Chapter 168

Multi-Arm Superiority by a Margin Tests for the Odds Ratio of Treatment and Control Proportions

Introduction

This module computes power and sample size for multi-arm, superiority by a margin 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 superiority by a margin 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%.

Example

Suppose that the current treatment for a disease works 60% of the time. Unfortunately, this treatment is expensive and occasionally exhibits serious side-effects. Two promising new treatments have been developed and are now ready to be tested. Hence, three groups are needed to complete this study. Two superiority by a margin hypotheses need to be tested in this study: whether each new treatment is better than the current treatment.

Because of the costs of switching to a new treatment, clinicians will only recommend it if it is definitely more effective than the current treatment. They must determine, however, how much more effective the new treatment must be to be adopted. Should it be adopted if it's odds ratio with the control group is 1.05? 1.1? 1.25? There is an odds ratio that is so low that decrease in response is no longer ignorable. In this example, after thoughtful discussion with several clinicians, it was decided that if the response odds ratio is at least 1.25, the new treatment will be adopted. The boundary odds ratio between these two percentages is called the *superiority odds ratio* (OR_0).

The developers must design an experiment to test the hypothesis that the response rate odds ratio is at least 1.25. The statistical hypotheses to be tested are

$$H_0: OR_A \leq OR_0 \quad \text{vs.} \quad H_1: OR_A > OR_0$$

$$H_0: OR_B \leq OR_0 \quad \text{vs.} \quad H_1: OR_B > OR_0$$

where $OR_0 = 1.25$.

Notice that when the null hypothesis is rejected, the conclusion is that the odds ratio is higher than 1.25.

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 $N = N_1 + N_2 + \ldots + N_k + N_C$.

Let $O_i = p_i/(1-p_i)$ represent the *odds* for group *i*. The *k* one-sided superiority tests are

$$H_{0i}: O_i/O_C \leq OR_0 \quad \text{vs.} \quad H_{1i}: O_i/O_C > OR_0 \quad \text{for } i = 1, 2, \ldots, k.$$

Note that if higher proportions are better, $OR_0 > 1$ and if lower proportions are better, $OR_0 < 1$.

If we define $OR_i = O_i/O_C$, these are equivalent to

$$H_{0i}: OR_i \leq OR_0 \quad \text{vs.} \quad H_{1i}: OR_i > OR_0 \quad \text{for } i = 1, 2, \ldots, k$$

For convenience, these hypotheses are collectively referred to as

$$H_0: OR \leq OR_0 \quad \text{vs.} \quad H_1: OR > OR_0$$

Test Statistics

Two test statistics are available in this procedure. 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* 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 compare three doses of a test compound with the current standard therapy using three superiority Miettinen and Nurminen Likelihood Scores tests. Suppose the standard therapy has a response rate of 0.6. 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 0.80. The superiority odds ratio is 1.15.

The response rates of treatment group 1 are set to 0.74, 0.76, 0.78. The response rate of group 2 is 0.8. The response rate of group 3 is 0.85.

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**
Higher Proportions Are	**Better (H1: OR > OR0)**
Test Type	**Likelihood Score (Miet. & Nurm.)**
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**
OR0 (Superiority Odds Ratio)	**1.15**
Control Proportion	**0.6**
Control Sample Size Allocation	**1.732**
Set A Number of Groups	**1**
Set A Proportion	**0.74 0.76 0.78**
Set A Sample Size Allocation	**1**
Set B Number of Groups	**1**
Set B Proportion	**0.8**
Set B Sample Size Allocation	**1**
Set C Number of Groups	**1**
Set C Proportion	**0.85**
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
Test Type:	Miettinen & Nurminen Likelihood Score Test
Higher Proportions Are:	Better
Hypotheses:	H0: OR ≤ OR0 vs. H1: OR > OR0
Number of Groups:	4
Bonferroni Adjustment:	Standard Bonferroni (Divisor = 3)

	Power		Sample Size		Proportion		Odds Ratio		Alpha	
Comparison	Target	Actual	Ni	Allocation	Pi|H0 Pi.0	Pi|H1 Pi.1	Superiority OR0	Actual ORi	Overall	Bonferroni-Adjusted
Control			471	1.732	0.60000	0.60				
vs A	0.8	0.80096	272	1.000	0.63303	0.74	1.15	1.89744	0.05	0.016667
vs B	0.8	0.99245	272	1.000	0.63303	0.80	1.15	2.66667	0.05	0.016667
vs C	0.8	0.99983	272	1.000	0.63303	0.85	1.15	3.77778	0.05	0.016667
Total			1287							
Control			333	1.732	0.60000	0.60				
vs A	0.8	0.80024	192	1.000	0.63303	0.76	1.15	2.11111	0.05	0.016667
vs B	0.8	0.95816	192	1.000	0.63303	0.80	1.15	2.66667	0.05	0.016667
vs C	0.8	0.99676	192	1.000	0.63303	0.85	1.15	3.77778	0.05	0.016667
Total			909							
Control			248	1.732	0.60000	0.60				
vs A	0.8	0.80023	143	1.000	0.63303	0.78	1.15	2.36364	0.05	0.016667
vs B	0.8	0.88964	143	1.000	0.63303	0.80	1.15	2.66667	0.05	0.016667
vs C	0.8	0.98208	143	1.000	0.63303	0.85	1.15	3.77778	0.05	0.016667
Total			677							

