Chapter 326

Multi-Arm Equivalence Tests for the Odds Ratio of Treatment and Control Proportions

Introduction

This module computes power and sample size for multi-arm, equivalence tests of the odds ratio of treatment and control proportions. This procedure is based on the results in Machin, Campbell, Tan, and Tan (2018). In this design, there are *k* treatment groups and one control group. The groups are independent and are sampled using simple random sampling. A proportion is measured in each group. A total of *k* hypothesis tests are anticipated each comparing a treatment group with the common control group using a simple equivalence test of the odds ratio of two proportions.

The Bonferroni multiplicity adjustment of the type I error rate may be optionally made because several tests are being constructed from the same data. Making a multiplicity adjustment is usually recommended, but not always. In fact, Saville (1990) advocates not applying it and Machin, Campbell, Tan, and Tan (2018) include omitting it as a possibility.

Whether you want to test several doses of a single treatment or several types of treatments, good research practice requires that each treatment be compared with a control. For example, a popular three-arm design consists of three groups: control, treatment A, and treatment B. Two tests are run: treatment A versus control and treatment B versus the same control. This avoids having to obtain a second control group for treatment B. Besides the obvious efficiency in subjects, it may be easier to recruit subjects if their chances of receiving a new treatment are better than 50%.

Odds Ratio

The odds ratio, $OR = \left(p_1/(1-p_1)\right)/\left(p_2/(1-p_2)\right)$, gives the relative change in the odds (*o*) of the response. Let $o_T = P_T/(1-P_T)$. Testing equivalence use the formulation

$$H_0: o_T/o_C \leq OR_L \quad \text{or} \quad o_T/o_C \geq OR_U \quad \text{vs.} \quad H_1: OR_L \leq o_T/o_C \leq OR_U$$

For equivalence tests, $OR_L < 1$ and $\mathrm{OR}_U > 1$. Usually, $\mathrm{OR}_L = 1 \,/\, OR_U$.

The equivalence test is usually carried out using the Two One-Sided Tests (TOST) method. This procedure computes power and sample size for the TOST equivalence test method.

Technical Details

Suppose you have *k* treatment groups with response probabilities P_i of size N_i and one control group with response probability P_C of size N_C. The total sample size is $N = N_1 + N_2 + \ldots + N_k + N_C$.

The *k* equivalence tests hypotheses are

$$H_{0i}: o_i/o_C \leq OR_L \quad \text{or} \quad o_i/o_C \geq OR_U \quad \text{vs.} \quad H_{1i}: OR_L < o_i/o_C < OR_U \quad \text{for } i = 1, 2, \ldots, k$$

where OR_L and OR_U are the equivalence limits (boundaries).

If we define $OR_i = o_i/o_C$, these are equivalent to

$$H_{0i}: OR_i \geq OR_U \quad \text{or} \quad OR_i \leq OR_L \quad \text{vs.} \quad H_{1i}: OR_L < OR_i < OR_U \quad \text{for } i = 1, 2, \ldots, k$$

For convenience, these hypotheses are collectively referred to as

$$H_0: OR \geq OR_U \quad \text{or} \quad OR \leq OR_L \quad \text{vs.} \quad H_1: OR_L < OR < OR_U$$

Test Statistics

Two test statistics are available in this procedure. Symmetric versions of these tests are presented below. These tests are both likelihood score tests.

Miettinen and Nurminen's Likelihood Score Test

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the odds ratio is equal to a specified value, OR_0. Because the approach they used with the difference and ratio does not easily extend to the odds ratio, they used a score statistic approach for the odds ratio. The regular MLE's are $\hat{p}_1$ and $\hat{p}_2$. The constrained MLE's are $\tilde{p}_1$ and $\tilde{p}_2$. These estimates are constrained so that $\widetilde{OR} = OR_0$. A correction factor of *N*/(*N*-1) is applied to make the variance estimate less biased. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing the test statistic is

