Chapter 219

Confidence Intervals for Vaccine Efficacy using a Cohort Design

Introduction

This routine calculates the group sample sizes necessary to achieve a specified confidence interval width of vaccine efficacy (VE) from data collected using a cohort design.

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.

This routine is partially based on O'Neill (1988). This paper provides a useful overview of the vaccine efficacy studies, the assumptions that are made, and discussion of when this design is useful. We highly recommend it.

Technical Details

This section will first review the details of calculating sample sizes for 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.

Confidence Intervals for the Ratio (Relative Risk)

Many methods have been devised for computing confidence intervals for the ratio of two proportions $\phi = p_1/p_2$. Six of these methods are available in this procedure. They are

1. Score (Farrington and Manning)
2. Score (Miettinen and Nurminen)
3. Score with Correction for Skewness (Gart and Nam)
4. Logarithm (Katz)
5. Logarithm + 1/2 (Walter)
6. Fleiss

Farrington and Manning's Score

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.

Here is the formula for computing the test

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

where

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

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

$$A = N\phi_0$$

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

$$C = m_1$$

as in the test of Miettinen and Nurminen (1985).

Farrington and Manning (1990) proposed inverting their score test to find the confidence interval. The lower limit is found by solving

$$z_{FMR} = |z_{\alpha/2}|$$

and the upper limit is the solution of

$$z_{FMR} = -|z_{\alpha/2}|$$

Miettinen and Nurminen's Score

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.

Here is the formula for computing the test

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

where

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

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

$$A = N\phi_0$$

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

$$C = m_1$$

Miettinen and Nurminen (1985) proposed inverting their score test to find the confidence interval. The lower limit is found by solving

$$z_{MNR} = |z_{\alpha/2}|$$

and the upper limit is the solution of

$$z_{MNR} = -|z_{\alpha/2}|$$

Gart and Nam's Score

Gart and Nam (1988) page 329 proposed a modification to the Farrington and Manning (1988) ratio test that corrected for skewness. Let $z_{FM}(\phi)$ stand for the Farrington and Manning ratio test statistic described above. The skewness corrected test statistic z_{GN} is the appropriate solution to the quadratic equation

$$(-\tilde{\varphi})z_{GNR}^2 + (-1)z_{GNR} + (z_{FMR}(\phi) + \tilde{\varphi}) = 0$$

where

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

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

Gart and Nam (1988) proposed inverting their score test to find the confidence interval. The lower limit is found by solving

$$z_{GNR} = |z_{\alpha/2}|$$

and the upper limit is the solution of

$$z_{GNR} = -|z_{\alpha/2}|$$

Logarithm (Katz)

This was one of the first methods proposed for computing confidence intervals for risk ratios.

For details, see Gart and Nam (1988), page 324.

$$L = \hat{\phi}\exp\left(-z\sqrt{\frac{\hat{q}_1}{n\hat{p}_1}+\frac{\hat{q}_2}{n\hat{p}_2}}\right)$$

$$U = \hat{\phi}\exp\left(z\sqrt{\frac{\hat{q}_1}{n\hat{p}_1}+\frac{\hat{q}_2}{n\hat{p}_2}}\right)$$

where

$$\hat{\phi} = \frac{\hat{p}_1}{\hat{p}_2}$$

Logarithm (Walters)

For details, see Gart and Nam (1988), page 324.

$$L = \hat{\phi}\exp\left(-z\sqrt{\hat{u}}\right)$$

$$U = \hat{\phi}\exp\left(z\sqrt{\hat{u}}\right)$$

where

$$\hat{\phi} = \exp\left(\ln\left(\frac{a+\frac{1}{2}}{m+\frac{1}{2}}\right)-\ln\left(\frac{b+\frac{1}{2}}{n+\frac{1}{2}}\right)\right)$$

$$\hat{u} = \frac{1}{a+\frac{1}{2}}-\frac{1}{m+\frac{1}{2}}+\frac{1}{b+\frac{1}{2}}-\frac{1}{n+\frac{1}{2}}$$

$$V = \left(\phi^2\left(\frac{\tilde{q}_1}{m\tilde{p}_1}+\frac{\tilde{q}_2}{n\tilde{p}_2}\right)\right)^{-1}$$

