

Chapter 334

Multi-Arm Equivalence Tests for the Ratio of Treatment and Control Means (Normal Data)

Introduction

This module computes power and sample size for multiple equivalence tests of treatment means versus a control mean when *the data are assumed to follow the normal distribution* (so the log transformation is not used). This routine deals with the case in which the statistical hypotheses are expressed in terms of *mean ratios* rather than mean differences.

The details of this analysis are given in Hauschke *et al.* (1999) and Kieser and Hauschke (1999). The multiplicity is based on the results in Machin, Campbell, Tan, and Tan (2018).

In this design, there are k treatment groups and one control group. A mean is measured in each group. A total of k hypothesis tests are anticipated each comparing a treatment group with the common control group using a test based on the ratio of the two means.

The Bonferroni adjustment of the type I error rate may be optionally made because several comparisons are being tested using the same data. Making a multiplicity adjustment is usually recommended, but not always. In fact, Saville (1990) advocates not applying it and Machin, Campbell, Tan, and Tan (2018) include omitting it as a possibility.

Background

Whether you want to test several doses of a single treatment or several types of treatments, good research practice requires that each treatment be compared with a control. For example, a popular three-arm design consists of three groups: control, treatment A, and treatment B. Two equivalence tests are run: treatment A versus control and treatment B versus the same control. This design avoids having to obtain a second control group for treatment B. Besides the obvious efficiency in subjects, it may be easier to recruit subjects if their chances of receiving the new treatment are better than 50-50.

Technical Details

Suppose you want to compare k treatment groups with means μ_i and sample sizes all equal to N_C and one control group with mean μ_C and sample size also N_C . The total sample size is $N = (k + 1)N_C$.

Equivalence Tests

Measurements are made on individuals that have been randomly assigned to the groups. This *parallel-groups* design may be analyzed by a set of TOST equivalence tests to show that the means of the treatment and control groups do not differ by more than a small amount, either positive or negative. To conduct an equivalence test, you must set upper and lower equivalence limits for the mean ratio of the treatment mean and the control mean. These limits, which will be called R_L and R_U , establish an interval of equivalence. When the sample mean ratio falls between these limits, the null hypothesis of non-equivalence is rejected and the equivalence of the two group means is concluded.

The statistical hypotheses are written as follows.

With $R_L < 1$ and $R_U > 1$, the null hypothesis of non-equivalence is

$$H_0: R \leq R_L \text{ or } R \geq R_U$$

and alternative hypothesis of equivalence is

$$H_1: R_L < R < R_U$$

where $R = \mu_i/\mu_C$.

Test Statistic

The response of interest is assumed to follow the normal distribution with (possibly different) means μ_T and μ_C and common variance σ^2 . To carry out the comparison, a random sample of $N_1 = N_2$ subjects is obtained from each group. The parameters of the study will be presented in terms of the mean ratio μ_T/μ_C .

The equivalence hypotheses are as given above. The null hypothesis H_0 is rejected in favor of the alternative if a two-sided $100(1 - 2\alpha)\%$ Fieller confidence interval is included completely between the two equivalence limits.

Power Approximation

As mentioned above, **PASS** uses the approximate formulas given by Kieser and Hauschke (1999). These formulas are stated in terms of the control group sample size N_C , but they can be rearranged to give formulas for power as well. Let $R1 = \mu_i/\mu_C$ and $COV = \sigma/\mu_C$.