$$z_{MNO} = \frac{\frac{(\hat{p}_i - \tilde{p}_i)}{\tilde{p}_i\tilde{q}_i} - \frac{(\hat{p}_C - \tilde{p}_C)}{\tilde{p}_C\tilde{q}_C}}{\sqrt{\left(\frac{1}{n_i\tilde{p}_i\tilde{q}_i} + \frac{1}{n_C\tilde{p}_C\tilde{q}_C}\right)\left(\frac{N}{N-1}\right)}}$$

where

$$\tilde{p}_i = \frac{\tilde{p}_C OR_0}{1 + \tilde{p}_C(OR_0 - 1)}$$

$$\tilde{p}_2 = \frac{-B + \sqrt{B^2 - 4AC}}{2A}$$

$$A = n_C(OR_0 - 1),$$

$$B = n_i OR_0 + n_C - m_1(OR_0 - 1),$$

$$C = -m_1$$

$$m_1 = \text{number of successes}$$

Farrington and Manning's Likelihood Score Test

Farrington and Manning (1990) indicate that the Miettinen and Nurminen statistic may be modified by removing the factor *N*/(*N*-1).

The formula for computing this test statistic is

$$z_{FMO} = \frac{\frac{(\hat{p}_i - \tilde{p}_i)}{\tilde{p}_i\tilde{q}_i} - \frac{(\hat{p}_C - \tilde{p}_C)}{\tilde{p}_C\tilde{q}_C}}{\sqrt{\left(\frac{1}{n_i\tilde{p}_i\tilde{q}_i} + \frac{1}{n_C\tilde{p}_C\tilde{q}_C}\right)}}$$

where the estimates $\tilde{p}_i$ and $\tilde{p}_C$ are computed as in the corresponding test of Miettinen and Nurminen (1985) given above.

Asymptotic Approximation to Power

A large sample approximation is used to compute power. The large sample approximation is made by replacing the values of $\hat{p}_i$ and $\hat{p}_C$ in the *z* statistic with the corresponding values of p_i and p_C, and then computing the results based on the normal distribution.

Multiplicity Adjustment

Because *k* z-tests between treatment groups and the control group are run when analyzing the results of this study, many statisticians recommend that the Bonferroni adjustment be applied. This adjustment is easy to apply: the value of alpha that is used in the test is found by dividing the original alpha by the number of tests. For example, if the original alpha is set at 0.05 and the number of treatment (not including the control) groups is five, the individual tests will be conducted using an alpha of 0.01.

The main criticism of this procedure is that if there are many tests, the value of alpha becomes very small. To mitigate against this complaint, some statisticians recommend separating the treatment groups into those that are of primary interest and those that are of secondary interest. The Bonferroni adjustment is made by the using the number of primary treatments rather than the total number of treatments.

There are some who advocate ignoring the adjustment entirely in the case of randomized clinical trials. See for example Saville (1990) and the discussion in chapter 14 of Machin, Campbell, Tan, and Tan (2018).

Size of the Control Group

Because the control group is used over and over, some advocate increasing the number of subjects in this group. The standard adjustment is to include $\sqrt{k}$ subjects in the control group for each subject in one of the treatment groups. See Machin, Campbell, Tan, and Tan (2018, pages 231-232). Note that often, the treatment groups all have the same size.

Example 1 – Finding the Sample Size

A parallel-group, clinical trial is being designed to establish that each of three doses of a test compound is equivalent to the standard therapy using three Farrington-Manning equivalence tests. Suppose the standard therapy has a response rate of 60%. The investigators would like a sample size large enough to find statistical significance at an overall 0.05 level and an individual-test power of at least 0.80. The response rates to be used for group 1 are 60%, 61%, or 62%. The response rate of group 2 is set to 60%. The response rate of group 3 is set to 60%. The equivalence limits on the odds ratio are 0.80 and 1.25.

Following common practice, the control-group sample-size multiplier will be set to $\sqrt{k} = \sqrt{3} = 1.732$ since there are three treatment groups in this design.

