

Chapter 530

Non-Inferiority Tests for the Difference of Two Means in a Higher-Order Cross-Over Design

Introduction

This procedure calculates power and sample size for non-inferiority tests which use the difference in the means of a higher-order cross-over design. Measurements are made on individuals that have been randomly assigned to one of several treatment sequences. Only a brief introduction to the subject will be given here. For a comprehensive discussion on the subject, refer to Chen et al. (1997) and Chow et al. (2003).

Cross-Over Designs

Senn (2002) defines a *cross-over* design as one in which each subject receives all treatments at least once and the object is to study differences among the treatments. The name *cross-over* comes from the most common case in which there are only two treatments. In this case, each subject *crosses over* from one treatment to another. It is assumed that there is a *washout* period between treatments during which the response returns to its baseline value. If this does not occur, there is said to be a *carryover* effect.

A 2x2 cross-over design refers to two treatments (periods) and two *sequences* (treatment orderings). One sequence of treatments is treatment A followed by treatment B. The other sequence is B and then A. The design includes a washout period between responses to make certain that the effects of the first drug do no carryover to the second. Thus, the groups of subjects in this design are defined by the sequence in which the two treatments are administered, not by the treatments they receive.

Higher-Order Cross-Over Designs

Chen et al. (1997) present the results for four cross-over designs that are more complicated than the 2x2 design. Assume that the two treatments are labeled A and B. The available designs are defined by the order and number of times the two treatments are administered.

Non-Inferiority Tests for the Difference of Two Means in a Higher-Order Cross-Over Design

Balaam's Design

Balaam's design has four sequences with two treatments each. It is popular because it allows the intrasubject variabilities to be estimated. The design is

<u>Sequence</u>	<u>Period 1</u>	<u>Period 2</u>
1	A	A
2	B	B
3	A	B
4	B	A

Two-Sequence Dual Design

This design has two sequences with three periods each. It is popular because it allows the intrasubject variabilities to be estimated. The design is

<u>Sequence</u>	<u>Period 1</u>	<u>Period 2</u>	<u>Period 3</u>
1	A	B	B
2	B	A	A

Four-Period Design with Two Sequences

This design has two sequences of four periods each. The design is

<u>Sequence</u>	<u>Period 1</u>	<u>Period 2</u>	<u>Period 3</u>	<u>Period 4</u>
1	A	B	B	A
2	B	A	A	B

Four-Period Design with Four Sequences

This design has four sequences of four periods each. The design is

<u>Sequence</u>	<u>Period 1</u>	<u>Period 2</u>	<u>Period 3</u>	<u>Period 4</u>
1	A	A	B	B
2	B	B	A	A
1	A	B	B	A
2	B	A	A	B

Advantages of Cross-Over Designs

A comparison of treatments on the same subject is expected to be more precise. The increased precision often translates into a smaller sample size. Also, patient enrollment may be easier to obtain because each patient will receive both treatments.

Disadvantages of Cross-Over Designs

The statistical analysis of a cross-over experiment is more complex than a parallel-group experiment and requires additional assumptions. In a cross-over experiment, it may be difficult to separate the treatment effect from the time effect and the carry-over effect of the previous treatment.

These cross-over designs cannot be used when the treatment (or the measurement of the response) alters the subject permanently. Hence, it cannot be used to compare treatments that are intended to provide a cure.

Because subjects must be measured at least twice, it may be more difficult to keep patients enrolled in the study. This is particularly true when the measurement process is painful, uncomfortable, embarrassing, or time consuming.

The Statistical Hypotheses

Both non-inferiority and superiority tests are examples of directional (one-sided) tests. Remember that in the usual t-test setting, the null (H_0) and alternative (H_1) hypotheses for one-sided tests are defined as

$$H_0: \delta \leq A \text{ versus } H_1: \delta > A$$

Rejecting H_0 implies that the mean is larger than the value A . This test is called an *upper-tailed test* because H_0 is rejected only in samples in which the difference in sample means is larger than A .

