50

Setup

To run this example, complete the following steps:

1 Specify the Binary Diagnostic Tests – Two Independent Samples procedure options

- Find and open the **Binary Diagnostic Tests – Two Independent Samples** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Data Tab	
T11	20
T10	5
F11	7
F10	18
T21	21
T20	4
F21	5
F20	20
Difference C.I. Method	Score w/ Skewness (Gart-Nam)
Ratio C.I. Method	Score w/ Skewness (Gart-Nam)
Max Equivalence Difference	0.25
Max Equivalence Ratio	1.25

2 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Binary Diagnostic Tests – Two Independent Samples

Data and Proportions Sections

Counts for Tests 1 and 2

True Condition	Counts for Test 1			Counts for Test 2		
	Diagnostic Test Result			Diagnostic Test Result		
	Positive	Negative	Total	Positive	Negative	Total
Present	20	5	25	21	4	25
Absent	7	18	25	5	20	25
Total	27	23	50	26	24	50

Table Proportions for Tests 1 and 2

True Condition	Table Proportions for Test 1			Table Proportions for Test 2		
	Diagnostic Test Result			Diagnostic Test Result		
	Positive	Negative	Total	Positive	Negative	Total
Present	0.4000	0.1000	0.5000	0.4200	0.0800	0.5000
Absent	0.1400	0.3600	0.5000	0.1000	0.4000	0.5000
Total	0.5400	0.4600	1.0000	0.5200	0.4800	1.0000

Row Proportions for Tests 1 and 2

True Condition	Row Proportions for Test 1			Row Proportions for Test 2		
	Diagnostic Test Result			Diagnostic Test Result		
	Positive	Negative	Total	Positive	Negative	Total
Present	0.8000	0.2000	1.0000	0.8400	0.1600	1.0000
Absent	0.2800	0.7200	1.0000	0.2000	0.8000	1.0000
Total	0.5400	0.4600	1.0000	0.5200	0.4800	1.0000

Column Proportions for Tests 1 and 2

True Condition	Column Proportions for Test 1			Column Proportions for Test 2		
	Diagnostic Test Result			Diagnostic Test Result		
	Positive	Negative	Total	Positive	Negative	Total
Present	0.7407	0.2174	0.5000	0.8077	0.1667	0.5000
Absent	0.2593	0.7826	0.5000	0.1923	0.8333	0.5000
Total	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

These reports display the counts that were entered for the two tables along with various proportions that make interpreting the table easier. Note that the sensitivity and specificity are displayed in the Row Proportions table.

Sensitivity Confidence Intervals Section

Sensitivity Confidence Intervals Section

Statistic	Test	Value	Lower 95.0% Conf. Limit	Upper 95.0% Conf. Limit
Sensitivity (Se1)	1	0.8000	0.6087	0.9114
Sensitivity (Se2)	2	0.8400	0.6535	0.9360
Difference (Se1-Se2)		-0.0400	-0.2618	0.1824
Ratio (Se1/Se2)		0.9524	0.7073	1.2659

Notes:

Sensitivity: proportion of those that actually have the condition for which the diagnostic test is positive.

Difference confidence limits based on Gart and Nam's score method with skewness correction.

Ratio confidence limits based on Gart and Nam's score method with skewness correction.

This report displays the sensitivity for each test as well as corresponding confidence interval. It also shows the value and confidence interval for the difference and ratio of the sensitivity. Note that for a perfect diagnostic test, the sensitivity would be one. Hence, the larger the values the better.

Also note that the type of confidence interval for the difference and ratio is specified on the Data panel. The Wilson score method is used to calculate the individual confidence intervals for the sensitivity of each test.

Specificity Confidence Intervals Section

Specificity Confidence Intervals Section

Statistic	Test	Value	Lower 95.0% Conf. Limit	Upper 95.0% Conf. Limit
Specificity (Sp1)	1	0.7200	0.5242	0.8572
Specificity (Sp2)	2	0.8000	0.6087	0.9114
Difference (Sp1-Sp2)		-0.0800	-0.3173	0.1616
Ratio (Sp1/Sp2)		0.9000	0.6326	1.2501

Notes:

Specificity: proportion of those that do not have the condition for which the diagnostic test is negative.

Difference confidence limits based on Gart and Nam's score method with skewness correction.

