Chapter 438

Tests for the Ratio of Two Negative Binomial Rates

Introduction

Count data arise from counting the number of events of a particular type that occur during a specified time interval. Examples include the number of accidents at an intersection during a year, the number of calls to a call center during an hour, or the number of meteors seen in the evening sky during the night. Clinical trial often result in count data, with examples including the number of patients who are admitted to the hospital or the number of patients who respond favorably (or unfavorably) to a particular treatment.

Traditionally, the Poisson distribution (e.g., Poisson regression) has been used to model count data. The Poisson model assumes that the mean and variance are equal, but in many clinical trials the variance is observed to be greater than the mean in a condition called *overdispersion*. When overdispersion occurs, the Poisson model provides a poor fit to the data. As an alternative, the negative binomial model is increasingly being used to model overdispersed count data. While the Poisson distribution is characterized by a single parameter which represents both the mean and the variance, the negative binomial distribution includes two parameters, allowing for greater flexibility in modeling the mean-variance relationship that is observed in overdispersed, heterogeneous count data.

This procedure is based on the formulas and results outlined in Zhu and Lakkis (2014) and calculates the power and sample size for testing whether the ratio of two negative binomial event rates is different from one. The test is often performed using the Wald (or likelihood ratio) test statistic in the context of generalized linear models. Such an analysis is available within SAS Proc GENMOD. These asymptotic tests are appropriate when the sample size is greater than 50 per group. When the sample size is less than 50 per group, the results from this procedure can be used to obtain a rough estimate of the power (see Zhu and Lakkis (2014), page 381).

Technical Details

The Negative Binomial Model

As in Zhu and Lakkis (2014), define y_{ij} as the number of events during time t_{ij} for subject *i* (*i* = 1 to n_j) in group *j* (*j* = 1, 2). Usually, group 1 is considered the control or reference group and group 2 is considered the treatment group. If y_{ij} follows a negative binomial distribution with mean μ_{ij} and dispersion parameter κ, the probability function for y_{ij} is

$$P(y_{ij}) = \frac{\Gamma(\kappa^{-1} + y_{ij})}{\Gamma(\kappa^{-1})y_{ij}!}\left(\frac{\kappa\mu_{ij}}{1+\kappa\mu_{ij}}\right)^{y_{ij}}\left(\frac{1}{1+\kappa\mu_{ij}}\right)^{1/\kappa}$$

where $\Gamma(.)$ is the gamma function. Using negative binomial regression, we can model μ_{ij} as

$$\log(\mu_{ij}) = \log(t_{ij}) + \beta_0 + \beta_1 x_{ij}$$

such that

$$\log\left(\frac{\mu_{ij}}{t_{ij}}\right) = \beta_0 + \beta_1 x_{ij}$$

where $x_{ij} = 0$ if the i^{th} subject is in group 1 and $x_{ij} = 1$ if the i^{th} subject is in group 2.

Further define λ_1 and λ_2 as the mean event rates per time unit for groups 1 and 2, respectively, and $RR = \lambda_2/\lambda_1$ as the ratio of event rates. Using the negative binomial model, it follows then that

$$\lambda_1 = e^{\beta_0}$$

$$\lambda_2 = e^{\beta_0+\beta_1}$$

$$RR = \frac{\lambda_2}{\lambda_1} = e^{\beta_1}$$

If we define $\hat{\beta}_1$ as the asymptotic maximum likelihood estimate of β_1, then the variance of $\hat{\beta}_1$ can be written as

$$\mathrm{Var}(\hat{\beta}_1) = \frac{1}{n_1}\left[\frac{1}{\mu_t}\left(\frac{1}{\lambda_1} + \frac{1}{R\lambda_2}\right) + \frac{(1+R)\kappa}{R}\right]$$

where n_1 and n_2 are the sample sizes and λ_1 and λ_2 are the event rates from groups 1 and 2, respectively, $R = n_2/n_1$ is the sample allocation ratio, κ is the negative binomial dispersion parameter (assumed to be constant for power calculations), and μ_t is the average exposure time across all subjects (i.e. t_{ij} = μ_t for all *i*, *j*).

Hypothesis Test

The two-sided null and alternative hypotheses for testing equality of the two event rates can be written as

$$H_0: \beta_1 = 0 \quad \text{vs.} \quad H_A: \beta_1 \neq 0$$

or equivalently in terms of $RR = \lambda_2 / \lambda_1$ as

$$H_0: RR = 1 \quad \text{vs.} \quad H_A: RR \neq 1$$

The upper and lower one-sided tests, respectively, are

$$H_0: \beta_1 \leq 1 \quad \text{vs.} \quad H_A: \beta_1 > 1$$

$$H_0: \beta_1 \geq 1 \quad \text{vs.} \quad H_A: \beta_1 < 1$$

or equivalently in terms of $RR = \lambda_2 / \lambda_1$ as

$$H_0: RR \leq 1 \quad \text{vs.} \quad H_A: RR > 1$$

$$H_0: RR \geq 1 \quad \text{vs.} \quad H_A: RR < 1$$

These hypotheses are most commonly tested using the Wald test statistic within generalized linear models. The likelihood ratio test statistic is also used. Such an analysis can be performed for the negative binomial distribution using SAS Proc GENMOD with a logarithmic link function and an indicator variable for group (1 or 2) as the single independent variable. For more information see Zhu and Lakkis (2014) or the SAS help manual.

