

Chapter 193

Superiority by a Margin Tests for Vaccine Efficacy using the Ratio of Two Proportions

Introduction

This module provides power analysis and sample size calculation for superiority by a margin 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.

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

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

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.

Superiority by a Margin Tests for Vaccine Efficacy using the Ratio of Two Proportions

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) given above.

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}_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

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

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

$$A = N\phi_0$$

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

$$C = m_1$$

$$\tilde{p}_1 = \frac{\tilde{p}_2 \psi_0}{1 + \tilde{p}_2(\psi_0 - 1)}$$

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

Superiority by a Margin Tests for Vaccine Efficacy using the Ratio of Two Proportions

$$A = n_2(\psi_0 - 1)$$

$$B = n_1\psi_0 + n_2 - m_1(\psi_0 - 1)$$

$$C = -m_1$$

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.

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 superiority by a margin 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 superiority of a new vaccine over a placebo. The researchers plan to use the Gart and Nam likelihood score test to analyze the data. They want to find the sample size required to guarantee a power of 0.9 when the superiority vaccine efficacy is set to 0.4 and the actual vaccine efficacy is set to values of 0.5, 0.6, 0.7, 0.8, and 0.9. The event probability of the control group is 0.04. The significance level will be 0.025. They want to make all calculations using the ‘Normal Approximation’.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the procedure window. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Power Calculation Method.....	Normal Approximation
Test Type.....	Likelihood Score (Gart & Nam)
Power.....	0.90
Alpha.....	0.025
Group Allocation	Equal (N1 = N2)
Vaccine Efficacy Input Type	Enter VE0, VE1, and P2
VE0 (Superiority Vaccine Efficacy).....	0.4
VE1 (Actual Vaccine Efficacy)	0.5 0.6 0.7 0.8 0.9
P2 (Control Group Event Probability)	0.04

Annotated Output

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

Numeric Results

Numeric Results

Test Statistic: Gart & Nam Likelihood Score Test
 Hypotheses: H0: VE ≤ VE0 vs. H1: VE > VE0

Target Power	Actual Power*	Event Probability					Vaccine Efficacy		Alpha	
		N1	N2	N	Cntl P2	Sup Vax P1.0	Act Vax P1.1	Sup VE0		Act VE1
0.9	0.90000	22577	22577	45154	0.04	0.024	0.020	0.4	0.5	0.025
0.9	0.90000	5168	5168	10336	0.04	0.024	0.016	0.4	0.6	0.025
0.9	0.90004	2083	2083	4166	0.04	0.024	0.012	0.4	0.7	0.025
0.9	0.90018	1050	1050	2100	0.04	0.024	0.008	0.4	0.8	0.025
0.9	0.90048	593	593	1186	0.04	0.024	0.004	0.4	0.9	0.025

* Power was computed using the normal approximation method.

References
 Blackwelder, William C. 1993. 'Sample Size and Power for Prospective Analysis of Relative Risk.' *Statistics in Medicine*, Vol. 12, 691-698.
 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.

Superiority by a Margin Tests for Vaccine Efficacy using the Ratio of Two Proportions

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.

Report Definitions

Target Power is the desired power value. Power is the probability of rejecting a false null hypothesis.
 Actual Power is 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 are the sample sizes of the vaccinated group and the control group, respectively.
 N is the total sample size, N1 + N2.
 P2 is the event probability (attack rate) of the control group.
 P1.0 is the largest value of the event probability for vaccinated group that still yields a superiority conclusion.
 P1.1 is the value of the event probability for vaccinated group that is assumed by the alternative hypothesis, H1.
 VE0 is the vaccine efficacy assumed by the null hypothesis, H0. This is the lower superiority boundary of VE. $VE0 = 1 - P1.0/P2$.
 VE1 is 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 is the probability of rejecting a true null hypothesis.