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	The number of subjects in the ith group. The total sample size shown below the groups is equal to the sum of all individual group sample sizes.
Allocation	The group sample size allocation ratio of the ith group. 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 = f(Pc, OR0), where Pc is the control group proportion.
Pi.1	The response proportion in the ith group at which the power is calculated.
OR0	The superiority odds ratio boundary determines whether the treatment group is to be considered superior to the control group or not.
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 superior to the control group proportion by a margin, with a superiority odds ratio of 1.15 (H0: OR ≤ 1.15 vs. H1: OR > 1.15, OR = [Pi / (1 - Pi)] / [Pc / (1 - Pc)]). In this study, higher proportions are considered to be better. The superiority-by-a-margin hypotheses will be evaluated using 3 one-sided, two-sample, Bonferroni-adjusted Miettinen & Nurminen Likelihood Score tests of the odds ratio, with an overall (experiment-wise) Type I error rate (α) of 0.05. The control group proportion is assumed to be 0.6. To detect the treatment proportions 0.74, 0.8, and 0.85 with at least 80% power for each test, the control group sample size needed will be 471 and the number of needed subjects for the treatment groups will be 272, 272, and 272 (totaling 1287 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%	471	589	118
2	20%	272	340	68
3	20%	272	340	68
4	20%	272	340	68
Total		1287	1609	322
1	20%	333	417	84
2	20%	192	240	48
3	20%	192	240	48
4	20%	192	240	48
Total		909	1137	228
1	20%	248	310	62
2	20%	143	179	36
3	20%	143	179	36
4	20%	143	179	36
Total		677	847	170

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 589, 340, 340, and 340 subjects should be enrolled to obtain final group sample sizes of 471, 272, 272, and 272 subjects.

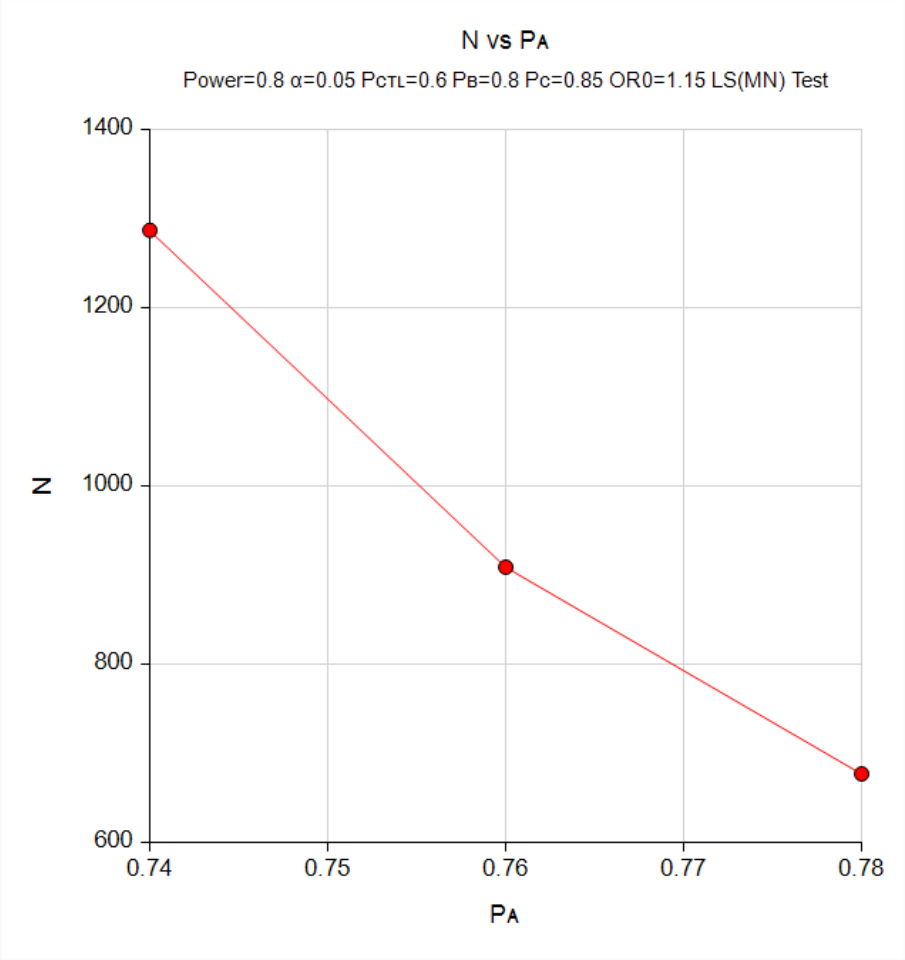
References

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.
D'Agostino, R.B., Chase, W., and Belanger, A. 1988. 'The Appropriateness of Some Common Procedures for Testing the Equality of Two Independent Binomial Populations', The American Statistician, August 1988, Volume 42 Number 3, pages 198-202.
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.