Estimating the Variance under the Null and Alternative Hypotheses

Asymptotically, the variance of $\hat{\beta}_1$ is

$$\text{Var}(\hat{\beta}_1) = \frac{1}{n_1}\left[\frac{1}{\mu_t}\left(\frac{1}{\lambda_1} + \frac{1}{R\lambda_2}\right) + \frac{(1+R)\kappa}{R}\right]$$

If we define V_A under the alternative hypothesis using the true rates λ_1 and λ_2 as

$$V_A = \frac{1}{\mu_t}\left(\frac{1}{\lambda_1} + \frac{1}{R\lambda_2}\right) + \frac{(1+R)\kappa}{R}$$

then the variance of $\hat{\beta}_1$ under the alternative hypothesis can be written as

$$\text{Var}_A(\hat{\beta}_1) = \frac{1}{n_1}\left[\frac{1}{\mu_t}\left(\frac{1}{\lambda_1} + \frac{1}{R\lambda_2}\right) + \frac{(1+R)\kappa}{R}\right]$$

$$= \frac{1}{n_1}[V_A]$$

If we define V_0 using the rates $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ estimated under the null hypothesis as

$$V_0 = \frac{1}{\mu_t}\left(\frac{1}{\tilde{\lambda}_1} + \frac{1}{R\tilde{\lambda}_2}\right) + \frac{(1+R)\kappa}{R}$$

then the variance of $\hat{\beta}_1$ under the null hypothesis is

$$\mathrm{Var}_0(\hat{\beta}_1) = \frac{1}{n_1}\left[\frac{1}{\mu_t}\left(\frac{1}{\tilde{\lambda}_1} + \frac{1}{R\tilde{\lambda}_2}\right) + \frac{(1+R)\kappa}{R}\right]$$

$$= \frac{1}{n_1}[V_0]$$

The portion V_0 (and therefore $\mathrm{Var}_0(\hat{\beta}_1)$) can be estimated in three different ways. Define $V_{0|M}$ as the estimate of V_0 given the chosen method M. It follows then that

$$\mathrm{Var}_{0|M}(\hat{\beta}_1) = \frac{1}{n_1}[V_{0|M}]$$

where $M = 1, 2, 3$ indicates the method used to estimate V_0.

The three methods for estimating V_0 are as follows:

- **Method 1: Use the Event Rate from Group 1 (λ_1)**
  Under H_0: $RR = 1$, the event rates are equal (i.e., $\lambda_2 = \lambda_1$), and the null variance is estimated using

  $$V_{0|1} = \frac{1+R}{\mu_t R\lambda_1} + \frac{(1+R)\kappa}{R}$$

- **Method 2: Use the True Event Rates (λ_1 and λ_2)**
  The true rates λ_1 and λ_2 are used, and the null variance is estimated using

  $$V_{0|2} = \frac{1}{\mu_t}\left(\frac{1}{\lambda_1} + \frac{1}{R\lambda_2}\right) + \frac{(1+R)\kappa}{R}$$

  This is equivalent to the estimation of V_A under the alternative hypothesis, i.e., $V_{0|2} = V_A$.

- **Method 3: Use Maximum Likelihood Estimation**
  The maximum likelihood estimate of λ under H_0: $RR = 1$, a weighted average of λ_1 and λ_2, is used, and the null variance is estimated using

  $$V_{0|3} = \frac{(1+R)^2}{\mu_t R(\lambda_1 + R\lambda_2)} + \frac{(1+R)\kappa}{R}$$

Simulation studies suggest that sample sizes calculated using methods 2 and 3 are more accurate than those calculated using method 1 (see Zhu and Lakkis (2014), page 385).

