Chapter 190

Non-Inferiority Tests for Vaccine Efficacy using the Ratio of Two Proportions in a Cluster-Randomized Design

Introduction

This module provides power analysis and sample size calculation for non-inferiority tests for vaccine efficacy (VE) using the ratio of two proportions in a two-sample, cluster-randomized design in which the outcome is binary.

VE is a traditional index of the protective efficacy of a vaccine. It is calculated as

$$VE = \frac{p_2 - p_1}{p_2} = 1 - \frac{p_1}{p_2}$$

where p_1 and p_2 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.

Note that because $p_1 < p_2$, the value of $VE < 1$.

Cluster-randomized designs are those in which whole clusters of subjects (classes, hospitals, communities, etc.) are placed into the vaccine group or the control group. The vaccine efficacy is tested using a *z* test or a logistic regression test. Generally speaking, the larger the cluster sizes and the higher the correlation among subjects within the same cluster, the larger will be the overall sample size necessary to detect an effect with the same power.

This routine is partially based on Blackwelder (1993). It is also based on the work about how to adapt two-sample formulas to cluster-randomized designs by Donner and Klar (2000) as well as Machin et al. (2018).

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 replace 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

$$rVE = \frac{p_2 - p_1}{p_2} = 1 - \frac{p_1}{p_2}$$

where now p_2 is the attack rate for those receiving the standard vaccine.

Technical Details

Our formulation comes from Donner and Klar (2000). Denote a binary observation by Y_{gkj} where g = 1 or 2 is the group, k = 1, 2, ..., K_g is a cluster within group g, and j = 1, 2, ..., M_g is an individual in cluster k of group g.

The statistical hypothesis that is tested concerns the ratio of the two group proportions, p_1 and p_2. We assume that group 1 is the vaccine group and group 2 is the control group. With a simple modification, all of the large-sample sample size formulas that are listed in the module for testing non-inferiority with two proportions using the ratio can be used here.

When the individual subjects are randomly assigned to one of the two groups, the variance of the sample proportion is

$$\sigma_{S,g}^2 = \frac{p_g(1-p_g)}{n_g}$$

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

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

$$= \sigma_{S,g}^2 DE$$

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, $\hat{p}_1$ and $\hat{p}_2$, are used in the numerator of the score statistic while MLE's $\tilde{p}_1$ and $\tilde{p}_2$, constrained so that $\tilde{p}_1 / \tilde{p}_2 = \phi_0$, 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

$$z_{FMR} = \frac{\hat{p}_1 / \hat{p}_2 - \phi_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \phi_0^2\frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)}}$$

where the estimates $\tilde{p}_1$ and $\tilde{p}_2$ 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

$$VE = \frac{p_2 - p_1}{p_2} = 1 - \frac{p_1}{p_2}$$

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.

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_1 can only be slightly larger than p_2. When $p_1 > p_2$, 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.

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.

Power Calculations

The power for the above test statistic can be computed exactly using two binomial distributions if the cluster-randomization is ignored. The following steps are taken to compute the power of these tests.

1. Find the critical value using the standard normal distribution. The critical value, $z_{critical}$, is that value of *z* that leaves exactly the target value of alpha in the appropriate tail of the normal distribution.
2. Compute the value of the test statistic, z_t, for every combination of x_{11} andx_{21}. Note that x_{11} ranges from 0 to n_1, and x_{21} ranges from 0 to n_2. A small value (around 0.0001) can be added to the zero-cell counts to avoid numerical problems that occur when the cell value is zero.
3. If $z_t > z_{critical}$, the combination is in the rejection region. Call all combinations of x_{11} and x_{21} that lead to a rejection the set *A*.
4. Compute the power for given values of $p_{1.1}$ and p_2 as

$$1-\beta = \sum_{A} \binom{n_1}{x_{11}} p_{1.1}^{x_{11}} q_{1.1}^{n_1-x_{11}} \binom{n_2}{x_{21}} p_2^{x_{21}} q_2^{n_2-x_{21}}.$$

5. Compute the actual value of alpha achieved by the design by substituting $p_{1.0}$ for $p_{1.1}$ to obtain

$$\alpha^* = \sum_{A} \binom{n_1}{x_{11}} p_{1.0}^{x_{11}} q_{1.0}^{n_1-x_{11}} \binom{n_2}{x_{21}} p_2^{x_{21}} q_2^{n_2-x_{21}}.$$

Asymptotic Approximations

In cluster-randomized designs, a large sample approximation can be used which uses the DE adjustment shown above. The large sample approximation is made by replacing the values of $\hat{p}_1$ and $\hat{p}_2$ in the *z* statistic with the corresponding values of $p_{1.1}$ and p_2, 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.