If $R1 = 1$, use

$$N_C \geq (1 + R_L^2)(t_{\alpha, 2N_C-2} + t_{\beta, 2N_C-2})^2 \left(\frac{COV}{1 - R_L} \right)^2$$

If $R_L < R1 < 1$, use

$$N_C \geq (1 + R_L^2)(t_{\alpha, 2N_C-2} + t_{\beta, 2N_C-2})^2 \left(\frac{COV}{R1 - R_L} \right)^2$$

If $1 < R1 < R_U$, use

$$N_C \geq (1 + R_U^2)(t_{\alpha, 2N_C-2} + t_{\beta, 2N_C-2})^2 \left(\frac{COV}{R1 - R_U} \right)^2$$

Multiplicity Adjustment

Because k tests between treatment groups and the control group are run when analyzing the results of this study, many statisticians recommend that the Bonferroni adjustment be applied. This adjustment is easy to apply: the value of alpha that is used in the test is found by dividing the original alpha by the number of tests. For example, if the original alpha is 0.05 and the number of treatment (not including the control) groups is five, the individual tests will be conducted using an alpha of $0.05 / 5 = 0.01$.

The main criticism of this procedure is that if there are many tests, the value of alpha becomes very small. To mitigate against this complaint, some statisticians recommend separating the treatment groups into those that are of primary interest and those that are of secondary interest. The Bonferroni adjustment is made by using the number of primary treatments rather than the total number of treatments.

There are some who advocate ignoring the adjustment entirely in the case of randomized clinical trials. See for example Saville (1990) and the discussion in chapter 14 of Machin, Campbell, Tan, and Tan (2018).

Size of the Control Group

Because the control group is used over and over, some advocate increasing the number of subjects in this group. The standard adjustment is to include \sqrt{k} subjects in the control group for each subject in one of the treatment groups. See Machin, Campbell, Tan, and Tan (2018, pages 231-232). Note that often, the treatment groups all have the same size.

Example 1 – Finding the Sample Size

A parallel-group clinical trial is being designed to determine if any or all of three treatment therapies are equivalent to the standard therapy. Suppose the standard therapy has mean response of 9.3 with a standard deviation of 2.5. The investigators would like a sample size large enough to find statistical significance at the 0.05 level if the actual mean responses of the three treatments are 9.1, 9.3, and 9.6. The power of each test is set to 0.80. The equivalence ratio limits are 0.8 and 1.25. The standard deviation ranges from 2.0 to 3.0. The group sample sizes are equal.

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 of Each Test	0.8
Overall Alpha	0.05
Bonferroni Adjustment	Standard Bonferroni
Group Allocation	Equal (Nc = N1 = N2 = ...)
Equivalence Limit Input Type.....	Enter Upper Equivalence Limit RU (RL = 1 / RU)
RU (Upper Equivalence Limit)	1.25
Control Mean	9.3
Set A Number of Groups.....	1
Set A Mean	9.1
Set B Number of Groups.....	1
Set B Mean	9.3
Set C Number of Groups	1
Set C Mean.....	9.6
Set D Number of Groups	0
More.....	Unchecked
σ (Standard Deviation).....	2 2.5 3

Output

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

Numeric Reports

Numeric Results

Solve For: **Sample Size**
 Group Allocation: Equal (Nc = N1 = N2 = ...)
 Test Type: Equal Variance T-Test
 Hypotheses: H0: $R \leq RL$ or $R \geq RU$ vs. H1: $RL < R < RU$
 Number of Groups: 4
 Bonferroni Adjustment: Standard Bonferroni (Divisor = 3)

Comparison	Power		Sample Size Ni	Mean μ_i	Mean Ratio Ri	Equivalence Limits		Standard Deviation σ	COVi	Alpha	
	Target	Actual				Lower RL	Upper RU			Overall	Bonferroni-Adjusted
Control			24	9.3				2.0	0.21505		
vs A	0.8	0.83420	24	9.1	0.97849	0.8	1.25	2.0	0.21978	0.05	0.01667
vs B	0.8	0.82010	24	9.3	1.00000	0.8	1.25	2.0	0.21505	0.05	0.01667
vs C	0.8	0.81484	24	9.6	1.03226	0.8	1.25	2.0	0.20833	0.05	0.01667
Total			96								
Control			36	9.3				2.5	0.26882		
vs A	0.8	0.82480	36	9.1	0.97849	0.8	1.25	2.5	0.27473	0.05	0.01667
vs B	0.8	0.80673	36	9.3	1.00000	0.8	1.25	2.5	0.26882	0.05	0.01667
vs C	0.8	0.80504	36	9.6	1.03226	0.8	1.25	2.5	0.26042	0.05	0.01667
Total			144								
Control			51	9.3				3.0	0.32258		
vs A	0.8	0.82213	51	9.1	0.97849	0.8	1.25	3.0	0.32967	0.05	0.01667
vs B	0.8	0.80297	51	9.3	1.00000	0.8	1.25	3.0	0.32258	0.05	0.01667
vs C	0.8	0.80228	51	9.6	1.03226	0.8	1.25	3.0	0.31250	0.05	0.01667
Total			204								