Computing Sample Size

From Zhu and Lakkis (2014), page 378, the sample size in group 1 required to achieve power of $1-\beta$ for the two-sided Wald or likelihood ratio test at significance level α can be calculated using calculation method M for V_0 as

$$n_1 \geq \frac{\left(z_{\alpha/2}\sqrt{V_{0|M}} + z_\beta\sqrt{V_A}\right)^2}{(\log(RR))^2}$$

with RR, V_A, and $V_{0|M}$ as defined earlier. The power of the one-sided Wald or likelihood ratio test at significance level α using calculation method M for V_0 is

$$n_1 \geq \frac{\left(z_{\alpha}\sqrt{V_{0|M}} + z_\beta\sqrt{V_A}\right)^2}{(\log(RR))^2}$$

Computing Power

From Zhu and Lakkis (2014), page 379, the power of the two-sided Wald or likelihood ratio test can be calculated at significance level α using calculation method M for V_0 as

$$\text{Power}_{2-\text{sided}} = 1-\beta = \Phi\left(\frac{\sqrt{n_1}|\log(RR)| - z_{\alpha/2}\sqrt{V_{0|M}}}{\sqrt{V_A}}\right)$$

with RR, V_A, and $V_{0|M}$ as defined earlier. The power of the one-sided Wald or likelihood ratio test at significance level α using calculation method M for V_0 is

$$\text{Power}_{1-\text{sided}} = 1-\beta = \Phi\left(\frac{\sqrt{n_1}|\log(RR)| - z_{\alpha}\sqrt{V_{0|M}}}{\sqrt{V_A}}\right)$$

The power calculations are accurate for the Wald and likelihood ratio tests when the group sample sizes are greater than 50. When the sample size is less than 50 per group the validity of the Wald and likelihood ratio tests is questionable, and these formulas should be used only to obtain a rough estimate of the power (see Zhu and Lakkis (2014), page 381).

Estimating the Negative Binomial Dispersion Parameter, κ, from a Previous Study that was Analyzed using Poisson Regression

Admittedly, the hardest value to determine among those required for these sample size and power calculations is the value for κ, the negative binomial dispersion parameter. If a suitable value for κ is not known, then you can estimate κ from a similar study that was analyzed using Poisson regression.

Given a Poisson mean event rate λ and an overdispersion factor ϕ, estimated from a similar Poisson regression study, the relationship between λ, ϕ, and κ is

$$\phi = 1 + \kappa\lambda$$

such that an estimate of κ can be calculated using overall estimates of λ and ϕ as

$$\hat{\kappa} = \frac{\hat{\phi} - 1}{\hat{\lambda}}$$

If $\hat{\phi}$ (the estimate for the overdispersion factor) is not directly reported in the previous study, Zhu and Lakkis (2014) suggest on page 384 that the Poisson overdispersion factor can be estimated from the total number of events, Y, the total exposure time, T, and the standard error, $\mathrm{SE}\left(\log\hat{\lambda}\right)$, as

$$\hat{\phi} = \frac{\mathrm{Var}(Y)}{\hat{\lambda}T} \approx T\hat{\lambda}\left(\mathrm{SE}\left(\log\hat{\lambda}\right)\right)^2$$

since for overdispersed Poisson

$$\mathrm{Var}(Y) = \phi\lambda T$$

The maximum likelihood estimator of the event rate of the Poisson distribution, λ, is

$$\hat{\lambda} = \frac{Y}{T}$$

The standard error, $\mathrm{SE}\left(\log\hat{\lambda}\right)$, can be determined by back-calculating from reported $100(1-\alpha)\%$ confidence interval endpoints for λ from Poisson regression as

$$\mathrm{SE}\left(\log\hat{\lambda}\right) = \frac{\left(\dfrac{\log(\text{Upper Bound}) - \log(\text{Lower Bound})}{2}\right)}{z_{1-\alpha/2}}$$

Finally, from a previous two-group Poisson regression study, κ can be calculated from the estimated Poisson event rates $\hat{\lambda}_1$ and $\hat{\lambda}_2$ and the estimated Poisson overdispersion factors $\hat{\phi}_1$ and $\hat{\phi}_2$ from each group as

$$\hat{\kappa} = \frac{\left(\dfrac{\hat{\phi}_1 + \hat{\phi}_2}{2}\right) - 1}{\left(\dfrac{\hat{\lambda}_1 + \hat{\lambda}_2}{2}\right)}$$

Example 1 – Finding the Sample Size (Validation 1 using Zhu and Lakkis (2014))

Zhu and Lakkis (2014) conducted numerous simulation studies to investigate the performance of the three null variance calculation methods. We'll use their example to demonstrate how to calculate sample size for various effect sizes, group 1 event rates, event rate ratios, and negative binomial dispersion values.

The settings for this example are similar to those that might be encountered when studying exacerbation events in COPD studies. A balanced, parallel study is designed to detect a 10-20% rate reduction from the control to the treatment group when the reference rate is about one exacerbation event per patient per year. They studied control rates of 0.8, 1.0, 1.2, and 1.4 events per patient-year and a rate ratio of 0.85, representing a 15% reduction in the treatment event rate relative to the control. They also included a rate ratio of 1.15 to represent a 15% increase in the treatment rate relative to the control for completeness. They assumed an average subject exposure time of 0.75 years since similar studies usually see discontinued participation by some patients for various reasons, resulting in an average exposure time that is less than the usual designed study length of 1 year. They studied dispersion parameter values of 0.4, 0.7, 1.0, and 1.5. They calculate the sample sizes required to achieve 80% power at a significance level of 0.05. They investigated all three null variance calculation methods, so this example will be presented in 3 parts, one for each method.