Example 1 – Finding Sample Size

A cluster-randomized study is being designed to establish the non-inferiority of vaccine A compared to vaccine B. The researchers plan to use the Farrington and Manning likelihood score test to analyze the data. They want to find the sample size required to guarantee a power of 0.9 when the non-inferiority vaccine efficacy is set to -0.1 and the actual vaccine efficacy is set to values of 0, 0.1, and 0.2. The event probability of the control group is 0.5. The significance level will be 0.025.

The researchers estimate that the average cluster size will be 100 and the COV of cluster sizes will be 0.65.

They want to consider two values of the intracluster correlation: 0.0 and 0.02. The value of zero lets them determine the sample size requirements if the clustering were ignored.

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 (Clusters)**
Power	**0.9**
Alpha	**0.025**
M1 (Average Cluster Size)	**100**
K2 (Clusters in Group 2)	**K1**
M2 (Average Cluster Size)	**M1**
COV of Cluster Sizes	**0.65**
Vaccine Efficacy Input Type	**Enter VE0, VE1, and P2**
VE0 (Non-Inferiority Vaccine Efficacy)	**-0.1**
VE1 (Actual Vaccine Efficacy)	**0 0.1 0.2**
P2 (Control Group Event Probability)	**0.5**
ρ (Intracluster Correlation, ICC)	**0 0.02**

Output

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

Numeric Reports

Numeric Results

Solve For: Sample Size (Clusters)
Test Statistic: Likelihood Score Test (Farrington & Manning)
Groups: 1 = Vaccine, 2 = Control
Hypotheses: H0: VE ≤ VE0 vs. H1: VE > VE0

	Number of Clusters			Cluster Size				Event Probability			Vaccine Efficacy			
							Total	Vaccine						
Power	Vaccine K1	Control K2	Total K	Vaccine M1	Control M2	COV	Sample Size N	Non-Inferiority P1.0	Actual P1.1	Control P2	Non-Inferiority VE0	Actual VE1	Intracluster Correlation ρ	Alpha
0.90950	24	24	48	100	100	0.65	4800	0.55	0.50	0.5	-0.1	0.0	0.00	0.025
0.90166	89	89	178	100	100	0.65	17800	0.55	0.50	0.5	-0.1	0.0	0.02	0.025
0.91049	6	6	12	100	100	0.65	1200	0.55	0.45	0.5	-0.1	0.1	0.00	0.025
0.90166	22	22	44	100	100	0.65	4400	0.55	0.45	0.5	-0.1	0.1	0.02	0.025
0.94099	3	3	6	100	100	0.65	600	0.55	0.40	0.5	-0.1	0.2	0.00	0.025
0.91397	10	10	20	100	100	0.65	2000	0.55	0.40	0.5	-0.1	0.2	0.02	0.025

Power	The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
K1, K2, and K	The number of clusters in groups 1, 2, and both, respectively.
M1 and M2	The average number of items (subjects) per cluster in groups 1 and 2, respectively.
COV	The coefficient of variation of the cluster sizes.
N	The total number of subjects in the study. N = (K1 x M1) + (K2 x M2).
P1.0	The largest value of the event probability for vaccinated group that still yields a non-inferiority conclusion.
P1.1	The value of the event probability for vaccinated group that is assumed by the alternative hypothesis, H1.
P2	The event probability (attack rate) of the control group.
VE0	The vaccine efficacy assumed by the null hypothesis, H0. This is the lower non-inferiority boundary of VE. VE0 = 1 - P1.0/P2.
VE1	The vaccine efficacy assumed by the alternative hypothesis, H1. This is the VE value at which the power is calculated. VE1 = 1 - P1.1/P2.
ρ	The intracluster correlation.
Alpha	The probability of rejecting a true null hypothesis.