Following is an example of a *lower-tailed test*.

$$H_0: \delta \geq A \text{ versus } H_1: \delta < A$$

Non-inferiority and *superiority* tests are special cases of the above directional tests. It will be convenient to adopt the following specialize notation for the discussion of these tests.

<u>Parameter</u>	<u>PASS Input/Output</u>	<u>Interpretation</u>
μ_T	Not used	<i>Treatment mean.</i> This is the treatment mean.
μ_R	Not used	<i>Reference mean.</i> This is the mean of a reference population.
M_{NI}	NIM	<i>Margin of non-inferiority.</i> This is a tolerance value that defines the magnitude of the amount that is not of practical importance. This may be thought of as the largest change from the baseline that is considered to be trivial. The absolute value is shown to emphasize that this is a magnitude. The sign of the value will be determined by the specific design that is being used.
δ	D	<i>True difference.</i> This is the value of $\mu_T - \mu_R$, the difference between the treatment and reference means. This is the value at which the power is calculated.

Note that the actual values of μ_T and μ_R are not needed. Only their difference is needed for power and sample size calculations.

Non-Inferiority Tests

A *non-inferiority test* tests that the treatment mean is not worse than the reference mean by more than the equivalence margin. The actual direction of the hypothesis depends on the response variable being studied.

Case 1: High Values Good, Non-Inferiority Test

In this case, higher 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 reference mean. The value of δ is often set to zero. The following are equivalent sets of hypotheses.

$$H_0: \mu_1 \leq \mu_2 - |M_{NI}| \text{ versus } H_1: \mu_1 > \mu_2 - |M_{NI}|$$

$$H_0: \mu_1 - \mu_2 \leq -|M_{NI}| \text{ versus } H_1: \mu_1 - \mu_2 > -|M_{NI}|$$

$$H_0: \delta \leq -|M_{NI}| \text{ versus } H_1: \delta > -|M_{NI}|$$

Non-Inferiority Tests for the Difference of Two Means in a Higher-Order Cross-Over Design

Case 2: High Values Bad, Non-Inferiority Test

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 reference mean. The value of δ is often set to zero. The following are equivalent sets of hypotheses.

$$H_0 : \mu_1 \geq \mu_2 + |M_{NI}| \quad \text{versus} \quad H_1 : \mu_1 < \mu_2 + |M_{NI}|$$

$$H_0 : \mu_1 - \mu_2 \geq |M_{NI}| \quad \text{versus} \quad H_1 : \mu_1 - \mu_2 < |M_{NI}|$$

$$H_0 : \delta \geq |M_{NI}| \quad \text{versus} \quad H_1 : \delta < |M_{NI}|$$

Test Statistics

The analysis for assessing equivalence (and thus non-inferiority) using higher-order cross-over designs is discussed in detail in Chapter 9 of Chow and Liu (2000). Unfortunately, their presentation is too lengthy to give here. Their method involves the computation of an analysis of variance to estimate the error variance. It also describes the construction of confidence limits for appropriate contrasts. One-sided confidence limits can be used for non-inferiority tests. Details of this approach are given in Chapter 3 of Chow et al. (2003). We refer you to these books for details.

Power Calculation

The power of the non-inferiority and superiority tests for the case in which higher values are better is given by

$$Power = T_V \left(\left(\frac{\delta - \varepsilon}{\sigma_w \sqrt{b/n}} \right) - t_{V,1-\alpha} \right)$$

where T represents the cumulative t distribution, V and b depend on the design, σ_w is the square root of the within mean square error from the ANOVA table used to analyze the cross-over design, and n is the average number of subjects per sequence. Note that the constants V and b depend on the design as follows.

The power of the non-inferiority and superiority tests for the case in which higher values are worse is given by

$$Power = 1 - T_V \left(t_{V,1-\alpha} - \left(\frac{\varepsilon - \delta}{\sigma_w \sqrt{b/n}} \right) \right)$$

The constants V and b depend on the design as follows:

<u>Design Type</u>	<u>Parameters (V,b)</u>
Balaam's Design	$V = 4n - 3, b = 2.$
Two-Sequence Dual Design	$V = 4n - 4, b = 3/4.$
Four-Period Design with Two Sequences	$V = 6n - 5, b = 11/20.$
Four-Period Design with Four Sequences	$V = 12n - 5, b = 1/4.$

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Design tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Design Tab

The Design tab contains the parameters associated with this test such as the means, sample sizes, alpha, and power.