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

$$\tilde{q}_1 = 1-\tilde{p}_1$$

$$\tilde{q}_2 = 1-\tilde{p}_2$$

$$\tilde{\mu}_3 = v^{3/2}\left(\frac{\tilde{q}_1(\tilde{q}_1 - \tilde{p}_1)}{(m\tilde{p}_1)^2} - \frac{\tilde{q}_2(\tilde{q}_2 - \tilde{p}_2)}{(n\tilde{p}_2)^2}\right)$$

$$v = \left(\frac{\tilde{q}_1}{m\tilde{p}_1} + \frac{\tilde{q}_2}{n\tilde{p}_2}\right)^{-1}$$

Iterated Method of Fleiss

Fleiss (1981) presents an improved confidence interval for the odds ratio and relative risk. This method forms the confidence interval as all those value of the odds ratio which would not be rejected by a chi-square hypothesis test. Fleiss gives the following details about how to construct this confidence interval. To compute the lower limit, do the following.

1. For a trial value of ψ, compute the quantities *X, Y, W, F, U,* and *V* using the formulas

$$X = \psi(m + s) + (n - s)$$

$$Y = \sqrt{X^2 - 4ms\psi(\psi - 1)}$$

$$A = \frac{X - Y}{2(\psi - 1)}$$

$$B = s - A$$

$$C = m - A$$

$$D = f - m + A$$

$$W = \frac{1}{A} + \frac{1}{B} + \frac{1}{C} + \frac{1}{D}$$

$$F = \left(a - A - \frac{1}{2}\right)^2 W - z_{\alpha/2}^2$$

$$T = \frac{1}{2(\psi - 1)^2}\left(Y - n - \frac{\psi - 1}{Y}[X(m + s) - 2ms(2\psi - 1)]\right)$$

$$U = \frac{1}{B^2} + \frac{1}{C^2} - \frac{1}{A^2} - \frac{1}{D^2}$$

$$V = T\left[\left(a - A - \frac{1}{2}\right)^2 U - 2W\left(a - A - \frac{1}{2}\right)\right]$$

Finally, use the updating equation below to calculate a new value for the odds ratio using the updating equation

$$\psi^{(k+1)} = \psi^{(k)} - \frac{F}{V}$$

2. Continue iterating until the value of *F* is arbitrarily close to zero.

The upper limit is found by substituting $+\frac{1}{2}$ for $-\frac{1}{2}$ in the formulas for *F* and *V*.

Confidence limits for the *relative risk* can be calculated using the expected counts *A, B, C,* and *D* from the last iteration of the above procedure. The lower limit of the relative risk

$$\phi_{lower} = \frac{A_{lower} n}{B_{lower} m}$$

$$\phi_{upper} = \frac{A_{upper} n}{B_{upper} m}$$

Sample Size Comparison

It is instructive to see the impact of the choice of calculation method on computed sample size. In the following table, we present the necessary sample size for the validation example for each method.

Method	N1 + N2	Increase Over Minimum
Score (Farrington & Manning)	27,686	280
Score (Miettinen & Nurminen)	27,688	282
Score w/Skewness (Gart & Nam)	27,406	0
Logarithm (Katz)	28,448	1,042
Logarithm + 1/2 (Walter)	29,010	1,604
Fleiss	31,488	4,082

From the table, we note that the three score intervals have nearly identical sample size requirements. We also note that the method originally chosen by O'Neill (1988), which was the logarithmic algorithm of Katz, requires 1,042 more subjects. The method by Fleiss, which is based on the odds ratio, requires an additional 4082 subjects to maintain the precision requirement.

Luckily, the method proposed by Gart and Nam is the method that is usually recommended today (in 2020).

Sample Size Estimation

Sample size estimation is relatively straight forward. For each method, anticipated (planned) values of the two proportions are substituted for the estimated values given by the above formulas, along with the sample sizes of each of the groups and the confidence level. The width between the two limits becomes the measure of precision. The narrower the confidence interval, the more precise it is.

