Chapter 347

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

Introduction

This module computes power and sample size for multi-arm, non-inferiority tests for vaccine efficacy (VE) using the ratio of treatment and control proportions. VE is a traditional index of the protective efficacy of a vaccine. It is calculated as

$$VE_i = \frac{P_C - P_i}{P_C} = 1 - \frac{P_i}{P_C}$$

where P_i and P_C are *attack rates* of the disease being studied among those vaccinated and those not vaccinated. An attack rate is the probability that a subject without the disease at the beginning of the study is infected by it during the duration of the study. Hence, an analysis of vaccine effectiveness reduces to an analysis of the ratio of two proportions.

The multi-arm nature of 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 (event probability) is measured for each group. A total of *k* hypothesis tests are anticipated, each comparing a treatment group with the common control group using a non-inferiority test based on 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%.

Relative Vaccine Efficacy

Often, the goal of the study is to show that the attack rate of a new vaccine is no worse than that of the current standard vaccine. For example, the standard vaccine might have serious side effects, be expensive to produce, etc. In this case, the trial is conducted to show that the new vaccine is an attractive replacement for the standard vaccine. In this case, the control group does not receive a placebo. Rather, it receives the standard vaccine. In this case, the quantity of interest is called the *relative vaccine efficacy (rVE).* It is calculated as

$$rVE_i = \frac{P_C - P_i}{P_C} = 1 - \frac{P_i}{P_C}$$

where now P_C is the attack rate for those receiving the standard vaccine.

Technical Details

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

If we define VE_0 to be the non-inferiority bound on vaccine efficacy, the *k* one-sided non-inferiority tests are

$$H_{0i}: VE_i \leq VE_0 \quad \text{vs.} \quad H_{1i}: VE_i > VE_0 \quad \text{for } i = 1, 2, \ldots, k$$

Note that if lower proportions are better, as is usually the case when the studying disease prevention, $VE_0 < 1$.

For convenience, these hypotheses are collectively referred to as

$$H_0: VE \leq VE_0 \quad \text{vs.} \quad H_1: VE > VE_0$$

These hypotheses may be restated in terms of proportions as

$$H_0: P_i \geq P_0 \quad \text{vs.} \quad H_1: P_i < P_0$$

where P_0 is the non-inferiority event probability.

Transforming from VE to P

It is often useful to transform a VE_i value into the corresponding P_i. The transformation formula is

$$P_i = P_C(1 - VE_i)$$

Test Statistics

Three common likelihood-score tests 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

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

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

$$A = NR_0$$

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

$$C = m_1$$

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

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

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

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

Adapting the Ratio of Two Proportions to Vaccine Efficacy Studies

A traditional index of the protective efficacy of a vaccine is called the vaccine efficacy (VE). It is calculated as

$$VE = \frac{p_C - p_i}{p_C} = 1 - \frac{p_i}{p_C}$$

Note that VE is a simple transformation of the ratio made by subtracting it from one. Thus, methods for the ratio of two proportions can be easily adapted for vaccine efficacy studies. Blackwelder (1993) gives the details. He recommends using the score test of Gart and Nam.

Non-Inferiority Bound

The idea of a non-inferiority test is that a new treatment is no worse than the treatment it is being compared to. To allow the comparison to be made, you must determine a non-inferiority boundary. In this procedure, that means that p_i can only be slightly larger than p_C. When $p_i > p_C$, the risk ratio will be greater than one so that the value of VE will be negative. See Nauta (2020) page 94 for a discussion and example of this.

Hence, one task that will have to be completed is to determine how much worse the new vaccine can be without causing it to be rejected.

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.

A Note on Setting the Significance Level, Alpha

Setting the significance level has always been somewhat arbitrary. For planning purposes, the standard has become to set alpha to 0.05 for two-sided tests. Almost universally, when someone states that a result is statistically significant, they mean statistically significant at the 0.05 level.

Although 0.05 may be the standard for two-sided tests, it is not always the standard for one-sided tests, such as non-inferiority tests. Statisticians often recommend that the alpha level for one-sided tests be set at 0.025 since this is the amount put in each tail of a two-sided test.

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 test the non-inferiority of three dose levels of a test compound against the standard therapy using three non-inferiority Gart and Nam Likelihood Scores tests. Suppose the standard therapy has an event probability of 0.45. The investigators would like a sample size large enough to detect statistical significance at an overall 0.025 level and an individual-test power of 0.80 when the event rates of treatment group 1 are 0.35, 0.38, 0.40. The event rate of group 2 is 0.35. The event rate of group 3 is 0.36. The non-inferiority ratio is 0.5. This corresponds to VE0 = -0.11111.