Solve For

Solve For

This option specifies the parameter to be solved for from the other parameters. Under most situations, you will select either *Power* for a power analysis or *Sample Size* for sample size determination.

Select *Sample Size* when you want to calculate the sample size needed to achieve a given power and alpha level. Note that there are two choices for finding *Sample Size*. Select *Sample Size (Equal Per Sequence)* when you want the design to have an equal number of subjects per sequence. Select *Sample Size (Exact)* when you want to find the exact sample size even though the number of subjects cannot be dividing equally among the sequences.

Select *Power* when you want to calculate the power of an experiment.

Test

Design Type

Specify which cross-over design that you want to analyze. These designs allow you to compare two means: a treatment (A) and a reference (B). The designs assume that each subject is measured two or more times (or periods) with ample time in between to wash-out the effect of the treatment.

A sequence is an ordering of the how A and B are applied. For example, 'ABA' means that the subjects are measured three times: first treatment A, then B, then A again. Other possible sequences are BAB, AAB, and BBA. Each design includes several sequences.

The design tries to balance the experiment so that each treatment occurs an equal number of times and after each other treatment an equal number of times.

Higher Means Are

This option defines whether higher values of the response variable are to be considered better or worse.

The choice here determines the direction of the non-inferiority test.

If Higher Means Are Better the null hypothesis is $\text{Diff} \leq \text{NIM}$ and the alternative hypothesis is $\text{Diff} > \text{NIM}$, where $\text{NIM} < 0$. If Higher Means Are Worse the null hypothesis is $\text{Diff} \geq \text{NIM}$ and the alternative hypothesis is $\text{Diff} < \text{NIM}$, where $\text{NIM} > 0$.

Non-Inferiority Tests for the Difference of Two Means in a Higher-Order Cross-Over Design

Power and Alpha

Power

This option specifies one or more values for power. Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as *0.8 to 0.95 by 0.05* may be entered.

Alpha

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as *0.01 0.05 0.10* or *0.01 to 0.10 by 0.01*.

Sample Size

N (Total Sample Size)

This option specifies one or more values of the sample size, the number of individuals in the study (total subjects in all sequences). This value must be an integer greater than one.

You may enter a list of values using the syntax *50,100,150,200,250* or *50 to 250 by 50*.

Effect Size – Mean Difference

|NIM| (Non-Inferiority Margin)

This is the magnitude of the margin of non-inferiority. It must be entered as a positive number.

When higher means are better, this value is the distance below the reference mean that is still considered non-inferior. When higher means are worse, this value is the distance above the reference mean that is still considered non-inferior.

D (Difference, $\mu_T - \mu_R$)

This is the actual difference between the treatment mean and the reference mean at which power is calculated.

For non-inferiority tests, this value is often set to zero. When this value is non-zero, care should be taken that this value is consistent with whether higher means are better or worse.

Effect Size – Standard Deviation

Specify σ as σ_w or σ_b and ρ

This option specifies how you will enter the within standard deviation, σ_w . You have two choices:

σ_w

Specify σ_w directly.

σ_b and ρ

Specify σ_b and ρ separately. The value of σ_w is calculated using the formula: $\sigma_w = \sigma_b \sqrt{(1 - \rho)}$.

Non-Inferiority Tests for the Difference of Two Means in a Higher-Order Cross-Over Design

σ_w (Within Std Dev)

Specify the value(s) of the standard deviation σ_w . This value can be estimated using $\sigma_w = \sqrt{\text{WMSE}}$ where WMSE is the within mean square error from the ANOVA table used to analyze the Cross-Over design.

If you were to compute the treatment difference for each subject, this would be the standard deviation of those differences. These values must be positive.