Comparison: The group that is involved in the comparison between the treatment and control displayed on this report line. The comparison is made using the ratio.

Target Power: The power desired. Power is probability of rejecting a false null hypothesis for this comparison. This power is of the comparison shown on this line only.

Actual Power: The power actually achieved.

Ni: Sample Size. The number of subjects in the ith group. The total sample size, N, is shown as the last row of the column.

μ_i : The mean of the ith group at which the power is computed. The first row contains μ_c , the control group mean.

Ri: The ratio of the means at which the power is calculated. $R_i = \mu_i / \mu_c$.

RL: The lower equivalence ratio. This is the smallest ratio of means of treatment groups and the control group that still results in the conclusion that the treatment group is equivalent to the control group.

RU: The upper equivalence ratio. This is the largest ratio of means of treatment groups and the control group that still results in the conclusion that the treatment group is equivalent to the control group.

σ : The standard deviation of the responses within each group.

COVi: The coefficient of variation of the ith group. $COV_i = \sigma / \mu_i$.

Overall Alpha: The probability of rejecting at least one of the comparisons in this experiment when each null hypothesis is true.

Bonferroni Alpha: The adjusted significance level at which each individual comparison is made.

Multi-Arm Equivalence Tests for the Ratio of Treatment and Control Means (Normal Data)

Summary Statements

A parallel, 4-group design (with one control group and 3 treatment groups) will be used to test whether the mean for each treatment group is equivalent to the control group mean, with equivalence ratio limits of 0.8 and 1.25 ($H_0: R \leq 0.8$ or $R \geq 1.25$ versus $H_1: 0.8 < R < 1.25, R = \mu_i / \mu_c$). Each of the 3 equivalence comparisons will be made using two one-sided, two-sample, Bonferroni-adjusted, equal-variance, ratio-based t-tests. The overall (experiment-wise) Type I error rate (α) is 0.05. The common standard deviation for all groups is assumed to be 2. The control group mean is assumed to be 9.3. To detect the treatment means 9.1, 9.3, and 9.6 with at least 80% power for each test, the (equal) group sample size needed for each of the 4 groups (control and treatments) will be 24 (totaling 96 subjects).

Dropout-Inflated Sample Size

Group	Dropout Rate	Sample Size Ni	Dropout-Inflated Enrollment Sample Size Ni'	Expected Number of Dropouts Di
1 - 4	20%	24	30	6
Total		96	120	24
1 - 4	20%	36	45	9
Total		144	180	36
1 - 4	20%	51	64	13
Total		204	256	52

- Group Lists the group numbers.
- 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.
- Ni The evaluable sample size for each group at which power is computed (as entered by the user). If Ni subjects are evaluated out of the Ni' subjects that are enrolled in the study, the design will achieve the stated power.
- Ni' The number of subjects that should be enrolled in each group in order to obtain Ni evaluable subjects, based on the assumed dropout rate. Ni' is calculated by inflating Ni using the formula $Ni' = Ni / (1 - DR)$, with Ni' 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.)
- Di The expected number of dropouts in each group. $Di = Ni' - Ni$.

Dropout Summary Statements

Anticipating a 20% dropout rate, group sizes of 30, 30, 30, and 30 subjects should be enrolled to obtain final group sample sizes of 24, 24, 24, and 24 subjects.