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	**Sample Size**
Test Type	**Likelihood Score (Farr. & Mann.)**
Power of Each Test	**0.8**
Overall Alpha	**0.05**
Bonferroni Adjustment	**Standard Bonferroni**
Group Allocation	**Enter Group Allocation Pattern, solve for group sample sizes**
ORU (Upper Equivalence Odds Ratio)	**1.25**
ORL (Lower Equivalence Odds Ratio)	**1/ORU**
Control Proportion	**0.60**
Control Sample Size Allocation	**1.723**
Set A Number of Groups	**1**
Set A Proportion	**0.60 0.61 0.62**
Set A Sample Size Allocation	**1**
Set B Number of Groups	**1**
Set B Proportion	**0.60**
Set B Sample Size Allocation	**1**
Set C Number of Groups	**1**
Set C Proportion	**0.60**
Set C Sample Size Allocation	**1**
Set D Number of Groups	**0**
More	**Unchecked**

Output

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

Numeric Reports

Numeric Results

Solve For:	Sample Size
Group Allocation:	Enter Group Allocation Pattern, solve for group sample sizes
Hypothesis:	H0: OR ≤ ORL or OR ≥ ORU vs. H1: ORL < OR < ORU
Test Type:	Farrington & Manning Likelihood Score Test
Number of Groups:	4
Bonferroni Adjustment:	Standard Bonferroni (Divisor = 3)

	Power		Sample Size		Proportion		Odds Ratio Equivalence Limits			Alpha	
Comparison	Target	Actual	Ni	Allocation	Pi|H0 Pi.0	Pi|H1 Pi.1	Lower ORL	Upper ORU	Actual ORi	Overall	Bonferroni-Adjusted
Control			2643	1.723	0.6	0.60					
vs A	0.8	0.80040	1534	1.000	0.6	0.60	0.8	1.25	1.00000	0.05	0.016667
vs B	0.8	0.80040	1534	1.000	0.6	0.60	0.8	1.25	1.00000	0.05	0.016667
vs C	0.8	0.80040	1534	1.000	0.6	0.60	0.8	1.25	1.00000	0.05	0.016667
Total			7245								
Control			3138	1.723	0.6	0.60					
vs A	0.8	0.80027	1821	1.000	0.6	0.61	0.8	1.25	1.01667	0.05	0.016667
vs B	0.8	0.88784	1821	1.000	0.6	0.60	0.8	1.25	1.00000	0.05	0.016667
vs C	0.8	0.88784	1821	1.000	0.6	0.60	0.8	1.25	1.00000	0.05	0.016667
Total			8601								
Control			5216	1.723	0.6	0.60					
vs A	0.8	0.80002	3027	1.000	0.6	0.62	0.8	1.25	1.03333	0.05	0.016667
vs B	0.8	0.99229	3027	1.000	0.6	0.60	0.8	1.25	1.00000	0.05	0.016667
vs C	0.8	0.99229	3027	1.000	0.6	0.60	0.8	1.25	1.00000	0.05	0.016667
Total			14297								

Comparison	The group that is involved in the comparison between the treatment and control displayed on this report line. The comparison is made using the odds ratio.
Target Power	The power desired. Power is probability of rejecting a false null hypothesis for this comparison. This power is of the comparison shown on this line only.
Actual Power	The power actually achieved.
Ni	Sample Size. The number of subjects in the ith group. The total sample size, N, is shown as the last row of the column.
Allocation	The group sample size allocation pattern. The value on each row represents the relative number of subjects assigned to the group.
Pi.0	The response proportion in the ith group assumed by the null hypothesis, H0. Note that Pi.0 = Pc, where Pc is the control group proportion.
Pi.1	The response proportion in the ith group at which the power is calculated.
ORL	The lower equivalence odds ratio limit. This is the lower equivalence bound of the odds ratio of treatment and control proportions that still results in the conclusion that the treatment group is equivalent to the control group.
ORU	The upper equivalence odds ratio limit. This is the largest equivalence bound of the odds ratio of treatment and control proportions that still results in the conclusion that the treatment group is equivalent to the control group.
ORi	The odds ratio of the ith treatment group proportion (Pi.1) and the control group proportion (Pc) at which the power is calculated.
Overall Alpha	The probability of rejecting at least one of the comparisons in this experiment when each null hypothesis is true.
Bonferroni Alpha	The adjusted significance level at which each individual comparison is made.