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	**0.80**
Overall Alpha	**0.025**
Bonferroni Adjustment	**Standard Bonferroni**
Group Allocation	**Enter Group Allocation Pattern, solve for group sample sizes**
P0 (Non-Inferiority Event Probability)	**0.5**
Control Event Probability	**0.45**
Control Sample Size Allocation	**1.732**
Set A Number of Groups	**1**
Set A Event Probability	**0.35 0.38 0.40**
Set A Sample Size Allocation	**1**
Set B Number of Groups	**1**
Set B Event Probability	**0.35**
Set B Sample Size Allocation	**1**
Set C Number of Groups	**1**
Set C Event Probability	**0.36**
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:	Gart & Nam Likelihood Score Test
Higher Proportions Are:	Worse
Hypotheses:	H0: Pi ≥ P0 vs. H1: Pi < P0 or H0: VEi ≤ VE0 vs. H1: VEi > VE0
Number of Groups:	4
Bonferroni Adjustment:	Standard Bonferroni (Divisor = 3)

	Power		Sample Size		Event Probability		Vaccine Efficacy		Alpha	
Comparison	Target	Actual	Ni	Allocation	Non-Inferiority P0	Actual Pi	Non-Inferiority VE0	Actual VEi	Overall	Bonferroni-Adjusted
Control			385	1.732	0.5	0.45				
vs A	0.8	0.86134	222	1.000	0.5	0.35	-0.11111	0.22222	0.025	0.008333
vs B	0.8	0.86134	222	1.000	0.5	0.35	-0.11111	0.22222	0.025	0.008333
vs C	0.8	0.80119	222	1.000	0.5	0.36	-0.11111	0.20000	0.025	0.008333
Total			1051							
Control			527	1.732	0.5	0.45				
vs A	0.8	0.80086	304	1.000	0.5	0.38	-0.11111	0.15556	0.025	0.008333
vs B	0.8	0.95398	304	1.000	0.5	0.35	-0.11111	0.22222	0.025	0.008333
vs C	0.8	0.91959	304	1.000	0.5	0.36	-0.11111	0.20000	0.025	0.008333
Total			1439							
Control			762	1.732	0.5	0.45				
vs A	0.8	0.80039	440	1.000	0.5	0.40	-0.11111	0.11111	0.025	0.008333
vs B	0.8	0.99412	440	1.000	0.5	0.35	-0.11111	0.22222	0.025	0.008333
vs C	0.8	0.98532	440	1.000	0.5	0.36	-0.11111	0.20000	0.025	0.008333
Total			2082							

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 vaccine efficacy.
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.
P0	The non-inferiority bound on the event probability determines when to conclude that a treatment is non-inferior or inferior to the control group.
Pi	The event probability of the ith group at which the power is calculated.
VE0	The non-inferiority bound on the vaccine efficacy determines when to conclude that a treatment is non-inferior or inferior to the control group. Note that VE0 = 1 - P0 / Pc.
VEi	The vaccine efficacy of the group reported on this line of the report. This is the value at which the power is calculated. The formula is VEi = 1 - Pi / 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 event probability for each treatment group (Pi) is non-inferior to the control group event probability (Pc), with a non-inferiority event probability limit of 0.5 (H0: Pi ≥ 0.5 versus H1: Pi < 0.5, or the corresponding hypotheses, H0: VEi ≤ -0.11111 versus H1: VEi > -0.11111, where VEi is the vaccine efficacy [1 - Pi / Pc] of treatment i, and -0.11111 is the vaccine efficacy non-inferiority limit). The non-inferiority hypotheses will be evaluated using 3 one-sided, two-sample, Bonferroni-adjusted Gart & Nam Likelihood Score tests of the ratio, with an overall (experiment-wise) Type I error rate (α) of 0.025. In this study, lower event probabilities are considered to be better. The control group event probability is assumed to be 0.45. To detect the treatment event probabilities 0.35, 0.35, and 0.36 with at least 80% power for each test, the control group sample size needed will be 385 and the number of needed subjects for the treatment groups will be 222, 222, and 222 (totaling 1051 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%	385	482	97
2	20%	222	278	56
3	20%	222	278	56
4	20%	222	278	56
Total		1051	1316	265
1	20%	527	659	132
2	20%	304	380	76
3	20%	304	380	76
4	20%	304	380	76
Total		1439	1799	360
1	20%	762	953	191
2	20%	440	550	110
3	20%	440	550	110
4	20%	440	550	110
Total		2082	2603	521

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 482, 278, 278, and 278 subjects should be enrolled to obtain final group sample sizes of 385, 222, 222, and 222 subjects.