Summary Statements

A two-group cluster-randomized design will be used to test whether the Group 1 (vaccine) proportion (P1) is non-inferior to the Group 2 (control) proportion (P2), by testing whether the vaccine efficacy (VE = 1 - P1 / P2) is greater than -0.1 (H0: VE ≤ -0.1 versus H1: VE > -0.1). The comparison will be made using a one-sided proportion-ratio score test with a Type I error rate (α) of 0.025. The coefficient of variation of the cluster sizes is assumed to be 0.65. The intracluster correlation is assumed to be 0. The control group proportion (event probability) is assumed to be 0.5. To detect a vaccine efficacy of 0 (or vaccine event probability of 0.5) with 90% power, the number of clusters needed will be 24 in the vaccine group, with 100 subjects per cluster (totaling 2400 subjects), and 24 clusters in the control group, with 100 subjects per cluster (totaling 2400 subjects).

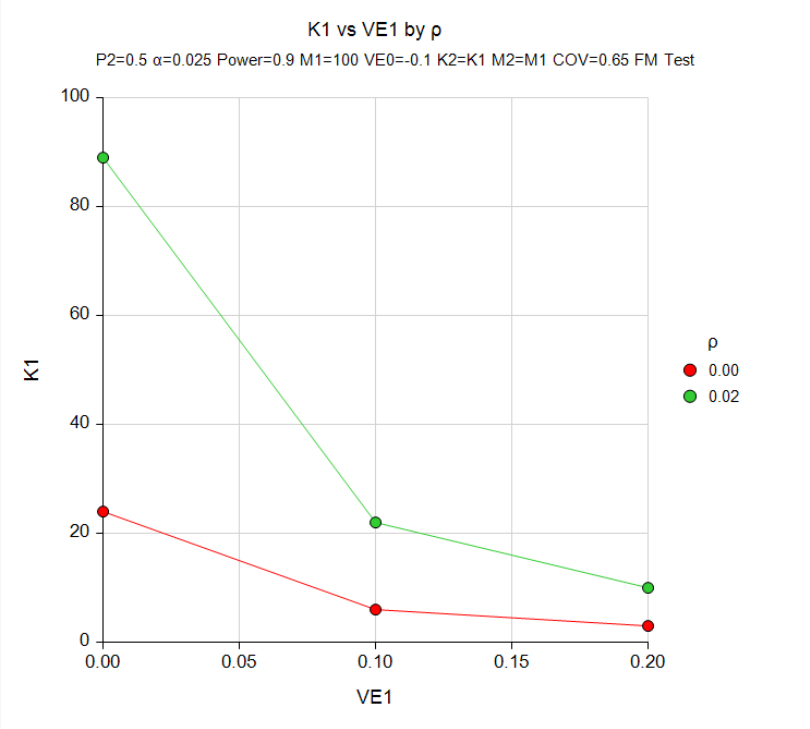
References

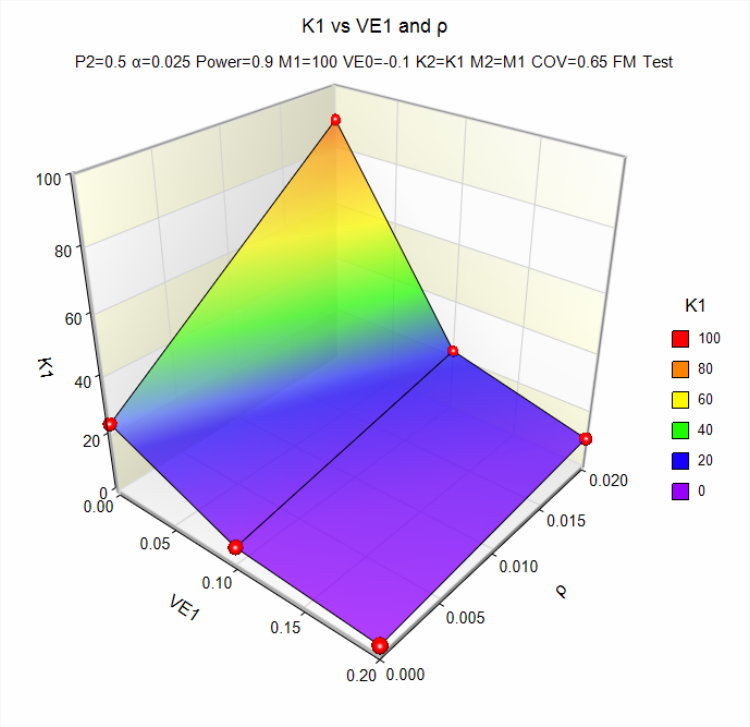
Blackwelder, W.C. 1998. 'Equivalence Trials.' In Encyclopedia of Biostatistics, John Wiley and Sons. New York. Volume 2, 1367-1372.
Campbell, M.J. and Walters, S.J. 2014. How to Design, Analyse and Report Cluster Randomised Trials in Medicine and Health Related Research. Wiley. New York.
Donner, A. and Klar, N. 2000. Design and Analysis of Cluster Randomization Trials in Health Research. Arnold. London.
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., Tan, S.B., and Tan, S.H. 2009. Sample Size Tables for Clinical Studies, 3rd Edition. Wiley-Blackwell. Chichester, UK.
Nauta, Jozef. 2020. Statistics in Clinical and Observational Vaccine Studies, 2nd Edition. Springer. Cham, Switzerland.

