Chapter 163

Multi-Arm Non-Inferiority Tests for the Ratio of Treatment and Control Proportions

Introduction

This module computes power and sample size for multi-arm, noninferiority tests of the 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 noninferiority test of the 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 70% 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 non-inferiority hypotheses need to be tested in this study: whether each new treatment is as good as the current treatment.

Because of the many benefits of the new treatment, clinicians are willing to adopt a new treatment even if it is slightly less effective than the current treatment. They must determine, however, how much less effective the new treatment can be and still be adopted. Should it be adopted if it's ratio with the control group is 0.95? 0.925? 0.9? 0.8? There is a 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 a response ratio of at least 0.80 is achieved, the new treatment will be adopted. The ratio between these two percentages is called the *non-inferiority ratio* (R_0).

The developers must design an experiment to test the hypothesis that the response rate ratio is at least 0.8. The statistical hypotheses to be tested are

$$H_0: P_A/P_C \leq R_0 \quad \text{vs.} \quad H_1: P_A/P_C > R_0$$

$$H_0: P_B/P_C \leq R_0 \quad \text{vs.} \quad H_1: P_B/P_C > R_0$$

where $R_0 = 0.8$.

Notice that when the null hypothesis is rejected, the conclusion is that the ratio is at least 0.8. Note that even though the response rate of the current treatment is 0.70, the hypothesis test is about a response rate ratio of 0.7. This results in a response rate boundary of 0.7(0.8) = 0.56.

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* one-sided non-inferiority tests are

$$H_{0i}: P_i/P_C \leq R_0 \quad \text{vs.} \quad H_{1i}: P_i/P_C > R_0 \quad \text{for } i = 1, 2, \ldots, k$$

Note that if higher proportions are better, $R_0 < 1$ and if lower proportions are better, $R_0 > 1$.

If we define $R_i = P_i/P_C$, these are equivalent to

$$H_{0i}: R_i \leq R_0 \quad \text{vs.} \quad H_{1i}: R_i > R_0 \quad \text{for } i = 1, 2, \ldots, k$$

For convenience, these hypotheses are collectively referred to as

$$H_0: R \leq R_0 \quad \text{vs.} \quad H_1: R > R_0$$

Test Statistics

Three test statistics are available in this procedure.

Miettinen and Nurminen's Likelihood Score Test

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the ratio is equal to a specified value R_0. The regular MLE's, $\hat{p}_i$ and $\hat{p}_C$, are used in the numerator of the score statistic while MLE's $\tilde{p}_i$ and $\tilde{p}_C$, constrained so that $\tilde{p}_i / \tilde{p}_C = R_0$, are used in the denominator. 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

$$z_{MNR} = \frac{\hat{p}_i / \hat{p}_C - R_0}{\sqrt{\left(\frac{\tilde{p}_i \tilde{q}_i}{N_i} + R_0^2 \frac{\tilde{p}_C \tilde{q}_C}{N_C}\right)\left(\frac{N}{N-1}\right)}}$$

where

$$\tilde{p}_i = \tilde{p}_C R_0$$

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

$$A = N R_0$$

$$B = -[N_i R_0 + x_{11} + N_C + x_{21} R_0]$$

$$C = m_1$$

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

Farrington and Manning's Likelihood Score Test

Farrington and Manning (1990) proposed a test statistic for testing whether the ratio is equal to a specified value R_0. The regular MLE's, $\hat{p}_i$ and $\hat{p}_C$, are used in the numerator of the score statistic while MLE's $\tilde{p}_i$ and $\tilde{p}_C$, constrained so that $\tilde{p}_i / \tilde{p}_C = R_0$, are used in the denominator. 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

$$z_{FMR} = \frac{\hat{p}_i / \hat{p}_C - R_0}{\sqrt{\left(\frac{\tilde{p}_i \tilde{q}_i}{N_i} + R_0^2 \frac{\tilde{p}_C \tilde{q}_C}{N_C}\right)}}$$

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

Gart and Nam's Likelihood Score Test

Gart and Nam (1988), page 329, proposed a modification to the Farrington and Manning (1988) ratio test that corrects for skewness. Let $z_{FMR}(R)$ stand for the Farrington and Manning ratio test statistic described above. The skewness corrected test statistic, z_{GNR}, is the appropriate solution to the quadratic equation

$$(-\tilde{\varphi})z_{GNR}^2 + (-1)z_{GNR} + (z_{FMR}(R) + \tilde{\varphi}) = 0$$

where

$$\tilde{\varphi} = \frac{1}{6\tilde{u}^{3/2}}\left(\frac{\tilde{q}_i(\tilde{q}_i - \tilde{p}_i)}{N_i^2 \tilde{p}_i^2} - \frac{\tilde{q}_C(\tilde{q}_C - \tilde{p}_C)}{N_C^2 \tilde{p}_C^2}\right)$$

$$\tilde{u} = \frac{\tilde{q}_i}{N_i \tilde{p}_i} + \frac{\tilde{q}_C}{N_C \tilde{p}_C}$$

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. Note that in large samples, the Farrington and Manning statistic is substituted for the Gart and Nam statistic.