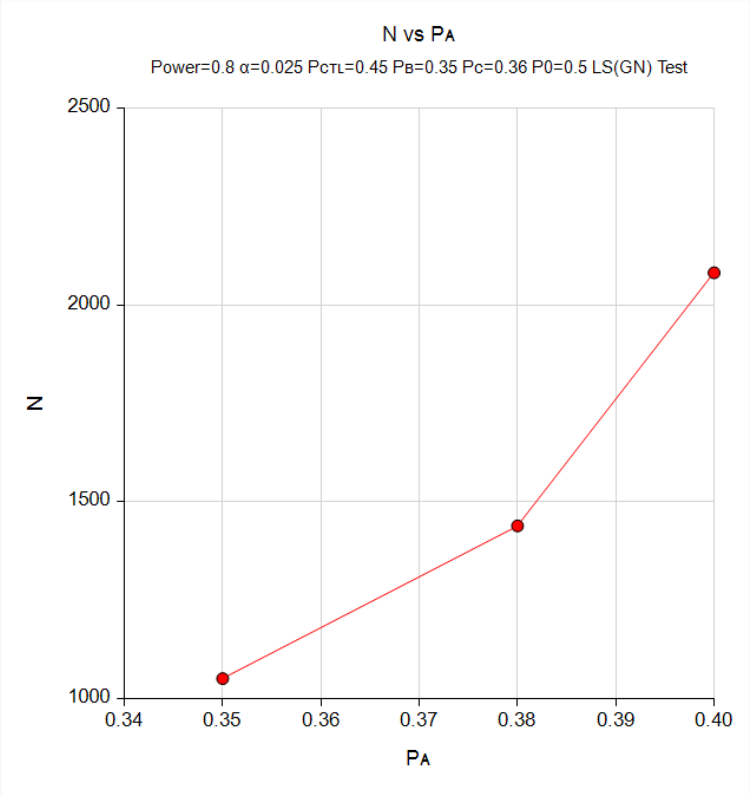
References

Blackwelder, William C. 1993. 'Sample Size and Power for Prospective Analysis of Relative Risk.' Statistics in Medicine, Vol. 12, 691-698.
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.
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.
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.
Nauta, Jozef. 2020. Statistics in Clinical and Observational Vaccine Studies, 2nd Edition. Springer. Cham, Switzerland.

This report shows the numeric results of this sample size study. Notice that the results are shown in blocks of five 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 event probability of the first treatment.

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 Vaccine Efficacy using the Ratio of Two Proportions**) to produce the results for the following example.

A parallel-group, clinical trial is being designed to test the non-inferiority of three dose levels of a test compound against the standard therapy using three non-inferiority Gart and Nam Likelihood Scores tests. Suppose the standard therapy has an event probability of 0.45. The investigators would like a sample size large enough to detect statistical significance at an overall 0.025 level and an individual-test power of 0.80 when the event rates of all three treatment groups are set to 0.38 because this is the rate that they want to be able to detect. The non-inferiority ratio is 0.5. Note that since there are three tests, the significance level will be set to 0.00833.

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

The **Non-Inferiority Tests for Vaccine Efficacy using 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	**0.00833** (which is Alpha / k)
Group Allocation	**Equal (N1 = N2)**
Vaccine Efficacy Input Type	**Enter P1.0, P1.1, and P2**
P1.0 (Non-Inferiority Vaccine Event Prob)	**0.5**
P1.1 (Actual Vaccine Event Prob)	**0.38**
P2 (Control Group Event Probability)	**0.45**

This set of options generates the following report.

Numeric Results

Solve For: Sample Size
Test Statistic: Gart & Nam Likelihood Score Test
Groups: 1 = Vaccine, 2 = Control
Hypotheses: H0: VE ≤ VE0 vs. H1: VE > VE0

					Event Probability					
						Vaccine		Vaccine Efficacy		
Power		Sample Size								
Target	Actual*	N1	N2	N	Control P2	Non-Inferiority P1.0	Actual P1.1	Non-Inferiority VE0	Actual VE1	Alpha
0.8	0.80033	393	393	786	0.45	0.5	0.38	-0.11111	0.15556	0.00833

* Power was computed using the normal approximation method.

In order to maintain a power of 80% for all groups, it is apparent that the groups will all need to have a sample size of 393.

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	**0.80**
Overall Alpha	**0.025**
Bonferroni Adjustment	**Standard Bonferroni**
Group Allocation	**Equal (Nc = N1 = N2 = ...)**
P0 (Non-Inferiority Event Probability)	**0.5**
Control Event Probability	**0.45**
Set A Number of Groups	**3**
Set A Event Probability	**0.38**
Set B Number of Groups	**0**
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:	Gart & Nam Likelihood Score Test
Higher Proportions Are:	Worse
Hypotheses:	H0: Pi ≥ P0 vs. H1: Pi < P0 or H0: VEi ≤ VE0 vs. H1: VEi > VE0
Number of Groups:	4
Bonferroni Adjustment:	Standard Bonferroni (Divisor = 3)

	Power			Event Probability		Vaccine Efficacy		Alpha	
Comparison	Target	Actual	Sample Size Ni	Non-Inferiority P0	Actual Pi	Non-Inferiority VE0	Actual VEi	Overall	Bonferroni-Adjusted
Control			393	0.5	0.45				
vs A1	0.8	0.80038	393	0.5	0.38	-0.11111	0.15556	0.025	0.008333
vs A2	0.8	0.80038	393	0.5	0.38	-0.11111	0.15556	0.025	0.008333
vs A3	0.8	0.80038	393	0.5	0.38	-0.11111	0.15556	0.025	0.008333
Total			1572						

The sample sizes and powers match, which validates this procedure.