Chapter 332

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

Introduction

This module computes power and sample size for multiple non-inferiority tests of treatment means versus a control mean when the data are assumed to follow the normal distribution and the statistical hypotheses are expressed in terms of mean ratios. Note that when the data follow a log-normal distribution rather than the normal distribution so that a log transformation is used, you should use other **PASS** procedures that assume a log-normal data distribution.

The details of this t-test are given in Rothmann, Wiens, and Chan (2012) and, to lesser extent, in Kieser and Hauschke (1999). The multiple comparison aspect of this procedure is based on the results in Machin, Campbell, Tan, and Tan (2018).

In this parallel-group 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 t-test based on the ratio of 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. A popular three-arm design consists of three groups: control, treatment A, and treatment B. Two 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 N_i and one control group with mean μ_C and sample size N_C. The total sample size is $N = N_1 + N_2 + \cdots + N_k + N_C$.

Non-Inferiority Tests

A *non-inferiority test* tests that the treatment mean is not worse than the control mean by more than the non-inferiority margin. The actual direction of the hypothesis depends on the response variable being studied. Define $R = \mu_i/\mu_C$.

Case 1: High Values Good

In this case, higher response values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is no less than a small amount below the control mean. This results in a non-inferiority boundary called R_L. The null and alternative hypotheses with are

$$H_0: R \leq R_L \quad \text{vs.} \quad H_1: R > R_L$$

where $R_L < 1$.

Case 2: High Values Bad

In this case, lower values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is no more than a small amount above the control mean. This results in a non-inferiority boundary called R_U. The null and alternative hypotheses with are

$$H_0: R \geq R_U \quad \text{vs.} \quad H_1: R < R_U$$

where $R_U > 1$.

Equal-Variances T-Test Statistic

The ratio hypotheses are rearranged as from

$$H_0: \frac{\mu_i}{\mu_C} \leq R_L \quad \text{vs.} \quad H_1: \frac{\mu_i}{\mu_C} > R_L$$

to

$$H_0: \mu_i - R_L\mu_C \leq 0 \quad \text{vs.} \quad H_1: \mu_i - R_L\mu_C > 0$$

The null hypothesis is tested using the test statistic

$$T_1 = \sqrt{\frac{\bar{X}_i - R_L\bar{X}_C}{S\left(\frac{1}{N_i} + \frac{R_L^2}{N_C}\right)}}$$

where $\bar{X}_i$ and $\bar{X}_C$ are the sample means of the treatment and control groups, and *S* is the pooled estimate of the standard deviation, σ which is given by

$$S^2 = \frac{(N_i - 1)s_i^2 + (N_C - 1)s_C^2}{N_i - N_C - 2}$$

It is assumed that T_1 is distributed as a central *t* distribution with degrees of freedom given by $N_i + N_C - 2$.

For a specified alternative R_1, T_1 follows the noncentral *t* distribution with $N_i + N_C - 2$ degrees of freedom and noncentrality

$$\left(\frac{R_1 - R_L}{CV}\right)\sqrt{\frac{N_C}{\frac{N_C}{N_i} + R_L^2}}$$

Hence, the power of this test is given by noncentral *t* distribution as follows

$$(1 - \beta) = \Pr\left(T_1 \geq t_{1-\alpha,N_i+N_C-2} | R_1, R_L, CV, N_C, N_i\right)$$

Multiplicity Adjustment

Because *k* t-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 set at 0.05 and the number of treatment (not including the control) groups is five, the individual tests will be conducted using an alpha of 0.01.

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 the 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 compare three treatment therapies against the standard therapy. Higher values of the response are desirable. Suppose the standard therapy has a 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 0.025 level (since this a one-sided test) if the actual mean responses of the three treatments are 9.1, 9.3, and 9.5, the power of each test is 0.80, and the non-inferiority limit is 0.8. They want to consider a range of standard deviations from 2.0 to 3.0.

Following standard procedure, the control group multiplier will be set to $\sqrt{k} = \sqrt{3} = 1.732$ since the control group is used for three comparisons in this design.

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**
Higher Means Are	**Better (H1: R > RL)**
Power of Each Test	**0.80**
Overall Alpha	**0.025**
Bonferroni Adjustment	**Standard Bonferroni**
Group Allocation	**Enter Group Allocation Pattern, solve for group sample sizes**
RL (Lower Non-Inferiority Limit)	**0.8**
Control Mean	**9.3**
Control Sample Size Allocation	**1.732**
Set A Number of Groups	**1**
Set A Mean	**9.1**
Set A Sample Size Allocation	**1**
Set B Number of Groups	**1**
Set B Mean	**9.3**
Set B Sample Size Allocation	**1**
Set C Number of Groups	**1**
Set C Mean	**9.5**
Set C Sample Size Allocation	**1**
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:	Enter Group Allocation Pattern, solve for group sample sizes
Test Type:	Equal-Variance T-Test
Higher Means Are:	Better
Hypotheses:	H0: R ≤ RL vs. H1: R > RL
Number of Groups:	4
Bonferroni Adjustment:	Standard Bonferroni (Divisor = 3)

