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 Diagnosis**		**Total**
	Yes	No	
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 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

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

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

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

and

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

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

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

where

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

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

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

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

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

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

$$c_X = \frac{z_{1-\alpha}\sqrt{\bar{V}_0(R_{0.X})} - E_1(R_{0.X})}{\sqrt{V_1(R_{0.X})}}$$

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

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

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

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

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

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}, $p_{10} + p_{01}$, $p_{11} + p_{00}$, the sensitivity of the experimental response, p_{11}/P_S, or the within-subject correlation, ρ.

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.

Estimating P11, P01, and P10 using Pt, Ps, and ρ

Sometimes, obtaining estimates of *P*11, *P*01, and/or *P*10 is problematic. This problem is solved by using the marginal probabilities and the within-subject correlation coefficient, which may be easier to estimate. As outlined in Zhang, Cao, and Ahn (2017), the relationship between *P*11, *Pt*, *Ps* and the correlation is

$$\rho = \frac{P11 - P_s P_t}{\sqrt{P_s P_t (1 - P_s)(1 - P_t)}}$$

Using this relationship, values of ρ can be entered and transformed to the corresponding value of *P*11 using the equation

$$P11 = \rho\sqrt{P_s P_t (1 - P_s)(1 - P_t)} + P_s P_t$$

The only concern is that values of ρ be used that limit *P*11, *P*01, *P*10, and *P*00 to be between 0 and 1. The lower and upper limits of the correlation are

$$\rho_L = \max\left\{-\sqrt{\frac{P_s P_t}{(1 - P_s)(1 - P_t)}}, -\sqrt{\frac{(1 - P_s)(1 - P_t)}{P_s P_t}}\right\}$$

$$\rho_U = \min\left\{\sqrt{\frac{P_s(1 - P_t)}{P_t(1 - P_s)}}, \sqrt{\frac{P_t(1 - P_s)}{P_s(1 - P_t)}}\right\}$$

*P*11, along with *Pt* and *Ps*, can then be used to calculate *P*01 and *P*10.

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

If the procedure window is not already open, use the PASS Home window to open it. The parameters 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.

Design Tab

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

Output

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

Numeric Reports

Numeric Results

Solve For: Power
Hypotheses: H0: Pt/Ps ≤ R0.L or Pt/Ps ≥ R0.U vs. H1: R0.L < Pt/Ps < R0.U

		Equivalence Ratios			Proportions			
Power*	Sample Size N	Lower R0.L	Upper R0.U	Actual Ratio R1	Treatment Pt	Standard Ps	Nuisance Parameter P01	Alpha
0.06511	200	0.95	1.053	1	0.8	0.8	0.05	0.05
0.00000	200	0.95	1.053	1	0.8	0.8	0.10	0.05
0.37821	300	0.95	1.053	1	0.8	0.8	0.05	0.05
0.00000	300	0.95	1.053	1	0.8	0.8	0.10	0.05
0.68145	450	0.95	1.053	1	0.8	0.8	0.05	0.05
0.21499	450	0.95	1.053	1	0.8	0.8	0.10	0.05
0.84510	600	0.95	1.053	1	0.8	0.8	0.05	0.05
0.43378	600	0.95	1.053	1	0.8	0.8	0.10	0.05
0.94427	800	0.95	1.053	1	0.8	0.8	0.05	0.05
0.64295	800	0.95	1.053	1	0.8	0.8	0.10	0.05
0.98100	1000	0.95	1.053	1	0.8	0.8	0.05	0.05
0.78019	1000	0.95	1.053	1	0.8	0.8	0.10	0.05

* Power was computed using the normal approximation method.

Power	The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
N	The number of subjects, the sample size.
R0.L and R0.U	The lower and upper ratio bounds, respectively, used to construct the equivalence hypotheses.
R1	The actual ratio at which the power is calculated. R1 = Pt/Ps.
Pt	The response proportion in the treatment (experimental) group.
Ps	The response proportion in the standard (baseline, reference, or control) group.
Nuisance Parameter	A value that is needed but is not a direct part of the hypotheses.
Alpha	The probability of rejecting a true null hypothesis.

Summary Statements

A paired design will be used to test whether the treatment proportion (Pt) is equivalent to the standard proportion (Ps), with equivalence ratio bounds of 0.95 and 1.053 (H0: Pt / Ps ≤ 0.95 or Pt / Ps ≥ 1.053 versus H1: 0.95 < Pt / Ps < 1.053). The comparison will be made using a constrained maximum likelihood asymptotic test, with a Type I error rate (α) of 0.05. The nuisance parameter (P01) is assumed to be 0.05. To detect a ratio of the proportions of 1 (Pt = 0.8, Ps = 0.8) with a sample size of 200 pairs, the power is 0.06511.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout-Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	200	250	50
20%	300	375	75
20%	450	563	113
20%	600	750	150
20%	800	1000	200
20%	1000	1250	250

Dropout Rate The percentage of subjects (or items) that are expected to be lost at random during the course of the study and for whom no response data will be collected (i.e., will be treated as "missing"). Abbreviated as DR.
N The evaluable sample size at which power is computed (as entered by the user). If N subjects are evaluated out of the N' subjects that are enrolled in the study, the design will achieve the stated power.
N' The total number of subjects that should be enrolled in the study in order to obtain N evaluable subjects, based on the assumed dropout rate. N' is calculated by inflating N using the formula N' = N / (1 - DR), with N' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. (2018) pages 32-33.)
D The expected number of dropouts. D = N' - N.