A list of values may be entered.

You can press the **σ button** to load the Standard Deviation Estimator window.

σ_b (Between Std Dev)

This is the between subject standard deviation of the response variable (Y) at a particular time point. It is assumed to be the same for all time points. This represents the variability from subject to subject that occurs when the subjects are treated identically.

As a standard deviation, the number(s) must be greater than zero.

You can enter a list of values separated by blanks or commas, in which case, a separate analysis will be calculated for each value.

You can press the σ button and select 'Covariance Matrix' to obtain help on estimating the standard deviation from an ANOVA table.

ρ (Correlation)

Specify the correlation between two observations made on the same subject. Negative values are not permitted, so it must be between 0 and 1. When no previous information is available, try values between 0.2 and 0.6.

You can press the **σ button** and select 'Covariance Matrix' to obtain help on estimating the standard deviation from an ANOVA table.

You can enter a single value or a list of values.

Non-Inferiority Tests for the Difference of Two Means in a Higher-Order Cross-Over Design

Example 1 – Finding Power

Researchers want to calculate the power of a non-inferiority test using data from a two-sequence, dual cross-over design. The margin of equivalence is either 5 or 10 at several sample sizes between 6 and 66. The true difference between the means under is assumed to be 0. Similar experiments have had a standard deviation (σ) of 10. The significance level is 0.025.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Non-Inferiority Tests for the Difference of Two Means in a Higher-Order Cross-Over Design** procedure window by expanding **Means**, then **Cross-Over (Higher-Order) Design**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for the Difference of Two Means in a Higher-Order Cross-Over Design**. 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	Power
Design Type.....	3x2 (Three-Period, Two-Sequence Dual: ABB BAA)
Higher Means Are.....	Better
Alpha.....	0.025
N (Total Sample Size).....	6 to 66 by 10
NIM (Non-Inferiority Margin)	5 10
D (Difference)	0
Specify σ as σ_w or σ_b and ρ	σ_w (Within Std Error)
σ_w (Within Std Dev).....	10

Annotated Output

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

Numeric Results

Numeric Results for a Non-Inferiority Test of the Mean Difference in a Three-Period, Two-Sequence Dual Design

Treatment Sequences: ABB | BAA

Hypotheses: $H_0: D = (\mu_T - \mu_R) \leq \text{NIM}$; $H_1: D = (\mu_T - \mu_R) > \text{NIM}$

Higher Means are Better

	Total Sample Size	Non- Inferiority Margin NIM	Mean Difference D ($\mu_T - \mu_R$)	Within Standard Deviation σ_w	Alpha
Power	N				
0.1139	6	-5.000	0.000	10.0000	0.025
0.3837	6	-10.000	0.000	10.0000	0.025
0.3405	16	-5.000	0.000	10.0000	0.025
0.8832	16	-10.000	0.000	10.0000	0.025
0.5282	26	-5.000	0.000	10.0000	0.025
0.9818	26	-10.000	0.000	10.0000	0.025
0.6744	36	-5.000	0.000	10.0000	0.025
0.9975	36	-10.000	0.000	10.0000	0.025
0.7817	46	-5.000	0.000	10.0000	0.025
0.9997	46	-10.000	0.000	10.0000	0.025
0.8571	56	-5.000	0.000	10.0000	0.025
1.0000	56	-10.000	0.000	10.0000	0.025
0.9084	66	-5.000	0.000	10.0000	0.025
1.0000	66	-10.000	0.000	10.0000	0.025

Non-Inferiority Tests for the Difference of Two Means in a Higher-Order Cross-Over Design

References

Chow, S.C. and Liu, J.P. 1999. Design and Analysis of Bioavailability and Bioequivalence Studies. Marcel Dekker. New York
 Chow, S.C.; Shao, J.; Wang, H. 2003. Sample Size Calculations in Clinical Research. Marcel Dekker. New York.
 Chen, K.W.; Chow, S.C.; and Li, G. 1997. 'A Note on Sample Size Determination for Bioequivalence Studies with Higher-Order Crossover Designs.' Journal of Pharmacokinetics and Biopharmaceutics, Volume 25, No. 6, pages 753-765.

