

Chapter 620

Multi-Arm Non-Inferiority Tests for Vaccine Efficacy using the Ratio of Treatment and Control Proportions in a Cluster-Randomized Design

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 when binary data are gathered from a cluster-randomized design. 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 formulas are based on results in Donner and Klar (2000) and Machin, Campbell, Tan, and Tan (2018).

A *cluster (group) randomized design* is one in which whole units, or clusters, of subjects are randomized to the groups rather than the individual subjects in those clusters. The conclusions of the study concern individual subjects rather than the clusters. Examples of clusters are families, school classes, neighborhoods, hospitals, and doctor's practices.

Cluster-randomized designs are often adopted when there is a high risk of contamination if cluster members were randomized individually. For example, it may be difficult for doctors to use two treatment methods in their practice. The price of randomizing by clusters is a loss of efficiency--the number of subjects needed to obtain a certain level of precision in a cluster-randomized trial is usually much larger than the number needed when the subjects are randomized individually. Hence, standard methods of sample size estimation cannot be used.

In this multi-arm design, there are G treatment groups and one control group. A proportion is measured in each group. A total of G hypothesis tests are anticipated each comparing a treatment group with the common control group using a z-test of the ratio between two proportions.

The Bonferroni adjustment of the type I error rate may be optionally made because several comparisons are being tested using 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.

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.

Multiple Treatments Versus a Single Control

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 design 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 the new treatment are better than 50-50.

Technical Details

Our formulation for cluster-randomized designs comes from Donner and Klar (2000). Suppose you have G treatment groups with response probabilities P_i that have samples of size N_i and one control group with response probability P_C that has a sample of size N_C . The total sample size is $N = N_1 + N_2 + \dots + N_G + N_C$.

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

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

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

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

Cluster-Randomized Designs

Denote a binary (0, 1) observation by Y_{ikj} where i is the group, $k = 1, 2, \dots, K_i$ is a cluster within group i , and $j = 1, 2, \dots, M_i$ is an item (often a subject) in cluster k of group i . The results that follow assume an equal number of items per cluster per group. When the number of items from cluster to cluster are about the same, the power and sample size values should be fairly accurate. In these cases, the average number of items per cluster can be used.

The statistical hypothesis that is tested concerns the ratio between a treatment group proportion and the control group proportion: P_i and P_C . With a simple modification, the large-sample sample size formulas that are listed in the module for testing two proportions can be used here.

When the items are randomly assigned to one of the $G + 1$ groups, the variance of the sample proportion is

$$\sigma_{S,i}^2 = \frac{P_i(1 - P_i)}{N_i}$$

When the randomization is by clusters of subjects, the variance of the sample proportion is

$$\begin{aligned}\sigma_{C,g}^2 &= \frac{p_g(1 - p_g)DE}{k_g m_g} \\ &= \sigma_{S,g}^2 DE\end{aligned}$$

where DE is the *design effect*. We use the following version of DE given by Machin et al. (2018) which allows for an adjustment for unequal cluster sizes

$$DE = 1 + \left\{ \left[COV(m)^2 \left(\frac{K-1}{K} \right) + 1 \right] \bar{m} - 1 \right\} \rho$$

where K is the total number of clusters. This formula assumes that the cluster sizes, m , are distributed with a mean of \bar{m} and a coefficient of variation of $COV(m)$.

The Greek letter ρ is used to represent the *intracluster correlation coefficient (ICC)*. This correlation may be thought of as the simple correlation between any two subjects within the same cluster. If we stipulate that ρ is positive, it may also be interpreted as the proportion of total variability that is attributable to differences between clusters. This value is critical to the sample size calculation.

The asymptotic formula for the Farrington and Manning Likelihood Score Test that was used in comparing two proportions (see Chapter 211, "Non-Inferiority Tests for the Ratio of Two Proportions") may be used with cluster-randomized designs as well, as long as an adjustment is made for the design effect.

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}_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 increase the variance estimate. 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). Note that in large samples, the Farrington and Manning statistic is substituted for the Gart and Nam statistic.

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.

Multiplicity Adjustment

Because G z-tests between treatment groups and the control group are run when analyzing the results of this study, many statisticians recommend that a 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 should 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 clusters in this group. The standard adjustment is to include \sqrt{G} clusters in the control group for each cluster 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 sample size.

