

## Chapter 166

# Equivalence Tests for the Ratio of Two Correlated Proportions

## Introduction

The procedure described in this chapter computes power and sample size for testing equivalence using ratios in designs in which two dichotomous responses are measured on each subject.

When one is interested in showing that two correlated proportions are different, the data are often analyzed with McNemar's test. However, the procedures discussed here are interested in showing equivalence rather than difference. For example, suppose a diagnostic procedure is accurate, but is expensive to apply or has serious side effects. A replacement procedure may be sought which is equally accurate, but is less expensive or has fewer side effects. In this case, we are not interested in showing that the two diagnostic procedures are different, but rather that they are the same. Equivalence tests were designed for this situation.

These tests are often divided into two categories: *equivalence* (two-sided) tests and *non-inferiority* (one-sided) tests. Here, the term *equivalence tests* means that we want to show that two diagnostic procedures are equivalent—that is, their accuracy is about the same. This requires a two-sided hypothesis test. On the other hand, *non-inferiority tests* are used when we want to show that a new (experimental) procedure is no worse than the existing (reference or gold-standard) one. This requires a one-sided hypothesis test.

## Technical Details

The results of a study in which two dichotomous responses are measured on each subject can be displayed in a 2-by-2 table in which one response is shown across the columns and the other is shown down the rows. In the discussion to follow, the columns of the table represent the standard (reference or control) response and the rows represent the treatment (experimental) response. The outcome probabilities can be classified into the following table.

Experimental Diagnosis	Standard Yes	Diagnosis No	Total
Yes	$p_{11}$	$p_{10}$	$P_T$
No	$p_{01}$	$p_{00}$	$1 - P_T$
Total	$P_S$	$1 - P_S$	1

In this table,  $p_{ij} = p_{Treatment, Standard}$ . That is, the first subscript represents the response of the new, experimental procedure while the second subscript represents the response of the standard procedure. Thus,  $p_{01}$  represents the proportion having a negative treatment response and a positive standard response.

## Sensitivity, Specificity, and Prevalence

To aid in interpretation, analysts have developed a few proportions that summarize the table. Three of the most popular ratios are *sensitivity*, *specificity*, and *prevalence*.

### Sensitivity

Sensitivity is the proportion of subjects with a positive standard response who also have a positive experimental response. In terms of proportions from the 2-by-2 table,

$$\text{Sensitivity} = \frac{p_{11}}{(p_{01} + p_{11})} = \frac{p_{11}}{P_S}$$

### Specificity

Specificity is the proportion of subjects with a negative standard response who also have a negative experimental response. In terms of proportions from the 2-by-2 table,

$$\text{Specificity} = \frac{p_{00}}{(p_{10} + p_{00})}$$

### Prevalence

Prevalence is the overall proportion of individuals with the disease (or feature of interest). In terms of proportions from the 2-by-2 table,

$$\text{Prevalence} = P_S$$

## Table Probabilities

The outcome counts from a sample of  $n$  subjects can be classified into the following table.

Experimental Diagnosis	Standard Diagnosis		Total
	Yes	No	
Yes	$n_{11}$	$n_{10}$	$n_T$
No	$n_{01}$	$n_{00}$	$n - n_T$
Total	$n_S$	$n - n_S$	$n$

Note that  $n_{11} + n_{00}$  is the number of matches (*concordant pairs*) and  $n_{01} + n_{10}$  is the number of *discordant pairs*.

The hypothesis of interest concerns the two marginal probabilities  $P_T$  and  $P_S$ .  $P_S$  represents the accuracy or success of the standard test and  $P_T$  represents the accuracy or success of the new, experimental test. Equivalence is defined in terms of either the difference of these two proportions,  $D = P_T - P_S$ , or the relative risk ratio,  $R = P_T/P_S$ . The choice between  $D$  and  $R$  will usually lead to different sample sizes to achieve the same power.

## Equivalence Tests for the Ratio of Two Correlated Proportions

## Equivalence Hypotheses using Ratios

If we define  $R_{0,L}$  and  $R_{0,U}$  as the lower and upper equivalence ratios, respectively, with  $0 < R_{0,L} < 1$  and  $R_{0,U} = 1/R_{0,L}$ , then the null and alternative hypotheses of equivalence in terms of the ratio are

$$H_0: P_T/P_S \leq R_{0,L} \text{ or } P_T/P_S \geq R_{0,U} \quad \text{versus} \quad H_1: R_{0,L} < P_T/P_S < R_{0,U},$$

These hypotheses can be decomposed into two sets of one-sided hypotheses

$$H_{0L}: P_T/P_S \leq R_{0,L} \quad \text{versus} \quad H_{1L}: P_T/P_S > R_{0,L}$$

and

$$H_{0U}: P_T/P_S \geq R_{0,U} \quad \text{versus} \quad H_{1U}: P_T/P_S < R_{0,U}.$$

Note that the first set of one-sided hypotheses,  $H_{0L}$  versus  $H_{1L}$ , is referred to as the hypotheses of non-inferiority.