Report Definitions

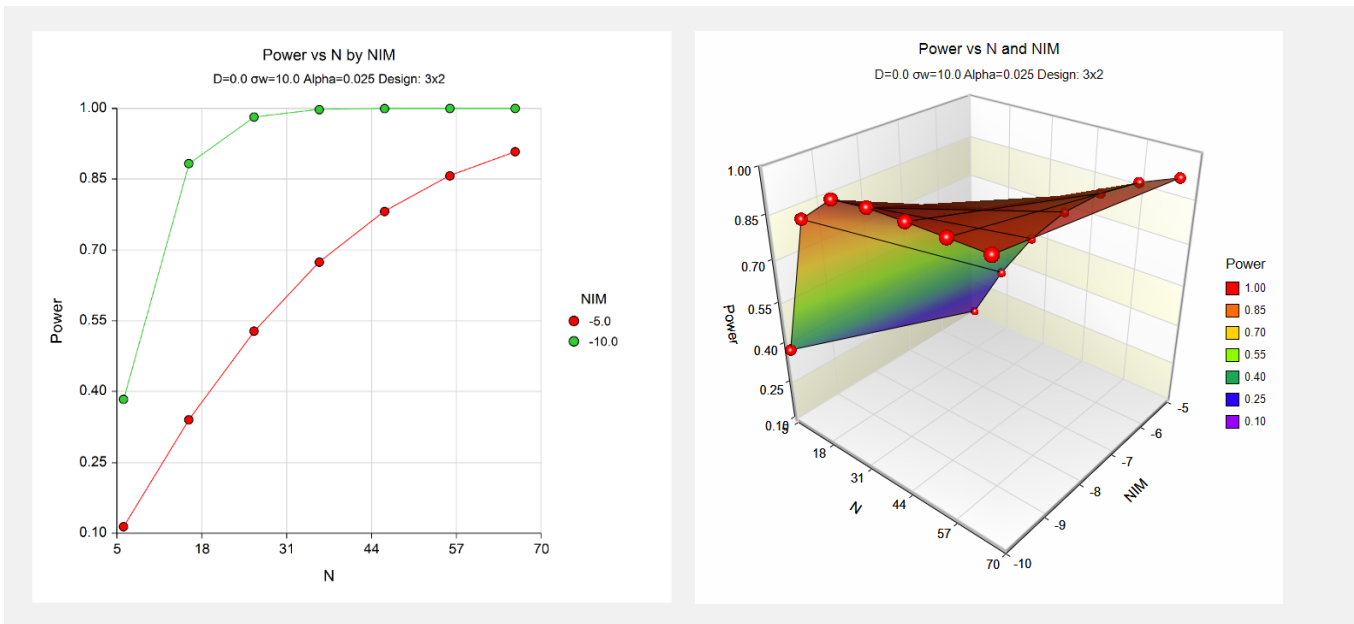
Power is the probability of rejecting H0 (concluding non-inferiority) when H0 is false.
 N is the total number of subjects. They are divided evenly among all sequences.
 μ_T is the treatment mean. It is usually associated with the letter 'A' in the design.
 μ_R is the reference mean. It is usually associated with the letter 'B' in the design.
 NIM is the magnitude and direction of the margin of non-inferiority. Since higher means are better, this value is negative. Non-inferiority means that $\mu_T > \mu_R + NIM$, where $NIM < 0$.
 D ($\mu_T - \mu_R$) is the actual difference between the treatment and reference means that is used in the power calculations.
 σ_w is the square root of the within mean square error from the ANOVA table.
 Alpha is the probability of falsely rejecting H0 (falsely concluding non-inferiority).

Summary Statements

In a non-inferiority test on data for which higher values are better drawn from a three-period, two-sequence dual cross-over design, a total sample size of 6 achieves 11% power at a 2.5% significance level when the true difference between the means is 0.000, the non-inferiority margin is -5.000, and the within-subject standard deviation is 10.0000.