Ratio confidence limits based on Gart and Nam's score method with skewness correction.

This report displays the specificity for each test as well as corresponding confidence interval. It also shows the value and confidence interval for the difference and ratio of the specificity. Note that for a perfect diagnostic test, the specificity would be one. Hence, the larger the values the better.

Also note that the type of confidence interval for the difference and ratio is specified on the Data panel. The Wilson score method is used to calculate the individual confidence intervals for the specificity of each test.

Sensitivity & Specificity Hypothesis Test Section

Sensitivity & Specificity Hypothesis Test Section

Hypothesis Test of	Value	Chi-Square	Prob Level	Decision at 5.0% Level
Se1 = Se2	-0.0400	0.1355	0.7128	Cannot Reject H0
Sp1 = Sp2	-0.0800	0.4386	0.5078	Cannot Reject H0

This report displays the results of hypothesis tests comparing the sensitivity and specificity of the two diagnostic tests. The Pearson chi-square test statistic and associated probability level is used.

Likelihood Ratio Section

Likelihood Ratio Section

Statistic	Test	Value	Lower 95.0% Conf. Limit	Upper 95.0% Conf. Limit
LR(Test=Positive)	1	2.8571	1.5948	6.1346
	2	4.2000	2.1020	10.8255
LR(Test=Negative)	1	0.2778	0.1064	0.5729
	2	0.2000	0.0669	0.4405

Notes:

$LR(\text{Test} = +) = P(\text{Test} = + \mid \text{True Condition} = +) / P(\text{Test} = + \mid \text{True Condition} = -)$.

$LR(\text{Test} = +) > 1$ indicates a positive test is more likely among those in which True Condition = +.

$LR(\text{Test} = -) = P(\text{Test} = - \mid \text{True Condition} = +) / P(\text{Test} = - \mid \text{True Condition} = -)$.

$LR(\text{Test} = -) < 1$ indicates a negative test is more likely among those in which True = -.

This report displays the positive and negative likelihood ratios with their corresponding confidence limits. You would want $LR(+)$ > 1 and $LR(-)$ < 1, so place close attention that the lower limit of $LR(+)$ is greater than one and that the upper limit of $LR(-)$ is less than one.

Note the $LR(+)$ means $LR(\text{Test}=\text{Positive})$. Similarly, $LR(-)$ means $LR(\text{Test}=\text{Negative})$.

Odds Ratio Section

Odds Ratio Section

Statistic	Test	Value	Lower 95.0% Conf. Limit	Upper 95.0% Conf. Limit
Odds Ratio (+ 1/2)	1	9.1939	2.5893	32.6452
	2	17.8081	4.4579	71.1386
Odds Ratio (Fleiss)	1	9.1939	2.3607	48.8331
	2	17.8081	4.1272	123.3154

Notes:
 Odds Ratio = Odds(True Condition = +) / Odds(True Condition = -)
 where
 Odds(Condition) = P(Positive Test | Condition) / P(Negative Test | Condition)

This report displays estimates of the odds ratio as well as its confidence interval. Because of the better coverage probabilities of the Fleiss confidence interval, we suggest that you use this interval.

Hypothesis Tests of the Equivalence of Sensitivity

Hypothesis Tests of the Equivalence of Sensitivity

Statistic	Prob Level	Lower 90.0% Conf. Limit	Upper 90.0% Conf. Limit	Lower Equiv. Bound	Upper Equiv. Bound	Reject H0 and Conclude Equivalence at the 5.0% Significance Level
Diff. (Se1-Se2)	0.0314	-0.2246	0.1446	-0.2500	0.2500	Yes
Ratio (Se1/Se2)	0.1109	0.7471	1.2021	0.8000	1.2500	No

Notes:
 Equivalence is concluded when the confidence limits fall completely inside the equivalence bounds.
 Difference confidence limits based on Gart and Nam's score method with skewness correction.
 Ratio confidence limits based on Gart and Nam's score method with skewness correction.

This report displays the results of the equivalence tests of sensitivity, one based on the difference and the other based on the ratio. Equivalence is concluded if the confidence limits are inside the equivalence bounds.

Prob Level

The probability level is the smallest value of alpha that would result in rejection of the null hypothesis. It is interpreted as any other significance level. That is, reject the null hypothesis when this value is less than the desired significance level.