Their sample size calculation results are given in Table I on page 381. By running this example, you'll see that **PASS** matches their sample size calculation results exactly for all parameter combinations and for all three variance calculation methods.

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 1a** 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**
Alternative Hypothesis .... **Two-Sided**
Null Variance Calculation Method .... **Use Maximum Likelihood Estimation**
Power.. **0.8**
Alpha.. **0.05**
μ(t) (Average Subject Exposure Time)........... **0.75**
Group Allocation .... **Equal (N1 = N2)**
λ1 (Event Rate of Group 1)............................ **0.8 1.0 1.2 1.4**
Enter λ2 or Ratio for Group 2........................ **RR (Ratio of Event Rates)**
RR (Ratio of Event Rates) **0.85 1.15**
κ (Negative Binomial Dispersion)................... **0.4 0.7 1.0 1.5**

Output

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

Numeric Reports

Numeric Results

Solve For: Sample Size
Alternative Hypothesis: Two-Sided (H0: RR = 1 vs. Ha: RR ≠ 1)
Null Variance Calculation Method: Use Maximum Likelihood Estimation

	Sample Size			Average Exposure Time	Average Event Rate		Event Rate Ratio	Negative Binomial Dispersion	
Power	N1	N2	N	μ(t)	λ1	λ2	RR	κ	Alpha
0.80008	1311	1311	2622	0.75	0.8	0.68	0.85	0.4	0.05
0.80025	1490	1490	2980	0.75	0.8	0.68	0.85	0.7	0.05
0.80016	1668	1668	3336	0.75	0.8	0.68	0.85	1.0	0.05
0.80010	1965	1965	3930	0.75	0.8	0.68	0.85	1.5	0.05
0.80019	1570	1570	3140	0.75	0.8	0.92	1.15	0.4	0.05
0.80015	1811	1811	3622	0.75	0.8	0.92	1.15	0.7	0.05
0.80011	2052	2052	4104	0.75	0.8	0.92	1.15	1.0	0.05
0.80012	2454	2454	4908	0.75	0.8	0.92	1.15	1.5	0.05
0.80031	1097	1097	2194	0.75	1.0	0.85	0.85	0.4	0.05
0.80017	1275	1275	2550	0.75	1.0	0.85	0.85	0.7	0.05
0.80007	1453	1453	2906	0.75	1.0	0.85	0.85	1.0	0.05
0.80002	1750	1750	3500	0.75	1.0	0.85	0.85	1.5	0.05
0.80010	1320	1320	2640	0.75	1.0	1.15	1.15	0.4	0.05
0.80006	1561	1561	3122	0.75	1.0	1.15	1.15	0.7	0.05
0.80003	1802	1802	3604	0.75	1.0	1.15	1.15	1.0	0.05
0.80006	2204	2204	4408	0.75	1.0	1.15	1.15	1.5	0.05
0.80038	954	954	1908	0.75	1.2	1.02	0.85	0.4	0.05
0.80022	1132	1132	2264	0.75	1.2	1.02	0.85	0.7	0.05
0.80010	1310	1310	2620	0.75	1.2	1.02	0.85	1.0	0.05
0.80004	1607	1607	3214	0.75	1.2	1.02	0.85	1.5	0.05
0.80024	1154	1154	2308	0.75	1.2	1.38	1.15	0.4	0.05
0.80017	1395	1395	2790	0.75	1.2	1.38	1.15	0.7	0.05
0.80012	1636	1636	3272	0.75	1.2	1.38	1.15	1.0	0.05
0.80013	2038	2038	4076	0.75	1.2	1.38	1.15	1.5	0.05
0.80006	851	851	1702	0.75	1.4	1.19	0.85	0.4	0.05
0.80031	1030	1030	2060	0.75	1.4	1.19	0.85	0.7	0.05
0.80017	1208	1208	2416	0.75	1.4	1.19	0.85	1.0	0.05
0.80009	1505	1505	3010	0.75	1.4	1.19	0.85	1.5	0.05
0.80020	1035	1035	2070	0.75	1.4	1.61	1.15	0.4	0.05
0.80013	1276	1276	2552	0.75	1.4	1.61	1.15	0.7	0.05
0.80009	1517	1517	3034	0.75	1.4	1.61	1.15	1.0	0.05
0.80011	1919	1919	3838	0.75	1.4	1.61	1.15	1.5	0.05

Power The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
N1 and N2 The number of subjects in groups 1 and 2, respectively.
N The total sample size. N = N1 + N2.
μ(t) The average exposure time across all subjects.
λ1 The mean event rate per time unit in group 1 (control). This is often the baseline event rate.
λ2 The mean event rate per time unit in group 2 (treatment).
RR The ratio of the two event rates. RR = λ2/λ1.
κ The negative binomial dispersion parameter.
Alpha The probability of rejecting a true null hypothesis.

