Chapter 192

# Non-Inferiority Tests for Vaccine Efficacy using the Ratio of Two Proportions

## Introduction

This module provides power analysis and sample size calculation for non-inferiority tests for vaccine efficacy (VE) using the ratio of two proportions.

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 $VE < 1$.

This routine is partially based on Blackwelder (1993). We highly recommend it.

### 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

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

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

# Technical Details

This section will first review the ratio of two proportions. It will then adapt those results to calculating sample sizes for vaccine efficacy.

## Comparing Two Proportions

Suppose you have two populations from which dichotomous (binary) responses will be recorded. The probability (or risk) of obtaining an event of interest (testing positive for a disease) in population 1 (the treatment group) is $p_1$ and in population 2 (the control group) is $p_2$. The corresponding failure proportions are given by $q_1 = 1 - p_1$ and $q_2 = 1 - p_2$.

The assumption is made that the responses from each group follow a binomial distribution. This means that the event probability $p_i$ is the same for all subjects within a population and that the responses from one subject to the next are independent of one another.

Random samples of $n_1$ and $n_2$ individuals are obtained from these two populations. The data from these samples can be displayed in a 2-by-2 contingency table as follows

| | Disease | No Disease | Total |
|---|---|---|---|
| **Vaccinees** | $x_{11}$ | $x_{12}$ | $n_1$ |
| **Controls** | $x_{21}$ | $x_{22}$ | $n_2$ |
| **Totals** | $m_1$ | $m_2$ | $N$ |

The binomial proportions are estimated from these data using the formulae

$$\hat{p}_1 = \frac{x_{11}}{n_1} \quad \text{and} \quad \hat{p}_2 = \frac{x_{21}}{n_2}$$

In this procedure, our attention will focus on the using the ratio (often called the risk ratio) to compare the two binomial proportions. The (risk) ratio $\phi = p_1/p_2$ gives the relative change in the disease risk due to the application of the treatment.

## Test Statistics

Three test statistics have been proposed for testing whether the ratio is different from a specified value. The main difference among the test statistics is in the formula used to compute the standard error used in the denominator. These tests are based on the following *z*-test

$$z_t = \frac{\hat{p}_1/\hat{p}_2 - \phi_0}{\hat{\sigma}}$$

In power calculations, the values of $\hat{p}_1$ and $\hat{p}_2$ are not known. The corresponding values of $p_{1.1}$ and $p_2$ may be reasonable substitutes.

Following is a list of the test statistics available in **PASS**. The availability of several test statistics begs the question of which test statistic one should use. The answer is simple: one should use the test statistic that will be used to analyze the data. You may choose a method because it is a standard in your industry, because it seems to have better statistical properties, or because your statistical package calculates it.

Whatever your reasons for selecting a certain test statistic, you should use the same test statistic when doing the analysis after the data have been collected.

## 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 $\phi_0$. 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 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

$$z_{MNR} = \frac{\hat{p}_1 / \hat{p}_2 - \phi_0}{\sqrt{\left(\frac{\tilde{p}_1\tilde{q}_1}{n_1} + \phi_0^2\frac{\tilde{p}_2\tilde{q}_2}{n_2}\right)\left(\frac{N}{N-1}\right)}}$$

where

$$\tilde{p}_1 = \tilde{p}_2\phi_0$$

$$\tilde{p}_2 = \frac{-B - \sqrt{B^2 - 4AC}}{2A}$$

$$A = N\phi_0$$

$$B = -[n_1\phi_0 + x_{11} + n_2 + x_{21}\phi_0]$$

$$C = m_1$$

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

## 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 $\phi_0$. 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. 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) 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

$$\tilde{\varphi} = \frac{1}{6\tilde{u}^{3/2}}\left(\frac{\tilde{q}_1(\tilde{q}_1 - \tilde{p}_1)}{n_1^2\tilde{p}_1^2} - \frac{\tilde{q}_2(\tilde{q}_2 - \tilde{p}_2)}{n_2^2\tilde{p}_2^2}\right)$$

$$\tilde{u} = \frac{\tilde{q}_1}{n_1\tilde{p}_1} + \frac{\tilde{q}_2}{n_2\tilde{p}_2}$$

## 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. 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 $p_1$ 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 Calculation

The power for a test statistic that is based on the normal approximation can be computed exactly using two binomial distributions. 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 $x_{11}$ and $x_{21}$. 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

When the values of $n_1$ and $n_2$ are large (say over 200), these formulas often take a long time to evaluate. In this case, a large sample approximation can be used. 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 study is being designed to establish the non-inferiority of vaccine A compared to vaccine B. The researchers plan to use the Gart and Nam likelihood score test to analyze the data. The event probability of the vaccine B is 0.05. They want to find the sample size required to guarantee a power of 0.8 when the non-inferiority vaccine efficacy bound is set to -0.1 and the actual vaccine efficacy is set to values of 0 0.1 0.2 0.4. Since this is a one-sided test, the significance level will be 0.025. They want to make all calculations using the 'Normal Approximation'.

