Chapter 325

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

Introduction

This module computes power and sample size for multi-arm, equivalence 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 simple equivalence 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

An equivalence test example will set the stage for the discussion of the terminology that follows. Suppose that the response rate of the standard treatment of a disease is 0.70. Unfortunately, this treatment is expensive and occasionally exhibits serious side-effects. A promising new treatment has been developed to the point where it can be tested. One of the first questions that must be answered is whether the new treatment is therapeutically equivalent to the standard treatment.

Because of the many benefits of the new treatment, clinicians are willing to adopt the new treatment even if its effectiveness is slightly different from the standard. After thoughtful discussion with several clinicians, it is decided that if the response rate ratio of the new treatment to the standard treatment is between 0.9 and 1.1, the new treatment would be adopted.

The developers must design an experiment to test the hypothesis that the response rate ratio of the new treatment to the standard is between 0.9 and 1.1. The statistical hypothesis to be tested is

$$H_0: p_T/p_C < 0.9 \quad \text{or} \quad p_T/p_C > 1.1 \quad \text{vs.} \quad H_1: 0.9 \le p_T/p_C \le 1.1$$

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

$$H_{0i}: P_i/P_C \geq R_U \quad \text{or} \quad P_i/P_C \leq R_L \quad \text{vs.} \quad H_{1i}: R_L < P_i/P_C < R_U \quad \text{for } i = 1, 2, \ldots, k$$

where R_L and R_U are the equivalence limits (boundaries). Note that usually $R_L = 1/R_U$.

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

$$H_{0i}: R_i \geq R_U \quad \text{or} \quad R_i \leq R_L \quad \text{vs.} \quad H_{1i}: R_L < R_i < R_U \quad \text{for } i = 1, 2, \ldots, k$$

For convenience, these hypotheses are collectively referred to as

$$H_0: R \geq R_U \quad \text{or} \quad R \leq R_L \quad \text{vs.} \quad H_1: R_L < R < R_U$$

Test Statistics

Three test statistics are available in this routine. Symmetric versions of these tests are presented below.

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, ϕ_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 $\tilde{p}_i \,/\, \tilde{p}_C = \phi_0$, 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

$$z_{MNR} = \frac{\hat{p}_i \,/\, \hat{p}_C - \phi_0}{\sqrt{\left(\frac{\tilde{p}_i \tilde{q}_i}{N_i} + \phi_0^2 \frac{\tilde{p}_C \tilde{q}_C}{N_C}\right)\left(\frac{N}{N-1}\right)}}$$

where

$$\tilde{p}_i = \tilde{p}_C \phi_0$$

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

$$A = N\phi_0$$

$$B = -[N_i\phi_0 + x_{11} + N_C + x_{21}\phi_0]$$

$$C = m_1$$

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

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, ϕ_0. The regular MLE's, $\hat{p}_T$ 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 $\tilde{p}_T / \tilde{p}_C = \phi_0$, 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

$$z_{FMR} = \frac{\hat{p}_i / \hat{p}_C - \phi_0}{\sqrt{\left(\frac{\tilde{p}_i \tilde{q}_i}{N_i} + \phi_0^2 \frac{\tilde{p}_C \tilde{q}_C}{N_C}\right)}}$$

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 $z_{FMR}(\phi)$ 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

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

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 establish that each of three doses of a test compound are equivalent to the standard therapy using three Gart-Nam 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 of group 1 are 60%, 62%, or 64%. 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 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 (Gart & Nam)**
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**
RU (Upper Equivalence Ratio)	**1.25**
RL (Lower Equivalence Ratio)	**1/RU**
Control Proportion	**0.6**
Control Sample Size Allocation	**1.723**
Set A Number of Groups	**1**
Set A Proportion	**0.6 0.62 0.64**
Set A Sample Size Allocation	**1**
Set B Number of Groups	**1**
Set B Proportion	**0.6**
Set B Sample Size Allocation	**1**
Set C Number of Groups	**1**
Set C Proportion	**0.6**
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: R ≤ RL or R ≥ RU vs. H1: RL < R < RU
Test Type:	Gart & Nam Likelihood Score Test
Number of Groups:	4
Bonferroni Adjustment:	Standard Bonferroni (Divisor = 3)