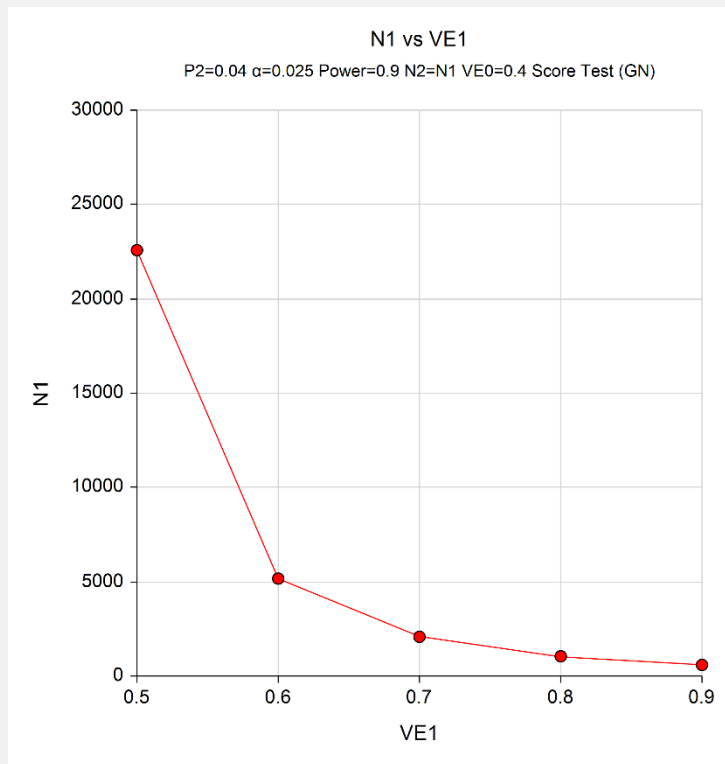
Summary Statements

Sample sizes of 22577 in the vaccine group and 22577 in the control group achieve 90% power to detect a vaccine efficacy of 0.5 when the superiority vaccine efficacy boundary is 0.4. The control group event probability is 0.04. The vaccine group event probability at the superiority limit is 0.024. The power was computed for the case when the actual vaccine group event probability is 0.02. The test statistic used is the one-sided Score test (Gart & Nam). The significance level of the test is 0.025.

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

Chart Section

Chart Section



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)

Blackwelder (1993), page 694, presents an example in which the significance level is 0.05, power is 0.8, P2 is 0.04, VE0 is 0.7, and VE1 is 0.9. The Miettinen and Nurminen likelihood score test is used. The calculations are based on the normal approximation to the binomial. His result is 2119. Notice that with this result N1 cannot be equal to N2 since the value of N is odd. **PASS** will find $N = 2120$.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the procedure window. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size
Power Calculation Method.....	Normal Approximation
Test Type.....	Likelihood Score (Miet. & Nurm.)
Power.....	0.8
Alpha.....	0.05
Group Allocation	Equal (N1 = N2)
Vaccine Efficacy Input Type	Enter VE0, VE1, and P2
VE0 (Superiority Vaccine Efficacy).....	0.7
VE1 (Actual Vaccine Efficacy)	0.9
P2 (Control Group Event Probability)	0.04

Output

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

Numeric Results

Numeric Results											
Test Statistic: Miettinen & Nurminen Likelihood Score Test											
Hypotheses: H0: $VE \leq VE0$ vs. H1: $VE > VE0$											
Target Power	Actual Power*	N1	N2	N	Event Probability				Vaccine Efficacy		
					Cntl P2	Sup Vax P1.0	Act Vax P1.1	Sup VE0	Act VE1	Alpha	
0.8	0.80004	1060	1060	2120	0.04	0.012	0.004	0.7	0.9	0.05	

* Power was computed using the normal approximation method.

PASS calculated the sample size to be 2120. This value is one different from Blackwelder's 2119. The difference occurs because **PASS** requires $N1 = N2$. Thus, the procedure is validated.