This report shows the power for the indicated scenarios.

Plots Section



These plots show the power versus the sample size.

Example 2 – Finding Sample Size

Continuing with Example 1, the researchers want to find the sample size needed to achieve both 80% power and 90% power when the number of subjects is equal for each sequence.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Non-Inferiority Tests for the Difference of Two Means in a Higher-Order Cross-Over Design** procedure window by expanding **Means**, then **Cross-Over (Higher-Order) Design**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for the Difference of Two Means in a Higher-Order Cross-Over Design**. 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 (Equal Per Sequence)
Design Type.....	3x2 (Three-Period, Two-Sequence Dual: ABB BAA)
Higher Means Are.....	Better
Power.....	0.80 0.90
Alpha.....	0.025
NIM (Non-Inferiority Margin)	5 10
D (Difference)	0
Specify σ as σ_w or σ_b and ρ	σ_w (Within Std Error)
σ_w (Within Std Dev).....	10

Output

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

Numeric Results

Numeric Results for a Non-Inferiority Test of the Mean Difference in a Three-Period, Two-Sequence Dual Design

Treatment Sequences: ABB | BAA

Hypotheses: H0: $D = (\mu_T - \mu_R) \leq \text{NIM}$; H1: $D = (\mu_T - \mu_R) > \text{NIM}$

Higher Means are Better

	Total Sample Size	Non- Inferiority Margin NIM	Mean Difference D ($\mu_T - \mu_R$)	Within Standard Deviation σ_w	Alpha
Power	N				
0.8153	50	-5.000	0.000	10.0000	0.025
0.8343	14	-10.000	0.000	10.0000	0.025
0.9084	66	-5.000	0.000	10.0000	0.025
0.9184	18	-10.000	0.000	10.0000	0.025

When the non-inferiority margin is -5, 66 subjects are needed to achieve 90% power and 50 subjects are needed to achieve at least 80% power.

Non-Inferiority Tests for the Difference of Two Means in a Higher-Order Cross-Over Design

Example 3 – Validation

We could not find a validation example for this procedure in the statistical literature, so we will have to generate a validated example from within **PASS**. To do this, we use the Equivalence Tests for the Difference of Two Means in a Higher-Order Cross-Over Design procedure which was validated. By setting the upper equivalence limit to a large value (we used 22), we obtain results for a non-inferiority test.

A Balaam design is being planned. Suppose the square root of the within mean square error is 0.10, the equivalence limit is 0.20, the difference between the means is 0.05, the power is 90%, and the significance level is 0.05 (see the Example4 template). **PASS** calculates a sample size of 16.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the **PASS** Home window, load the **Non-Inferiority Tests for the Difference of Two Means in a Higher-Order Cross-Over Design** procedure window by expanding **Means**, then **Cross-Over (Higher-Order) Design**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for the Difference of Two Means in a Higher-Order Cross-Over Design**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size (Equal Per Sequence)
Design Type.....	2x4 (Balaam: AA BB AB BA)
Higher Means Are.....	Better
Power.....	0.90
Alpha.....	0.05
NIM (Non-Inferiority Margin)	0.2
D (Difference)	0.05
Specify σ as σ_w or σ_b and ρ	σ_w (Within Std Error)
σ_w (Within Std Dev).....	0.10

Output

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

Numeric Results

Numeric Results for a Non-Inferiority Test of the Mean Difference in a Two-Period, Four-Sequence (Balaam) Design

Treatment Sequences: AA | BB | AB | BA

Hypotheses: $H_0: D = (\mu_T - \mu_R) \leq \text{NIM}$; $H_1: D = (\mu_T - \mu_R) > \text{NIM}$

Higher Means are Better

	Total Sample Size	Non- Inferiority Margin	Mean Difference	Within Standard Deviation	Alpha
Power	N	NIM	D ($\mu_T - \mu_R$)	σ_w	
0.9495	16	-0.200	0.050	0.1000	0.050

PASS has also obtained a sample size of 16 using the non-inferiority procedure.