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 compare three doses of a test compound against the standard therapy using three non-inferiority 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 response rates of treatment group 1 are 0.62, 0.65, 0.68. The response rate of group 2 is 0.7. The response rate of group 3 is 0.75. The non-inferiority ratio is 0.8.

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: R > R0)**
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**
R0 (Non-Inferiority Ratio)...............................**0.8**
Control Proportion..**0.6**
Control Sample Size Allocation......................**1.73**
Set A Number of Groups................................**1**
Set A Proportion **0.62 0.65 0.68**
Set A Sample Size Allocation **1**
Set B Number of Groups................................**1**
Set B Proportion **0.7**
Set B Sample Size Allocation **1**
Set C Number of Groups **1**
Set C Proportion **0.75**
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: R ≤ R0 vs. H1: R > R0
Number of Groups:	4
Bonferroni Adjustment:	Standard Bonferroni (Divisor = 3)

	Power		Sample Size		Proportion		Ratio		Alpha	
							Non-Inferiority	Actual		
Comparison	Target	Actual	Ni	Allocation	Pi|H0 Pi.0	Pi|H1 Pi.1	R0	Ri	Overall	Bonferroni-Adjusted
Control			260	1.73	0.60	0.60				
vs A	0.8	0.80165	150	1.00	0.48	0.62	0.8	1.03333	0.05	0.016667
vs B	0.8	0.99641	150	1.00	0.48	0.70	0.8	1.16667	0.05	0.016667
vs C	0.8	0.99996	150	1.00	0.48	0.75	0.8	1.25000	0.05	0.016667
Total			710							
Control			175	1.73	0.60	0.60				
vs A	0.8	0.80236	101	1.00	0.48	0.65	0.8	1.08333	0.05	0.016667
vs B	0.8	0.96425	101	1.00	0.48	0.70	0.8	1.16667	0.05	0.016667
vs C	0.8	0.99775	101	1.00	0.48	0.75	0.8	1.25000	0.05	0.016667
Total			478							
Control			125	1.73	0.60	0.60				
vs A	0.8	0.80148	72	1.00	0.48	0.68	0.8	1.13333	0.05	0.016667
vs B	0.8	0.87934	72	1.00	0.48	0.70	0.8	1.16667	0.05	0.016667
vs C	0.8	0.97912	72	1.00	0.48	0.75	0.8	1.25000	0.05	0.016667
Total			341							

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 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 $Pi.0 = Pc \times R0$, where Pc is the control group proportion.
Pi.1	The response proportion in the ith group at which the power is calculated.
R0	The non-inferiority ratio is the ratio boundary between concluding that a treatment is non-inferior or inferior.
Ri	The ratio of the ith group proportion (Pi.1) and the control group proportion (Pc) at which the power is calculated. The formula is $Ri = Pi.1 / Pc$.
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 non-inferior to the control group proportion, with a non-inferiority ratio of 0.8 (H0: R ≤ 0.8 vs. H1: R > 0.8, R = Pi / Pc). In this study, higher proportions are considered to be better. The non-inferiority hypotheses will be evaluated using 3 one-sided, two-sample, Bonferroni-adjusted Miettinen & Nurminen Likelihood Score tests of the 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 0.62, 0.7, and 0.75 with at least 80% power for each test, the control group sample size needed will be 260 and the number of needed subjects for the treatment groups will be 150, 150, and 150 (totaling 710 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%	260	325	65
2	20%	150	188	38
3	20%	150	188	38
4	20%	150	188	38
Total		710	889	179
1	20%	175	219	44
2	20%	101	127	26
3	20%	101	127	26
4	20%	101	127	26
Total		478	600	122
1	20%	125	157	32
2	20%	72	90	18
3	20%	72	90	18
4	20%	72	90	18
Total		341	427	86

