Chapter 536

Binary Diagnostic Tests – Paired 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 noninferiority 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.

Experimental Design

Suppose you are interested in comparing the sensitivities of two diagnostic tests for a particular disease (or condition). Each test provides a binary (yes or no) response. Further suppose you draw a random sample of subjects from the population with the disease and administered both diagnostic tests to each subject in random order. Assume that Test 1 is a new (experimental or treatment) test that will replace Test 2, the existing (standard or reference) test, if it is found to be better.

The results of such a study can be displayed in a 2-by-2 table in which the Test 1 result is shown as the rows and the Test 2 result is shown as the columns.

	Test 2 Result		
Test 1 Result	Positive	Negative	Total
Positive	X_{11}	X_{10}	m_1
Negative	X_{01}	X_{00}	m_0
Total	n_1	n_0	N

Data such as this can be analyzed using standard techniques for comparing two correlated proportions which are presented in the chapter on Two Correlated Proportions. Such a table was originally analyzed using McNemar's Test. However, procedures with better statistical properties have recently been proposed. See for example Nam and Blackwelder (2002).

Sensitivity

Sensitivity is the proportion of those that have the condition for which the diagnostic test is positive. Since this design assumes that the subjects come from the population of individuals with the disease, the sensitivity can be calculated.

Specificity

Specificity is the proportion of those that do not have the condition for which the diagnostic test is negative. To study specificity, a separate study would have to be conducted in which subjects were drawn from the population of individuals without the disease. The data from such a study could be analyzed with this procedure by changing the meaning of *positive* and *negative*. Instead of positive meaning that the person had the disease, positive would mean that the diagnostic test result matched the true condition of the subject. Likewise, negative would mean that the diagnostic test result did not match the true condition. In the procedure printouts, you would substitute specificity for sensitivity.

Comparing Sensitivity and Specificity

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

	Test 2 Result		
Test 1 Result	Positive	Negative	Total
Positive	X_{11}	X_{10}	m_1
Negative	X_{01}	X_{00}	m_0
Total	n_1	n_0	N

Hence, the study design includes $N = n_1 + n_0$ 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

$$\hat{S}e_1 = \frac{m_1}{N} \text{ and } \hat{S}e_2 = \frac{n_1}{N}$$

The sensitivities of the two diagnostic tests may be compared using either their differences or their ratios. Hence, the comparison of the sensitivity reduces to the problem of comparing two correlated binomial proportions. The formulas used for hypothesis testing and confidence intervals are the same as presented in the chapter on testing two correlated 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 2-by-2 table on the panel.

Example 1 – Binary Diagnostic Test of Paired Samples

This section presents an example of how to enter data and run an analysis. In this example, a sample of 50 individuals known to have a certain disease was selected. For this study, Test 1 refers to a new, cheaper, less-invasive diagnostic test and Test 2 refers to the standard diagnostic test that is currently being used. The results are summarized into the following table:

	Test 2 Result		
Test 1 Result	Positive	Negative	Total
Positive	31	5	36
Negative	4	10	14
Total	35	15	50

Setup

To run this example, complete the following steps:

1 Specify the Binary Diagnostic Tests – Paired Samples procedure options

- Find and open the **Binary Diagnostic Tests – Paired 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

X11 Count..... **31**
X10 Count..... **5**
X01 Count..... **4**
X00 Count..... **10**
Difference C.I. Method..... **Score (Nam RMLE)**
Ratio C.I. Method..... **Score (Nam Blackwelder)**
Max Equivalence Difference..... **0.2**
Max Equivalence Ratio..... **1.25**

2 Run the procedure

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

Data and Proportions

Counts and Table Proportions

Test 1 (New) Result	Counts			Table Proportions		
	Test 2 (Standard) Result			Test 2 (Standard) Result		
	Positive	Negative	Total	Positive	Negative	Total
Positive	31	5	36	0.6200	0.1000	0.7200
Negative	4	10	14	0.0800	0.2000	0.2800
Total	35	15	50	0.7000	0.3000	1.0000

Row Proportions and Column Proportions

Test 1 (New) Result	Row Proportions			Column Proportions		
	Test 2 (Standard) Result			Test 2 (Standard) Result		
	Positive	Negative	Total	Positive	Negative	Total
Positive	0.8611	0.1389	1.0000	0.8857	0.3333	0.7200
Negative	0.2857	0.7143	1.0000	0.1143	0.6667	0.2800
Total	0.7000	0.3000	1.0000	1.0000	1.0000	1.0000

These reports display the counts that were entered along with various proportions that make interpreting the table easier. Note that Test 1's sensitivity of 0.7200 and Test 2's sensitivity of 0.7000 are displayed in the margins of the Table Proportions table.

Sensitivity Confidence Intervals

Sensitivity Confidence Interval Section

Statistic	Test	Value	Lower 95.0% Conf. Limit	Upper 95.0% Conf. Limit
Sensitivity (Se1)	1	0.7200	0.5833	0.8253
Sensitivity (Se2)	2	0.7000	0.5625	0.8090
Difference (Se1-Se2)		0.0200	-0.1094	0.1511
Ratio (Se1/Se2)		1.0286	0.8524	1.2491

Notes:
Sensitivity: proportion of those that actually have the condition for which the diagnostic test is positive.
Difference confidence limits based on Nam's RMLE method.
Ratio confidence limits based on Blackwelder and Nam's method.

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.

Note that the type of confidence interval for the difference and ratio is specified on the Data panel.

Confidence Intervals for the Odds Ratio

Confidence Intervals for the Odds Ratio (Se12/Se21)

Confidence Interval Method	Value	Lower 95.0% Conf. Limit	Upper 95.0% Conf. Limit
Exact Conditional Binomial	1.2500	0.2690	6.2995
Maximum Likelihood	1.2500	0.3357	4.6549

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.

Hypothesis Tests about Sensitivity Difference

Hypothesis Tests about Sensitivity Difference

Test Name	Test Sides	Null Hypothesis (H0)	Test Statistic Value	Prob Level	Conclusion at the 5.0% Level
Nam	2	Se1-Se2=0	0.1111	0.7389	Cannot Reject H0
Nam Lower	1	Se1-Se2<=0	0.3333	0.3694	Cannot Reject H0
Nam Upper	1	Se1-Se2>=0	0.3333	0.6306	Cannot Reject H0

This report displays the results of hypothesis tests comparing the sensitivities of the two diagnostic tests using Nam's test. Note that for this test, identical test results are obtained from either the test of differences or test of ratios.

Tests of Equivalence

Tests of Equivalence

Parameter Tested	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
Difference (Se1-Se2)	0.0051	-0.0859	0.1274	-0.2000	0.2000	Yes
Ratio (Se1/Se2)	0.0247	0.8833	1.2039	0.8000	1.2500	Yes

Notes:
Equivalence is concluded when the confidence limits fall completely inside the equivalence bounds.
Difference confidence limits based on Nam's RMLE method.
Ratio confidence limits based on Blackwelder and Nam's method.

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.

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.

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

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

Parameter Tested	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)	0.0011	-0.0859	0.1274	-0.2000	0.2000	Yes
Ratio (Se1/Se2)	0.0067	0.8833	1.2039	0.8000	1.2500	Yes

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 Nam's RMLE method.
Ratio confidence limits based on Blackwelder and Nam's method.

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