To find an appropriate sample size for a given set of parameters, a binary search is conducted for the smallest sample size that meets the width requirements.

Confidence Level

The confidence level, 1 – α, has the following interpretation. If thousands of random samples of size n_1 and n_2 are drawn from populations 1 and 2, respectively, and a confidence interval for the true difference/ratio/odds ratio of proportions is calculated for each pair of samples, the proportion of those intervals that will include the true difference/ratio/odds ratio of proportions is 1 – α.

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, the confidence interval methods described above can be adapted for computing sample sizes for vaccine efficacy studies. O'Neill (1988) gives the details using the logarithmic transformation method of Katz (1979).

Since O'Neill (1988), more accurate methods for the computing the confidence interval of the risk ratio have been suggested. Of these, the skewness corrected interval published in Gart and Nam (1988) is often recommended. Luckily, the method of Gart and Nam often yields a significant reduction in the necessary sample size.

Relative Width

Because of the assumption that $p_1 < p_2$, VE is bounded above by one. Although negative values are possible, *VE* is usually restricted to be greater than zero for planning purposes.

Some authors prefer to use the following relative width (*RW*) in planning. The formula for *RW* is

$$RW = \frac{Width}{VE}$$

If *RW* is specified, the above relationship can be used to find the corresponding value of the *width*.

Example 1 – Calculating Sample Size

Suppose a study is planned in which the researcher wishes to construct a two-sided 95% confidence interval for vaccine efficacy (VE) such that the width of the interval is no wider than 0.2. Additional widths of 0.15 and 0.25 are also to be investigated. The confidence interval method to be used is the Score w/Skewness (Gart & Nam) method. The confidence level is set to 0.95. VE is set at 0.7 and 0.8. P2 is 0.06.

The goal is to determine the necessary sample size.

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**
Method	**Score w/Skewness (Gart & Nam)**
Interval Type	**Two-Sided**
Confidence Level (1 - α)	**0.95**
Group Allocation	**Equal (N1 = N2)**
Precision Input Type	**Absolute**
W (Confidence Interval Width)	**0.15 0.2 0.25**
Vaccine Efficacy Input Type	**Enter VE and P2**
VE (Vaccine Efficacy)	**0.7 0.8**
P2 (Control Group Event Probability)	**0.06**

Output

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

Numeric Reports

Numeric Results

Solve For: Sample Size
Interval Type: Two-Sided
Confidence Interval Method: Score with Correction for Skewness (Gart and Nam)

Confidence Level 1 - α	Sample Size: Vaccine N1	Control N2	Total N	Confidence Interval Width: Target W_T	Actual W_A	Relative RW	Event Probability: Vaccine P1	Control P2	Vaccine Efficacy VE	Confidence Interval Limits of VE: Lower LCL	Upper UCL
0.95	4379	4379	8758	0.15	0.14999	0.21427	0.018	0.06	0.7	0.61705	0.76704
0.95	2490	2490	4980	0.20	0.19998	0.28569	0.018	0.06	0.7	0.58599	0.78597
0.95	1616	1616	3232	0.25	0.24992	0.35703	0.018	0.06	0.7	0.55336	0.80328
0.95	2752	2752	5504	0.15	0.14998	0.18748	0.012	0.06	0.8	0.71363	0.86361
0.95	1580	1580	3160	0.20	0.19995	0.24994	0.012	0.06	0.8	0.68012	0.88007
0.95	1037	1037	2074	0.25	0.24988	0.31235	0.012	0.06	0.8	0.64458	0.89446

1 - α The Confidence Level. The proportion of confidence intervals (constructed with this same confidence level, sample size, etc.) that would contain the true value of VE.
N1 The number of subjects sampled from the vaccinated population.
N2 The number of subjects sampled from the control population.
N The total sample size. N = N1 + N2.
W_T The target width of the confidence interval of VE.
W_A The actual width of the confidence interval of VE that was computed by the procedure.
RW The relative width of the confidence interval. RW = Width / VE.
P1 The probability of an event (attack rate) for each member of the vaccine group during the fixed duration of the study.
P2 The probability of an event (attack rate) for each member of the control group during the fixed duration of the study.
VE The index of vaccine efficacy. It represents the proportion of cases of disease prevented by the vaccine. It is calculated using VE = 1 - P1 / P2.
LCL The lower confidence interval limit of VE.
UCL The upper confidence interval limit of VE.