## 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 Calculation Method | **Normal Approximation** |
| Test Type | **Likelihood Score (Gart & Nam)** |
| Power | **0.8** |
| Alpha | **0.025** |
| Group Allocation | **Equal (N1 = N2)** |
| 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 0.4** |
| P2 (Control Group Event Probability) | **0.05** |

# Output

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

## Numeric Reports

**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 | 0.80001 | 32854 | 32854 | 65708 | 0.05 | 0.055 | 0.050 | -0.1 | 0.0 | 0.025 |
| 0.8 | 0.80003 | 7834 | 7834 | 15668 | 0.05 | 0.055 | 0.045 | -0.1 | 0.1 | 0.025 |
| 0.8 | 0.80006 | 3312 | 3312 | 6624 | 0.05 | 0.055 | 0.040 | -0.1 | 0.2 | 0.025 |
| 0.8 | 0.80021 | 1069 | 1069 | 2138 | 0.05 | 0.055 | 0.030 | -0.1 | 0.4 | 0.025 |

* Power was computed using the normal approximation method.

| | |
|---|---|
| Target Power | The desired power value. Power is the probability of rejecting a false null hypothesis. |
| Actual Power | The calculated power obtained for the scenario on this row. Because N1 and N2 are discrete, this value is often (slightly) larger than the target power. |
| N1 and N2 | The sample sizes of the vaccinated group and the control group, respectively. |
| N | The total sample size. N = N1 + N2. |
| P2 | The event probability (attack rate) of the control group. |
| 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. |
| 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. |
| Alpha | The probability of rejecting a true null hypothesis. |

**Summary Statements**

A parallel two-group 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, two-sample Score test (Gart & Nam) with a Type I error rate (α) of 0.025. The control group proportion (event probability) is assumed to be 0.05. To detect a vaccine efficacy of 0 (or vaccine event probability of 0.05) with 80% power, the number of subjects needed will be 32854 in the vaccine group and 32854 in the control group.

**Dropout-Inflated Sample Size**

| | Sample Size | | | Dropout-Inflated Enrollment Sample Size | | | Expected Number of Dropouts | | |
|---|---|---|---|---|---|---|---|---|---|
| **Dropout Rate** | **N1** | **N2** | **N** | **N1'** | **N2'** | **N'** | **D1** | **D2** | **D** |
| 20% | 32854 | 32854 | 65708 | 41068 | 41068 | 82136 | 8214 | 8214 | 16428 |
| 20% | 7834 | 7834 | 15668 | 9793 | 9793 | 19586 | 1959 | 1959 | 3918 |
| 20% | 3312 | 3312 | 6624 | 4140 | 4140 | 8280 | 828 | 828 | 1656 |
| 20% | 1069 | 1069 | 2138 | 1337 | 1337 | 2674 | 268 | 268 | 536 |

| | |
|---|---|
| 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. |
| N1, N2, and N | The evaluable sample sizes at which power is computed. If N1 and N2 subjects are evaluated out of the N1' and N2' subjects that are enrolled in the study, the design will achieve the stated power. |
| N1', N2', and N' | The number of subjects that should be enrolled in the study in order to obtain N1, N2, and N evaluable subjects, based on the assumed dropout rate. After solving for N1 and N2, N1' and N2' are calculated by inflating N1 and N2 using the formulas N1' = N1 / (1 - DR) and N2' = N2 / (1 - DR), with N1' and N2' 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.) |
| D1, D2, and D | The expected number of dropouts. D1 = N1' - N1, D2 = N2' - N2, and D = D1 + D2. |