The following is based on Nam and Blackwelder (2002). We refer you to this paper for the complete details of which we will only provide a brief summary here.

## Test Statistics

The test statistic for an asymptotic test of  $H_{0L}$  versus  $H_{1L}$  based on constrained maximum likelihood for large  $n$  is given by

$$Z(R_{0,L}) = \frac{n(\hat{P}_T - R_{0,L}\hat{P}_S)}{\sqrt{R_{0,L}(\tilde{p}_{10} + \tilde{p}_{01})}}$$

where

$$\tilde{p}_{10} = \frac{-\hat{P}_T + R_{0,L}^2(\hat{P}_S + 2\hat{p}_{10}) + \sqrt{(\hat{P}_T - R_{0,L}^2\hat{P}_S)^2 + 4R_{0,L}^2\hat{p}_{10}\hat{p}_{01}}}{2R_{0,L}(R_{0,L} + 1)}$$

$$\tilde{p}_{10} = R_{0,L}\tilde{p}_{10} - (R_{0,L} - 1)(1 - \hat{p}_{00})$$

$$\hat{p}_{01} = \frac{n_{01}}{n}, \hat{p}_{10} = \frac{n_{10}}{n}, \hat{P}_T = \frac{n_{10} + n_{11}}{n}, \hat{P}_S = \frac{n_{01} + n_{11}}{n}$$

Similarly, an asymptotic test for testing  $H_{0U}$  versus  $H_{1U}$  is given by

$$Z(R_{0,U}) = \frac{n(\hat{P}_T - R_{0,U}\hat{P}_S)}{\sqrt{R_{0,U}(\tilde{p}_{10} + \tilde{p}_{01})}}$$

Equivalence is concluded if both the tests on  $Z_L$  and  $Z_U$  are rejected.

## Power Formula

The power of the one-sided procedure when the true value of the relative risk ratio is  $R_E$  can be evaluated exactly using the multinomial distribution. When  $n$  is large, we use a normal approximation to the multinomial distribution which leads to

$$\text{Power} = \Phi(c_U) - \Phi(c_L)$$

where

$$c_X = \frac{z_{1-\alpha}\sqrt{V_0(R_{0,X})} - E_1(R_{0,X})}{\sqrt{V_1(R_{0,X})}}$$

## Equivalence Tests for the Ratio of Two Correlated Proportions

$$\bar{V}_0(R_{0.X}) = \frac{R_{0.X}(\bar{p}_{10} + \bar{p}_{01})}{n}$$

$$E_1(R_{0.X}) = P_S(R_1 - R_{0.X})$$

$$V_1(R_0) = \frac{P_S(R_1 + R_{0.X}^2) - 2R_{0.X}p_{11} - P_S^2(R_1 - R_{0.X})^2}{n}$$

$$\bar{p}_{10} = \frac{-P_T + R_{0.X}^2(P_S + 2p_{10}) + \sqrt{(P_T - R_{0.X}^2P_S)^2 + 4R_{0.X}^2p_{10}p_{01}}}{2R_{0.X}(R_{0.X} + 1)}$$

$$\bar{p}_{01} = R_{0.X}\bar{p}_{10} - (R_{0.X} - 1)(1 - p_{00})$$

## Nuisance Parameter

The 2-by-2 table includes four parameters,  $p_{11}$ ,  $p_{10}$ ,  $p_{01}$ , and  $p_{00}$ , but the power calculations only require two parameters:  $P_S$  and  $R_1$ . A third parameter is defined implicitly since the sum of the four parameters is one. Thus, one parameter (known as a nuisance parameter) remains unaccounted for. This parameter must be addressed to fully specify the problem. This fourth parameter can be specified using any one of the following:  $p_{11}$ ,  $p_{10}$ ,  $p_{01}$ ,  $p_{00}$ ,  $p_{10} + p_{01}$ ,  $p_{11} + p_{00}$ , or the sensitivity of the experimental response,  $p_{11}/P_S$ .