Example 1 – Finding the Sample Size

A cluster-randomized, multi-arm trial is being designed to compare two treatments against the standard drug in patients with a specific type of disease. The plan is to analyze the binary data using the Farrington and Manning likelihood score test based on the response ratio. The non-inferiority boundary will be set in terms of vaccine efficacy.

Historically, the standard treatment has had a 60% failure rate. The new treatments are expected to reduce the seriousness of side effects and are cheaper to produce. The new treatments will be adopted even if they are slightly less effective than the standard treatment. The researchers will recommend adoption of the either of the new treatments that exhibit a vaccine efficacy of -0.1 .

The researchers will recruit patients from various hospitals. All patients at a particular hospital will receive the same treatment. They anticipate an average of 10 patients per hospital. They want to see the impact on cluster count of cluster sizes ranging for 5 to 15. The COV of cluster size is 0.65. Based on similar studies, they estimate the intracluster correlation to be 0.002.

The investigators would like a sample size large enough to find statistical significance at the 0.025 level and a power of 0.90 in each test.

Since the control group will be used twice, they set the control group cluster allocation to $\sqrt{G} = \sqrt{2} = 1.414$. The two treatment allocations are set to 1.0.

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
Power of Each Test	0.90
Overall Alpha	0.025
Bonferroni Adjustment	Standard Bonferroni
Group Allocation	Enter Group Allocation Pattern, solve for group numbers of clusters
M (Average Cluster Size).....	5 10 15
COV of Cluster Sizes.....	0.65
VE0 (Non-Inferiority Vaccine Efficacy).....	-0.1
Control Event Probability	0.6
Control Cluster Allocation	1.414
Set A Number of Groups.....	2
Set A Event Probability	0.55
Set A Cluster Allocation	1
Set B Number of Groups.....	0
Set C Number of Groups	0
Set D Number of Groups	0
More.....	Unchecked
ρ (Intracluster Correlation).....	0.002

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 numbers of clusters
 Test Type: Farrington and Manning Likelihood Score Test
 Hypotheses: $H_0: VE_i \leq VE_0$ vs. $H_1: VE_i > VE_0$
 Number of Groups: 3
 Bonferroni Adjustment: Standard Bonferroni (Divisor = 2)

Comparison	Power		Number of Clusters Ki	Cluster Size		Sample Size Ni	Event Probability		Vaccine Efficacy		ICC p	Alpha		
	Target	Actual		Cluster Allocation	Average Mi		COV	Non-Inferiority P0	Actual Pi	Non-Inferiority VE0		Actual VEi	Overall	Bonferroni-Adjusted
Control vs A1	0.9	0.90130	133	1.414	5	0.65	665	0.66	0.60	-0.1	0.00000	0.002		
vs A2	0.9	0.90130	94	1.000	5	0.65	470	0.66	0.55	-0.1	0.08333	0.002	0.025	0.0125
Total			321				1605							
Control vs A1	0.9	0.90367	68	1.414	10	0.65	680	0.66	0.60	-0.1	0.00000	0.002		
vs A2	0.9	0.90367	48	1.000	10	0.65	480	0.66	0.55	-0.1	0.08333	0.002	0.025	0.0125
Total			164				1640							
Control vs A1	0.9	0.90952	47	1.414	15	0.65	705	0.66	0.60	-0.1	0.00000	0.002		
vs A2	0.9	0.90952	33	1.000	15	0.65	495	0.66	0.55	-0.1	0.08333	0.002	0.025	0.0125
Total			113				1695							

- 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.
- Ki** The number of clusters in the ith group. The total number of clusters is reported in the last row of the column.
- Allocation** The cluster allocation ratio of the ith group. The value on each row represents the relative number of clusters assigned to the group.
- Mi** The average number of items per cluster (or average cluster size) in the ith group.
- COV** The coefficient of variation of the cluster sizes within the group.
- Ni** The number of items in the ith group. The total sample size is shown as the last row of the column.
- 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 $VE_0 = 1 - P_0 / P_c$.
- 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 $VE_i = 1 - P_i / P_c$.
- p** The intracluster correlation (ICC). The correlation between subjects within a cluster.
- 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, 3-group cluster-randomized design (with one control group and 2 treatment groups) will be used to test whether the proportion for each treatment group is non-inferior to the control group proportion, by testing whether the vaccine efficacy ($VE = 1 - \pi_i / \pi_c$) is greater than the non-inferiority boundary of -0.1 . The hypotheses will be evaluated using 2 one-sided, two-sample, Bonferroni-adjusted (divisor = 2) Farrington and Manning likelihood score tests, with an overall (experiment-wise) Type I error rate (α) of 0.025. The control group proportion is assumed to be 0.6. The intracluster correlation is assumed to be 0.002. The average cluster size (number of subjects or items per cluster) for the control group is assumed to be 5, and the average cluster size for each of the treatment groups is assumed to be 5 and 5. To detect the treatment proportions 0.55 and 0.55 with at least 90% power for each test, the control group cluster count needed will be 133 and the number of needed clusters for the treatment groups will be 94 and 94 (totaling 321 clusters overall).