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 325, 188, 188, and 188 subjects should be enrolled to obtain final group sample sizes of 260, 150, 150, and 150 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
_____
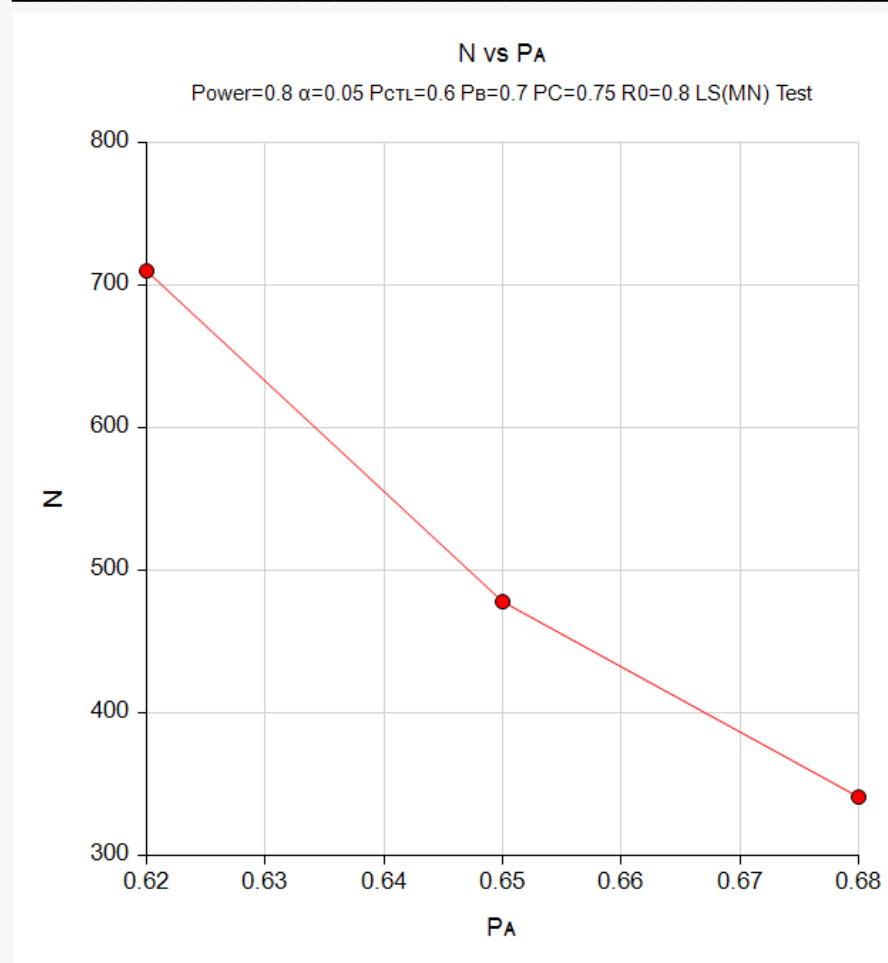
This plot gives a visual presentation of the results in the Numeric Report. We can quickly see the impact on the sample size of decreasing the difference between the treatment and control proportions.

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 (**Non-Inferiority Tests for the 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 non-inferiority 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 0.65 and 0.7, respectively. The non-inferiority ratio is 0.8.

In this example, the group sample sizes will be kept equal.

The **Non-Inferiority Tests for the 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: P1/P2 > R0)**

Test Type...**Likelihood Score (Miet. & Nurm.)**

Power...**0.8**

Alpha...**0.025** (which is Alpha / k)

Group Allocation ....**Equal (N1 = N2)**

R0 (Non-Inferiority Ratio) **0.8**

R1 (Actual Ratio) **1.083333 1.16667**

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: P1 / P2 ≤ R0   vs.   H1: P1 / P2 > R0
_____

Power		Sample Size			Proportions			Ratio		
Target	Actual*	N1	N2	N	Non-Inferiority P1.0	Actual P1.1	Reference P2	Non-Inferiority R0	Actual R1	Alpha
0.8	0.80290	106	106	212	0.48	0.65	0.6	0.8	1.08333	0.025
0.8	0.80412	62	62	124	0.48	0.70	0.6	0.8	1.16667	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 the groups will all need to have a sample size of 106. 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: P1 / P2 ≤ R0   vs.   H1: P1 / P2 > R0
_____

				Proportions			Ratio		
	Sample Size			Non-Inferiority	Actual	Reference	Non-Inferiority	Actual	
Power*	N1	N2	N	P1.0	P1.1	P2	R0	R1	Alpha
0.80290	106	106	212	0.48	0.65	0.6	0.8	1.08333	0.025
0.95936	106	106	212	0.48	0.70	0.6	0.8	1.16667	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: R > R0)**

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 = ...)**

R0 (Non-Inferiority Ratio)..............................**0.8**

Control Proportion..**0.6**

Set A Number of Groups................................**1**

Set A Proportion **0.65**

Set B Number of Groups................................**1**

Set B Proportion **0.7**

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: R ≤ R0   vs.   H1: R > R0
Number of Groups: 3
Bonferroni Adjustment: Standard Bonferroni (Divisor = 2)
_____

	Power		Sample Size	Proportion		Ratio Non-Inferiority	Ratio Actual	Alpha	
Comparison	Target	Actual	Ni	Pi|H0 Pi.0	Pi|H1 Pi.1	R0	Ri	Overall	Bonferroni-Adjusted
Control			106	0.60	0.60				
vs A	0.8	0.80291	106	0.48	0.65	0.8	1.08333	0.05	0.025
vs B	0.8	0.95936	106	0.48	0.70	0.8	1.16667	0.05	0.025
Total			318						

The sample sizes and powers match which validates this procedure.