Summary Statements

A parallel, 4-group design (with one control group and 3 treatment groups) will be used to test whether the proportion for each treatment group is equivalent to the control group proportion, with equivalence odds ratio bounds of 0.8 and 1.25 (H0: OR ≤ 0.8 or OR ≥ 1.25 versus H1: 0.8 < OR < 1.25, OR = [Pi / (1 - Pi)] / [Pc / (1 - Pc)]). Each of the 3 equivalence comparisons will be made using two one-sided, two-sample, Bonferroni-adjusted Farrington & Manning Likelihood Score tests of the odds ratio. The overall (experiment-wise) Type I error rate (α) is 0.05. The control group proportion is assumed to be 0.6. To detect the treatment proportions 0.6, 0.6, and 0.6 with at least 80% power for each test, the control group sample size needed will be 2643 and the number of needed subjects for the treatment groups will be 1534, 1534, and 1534 (totaling 7245 subjects overall).

Dropout-Inflated Sample Size

Group	Dropout Rate	Sample Size Ni	Dropout-Inflated Enrollment Sample Size Ni'	Expected Number of Dropouts Di
1	20%	2643	3304	661
2	20%	1534	1918	384
3	20%	1534	1918	384
4	20%	1534	1918	384
Total		7245	9058	1813
1	20%	3138	3923	785
2	20%	1821	2277	456
3	20%	1821	2277	456
4	20%	1821	2277	456
Total		8601	10754	2153
1	20%	5216	6520	1304
2	20%	3027	3784	757
3	20%	3027	3784	757
4	20%	3027	3784	757
Total		14297	17872	3575

Group	Lists the group numbers.
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.
Ni	The evaluable sample size for each group at which power is computed (as entered by the user). If Ni subjects are evaluated out of the Ni' subjects that are enrolled in the study, the design will achieve the stated power.
Ni'	The number of subjects that should be enrolled in each group in order to obtain Ni evaluable subjects, based on the assumed dropout rate. Ni' is calculated by inflating Ni using the formula Ni' = Ni / (1 - DR), with Ni' 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.)
Di	The expected number of dropouts in each group. Di = Ni' - Ni.

Dropout Summary Statements

Anticipating a 20% dropout rate, group sizes of 3304, 1918, 1918, and 1918 subjects should be enrolled to obtain final group sample sizes of 2643, 1534, 1534, and 1534 subjects.