	Power		Sample Size		Proportion		Ratio			Alpha	
							Equivalence				
Comparison	Target	Actual	Ni	Allocation	Pi|H0 Pi.0	Pi|H1 Pi.1	Lower RL	Upper RU	Actual Ri	Overall	Bonferroni-Adjusted
Control			434	1.723	0.6	0.60					
vs A	0.8	0.80159	252	1.000	0.6	0.60	0.8	1.25	1.00000	0.05	0.016667
vs B	0.8	0.80159	252	1.000	0.6	0.60	0.8	1.25	1.00000	0.05	0.016667
vs C	0.8	0.80159	252	1.000	0.6	0.60	0.8	1.25	1.00000	0.05	0.016667
Total			1190								
Control			441	1.723	0.6	0.60					
vs A	0.8	0.80063	256	1.000	0.6	0.62	0.8	1.25	1.03333	0.05	0.016667
vs B	0.8	0.81080	256	1.000	0.6	0.60	0.8	1.25	1.00000	0.05	0.016667
vs C	0.8	0.81080	256	1.000	0.6	0.60	0.8	1.25	1.00000	0.05	0.016667
Total			1209								
Control			555	1.723	0.6	0.60					
vs A	0.8	0.80012	322	1.000	0.6	0.64	0.8	1.25	1.06667	0.05	0.016667
vs B	0.8	0.91570	322	1.000	0.6	0.60	0.8	1.25	1.00000	0.05	0.016667
vs C	0.8	0.91570	322	1.000	0.6	0.60	0.8	1.25	1.00000	0.05	0.016667
Total			1521								

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, where Pc is the control group proportion.
Pi.1	The response proportion in the ith group at which the power is calculated.
RL	The lower equivalence ratio. This is the lower equivalence bound of the ratio of treatment and control proportions that still results in the conclusion that the treatment group is equivalent to the control group.
RU	The upper equivalence ratio. This is the largest equivalence bound of the ratio of treatment and control proportions that still results in the conclusion that the treatment group is equivalent to the control group.
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 equivalent to the control group proportion, with equivalence ratio bounds of 0.8 and 1.25 (H0: R ≤ 0.8 or R ≥ 1.25 versus H1: 0.8 < R < 1.25, R = Pi / Pc). Each of the 3 equivalence comparisons will be made using two one-sided, two-sample, Bonferroni-adjusted Gart & Nam Likelihood Score tests of the 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 434 and the number of needed subjects for the treatment groups will be 252, 252, and 252 (totaling 1190 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%	434	543	109
2	20%	252	315	63
3	20%	252	315	63
4	20%	252	315	63
Total		1190	1488	298
1	20%	441	552	111
2	20%	256	320	64
3	20%	256	320	64
4	20%	256	320	64
Total		1209	1512	303
1	20%	555	694	139
2	20%	322	403	81
3	20%	322	403	81
4	20%	322	403	81
Total		1521	1903	382

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 543, 315, 315, and 315 subjects should be enrolled to obtain final group sample sizes of 434, 252, 252, and 252 subjects.