Summary Statements

A parallel two-group design will be used to test whether the Group 2 (treatment) Negative Binomial rate (λ2) is different from the Group 1 (control) Negative Binomial rate (λ1) (H0: RR = 1 versus Ha: RR ≠ 1, RR = λ2 / λ1). The comparison will be made using a two-sample, two-sided, Wald or likelihood ratio test, where the null variance is calculated using maximum likelihood estimation, and with a Type I error rate (α) of 0.05. The event rate of Group 1 is assumed to be 0.8. The Negative Binomial dispersion parameter (κ) is assumed to be 0.4. The average exposure time in both groups is assumed to be 0.75. To detect a Group 2 event rate of 0.68 (or a ratio, λ2 / λ1, of 0.85) with 80% power, the number of subjects needed will be 1311 in Group 1 (control) and 1311 in Group 2 (treatment).

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%	1311	1311	2622	1639	1639	3278	328	328	656
20%	1490	1490	2980	1863	1863	3726	373	373	746
20%	1668	1668	3336	2085	2085	4170	417	417	834
20%	1965	1965	3930	2457	2457	4914	492	492	984
20%	1570	1570	3140	1963	1963	3926	393	393	786
20%	1811	1811	3622	2264	2264	4528	453	453	906
20%	2052	2052	4104	2565	2565	5130	513	513	1026
20%	2454	2454	4908	3068	3068	6136	614	614	1228
20%	1097	1097	2194	1372	1372	2744	275	275	550
20%	1275	1275	2550	1594	1594	3188	319	319	638
20%	1453	1453	2906	1817	1817	3634	364	364	728
20%	1750	1750	3500	2188	2188	4376	438	438	876
20%	1320	1320	2640	1650	1650	3300	330	330	660
20%	1561	1561	3122	1952	1952	3904	391	391	782
20%	1802	1802	3604	2253	2253	4506	451	451	902
20%	2204	2204	4408	2755	2755	5510	551	551	1102
20%	954	954	1908	1193	1193	2386	239	239	478
20%	1132	1132	2264	1415	1415	2830	283	283	566
20%	1310	1310	2620	1638	1638	3276	328	328	656
20%	1607	1607	3214	2009	2009	4018	402	402	804
20%	1154	1154	2308	1443	1443	2886	289	289	578
20%	1395	1395	2790	1744	1744	3488	349	349	698
20%	1636	1636	3272	2045	2045	4090	409	409	818
20%	2038	2038	4076	2548	2548	5096	510	510	1020
20%	851	851	1702	1064	1064	2128	213	213	426
20%	1030	1030	2060	1288	1288	2576	258	258	516
20%	1208	1208	2416	1510	1510	3020	302	302	604
20%	1505	1505	3010	1882	1882	3764	377	377	754
20%	1035	1035	2070	1294	1294	2588	259	259	518
20%	1276	1276	2552	1595	1595	3190	319	319	638
20%	1517	1517	3034	1897	1897	3794	380	380	760
20%	1919	1919	3838	2399	2399	4798	480	480	960

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, 1639 subjects should be enrolled in Group 1, and 1639 in Group 2, to obtain final group sample sizes of 1311 and 1311, respectively.