	Power		Sample Size				Lower Non-Inferiority			Alpha	
Comparison	Target	Actual	Ni	Allocation	Mean μi	Mean Ratio Ri	Limit RL	Standard Deviation σ	Coefficient of Variation COVi	Overall	Bonferroni-Adjusted
Control			38	1.732	9.3			2.0	0.21505		
vs A	0.8	0.80201	22	1.000	9.1	0.97849	0.8	2.0	0.21978	0.025	0.00833
vs B	0.8	0.89238	22	1.000	9.3	1.00000	0.8	2.0	0.21505	0.025	0.00833
vs C	0.8	0.94841	22	1.000	9.5	1.02151	0.8	2.0	0.21053	0.025	0.00833
Total			104								
Control			59	1.732	9.3			2.5	0.26882		
vs A	0.8	0.80593	34	1.000	9.1	0.97849	0.8	2.5	0.27473	0.025	0.00833
vs B	0.8	0.89532	34	1.000	9.3	1.00000	0.8	2.5	0.26882	0.025	0.00833
vs C	0.8	0.95028	34	1.000	9.5	1.02151	0.8	2.5	0.26316	0.025	0.00833
Total			161								
Control			83	1.732	9.3			3.0	0.32258		
vs A	0.8	0.80085	48	1.000	9.1	0.97849	0.8	3.0	0.32967	0.025	0.00833
vs B	0.8	0.89155	48	1.000	9.3	1.00000	0.8	3.0	0.32258	0.025	0.00833
vs C	0.8	0.94791	48	1.000	9.5	1.02151	0.8	3.0	0.31579	0.025	0.00833
Total			227								

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	The number of subjects in the ith group. The total sample size shown below the groups is equal to the sum of all individual group sample sizes.
Allocation	The group sample size allocation ratio of the ith group. The value on each row represents the relative number of subjects assigned to the group.
μi	The mean of the ith group at which the power is computed. The first row contains μc, the control group mean.
Ri	The ratio of the means at which the power is calculated. Ri = μi / μc.
RL	The lower non-inferiority boundary for the mean ratio. RL < 1.
σ	The standard deviation of the responses within each group.
COVi	The coefficient of variation of the ith group. COVi = σ / μi.
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.

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 non-inferior to the control group mean, with a non-inferiority ratio limit of 0.8 (H0: R ≤ 0.8 versus H1: R > 0.8, R = μi / μc). In this study, higher means are considered to be better. The non-inferiority hypotheses will be evaluated using 3 one-sided, two-sample, Bonferroni-adjusted, equal-variance, ratio-based t-tests, with an overall (experiment-wise) Type I error rate (α) of 0.025. 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 9.5 with at least 80% power for each test, the control group sample size needed will be 38 and the number of needed subjects for the treatment groups will be 22, 22, and 22 (totaling 104 subjects overall).

Dropout-Inflated Sample Size

Group	Dropout Rate	Sample Size Ni	Dropout-Inflated Enrollment Sample Size Ni'	Expected Number of Dropouts Di
1	20%	38	48	10
2	20%	22	28	6
3	20%	22	28	6
4	20%	22	28	6
Total		104	132	28
1	20%	59	74	15
2	20%	34	43	9
3	20%	34	43	9
4	20%	34	43	9
Total		161	203	42
1	20%	83	104	21
2	20%	48	60	12
3	20%	48	60	12
4	20%	48	60	12
Total		227	284	57

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 48, 28, 28, and 28 subjects should be enrolled to obtain final group sample sizes of 38, 22, 22, and 22 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.