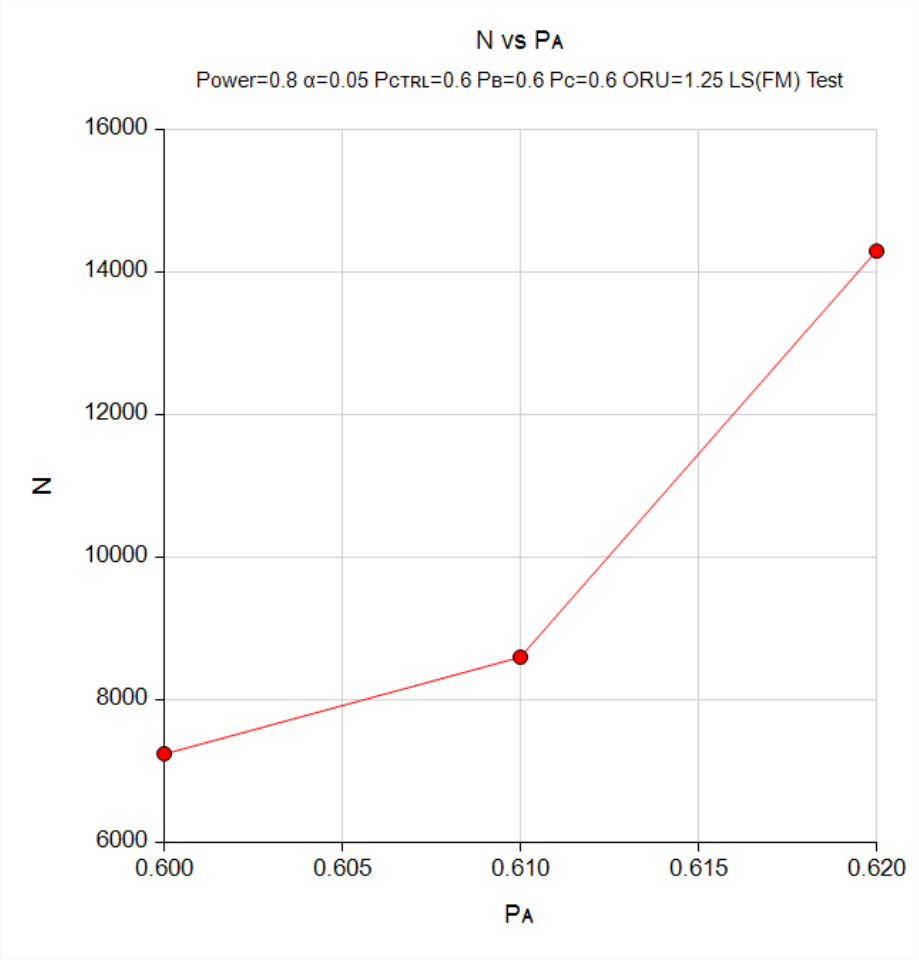
References

Blackwelder, W.C. 1998. 'Equivalence Trials.' In Encyclopedia of Biostatistics, John Wiley and Sons. New York. Volume 2, 1367-1372.
Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. 2018. Sample Size Calculations in Clinical Research, 3rd Edition. Chapman & Hall/CRC. Boca Raton, FL. Pages 86-88.
Farrington, C. P. and Manning, G. 1990. 'Test Statistics and Sample Size Formulae for Comparative Binomial Trials with Null Hypothesis of Non-Zero Risk Difference or Non-Unity Relative Risk.' Statistics in Medicine, Vol. 9, pages 1447-1454.
Fleiss, J. L., Levin, B., Paik, M.C. 2003. Statistical Methods for Rates and Proportions. Third Edition. John Wiley & Sons. New York.
Gart, John J. and Nam, Jun-mo. 1988. 'Approximate Interval Estimation of the Ratio in Binomial Parameters: A Review and Corrections for Skewness.' Biometrics, Volume 44, Issue 2, 323-338.
Gart, John J. and Nam, Jun-mo. 1990. 'Approximate Interval Estimation of the Difference in Binomial Parameters: Correction for Skewness and Extension to Multiple Tables.' Biometrics, Volume 46, Issue 3, 637-643.
Julious, S. A. and Campbell, M. J. 2012. 'Tutorial in biostatistics: sample sizes for parallel group clinical trials with binary data.' Statistics in Medicine, 31:2904-2936.
Lachin, J.M. 2000. Biostatistical Methods. John Wiley & Sons. New York.
Machin, D., Campbell, M.J., Tan, S.B, and Tan, S.H. 2018. Sample Sizes for Clinical, Laboratory, and Epidemiology Studies, 4th Edition. Wiley Blackwell.
Miettinen, O.S. and Nurminen, M. 1985. 'Comparative analysis of two rates.' Statistics in Medicine 4: 213-226.
Tubert-Bitter, P., Manfredi,R., Lellouch, J., Begaud, B. 2000. 'Sample size calculations for risk equivalence testing in pharmacoepidemiology.' Journal of Clinical Epidemiology 53, 1268-1274.

This report shows the numeric results of this power study. Notice that the results are shown in blocks of four rows at a time. Each block represents a single design. Also note that the column headed ‘Actual ORi' shows the odds ratios that result from the choices of the treatment group probabilities.

Plots Section

Plots



This plot gives a visual presentation to the results in the Numeric Report. We can quickly see the impact on the sample size of increasing the treatment response probabilities only slightly from the value of the control probability.