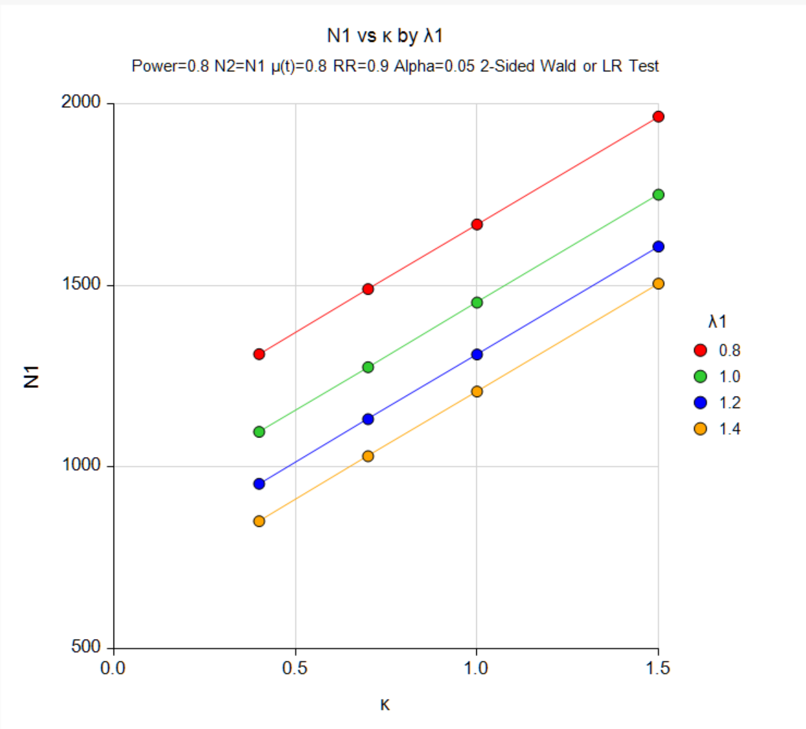
References

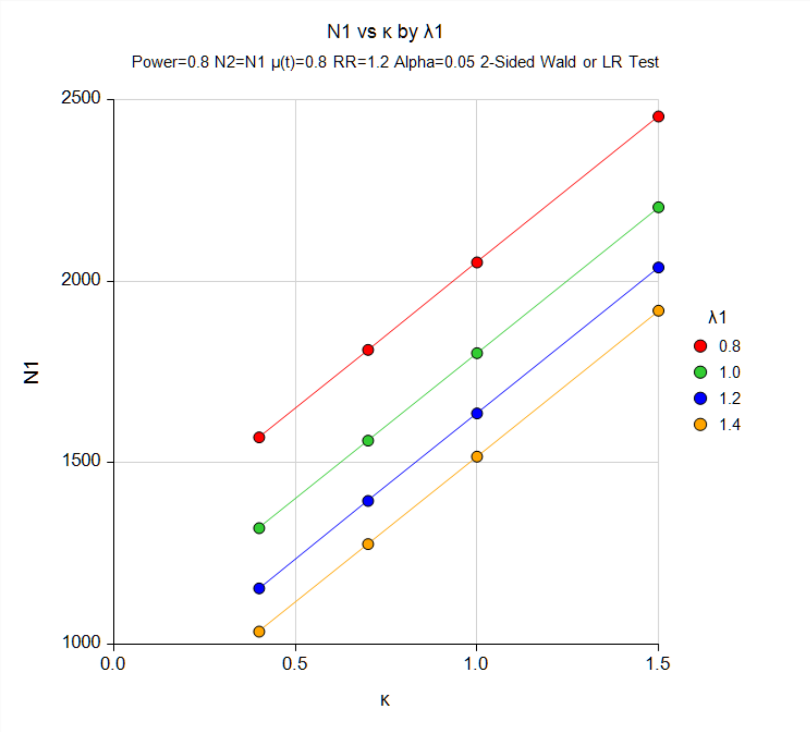
Zhu, H. and Lakkis, H. 2014. 'Sample Size Calculation for Comparing Two Negative Binomial Rates.' Statistics in Medicine, Volume 33, Pages 376-387.

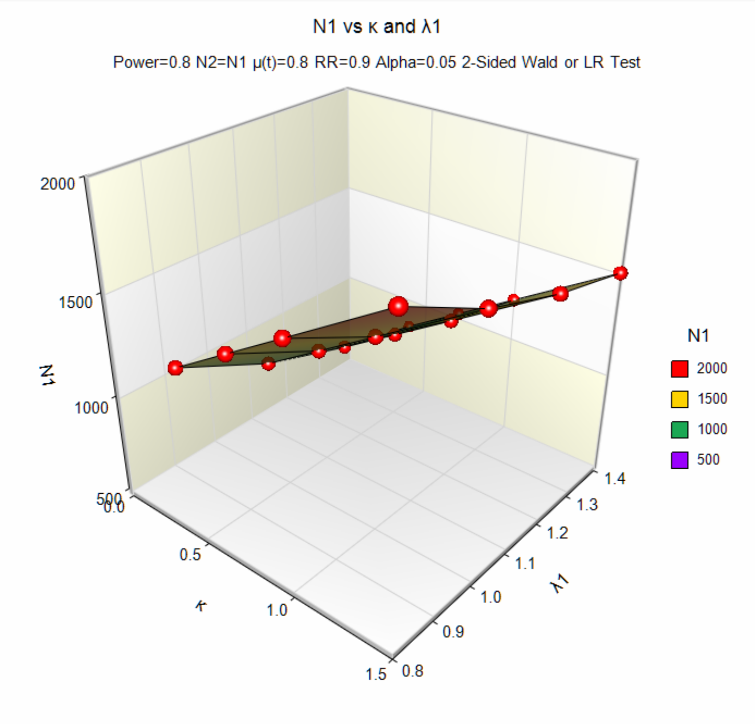
This report shows the sample size for each combination of the input parameters, one scenario per row. If you look at the results in the column labeled M3 under "Calculated N/group" in Table I on page 381 of Zhu and Lakkis (2014), you'll see that the group sample sizes calculated by **PASS** match those exactly in all cases (though they are presented in a different order).