References

Blackwelder, W.C. 1998. 'Equivalence Trials.' In Encyclopedia of Biostatistics, John Wiley and Sons. New York. Volume 2, 1367-1372.

Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. 2018. Sample Size Calculations in Clinical Research, 3rd Edition. Chapman & Hall/CRC. Boca Raton, FL. Pages 86-88.

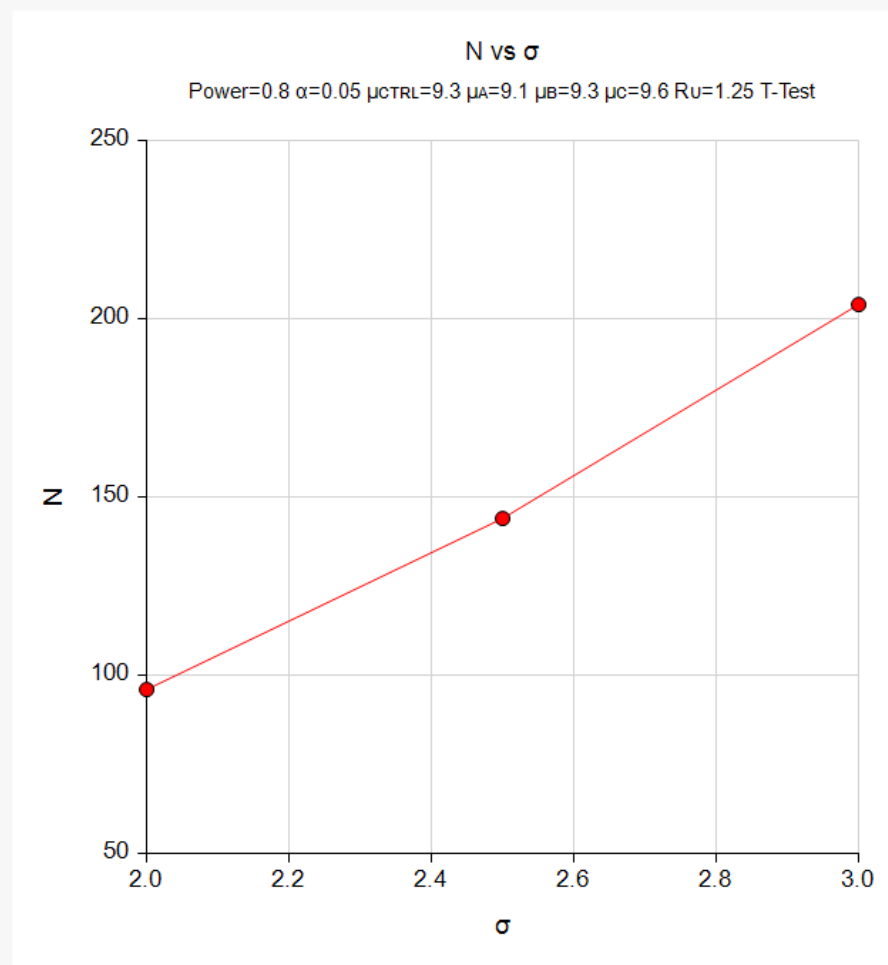
Julious, Steven A. 2004. 'Tutorial in Biostatistics. Sample sizes for clinical trials with Normal data.' Statistics in Medicine, 23:1921-1986.

Machin, D., Campbell, M.J., Tan, S.B, and Tan, S.H. 2018. Sample Sizes for Clinical, Laboratory, and Epidemiology Studies, 4th Edition. Wiley Blackwell.

This report shows the numeric results of this power study. Notice that the results are shown in blocks of three rows at a time. Each block represents a single design.

Plots Section

Plots



This plot gives a visual representation of the results in the Numeric Report. We can quickly see the impact on the sample size of varying the standard deviation.

Example 2 – Validation using a Previously Validated Procedure

We could not find a validation result in the statistical literature, so we will use a previously validated **PASS** procedure (**Equivalence Tests for the Ratio of Two Means (Normal Data)**) to produce the results for the following example.