This report shows the values of each of the parameters, one scenario per row.

Plots Section

Plots





The values from the table are displayed on the above charts. These charts give a quick look at the sample sizes that will be required for various values of VE1.

Example 2 – Validation using Blackwelder (1993)

We could not find a direct validation example, so we will mix an example from the literature with some hand calculations to obtain a validation example.

Blackwelder (1993), page 695, presents a table of power values for several scenarios using the risk ratio. The second line of the table presents the results for the following scenario: P2 = 0.04, R0 = 0.3 (VE0 = 0.7), R1=0.1 (VE1 = 0.9), N1 = N2=1044, one-sided alpha = 0.05, and beta = 0.20. Using the Farrington and Manning likelihood-score test statistic and the asymptotic formula, he found the approximate power to be 0.794. When we ran it, we obtained a power of 0.79373.

In order to duplicate the above result, we set the power to 0.7937 and M1 to 10.44. This resulted in K1 = K2 = 100.

This example did not include adjustments for cluster randomizations, so we will add those manually. The basic adjustment is to multiply the 1044 by the design effect, DE, to obtain the adjusted sample size.

Suppose we set M1 = M2 = 10.44, ρ = 0.1, and COV = 0.5. Using the relationship N1 = K1 x M1, we find K1 = 100 before the other CR adjustments are made. The value of DE is computed as

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

$$= 1 + \left\{\left[0.25\left(\frac{99}{100}\right) + 1\right]10.44 - 1\right\}0.1$$

$$= 2.20239$$

Hence, the corresponding cluster-randomized design requires $K1' = 100(2.20239) = 220.239 \approx 221$ clusters per group.

Note also that Blackwelder uses the Miettinen and Nurminen score test while this procedure uses the Farrington and Manning score test. It turns out that these two procedures give nearly identical results for large sample sizes.

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 (Clusters)**
Power	**0.7937**
Alpha	**0.05**
M1 (Average Cluster Size)	**10.44**
K2 (Clusters in Group 2)	**K1**
M2 (Average Cluster Size)	**M1**
COV of Cluster Sizes	**0.5**
Vaccine Efficacy Input Type	**Enter VE0, VE1, and P2**
VE0 (Non-Inferiority Vaccine Efficacy)	**0.7**
VE1 (Actual Vaccine Efficacy)	**0.9**
P2 (Control Group Event Probability)	**0.04**
ρ (Intracluster Correlation, ICC)	**0.1**

Output

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

Numeric Results

Solve For: Sample Size (Clusters)
Test Statistic: Likelihood Score Test (Farrington & Manning)
Groups: 1 = Vaccine, 2 = Control
Hypotheses: H0: VE ≤ VE0 vs. H1: VE > VE0

	Number of Clusters			Cluster Size			Total Sample	Event Probability			Vaccine Efficacy			
								Vaccine						
Power	Vaccine K1	Control K2	Total K	Vaccine M1	Control M2	COV	Size N	Non-Inferiority P1.0	Actual P1.1	Control P2	Non-Inferiority VE0	Actual VE1	Intracluster Correlation ρ	Alpha
0.79492	221	221	442	10.44	10.44	0.5	4614.48	0.012	0.004	0.04	0.7	0.9	0.1	0.05

PASS also calculated the number of clusters per group to be 221. Thus, the procedure is validated.