Note that for many types of confidence limits, a closed form solution for this value does not exist and it must be searched for.

Confidence Limits

These are the lower and upper confidence limits calculated using the method you specified. Note that for equivalence tests, these intervals use twice the alpha. Hence, for a 5% equivalence test, the confidence coefficient is 0.90, not 0.95.

Binary Diagnostic Tests – Two Independent Samples

Lower and Upper Bounds

These are the equivalence bounds. Values of the difference (ratio) inside these bounds are defined as being equivalent. Note that this value does not come from the data. Rather, you have to set it. These bounds are crucial to the equivalence test and they should be chosen carefully.

Reject H0 and Conclude Equivalence at the 5% Significance Level

This column gives the result of the equivalence test at the stated level of significance. Note that when you reject H0, you can conclude equivalence. However, when you do not reject H0, you cannot conclude nonequivalence. Instead, you conclude that there was not enough evidence in the study to reject the null hypothesis.

Hypothesis Tests of the Equivalence of Specificity

Hypothesis Tests of the Equivalence of Specificity						
Statistic	Prob Level	Lower 90.0% Conf. Limit	Upper 90.0% Conf. Limit	Lower Equiv. Bound	Upper Equiv. Bound	Reject H0 and Conclude Equivalence at the 5.0% Significance Level
Diff. (Sp1-Sp2)	0.0793	-0.2787	0.1216	-0.2500	0.2500	No
Ratio (Sp1/Sp2)	0.2385	0.6744	1.1799	0.8000	1.2500	No

Notes:
 Equivalence is concluded when the confidence limits fall completely inside the equivalence bounds.
 Difference confidence limits based on Gart and Nam's score method with skewness correction.
 Ratio confidence limits based on Gart and Nam's score method with skewness correction.

This report displays the results of the equivalence tests of specificity, one based on the difference and the other based on the ratio. Report definitions are identical with those above for sensitivity.

Hypothesis Tests of the Non-inferiority of Sensitivity

Tests Showing the Sensitivity Non-inferiority of Test 1 Compared to Test 2

Statistic	Prob Level	Lower 90.0% Conf. Limit	Upper 90.0% Conf. Limit	Lower Equiv. Bound	Upper Equiv. Bound	Reject H0 and Conclude Non-inferiority at the 5.0% Significance Level
Diff. (Se1-Se2)	0.0314	-0.2246	0.1446	-0.2500	0.2500	Yes
Ratio (Se1/Se2)	0.1109	0.7471	1.2021	0.8000	1.2500	No

Notes:

H0: The sensitivity of Test 1 is inferior to Test 2.

Ha: The sensitivity of Test 1 is non-inferior to Test 2.

The non-inferiority of Test 1 compared to Test 2 is concluded when the lower c.l. > lower bound.

Difference confidence limits based on Gart and Nam's score method with skewness correction.

Ratio confidence limits based on Gart and Nam's score method with skewness correction.

This report displays the results of two non-inferiority tests of sensitivity, one based on the difference and the other based on the ratio. The non-inferiority of test 2 as compared to test 1 is concluded if the upper confidence limit is less than the upper bound. The columns are as defined above for equivalence tests.

Hypothesis Tests of the Non-inferiority of Specificity

Tests Showing the Specificity Non-inferiority of Test 1 Compared to Test 2

Statistic	Prob Level	Lower 90.0% Conf. Limit	Upper 90.0% Conf. Limit	Lower Equiv. Bound	Upper Equiv. Bound	Reject H0 and Conclude Non-inferiority at the 5.0% Significance Level
Diff. (Sp1-Sp2)	0.0793	-0.2787	0.1216	-0.2500	0.2500	No
Ratio (Sp1/Sp2)	0.2385	0.6744	1.1799	0.8000	1.2500	No

Notes:

H0: The specificity of Test 1 is inferior to Test 2.

Ha: The specificity of Test 1 is non-inferior to Test 2.

The non-inferiority of Test 1 compared to Test 2 is concluded when the lower c.l. > lower bound.

Difference confidence limits based on Gart and Nam's score method with skewness correction.

Ratio confidence limits based on Gart and Nam's score method with skewness correction.

This report displays the results of the non-inferiority tests of specificity. Report definitions are identical with those above for sensitivity.