This report shows the numeric results of this power study. Notice that the results are shown in blocks of three rows at a time. Each block represents a single design.

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 changing the expected value of the response rate for treatment 1.

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 (**Superiority by a Margin Tests for the Odds Ratio of Two Proportions**) to produce the results for the following example.

Suppose a parallel-group, clinical trial is being designed to compare two doses of a test compound against the standard therapy using two superiority by a margin Miettinen and Nurminen Likelihood Scores tests. Suppose the standard therapy has a response rate of 0.60. 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 0.80. The response rates of groups 1 and 2 are 0.75 and 0.81, respectively. The superiority odds ratio is 1.15.

The **Superiority by a Margin 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**
Higher Proportions Are	**Better (H1: OR > OR0)**
Test Type	**Likelihood Score (Miet. & Nurm.)**
Power	**0.8**
Alpha	**0.025** (which is Alpha / k)
Group Allocation	**Equal (N1 = N2)**
OR0 (Superiority Odds Ratio)	**1.15**
OR1 (Actual Odds Ratio)	**2 2.84211**
P2 (Group 2 Proportion)	**0.6**

This set of options generates the following report.

Numeric Results

Solve For: Sample Size
Groups: 1 = Treatment, 2 = Reference
Test Statistic: Miettinen & Nurminen Likelihood Score Test
Hypotheses: H0: OR ≤ OR0 vs. H1: OR > OR0

Power		Sample Size			Proportions			Odds Ratio		
Target	Actual*	N1	N2	N	Superiority P1.0	Actual P1.1	Reference P2	Superiority OR0	Actual OR1	Alpha
0.8	0.80067	245	245	490	0.63303	0.75	0.6	1.15	2.00000	0.025
0.8	0.80363	104	104	208	0.63303	0.81	0.6	1.15	2.84211	0.025

* Power was computed using the normal approximation method.

In order to maintain a power of 80% for both groups, it is apparent that all groups will need to have a sample size of 245. We next calculate the powers of the two groups using these sample sizes. The results are displayed in the following table.

Numeric Results

Solve For: Power
Groups: 1 = Treatment, 2 = Reference
Test Statistic: Miettinen & Nurminen Likelihood Score Test
Hypotheses: H0: OR ≤ OR0 vs. H1: OR > OR0

	Sample Size			Proportions			Odds Ratio		
Power*	N1	N2	N	Superiority P1.0	Actual P1.1	Reference P2	Superiority OR0	Actual OR1	Alpha
0.80067	245	245	490	0.63303	0.75	0.6	1.15	2.00000	0.025
0.98964	245	245	490	0.63303	0.81	0.6	1.15	2.84211	0.025

* Power was computed using the normal approximation method.

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**
Higher Proportions Are	**Better (H1: OR > OR0)**
Test Type	**Likelihood Score (Miet. & Nurm.)**
Power of Each Test	**0.8**
Overall Alpha	**0.05**
Bonferroni Adjustment	**Standard Bonferroni**
Group Allocation	**Equal (Nc = N1 = N2 = ...)**
OR0 (Superiority Odds Ratio)	**1.15**
Control Proportion	**0.6**
Set A Number of Groups	**1**
Set A Proportion	**0.75**
Set B Number of Groups	**1**
Set B Proportion	**0.81**
Set C Number of Groups	**0**
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 = ...)
Test Type:	Miettinen & Nurminen Likelihood Score Test
Higher Proportions Are:	Better
Hypotheses:	H0: OR ≤ OR0 vs. H1: OR > OR0
Number of Groups:	3
Bonferroni Adjustment:	Standard Bonferroni (Divisor = 2)

	Power		Sample Size	Proportion		Odds Ratio		Alpha	
Comparison	Target	Actual	Ni	Pi|H0 Pi.0	Pi|H1 Pi.1	Superiority OR0	Actual ORi	Overall	Bonferroni-Adjusted
Control			245	0.60000	0.60				
vs A	0.8	0.80067	245	0.63303	0.75	1.15	2.00000	0.05	0.025
vs B	0.8	0.98964	245	0.63303	0.81	1.15	2.84211	0.05	0.025
Total			735						

The sample sizes and powers match which validates this procedure.