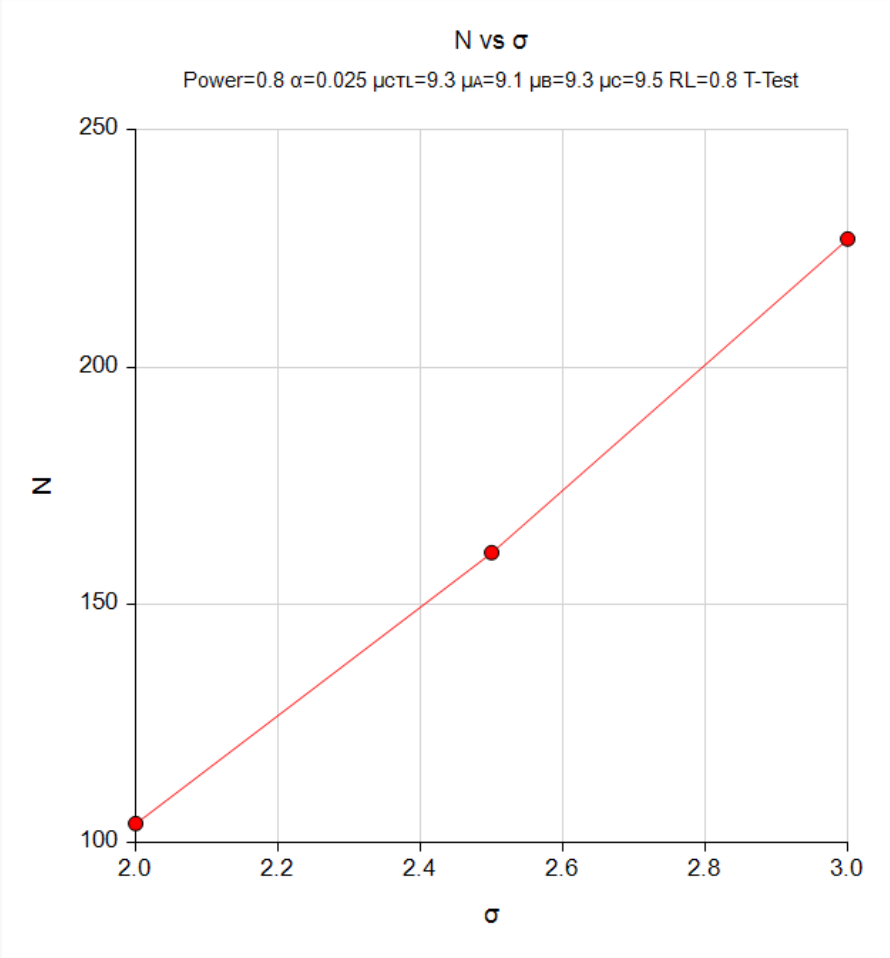
Julious, Steven A. 2004. 'Tutorial in Biostatistics. Sample sizes for clinical trials with Normal data.' Statistics in Medicine, 23:1921-1986.

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 presentation to the results in the Numeric Report. We can quickly see the impact on the sample size of changing 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 the previously validated **PASS** procedure **Non-Inferiority 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 compare three treatment therapies against the standard therapy. Higher values of the response are desirable. Suppose the standard therapy has a 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 0.025 level if the actual mean responses of the three treatments are 9.1, 9.2 and 9.3, the power of each test is 0.80, and the non-inferiority margin is -10% of 9.3 = -0.93.

The sample sizes of all groups will be equal.

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

Design Tab

Solve For	**Sample Size**
Higher Means Are	**Better (H1: R > RL, where RL < 1)**
Power	**0.8**
Alpha	**0.00833** (which is Alpha / k)
Group Allocation	**Equal (N1 = N2)**
Test Statistic	**Equal Variances T-Test**
RL (Lower Non-Inferiority Limit)	**0.8**
R1 (Actual Mean Ratio, μ1 / μ2)	**0.97849 1 1.02151**
CV (Coef of Variation, σ2 / μ2)	**0.26882**

This set of options generates the following report.

Numeric Results

Solve For:	Sample Size
Groups:	1 = Treatment, 2 = Control
Ratio:	R = μ1 / μ2
Higher Means Are:	Better
Hypotheses:	H0: R ≤ RL vs. H1: R > RL
Test:	Equal Variances T-Test

Power		Sample Size			Mean Ratio				
Target	Actual	N1	N2	N	Lower Non-Inferiority Limit RL	Actual R1	Control Group Coefficient of Variation CV	Standard Deviation Ratio λ	Alpha
0.8	0.80676	41	41	82	0.8	0.97849	0.26882	1	0.00833
0.8	0.80733	33	33	66	0.8	1.00000	0.26882	1	0.00833
0.8	0.80416	27	27	54	0.8	1.02151	0.26882	1	0.00833

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 41. 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**
Higher Means Are	**Better (H1: R > RL)**
Power of Each Test	**0.80**
Overall Alpha	**0.025**
Bonferroni Adjustment	**Standard Bonferroni**
Group Allocation	**Equal (Nc = N1 = N2 = ...)**
RL (Lower Non-Inferiority Limit)	**0.8**
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	**9.5**
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
Higher Means Are:	Better
Hypotheses:	H0: R ≤ RL vs. H1: R > RL
Number of Groups:	4
Bonferroni Adjustment:	Standard Bonferroni (Divisor = 3)

	Power					Lower Non-Inferiority			Alpha	
Comparison	Target	Actual	Sample Size Ni	Mean μi	Mean Ratio Ri	Limit RL	Standard Deviation σ	Coefficient of Variation COVi	Overall	Bonferroni-Adjusted
Control			41	9.3			2.5	0.26882		
vs A	0.8	0.80684	41	9.1	0.97849	0.8	2.5	0.27473	0.025	0.00833
vs B	0.8	0.89598	41	9.3	1.00000	0.8	2.5	0.26882	0.025	0.00833
vs C	0.8	0.95069	41	9.5	1.02151	0.8	2.5	0.26316	0.025	0.00833
Total			164							

As you can see, the sample sizes are all 41, which match the largest sample size found in the validation run above. The procedure is validated.