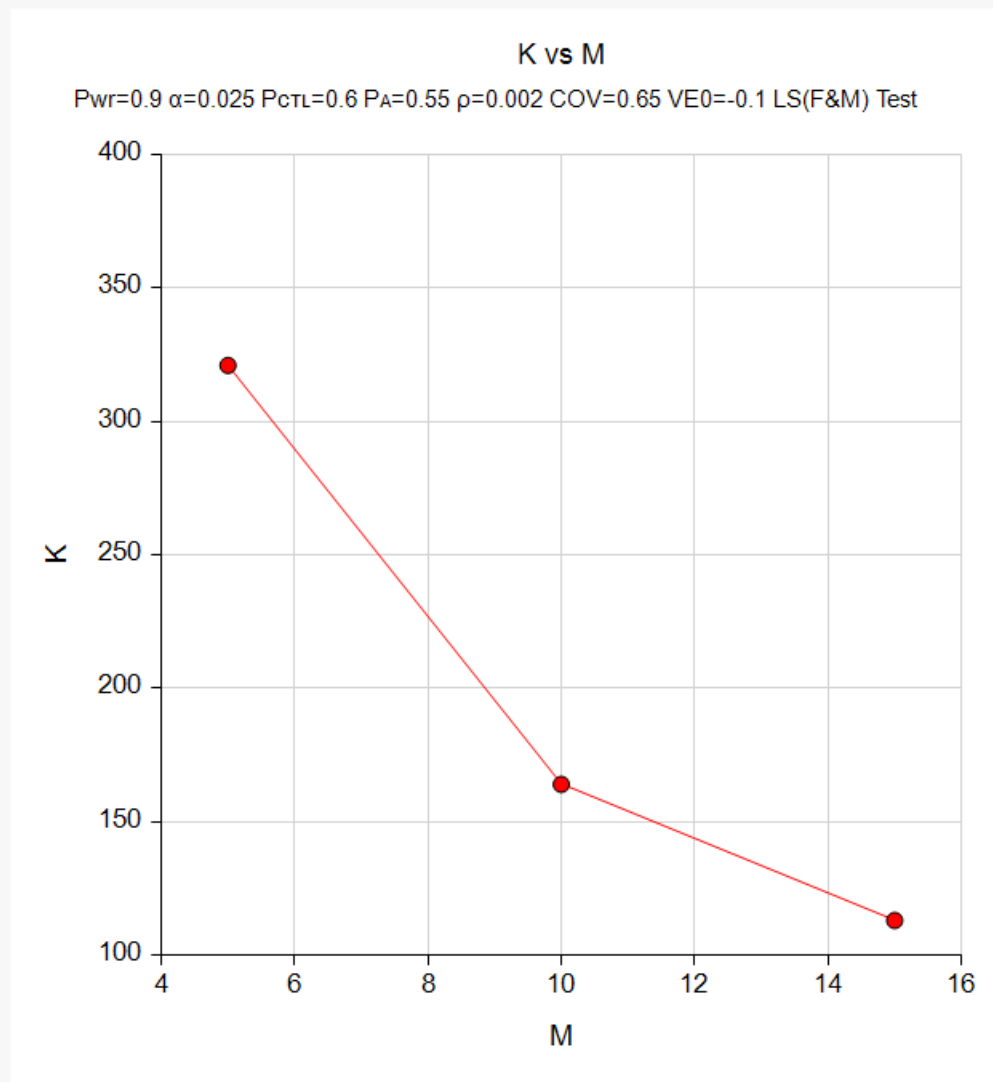
References

- Blackwelder, W.C. 1998. 'Equivalence Trials.' In Encyclopedia of Biostatistics, John Wiley and Sons. New York. Volume 2, 1367-1372.
- Campbell, M.J. and Walters, S.J. 2014. How to Design, Analyse and Report Cluster Randomised Trials in Medicine and Health Related Research. Wiley. New York.
- Donner, A. and Klar, N. 2000. Design and Analysis of Cluster Randomization Trials in Health Research. Arnold. London.
- Chow, S.C., Shao, J., Wang, H., and Lohknygina, 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.
- 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.
- Nauta, Jozef. 2020. Statistics in Clinical and Observational Vaccine Studies, 2nd Edition. Springer. Cham, Switzerland.
-

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 total cluster count, K, of increasing the cluster size, M.