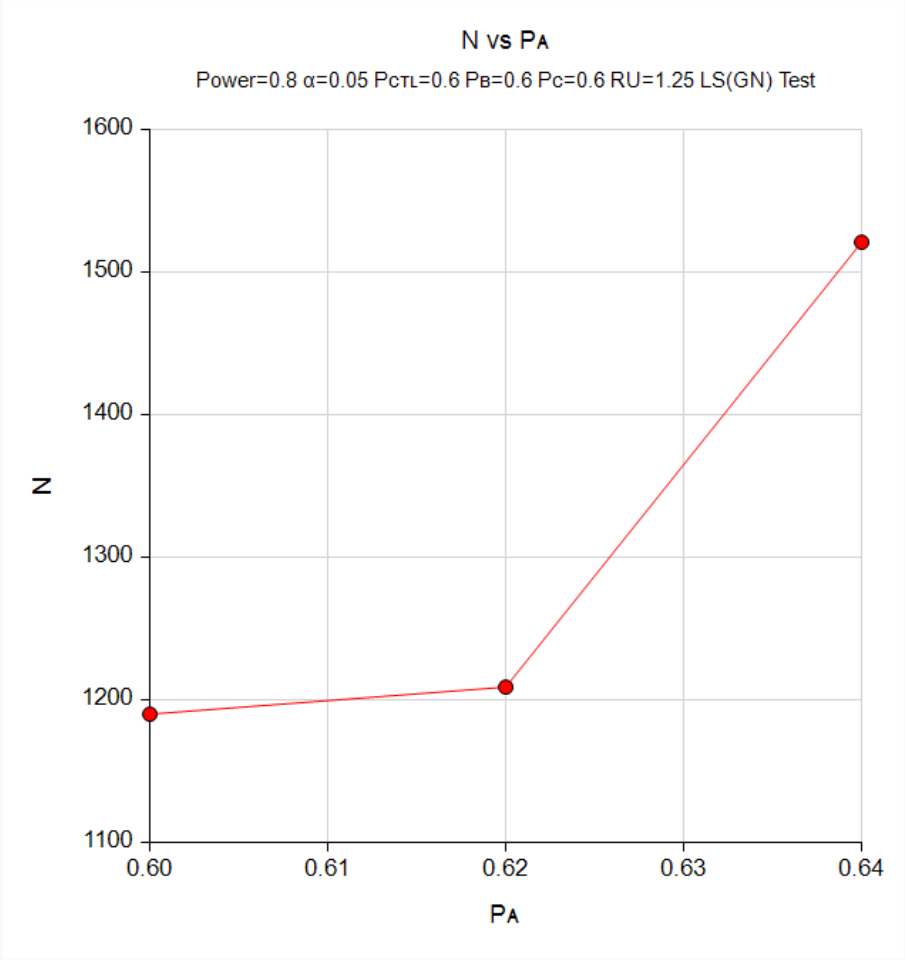
References

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.

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 ratio of 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 (**Equivalence Tests for the Ratio of Two Proportions**) to produce the results for the following example.

A parallel-group, clinical trial is being designed to compare three doses of a test compound against the standard therapy using three Gart-Nam 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 0.80. The response rate of group 1 is 60%. The response rate of group 2 is 60%. The response rate of group 3 is 60%. The upper equivalence limit is 1.25. The lower equivalence limit is 0.8 (1/1.25). The sample sizes of all groups will be equal.

The **Equivalence Tests for the 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 (Gart & Nam)**
Power	**0.8**
Alpha	**0.016667** (which is Alpha / k)
Group Allocation	**Equal (N1 = N2)**
R0.U (Upper Equivalence Ratio)	**1.25**
R0.L (Lower Equivalence Ratio)	**1/R0.U**
R1 (Actual Ratio)	**1.0**
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: Gart & Nam Likelihood Score Test
Hypotheses: H0: P1 / P2 ≤ R0.L or P1 / P2 ≥ R0.U vs. H1: R0.L < P1 / P2 < R0.U

Power		Sample Size			Proportions				Ratio			
					Equivalence				Equivalence			
Target	Actual*	N1	N2	N	Lower P1.0L	Upper P1.0U	Actual P1.1	Reference P2	Lower R0.L	Upper R0.U	Actual R1	Alpha
0.8	0.80004	318	318	636	0.48	0.75	0.6	0.6	0.8	1.25	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 318. 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 (Gart & Nam)**
Power of Each Test	**0.8**
Overall Alpha	**0.05**
Bonferroni Adjustment	**Standard Bonferroni**
Group Allocation	**Equal (Nc = N1 = N2 = ...)**
RU (Upper Equivalence Ratio)	**1.25**
RL (Lower Equivalence Ratio)	**1/RU**
Control Proportion	**0.6**
Set A Number of Groups	**1**
Set A Proportion	**0.6**
Set B Number of Groups	**1**
Set B Proportion	**0.6**
Set C Number of Groups	**1**
Set C Proportion	**0.6**
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: R ≤ RL or R ≥ RU vs. H1: RL < R < RU
Test Type:	Gart & Nam Likelihood Score Test
Number of Groups:	4
Bonferroni Adjustment:	Standard Bonferroni (Divisor = 3)

						Ratio				
	Power		Sample Size	Proportion		Equivalence			Alpha	
Comparison	Target	Actual	Ni	Pi|H0 Pi.0	Pi|H1 Pi.1	Lower RL	Upper RU	Actual Ri	Overall	Bonferroni-Adjusted
Control			318	0.6	0.6					
vs A	0.8	0.80004	318	0.6	0.6	0.8	1.25	1	0.05	0.016667
vs B	0.8	0.80004	318	0.6	0.6	0.8	1.25	1	0.05	0.016667
vs C	0.8	0.80004	318	0.6	0.6	0.8	1.25	1	0.05	0.016667
Total			1272							

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