It may be difficult to specify a reasonable value for the nuisance parameter since its value may not be even approximately known until after the study is conducted. Because of this, we suggest that you calculate power or sample size for a range of values of the nuisance parameter. This will allow you to determine how sensitive the results are to its value.

## Procedure Options

This section describes the options that are specific to this procedure. These are located on the Design tab. For more information about the options of other tabs, go to the Procedure Window chapter.

## Design Tab

The Design tab contains the parameters associated with this test such as the proportions, sample sizes, alpha, and power.

### Solve For

#### Solve For

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *Power* or *Sample Size*.

## Equivalence Tests for the Ratio of Two Correlated Proportions

---

### Power Calculation

#### Power Calculation Method

Select the method to be used to calculate power.

The choices are

- **Multinomial Enumeration**

Power is computed using multinomial enumeration of all possible outcomes when  $N \leq \text{Max } N$  for Multinomial Enumeration (otherwise, the normal approximation is used). Multinomial enumeration of all outcomes is possible because of the discrete nature of the data.

- **Normal Approximation**

Approximate power is computed using the normal approximation to the multinomial distribution.

The exact calculation using the multinomial distribution becomes very time consuming for  $N > 500$ . When  $N > 500$ , the difference between the multinomial and approximate calculations is small.

For small values of  $N$  (less than 100), the Multinomial Enumeration power may be overly optimistic because the discrete nature of the trinomial distribution results in the actual alpha value being higher than its target. To be on the safe side, we recommend that you use the approximate calculation.

#### Max N for Multinomial Enumeration

*Only shown when Power Calculation Method = "Multinomial Enumeration"*

Specify the maximum value of  $N$  (sample size) that uses the exact power calculation based on the multinomial distribution.  $N$ 's greater than this value will use the asymptotic approximation.

---

### Power and Alpha

#### Power

This option specifies one or more values for power. Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when you fail to conclude equivalence when in fact it is true.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as *0.8 to 0.95 by 0.05* may be entered.

#### Alpha

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. Here, a type-I error occurs when you falsely conclude equivalence.

---

### Sample Size

#### N (Sample Size)

Enter a value for the sample size. This value must be greater than two. You may enter a range of values such as *10 to 100 by 10*.

## Equivalence Tests for the Ratio of Two Correlated Proportions

### Effect Size – Ratios

#### R0.L (Lower Equivalence Ratio)

Enter R0.L, the lower equivalence bound for the ratio, Pt/Ps. The upper equivalence bound is calculated as

$$R0.U = 1/R0.L$$

You can enter a single value such as *0.8* or a series of values such as *0.8 0.85 0.9* or *0.8 to 0.9 by 0.05* in the range  $0 < R0.L < 1$  and  $R0.L < R1 < R0.U$ .

#### R1 (Actual Ratio)

Specify the actual ratio, Pt/Ps, at which power will be calculated. This value is used to calculate Pt using the formula

$$Pt = Ps \times D1.$$

Often this value is set equal to 1, but this is not required. Enter values in the range  $R0.L < R1 < R0.U$ .

Be careful that the value you enter here along with other values in this procedure do not result in an undefined 2×2 table. The values entered in this procedure must result in a 2×2 table whose cell values and marginal totals each sum to 1.

### Effect Size – Standard Proportion

#### Ps (Standard Proportion)

The proportion of Yes's (positive responses or successes) when subjects receive the standard procedure.

You may enter a set of values separated by blanks. For example, you could enter *0.50 0.60 0.70*. Enter values in the range  $0 < Ps < 1$ .

Be careful that the value you enter here along with other values in this procedure do not result in an undefined 2×2 table. The values entered in this procedure must result in a 2×2 table whose cell values and marginal totals each sum to 1.

### Effect Size – Nuisance Parameter

#### Nuisance Parameter Type

Enter the type of nuisance parameter here. Unfortunately, the 2-by-2 table cannot be completely specified by using only the parameters Ps and Da or Ps and Ra. One other parameter must be specified. This additional parameter is called a “nuisance” parameter. It will be assumed to be a known quantity. Several possible choices are available. This option lets you specify which parameter you want to use. In all cases, the value you specify is a proportion.

- **P11 (% Positive Matches)**  
The proportion of subjects that are positive on both tests.
- **P00 (% Negative Matches)**  
The proportion of subjects that are negative on both tests.
- **P01 (% -Trt +Std)**  
The proportion of subjects that are negative on the treatment, but positive on the standard.
- **P10 (% +Trt -Std)**  
The proportion of subjects that are positive on the treatment, but negative on the standard.