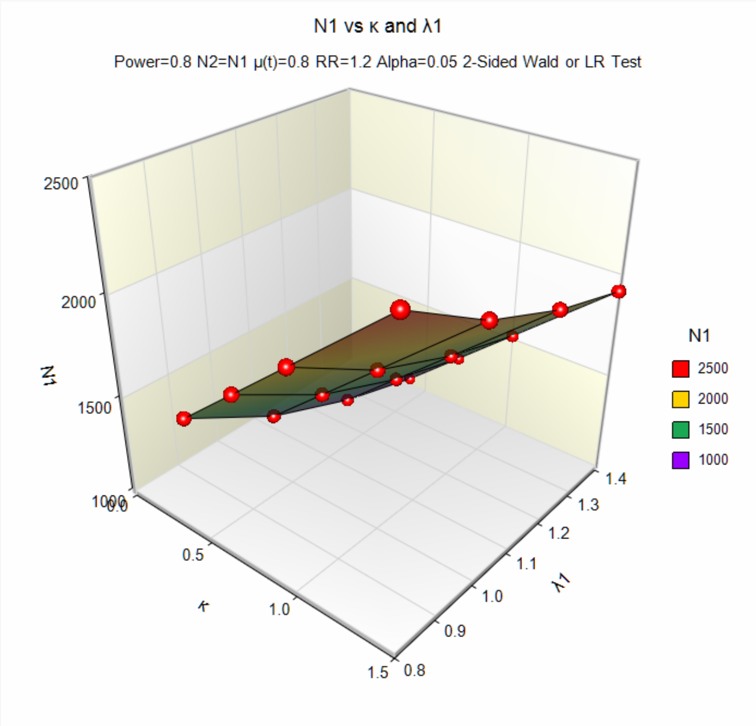
Plots Section

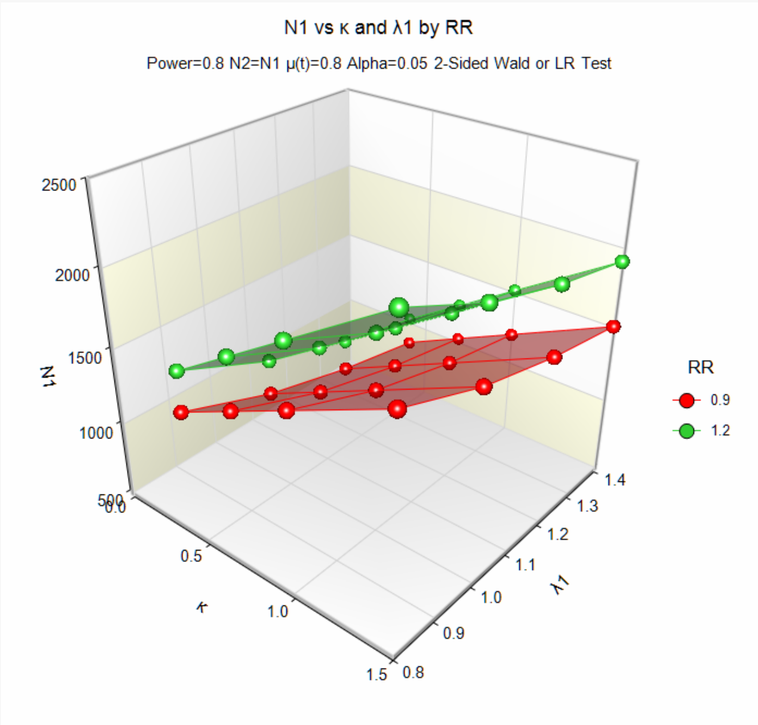
Plots




N1 vs κ by λ1
Power=0.8 N2=N1 μ(t)=0.8 RR=1.2 Alpha=0.05 2-Sided Wald or LR Test
N1
κ
2500
2000
1500
1000
0.0
0.5
1.0
1.5
λ1
0.8
1.0
1.2
1.4


N1 vs κ and λ1
Power=0.8 N2=N1 μ(t)=0.8 RR=0.9 Alpha=0.05 2-Sided Wald or LR Test
N1
κ
λ1
2000
1500
1000
500
0.0
0.5
1.0
1.5
0.8
0.9
1.0
1.1
1.2
1.3
1.4
N1
2000
1500
1000
500





These plots show the relationship between sample size, κ, λ1, and the event rate ratio, RR.

To match the results for Method 1 (M1) in Table I of Zhu and Lakkis (2014) on page 381, change Null Variance Calculation Method to "Use Event Rate from Group 1 (λ1)" in **PASS** and then re-calculate the sample size (or load the example template **Example 1b**). The results are not displayed here.

To match the results for Method 2 (M2) in Table I of Zhu and Lakkis (2014) on page 381, change Null Variance Calculation Method to "Use True Event Rates (λ1 and λ2)" in **PASS** and then re-calculate the sample size (or load the example template **Example 1c**). The results are not displayed here.