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 in a Cluster-Randomized Design**) to produce the results for the following example.

We will use a slightly modified version of the settings from Example 1. Specifically we will set $K_i = 60$, $M = 10$, sample size allocations equal, and solve for power. The Bonferroni adjustment changes the significance level from 0.025 to 0.0125.

The **Non-Inferiority Tests for Vaccine Efficacy using the Ratio of Two Proportions in a Cluster-Randomized Design** procedure is set up as follows.

Design Tab

Solve For **Power**
 Alpha..... **0.0125**
 K1 (Clusters in Group 1) **60**
 M1 (Average Cluster Size)..... **10**
 K2 (Clusters in Group 2) **60**
 M2 (Average Cluster Size)..... **10**
 COV of Cluster Sizes **0.65**
 Vaccine Efficacy Input Type..... **Enter P1.0, P1.1, and P2**
 P1.0 (Non-Inferiority Vaccine Event Prob) **0.66**
 P1.1 (Actual Vaccine Event Prob)..... **0.55**
 P2 (Control Group Event Probability)..... **0.6**
 ρ (Intracluster Correlation, ICC)..... **0.002**

This set of options generates the following report.

Numeric Results

Solve For: **Power**
 Test Statistic: Likelihood Score Test (Farrington & Manning)
 Groups: 1 = Vaccine, 2 = Control
 Hypotheses: $H_0: VE \leq VE_0$ vs. $H_1: VE > VE_0$

Power	Number of Clusters			Cluster Size			Total Sample Size N	Event Probability			Vaccine Efficacy		Intracluster Correlation ρ	Alpha
	Vaccine	Control	Total	Vaccine	Control	COV		Vaccine		Control	Non-Inferiority	Actual		
	K1	K2	K	M1	M2			Non-Inferiority P1.0	Actual P1.1	P2	VE0	VE1		
0.9162	60	60	120	10	10	0.65	1200	0.66	0.55	0.6	-0.1	0.08333	0.002	0.0125

The power is computed to be 0.9162.

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	Power
Overall Alpha	0.025
Bonferroni Adjustment	Standard Bonferroni
Group Allocation	Equal (Kc = K1 = K2 = ...)
Ki (Group Number of Clusters)	60
M (Average Cluster Size).....	10
COV of Cluster Sizes.....	0.65
VE0 (Non-Inferiority Vaccine Efficacy).....	-0.1
Control Event Probability	0.6
Set A Number of Groups.....	2
Set A Event Probability	0.55
Set B Number of Groups.....	0
Set C Number of Groups	0
Set D Number of Groups	0
More.....	Unchecked
ρ (Intracluster Correlation).....	0.002

Output

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

Numeric Results												
Solve For:		Power										
Test Type:		Farrington and Manning Likelihood Score Test										
Hypotheses:		H0: VEi ≤ VE0 vs. H1: VEi > VE0										
Number of Groups:		3										
Bonferroni Adjustment:		Standard Bonferroni (Divisor = 2)										
Comparison	Power	Number of Clusters Ki	Cluster Size		Sample Size Ni	Event Probability		Vaccine Efficacy			Alpha	
			Average Mi	COV		Non-Inferiority P0	Actual Pi	Non-Inferiority VE0	Actual VEi	ICC ρ	Overall	Bonferroni-Adjusted
Control		60	10	0.65	600	0.66	0.60	-0.1	0.00000	0.002		
vs A1	0.9162	60	10	0.65	600	0.66	0.55	-0.1	0.08333	0.002	0.025	0.0125
vs A2	0.9162	60	10	0.65	600	0.66	0.55	-0.1	0.08333	0.002	0.025	0.0125
Total		180			1800							

As you can see, the power is 0.9162 for both treatment groups which match the power found in the validation run above. The procedure is validated.