## Equivalence Tests for the Ratio of Two Correlated Proportions

- **P11+P00 (% Matches)**  
The proportion of matches (concordant pairs).
- **P01+P10 (% Disagree)**  
The proportion of non-matches (discordant pairs).
- **P11/Ps (Sensitivity)**  
The sensitivity.

### Nuisance Parameter Value

Enter the value of the nuisance parameter that you specified in the “Nuisance Parameter Type” box. This value is a proportion, so it must be between 0 and 1.

## Equivalence Tests for the Ratio of Two Correlated Proportions

### Example 1 – Finding Power

A clinical trial will be conducted to show that a non-invasive MRI test is equivalent to the invasive CTAP reference test. Historical data suggest that the CTAP test is 80% accurate. After careful discussion, the researchers decide that if the ratio of the MRI test accuracy to the CTAP accuracy is at least 95% it will be considered equivalent. They decide to use a ratio test statistic. Thus, the equivalence ratio is 0.95. They want to study the power for various sample sizes between 200 and 1000 at the 5% significance level.

They use P01 as the nuisance parameter and look at two values: 0.05 and 0.10.

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for the Ratio of Two Correlated Proportions** procedure window by expanding **Proportions**, then **Two Correlated Proportions**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for the Ratio of Two Correlated Proportions**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
<b>Design Tab</b>	
Solve For .....	<b>Power</b>
Power Calculation Method.....	<b>Normal Approximation</b>
Alpha.....	<b>0.05</b>
N (Sample Size).....	<b>200 300 450 600 800 1000</b>
R0.L (Lower Equivalence Ratio).....	<b>0.95</b>
R1 (Actual Ratio) .....	<b>1.0</b>
Ps (Standard Proportion).....	<b>0.80</b>
Nuisance Parameter Type.....	<b>P01 (% -Trt +Std)</b>
Nuisance Parameter Value.....	<b>0.05 0.10</b>

### Annotated Output

Click the Calculate button to perform the calculations and generate the following output.

### Numeric Results

<b>Numeric Results</b>								
Hypotheses: $H_0: P_t/P_s \leq R_{0.L} \text{ or } P_t/P_s \geq R_{0.U}$ vs. $H_1: R_{0.L} < P_t/P_s < R_{0.U}$								
Power	Sample Size N	Lower Equiv. Ratio R0.L	Upper Equiv. Ratio R0.U	Actual Ratio R1	Treatment Proportion Pt	Standard Proportion Ps	Nuisance Parameter P01	Alpha
0.06511	200	0.950	1.053	1.000	0.800	0.800	0.050	0.050
0.00000	200	0.950	1.053	1.000	0.800	0.800	0.100	0.050
0.37821	300	0.950	1.053	1.000	0.800	0.800	0.050	0.050
0.00000	300	0.950	1.053	1.000	0.800	0.800	0.100	0.050
0.68145	450	0.950	1.053	1.000	0.800	0.800	0.050	0.050
0.21499	450	0.950	1.053	1.000	0.800	0.800	0.100	0.050
(report continues)								
* Power was computed using the normal approximation method.								



## Equivalence Tests for the Ratio of Two Correlated Proportions

### Report Definitions

Power is the probability of rejecting the null hypothesis when it is false.

N is the number of subjects, the sample size.

R0.L and R0.U are the lower and upper ratio bounds, respectively, used to construct the equivalence hypotheses.

$R1 = Pt/Ps$  is the actual ratio at which the power is calculated.

Pt is the response proportion in the treatment (experimental) group.

Ps is the response proportion in the standard (baseline, reference, or control) group.

The Nuisance Parameter is a value that is needed, but is not a direct part of the hypotheses.