Dropout Summary Statements

Anticipating a 20% dropout rate, 250 subjects should be enrolled to obtain a final sample size of 200 subjects.

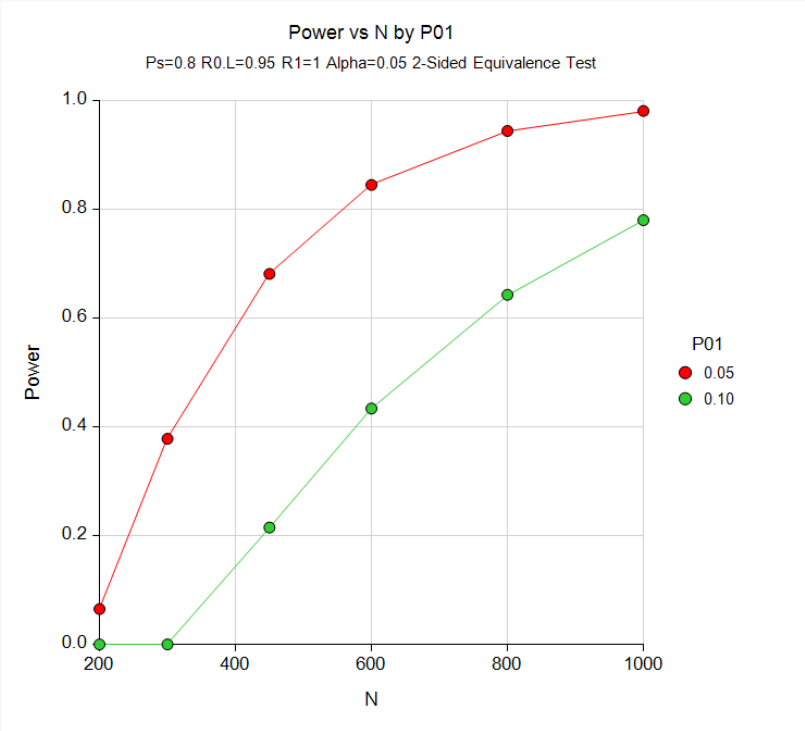
References

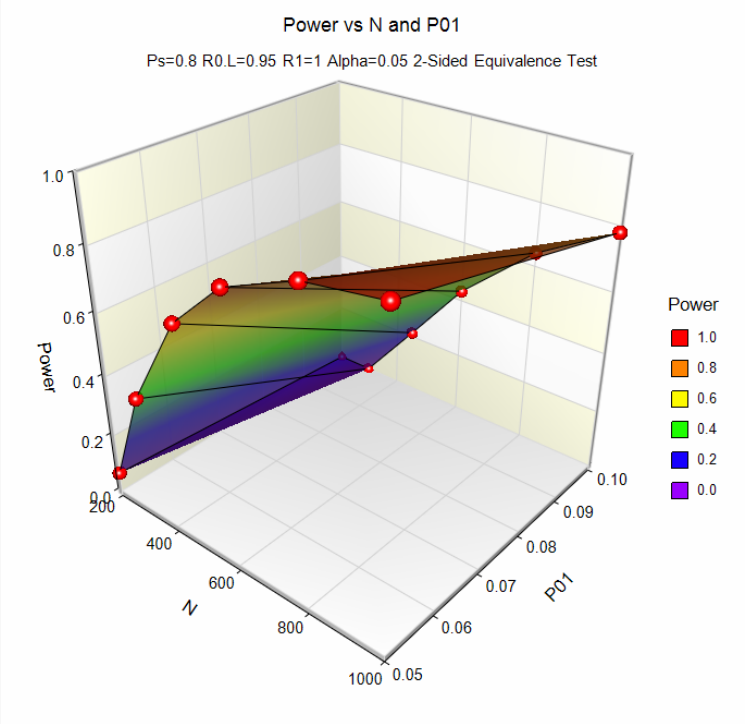
Lewis, J.A. 1999. 'Statistical principles for clinical trials (ICH E9) an introductory note on an international guideline.' Statistics in Medicine, 18, pages 1903-1942.
Liu, J., Hsueh, H., Hsieh, E., and Chen, J.J. 2002. 'Tests for equivalence or non-inferiority for paired binary data', Statistics in Medicine, Volume 21, pages 231-245.
Nam, Jun-mo. 1997. 'Establishing equivalence of two treatments and sample size requirements in matched-pairs design', Biometrics, Volume 53, pages 1422-1430.
Nam, Jun-mo and Blackwelder, W.C. 2002. 'Analysis of the ratio of marginal probabilities in a matched-pair setting', Statistics in Medicine, Volume 21, pages 689-699.
Zhang, S., Cao, J., Ahn, C. 2017. 'Inference and sample size calculation for clinical trials with incomplete observations of paired binary outcomes'. Statistics in Medicine. Volume 36. Pages 581-591.

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

Plots Section

Plots





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.

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

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

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

Output

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

Numeric Results

Solve For: Sample Size
Hypotheses: H0: Pt/Ps ≤ R0.L or Pt/Ps ≥ R0.U vs. H1: R0.L < Pt/Ps < R0.U

		Equivalence Ratios			Proportions			
Power*	Sample Size N	Lower R0.L	Upper R0.U	Actual Ratio R1	Treatment Pt	Standard Ps	Nuisance Parameter P01	Alpha
0.90046	688	0.95	1.053	1	0.8	0.8	0.05	0.05
0.90025	1310	0.95	1.053	1	0.8	0.8	0.10	0.05

* 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.