A parallel-group clinical trial is being designed to determine if any or all of three treatment therapies are equivalent to the standard therapy. Suppose the standard therapy has mean response of 9.3 with a standard deviation of 2.5. The investigators would like a sample size large enough to find statistical significance at the 0.05 level if the actual mean responses of the three treatments are 9.1, 9.3, and 9.6. This results in values of R1 of 0.97849, 1, and 1.03226. The power of each test is set to 0.80. The equivalence ratio limits are 0.8 and 1.25. The standard deviation is 2.5 so the COV = 2.5 / 9.3 = 0.26882. The group sample sizes are equal.

The **Equivalence Tests for the Ratio of Two Means (Normal Data)** procedure is set up as follows.

Design Tab

Solve For **Sample Size**
 Power..... **0.8**
 Alpha..... **0.01667** (which is Alpha / k)
 Equivalence Limit Input Type..... **Enter Lower Equivalence Limit RL (RU = 1 / RL)**
 RL (Lower Equivalence Limit)..... **0.80**
 R1 (Actual Ratio) **0.97849 1 1.03226**
 COV (Coefficient of Variation)..... **0.26882**

This set of options generates the following report.

Numeric Results

Solve For: [Sample Size](#)
 Ratio: R = Treatment Mean / Reference Mean
 Hypotheses: H0: $R \leq RL$ or $R \geq RU$ vs. H1: $RL < R < RU$

Target Power	Actual Power	N1	N2	N	Lower Equiv Limit RL	Upper Equiv Limit RU	R1	COV	Alpha
0.8	0.80074	34	34	68	0.8	1.25	0.97849	0.26882	0.01667
0.8	0.80674	36	36	72	0.8	1.25	1.00000	0.26882	0.01667
0.8	0.80505	36	36	72	0.8	1.25	1.03226	0.26882	0.01667

In order to maintain a power of 80% for all three groups, it is apparent that the groups will all need to have a sample size of 36. This table contains the validation values. We will now run these values through the current procedure and compare the results with these values.

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 of Each Test	0.8
Overall Alpha	0.05
Bonferroni Adjustment	Standard Bonferroni
Group Allocation	Equal (Nc = N1 = N2 = ...)
Equivalence Limit Input Type.....	Enter Upper Equivalence Limit RU (RL = 1 / RU)
RU (Upper Equivalence Limit)	1.25
Control Mean	9.3
Set A Number of Groups.....	1
Set A Mean	9.1
Set B Number of Groups.....	1
Set B Mean	9.3
Set C Number of Groups	1
Set C Mean.....	9.6
Set D Number of Groups	0
More.....	Unchecked
σ (Standard Deviation).....	2.5

Output

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

Numeric Results											
Solve For:		Sample Size									
Group Allocation:		Equal (Nc = N1 = N2 = ...)									
Test Type:		Equal Variance T-Test									
Hypotheses:		H0: R ≤ RL or R ≥ RU vs. H1: RL < R < RU									
Number of Groups:		4									
Bonferroni Adjustment:		Standard Bonferroni (Divisor = 3)									
Comparison	Power		Sample Size Ni	Mean μ_i	Mean Ratio Ri	Equivalence Limits		Standard Deviation σ	COVi	Alpha	
	Target	Actual				Lower RL	Upper RU			Overall	Bonferroni-Adjusted
Control			36	9.3				2.5	0.26882		
vs A	0.8	0.82480	36	9.1	0.97849	0.8	1.25	2.5	0.27473	0.05	0.01667
vs B	0.8	0.80673	36	9.3	1.00000	0.8	1.25	2.5	0.26882	0.05	0.01667
vs C	0.8	0.80504	36	9.6	1.03226	0.8	1.25	2.5	0.26042	0.05	0.01667
Total			144								

As you can see, the sample sizes are all 36, which match the largest sample size found in the validation run above. The procedure is validated.