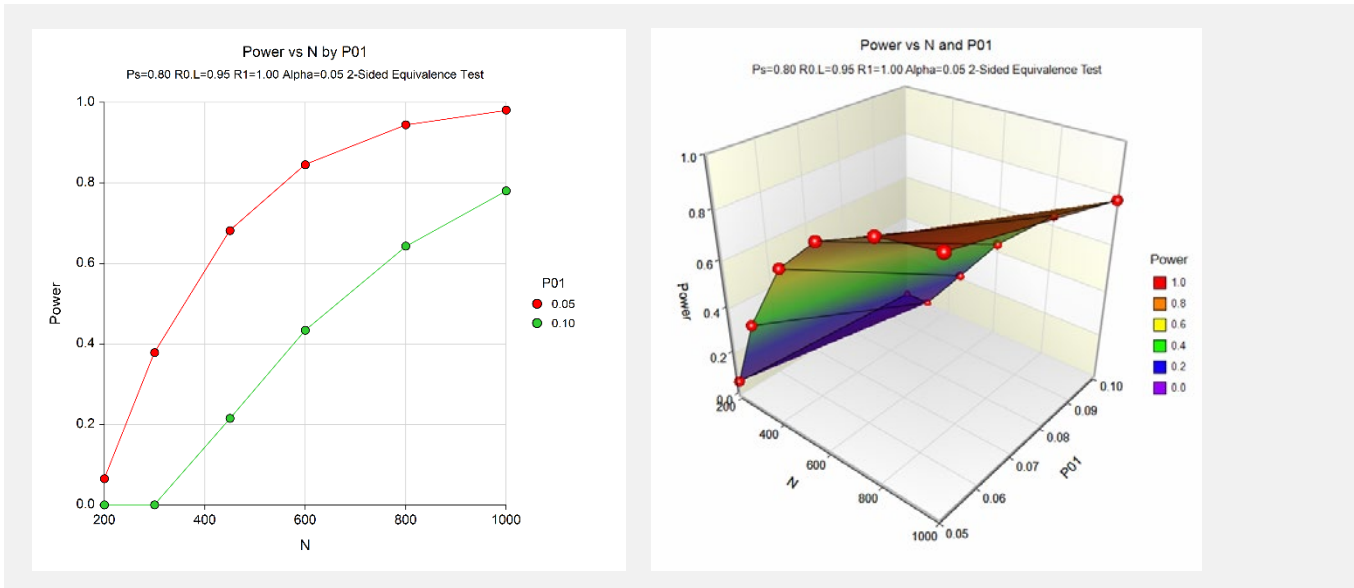
Alpha is the probability of rejecting the null hypothesis when it is true.

### Summary Statements

For an equivalence test of the ratio of two correlated proportions, a sample size of 200 subjects achieves 6.511% power at a significance level of 0.050 when the lower equivalence ratio is 0.950, the upper equivalence ratio is 1.053, the treatment proportion is 0.800, the standard proportion is 0.800, and the actual ratio of the proportions is 1.000. These results assume that P01 is equal to 0.050.

This report shows the power for the indicated scenarios. All of the columns are defined in the “Report Definitions” section.

### Plots Section



These plots show the power versus the sample size for the two values of P01. In this example, we see that the value of the nuisance parameter has a large effect on the calculated sample size.

## Equivalence Tests for the Ratio of Two Correlated Proportions

## Example 2 – Finding Sample Size

Continuing with Example 1, the analysts want to determine the exact sample size necessary to achieve 90% power for both values of the nuisance parameter.

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for the Ratio of Two Correlated Proportions** procedure window by expanding **Proportions**, then **Two Correlated Proportions**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for the Ratio of Two Correlated Proportions**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
<b>Design Tab</b>	
Solve For .....	<b>Sample Size</b>
Power Calculation Method.....	<b>Normal Approximation</b>
Power.....	<b>0.90</b>
Alpha.....	<b>0.05</b>
R0.L (Lower Equivalence Ratio).....	<b>0.95</b>
R1 (Actual Ratio) .....	<b>1.0</b>
Ps (Standard Proportion) .....	<b>0.80</b>
Nuisance Parameter Type .....	<b>P01 (% -Trt +Std)</b>
Nuisance Parameter Value.....	<b>0.05 0.10</b>

### Output

Click the Calculate button to perform the calculations and generate the following output.

### Numeric Results

<b>Numeric Results</b>								
Hypotheses: $H_0: Pt/Ps \leq R0.L \text{ or } Pt/Ps \geq R0.U$ vs. $H_1: R0.L < Pt/Ps < R0.U$								
	Sample Size	Lower Equiv. Ratio	Upper Equiv. Ratio	Actual Ratio	Treatment Proportion	Standard Proportion	Nuisance Parameter	Alpha
Power	N	R0.L	R0.U	R1	Pt	Ps	P01	
0.90046	688	0.950	1.053	1.000	0.800	0.800	0.050	0.050
0.90025	1310	0.950	1.053	1.000	0.800	0.800	0.100	0.050

\* Power was computed using the normal approximation method.

This report shows that the sample size required nearly doubles when P01 is changed from 0.05 to 0.10.

### Example 3 – Validation

We could not find a validation example for an equivalence test for the ratio of two correlated proportions. The calculations are basically the same as those for a non-inferiority test of the ratio of two correlated proportions, which has been validated using Nam and Blackwelder (2002). We refer you to Example 3 of Chapter 161, “Non-Inferiority Tests for the Ratio of Two Correlated Proportions,” for a validation example.