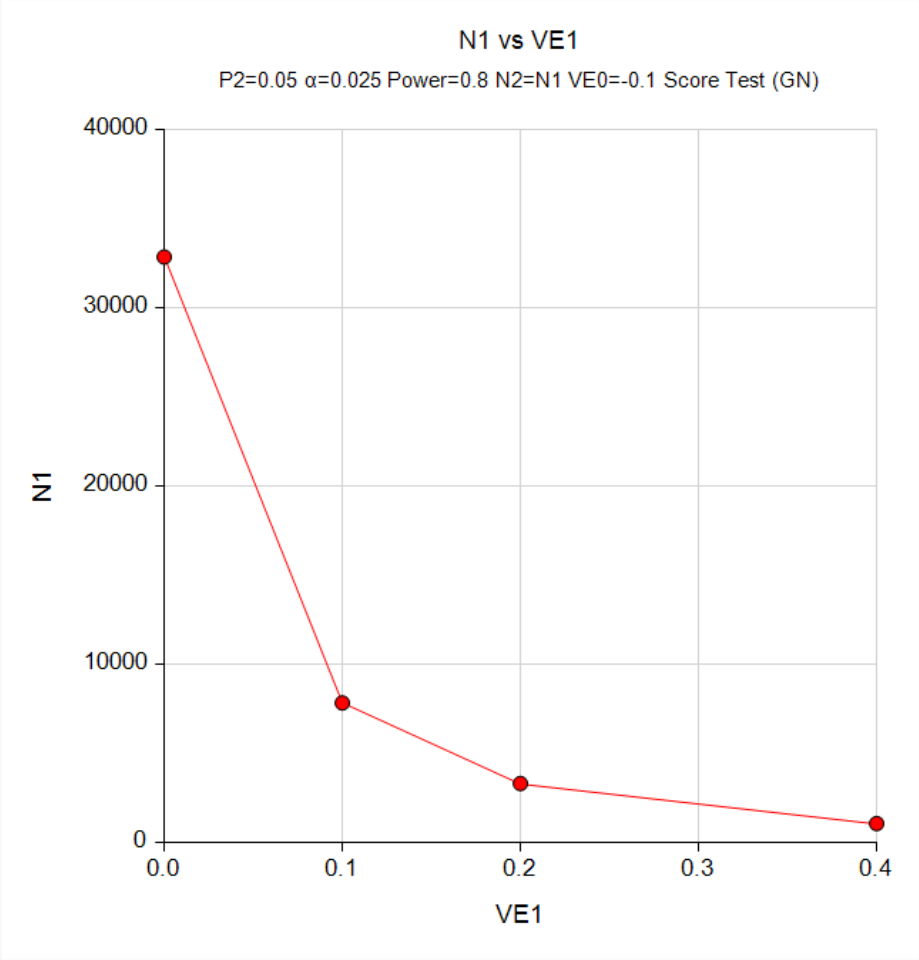
**Dropout Summary Statements**

Anticipating a 20% dropout rate, 41068 subjects should be enrolled in Group 1, and 41068 in Group 2, to obtain final group sample sizes of 32854 and 32854, respectively.

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

## Plots Section

**Plots**




The values from the table are displayed in the above chart. This chart gives a quick look at the sample sizes that are required for various values of VE1.

# Example 2 – Validation of Power Calculations using Blackwelder (1993)

We could not find a validation example in the literature for this procedure, so we will generate a validation example using the “Non-Inferiority Tests for the Ratio of Two Proportions” which theoretically gives the same results as this procedure.

In that procedure we set Solve For to “Sample Size”, Higher Proportion Are to “Worse”, Test Type to “Likelihood Score (Gart & Nam), Power to 0.8, Alpha to 0.025, Group Allocation to “Equal”, R0 to 1.1, R1 to 0.6, and P2 to 0.05. The sample size is found to be 1069 in both groups.

In this procedure we set Solve For to “Sample Size”, Power Calculation Method to “Normal Approximation”, Test Type to “Likelihood Score (Gart & Nam), Power to 0.8, Alpha to 0.025, Group Allocation to “Equal”, VE0 to -0.1, VE1 to 0.4, and P2 to 0.05.

## 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** |
| Power Calculation Method | **Normal Approximation** |
| Test Type | **Likelihood Score (Gart & Nam)** |
| Power | **0.8** |
| Alpha | **0.025** |
| Group Allocation | **Equal (N1 = N2)** |
| Vaccine Efficacy Input Type | **Enter VE0, VE1, and P2** |
| VE0 (Non-Inferiority Vaccine Efficacy) | **-0.1** |
| VE1 (Actual Vaccine Efficacy) | **0.4** |
| P2 (Control Group Event Probability) | **0.05** |

# Output

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

**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 | | | Control | Non-Inferiority | Actual | Non-Inferiority | Actual | |
| Target | Actual* | N1 | N2 | N | P2 | P1.0 | P1.1 | VE0 | VE1 | Alpha |
| 0.8 | 0.80021 | 1069 | 1069 | 2138 | 0.05 | 0.055 | 0.03 | -0.1 | 0.4 | 0.025 |

* Power was computed using the normal approximation method.

**PASS** also calculated the sample size to be 1069 in both groups. Thus, the procedure is validated.