Summary Statements

A two-group cohort design will be used to obtain a two-sided 95% confidence interval for the vaccine efficacy (1 - P1 / P2). The vaccine group event probability (attack rate) is assumed to be 0.018 and the control group event probability (attack rate) is assumed to be 0.06, corresponding to a vaccine efficacy of 0.7. The Score (Gart, Nam) method will be used to compute the confidence interval limits. To produce a confidence interval width of 0.15, the number of subjects needed will be 4379 in the vaccine group and 4379 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%	4379	4379	8758	5474	5474	10948	1095	1095	2190
20%	2490	2490	4980	3113	3113	6226	623	623	1246
20%	1616	1616	3232	2020	2020	4040	404	404	808
20%	2752	2752	5504	3440	3440	6880	688	688	1376
20%	1580	1580	3160	1975	1975	3950	395	395	790
20%	1037	1037	2074	1297	1297	2594	260	260	520

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, 5474 subjects should be enrolled in Group 1, and 5474 in Group 2, to obtain final group sample sizes of 4379 and 4379, respectively.

References

O'Neill, Robert T. 1988. 'On Sample Sizes to Estimate the Protective Efficacy of a Vaccine'. Statistics in Medicine, Volume 7, Pages 1279-1288.

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.

Fleiss, J. L. 1981. Statistical Methods for Rates and Proportions. John Wiley & Sons. New York.

Gart, John J. and Nam, Jun-mo. 1988. 'Approximate Interval Estimation of the Ratio of Binomial Parameters: A Review and Corrections for Skewness.' Biometrics, Volume 44, 323-338.

Koopman, P. A. R. 1984. 'Confidence Intervals for the Ratio of Two Binomial Proportions.' Biometrics, Volume 40, Issue 2, 513-517.

Katz, D., Baptista, J., Azen, S. P., and Pike, M. C. 1978. 'Obtaining Confidence Intervals for the Risk Ratio in Cohort Studies.' Biometrics, Volume 34, 469-474.

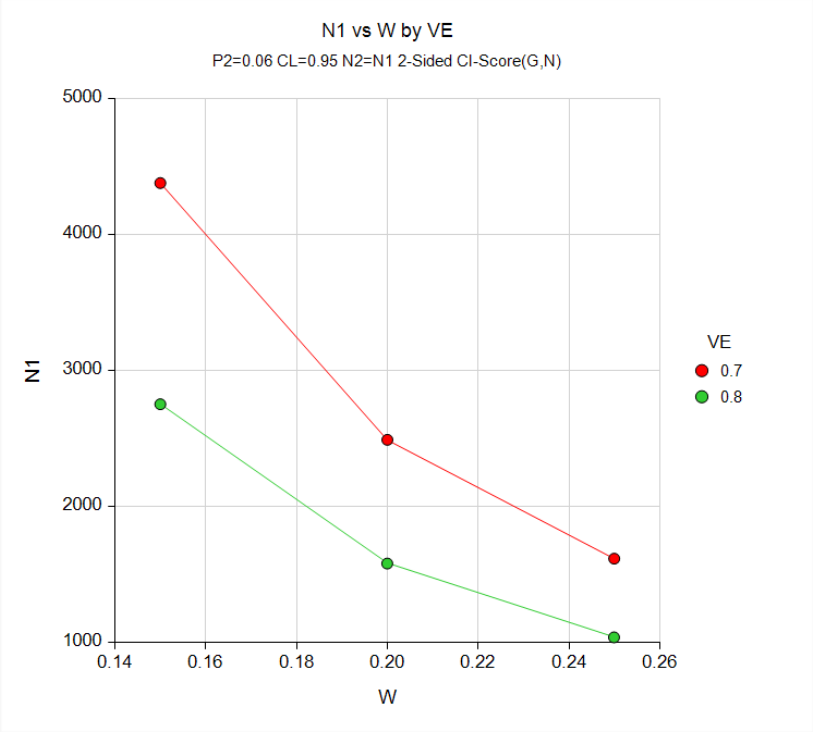
Miettinen, O.S. and Nurminen, M. 1985. 'Comparative analysis of two rates.' Statistics in Medicine 4: 213-226.