Example 2 – Validation using a Previously Validated Procedure

We could not find a validation result in the statistical literature, so we will use a previously validated **PASS** procedure (**Equivalence Tests for the Odds Ratio of Two Proportions**) to produce the results for the following example.

A parallel-group, clinical trial is being designed to establish that each of three doses of a test compound is equivalent to the standard therapy using three Farrington-Manning equivalence tests. Suppose the standard therapy has a response rate of 60%. The investigators would like a sample size large enough to find statistical significance at an overall 0.05 level and an individual-test power of at least 0.80. The response rates to be used for group 1, 2, and 3 are all set to 60%. The equivalence limits on the odds ratio are 0.50 and 2.0. The sample sizes are set equal.

The **Equivalence Tests for the Odds Ratio of Two Proportions** procedure is set up as follows

Design Tab

Solve For	**Sample Size**
Power Calculation Method	**Normal Approximation**
Test Type	**Likelihood Score (Farr. & Mann.)**
Power	**0.8**
Alpha	**0.016667** (which is Alpha / k)
Group Allocation	**Equal (N1 = N2)**
OR0.U (Upper Equivalence Odds Ratio)	**2**
OR0.L (Lower Equivalence Odds Ratio)	**1/OR0.U**
OR1 (Actual Odds Ratio)	**1.0**
P2 (Group 2 Proportion)	**0.6**

This set of options generates the following report.

Numeric Results

Solve For: Sample Size
Test Statistic: Farrington & Manning Likelihood Score Test
Hypotheses: H0: OR ≤ OR0.L or OR ≥ OR0.U vs. H1: OR0.L < OR < OR0.U

Target Power	Actual Power*	N1	N2	N	Ref. P2	P1.1	P1.0L	P1.0U	OR0.L	OR0.U	OR1	Alpha
0.8	0.80034	197	197	394	0.6	0.6	0.42857	0.75	0.5	2	1	0.01667

* Power was computed using the normal approximation method.

In order to maintain a power of 80% for all three groups, it is apparent that the groups will all need to have a sample size of 197. This table contains the validation values. We will now run these values through the current procedure and compare the results with these values.

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**
Test Type	**Likelihood Score (Farr. & Mann.)**
Power of Each Test	**0.8**
Overall Alpha	**0.05**
Bonferroni Adjustment	**Standard Bonferroni**
Group Allocation	**Equal (Nc = N1 = N2 = ...)**
ORU (Upper Equivalence Odds Ratio)	**2**
ORL (Lower Equivalence Odds Ratio)	**1/ORU**
Control Proportion	**0.60**
Set A Number of Groups	**1**
Set A Proportion	**0.60**
Set B Number of Groups	**1**
Set B Proportion	**0.60**
Set C Number of Groups	**1**
Set C Proportion	**0.60**
Set D Number of Groups	**0**
More	**Unchecked**

Output

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

Numeric Results

Solve For:	Sample Size
Group Allocation:	Equal (Nc = N1 = N2 = ...)
Hypothesis:	H0: OR ≤ ORL or OR ≥ ORU vs. H1: ORL < OR < ORU
Test Type:	Farrington & Manning Likelihood Score Test
Number of Groups:	4
Bonferroni Adjustment:	Standard Bonferroni (Divisor = 3)

						Odds Ratio				
	Power		Sample Size	Proportion		Equivalence Limits			Alpha	
Comparison	Target	Actual	Ni	Pi|H0 Pi.0	Pi|H1 Pi.1	Lower ORL	Upper ORU	Actual ORi	Overall	Bonferroni-Adjusted
Control			197	0.6	0.6					
vs A	0.8	0.80034	197	0.6	0.6	0.5	2	1	0.05	0.016667
vs B	0.8	0.80034	197	0.6	0.6	0.5	2	1	0.05	0.016667
vs C	0.8	0.80034	197	0.6	0.6	0.5	2	1	0.05	0.016667
Total			788							

As you can see, the sample sizes and powers match thus validating this procedure.