Example 2 – Estimating the Negative Binomial Dispersion Parameter from a Previous Study (Validation 2 using Zhu and Lakkis (2014))

Zhu and Lakkis (2014) suggests that the hardest parameter to obtain in these sample size and power calculations is usually the negative binomial dispersion parameter, κ. When a suitable value for κ is not known, then you can estimate κ from a similar study that was analyzed using Poisson regression. Given a Poisson mean event rate λ and an overdispersion factor ϕ, estimated from a similar Poisson regression study, κ can be estimated as

$$\hat{\kappa} = \frac{\hat{\phi} - 1}{\hat{\lambda}}$$

Zhu and Lakkis (2014) go even further and demonstrate how to obtain the Poisson overdispersion factor estimate when it is not reported directly. (See the section "Estimating the Negative Binomial Dispersion Parameter, κ, from a Previous Study that was Analyzed using Poisson Regression" above for details).

In their example (pages 382 through 384), a one-year study is being designed for a drug that is intended to reduce asthma exacerbations in a particular set of patients. They found a similar study to that being designed that would provide a basis for the sample size calculations of the present study. The previous study was analyzed using Poisson regression but did not report the estimated overdispersion factor directly. It did report total exposure times, estimated exacerbation rates, and asymmetric 95% confidence limits for both groups:

Group 1 (Placebo)
Exposure Time = 397 patient-years, $\hat{\lambda}_1 = 0.663$, 95% Confidence Interval = (0.573, 0.768)

Group 2 (Tiotropium):
Exposure Time = 399 patient-years, $\hat{\lambda}_2 = 0.530$, 95% Confidence Interval = (0.450, 0.625)

Using this information and the equations presented above, they back-calculate $\text{SE}(\log \hat{\lambda}_1)$ and $\text{SE}(\log \hat{\lambda}_2)$ from the 95% confidence limits as

$$\text{SE}(\log \hat{\lambda}_1) = \frac{\left(\frac{\log(\text{Upper Bound}) - \log(\text{Lower Bound})}{2}\right)}{z_{1-\alpha/2}} = \frac{\left(\frac{\log(0.768) - \log(0.573)}{2}\right)}{1.96} = 0.0747$$

$$\text{SE}(\log \hat{\lambda}_2) = \frac{\left(\frac{\log(\text{Upper Bound}) - \log(\text{Lower Bound})}{2}\right)}{z_{1-\alpha/2}} = \frac{\left(\frac{\log(0.625) - \log(0.450)}{2}\right)}{1.96} = 0.0838$$

From this they calculate $\hat{\phi}_1$ and $\hat{\phi}_2$ as

$$\hat{\phi}_1 = T_1 \hat{\lambda}_1 \left(\text{SE}(\log \hat{\lambda}_1)\right)^2 = 397 \times 0.663 \times (0.0747)^2 = 1.47$$

$$\hat{\phi}_2 = T_2 \hat{\lambda}_2 \left(\text{SE}(\log \hat{\lambda}_2)\right)^2 = 399 \times 0.530 \times (0.0838)^2 = 1.49$$

Finally, an estimate for the negative binomial dispersion κ is calculated as

$$\hat{\kappa} = \frac{\left(\frac{\hat{\phi}_1 + \hat{\phi}_2}{2}\right) - 1}{\left(\frac{\hat{\lambda}_1 + \hat{\lambda}_2}{2}\right)} = \frac{\left(\frac{1.47 + 1.49}{2}\right) - 1}{\left(\frac{0.663 + 0.530}{2}\right)} = 0.8$$

Using this estimated negative binomial dispersion of 0.8, the placebo rate of 0.66 events per patient-year, an event rate ratio of 0.8 (representing a 20% reduction in exacerbations with the new drug), an average exposure time of 0.9 years (since not everybody is expected to be followed for the full year), and using method 3 to estimate the null variance, they calculate a per-group sample size of 1131 for the new study to achieve 90% power at a significance level of 0.05. **PASS** matches this result exactly as follows.

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**
Alternative Hypothesis	**Two-Sided**
Null Variance Calculation Method	**Use Maximum Likelihood Estimation**
Power	**0.90**
Alpha	**0.05**
μ(t) (Average Subject Exposure Time)	**0.9**
Group Allocation	**Equal (N1 = N2)**
λ1 (Event Rate of Group 1)	**0.66**
Enter λ2 or Ratio for Group 2	**RR (Ratio of Event Rates)**
RR (Ratio of Event Rates)	**0.8**
κ (Negative Binomial Dispersion)	**0.8**

Output

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

Numeric Results

Solve For: Sample Size
Alternative Hypothesis: Two-Sided (H0: RR = 1 vs. Ha: RR ≠ 1)
Null Variance Calculation Method: Use Maximum Likelihood Estimation

	Sample Size			Average Exposure Time	Average Event Rate		Event Rate Ratio	Negative Binomial Dispersion	
Power	N1	N2	N	μ(t)	λ1	λ2	RR	κ	Alpha
0.9	1131	1131	2262	0.9	0.66	0.53	0.8	0.8	0.05

The group sample size of 1131 calculated by **PASS** matches the result obtained by Zhu and Lakkis (2014) exactly.