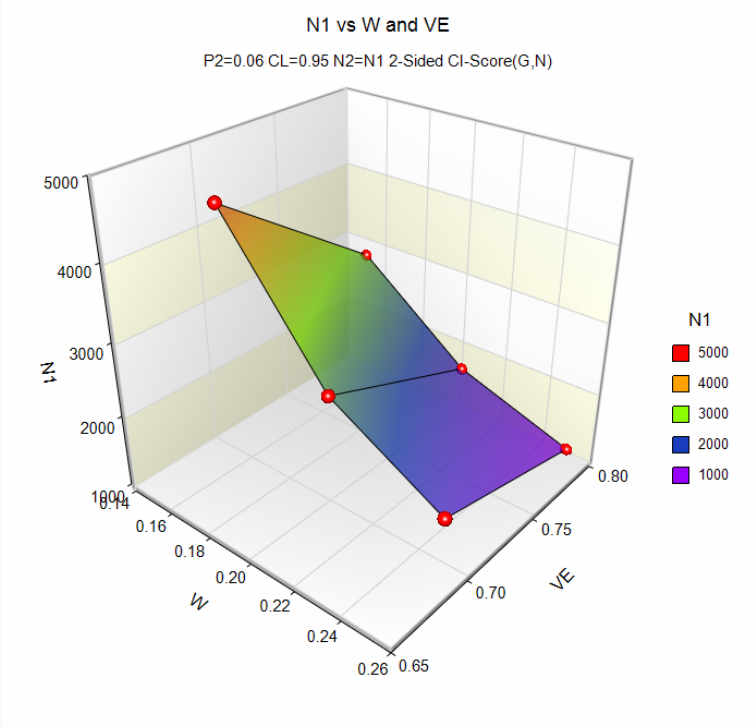
Walter, S. D. 1976. 'The Distribution of Levin's Measure of Attributable Risk.' Biometrika, Volume 62, 371-375.

This report shows the calculated sample sizes for each of the scenarios.

Plots Section

Plots





These plots show the group sample size versus the confidence interval width for the two VE values.

Example 2 – Validation using O'Neill (1988)

O'Neill (1988) page 1284 gives an example of a sample size calculation for a confidence interval for VE when the confidence level is 95%, P1 is 0.001, P2 is 0.005, and the desired width is 0.24. The confidence interval method is Logarithm (Katz). The sample size in each group is 14224.

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**
Method	**Logarithm (Katz)**
Interval Type	**Two-Sided**
Confidence Level (1 - α)	**0.95**
Group Allocation	**Equal (N1 = N2)**
Precision Input Type	**Absolute**
W (Confidence Interval Width)	**0.24**
Vaccine Efficacy Input Type	**Enter P1 and P2**
P1 (Vaccine Group Event Probability)	**0.001**
P2 (Control Group Event Probability)	**0.005**

Output

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

Numeric Results

Solve For:	Sample Size
Interval Type:	Two-Sided
Confidence Interval Method:	Logarithm (Katz)

Confidence Level 1 - α	Sample Size			Confidence Interval Width			Event Probability		Vaccine Efficacy VE	Confidence Interval Limits of VE	
	Vaccine N1	Control N2	Total N	Target W_T	Actual W_A	Relative RW	Vaccine P1	Control P2		Lower LCL	Upper UCL
0.95	14224	14224	28448	0.24	0.23999	0.29999	0.001	0.005	0.8	0.64677	0.88676

PASS has also computed a group sample size of 14224, so the procedure is validated.