Chapter 478

Group-Sequential Non-Inferiority Tests for Two Means (Simulation)

Introduction

This procedure can be used to determine power, sample size and/or boundaries for group sequential non-inferiority tests comparing the means of two groups. The common two-sample T-Test and the Mann-Whitney U test can be simulated in this procedure. Significance and futility boundaries can be produced. The spacing of the looks can be equal or custom specified. Boundaries can be computed based on popular alpha- and beta-spending functions (O'Brien-Fleming, Pocock, Hwang-Shih-DeCani Gamma family, linear) or custom spending functions. Boundaries can also be input directly to verify alpha- and/or beta-spending properties. Futility boundaries can be binding or non-binding. Maximum and average (expected) sample sizes are reported as well as the alpha and/or beta spent and incremental power at each look. Corresponding P-Value boundaries are also given for each boundary statistic. Plots of boundaries are also produced.

Technical Details

This section outlines many of the technical details of the techniques used in this procedure including the simulation summary, the test statistic details, and the use of spending functions.

An excellent text for the background and details of many group-sequential methods is Jennison and Turnbull (2000).

Simulation Procedure

In this procedure, a large number of simulations are used to calculate boundaries and power using the following steps

1. Based on the specified distributions, random samples of size N1 and N2 are generated under the null distribution and under the alternative distribution. These are simulated samples as though the final look is reached.

2. For each sample, test statistics for each look are produced. For example, if N1 and N2 are 100 and there are 5 equally spaced looks, test statistics are generated from the random samples at N1 = N2 = 20, N1 = N2 = 40, N1 = N2 = 60, N1 = N2 = 80, and N1 = N2 = 100 for both null and alternative samples.

3. To generate the first significance boundary, the null distribution statistics of the first look (e.g., at N1 = N2 = 20) are ordered and the percent of alpha to be spent at the first look is determined (using either the alpha-spending function or the input value). The statistic for which the percent of statistics above (or below, as the case may be) that value is equal to the percent of alpha to be spent at the first look is the boundary statistic. It is seen here how important a large number of simulations is to the precision of the boundary estimates.

4. All null distribution samples that are outside the first significance boundary at the first look are removed from consideration for the second look. If binding futility boundaries are also being computed, all null distribution samples with statistics that are outside the first futility boundary are also removed from consideration for the second look. If non-binding futility boundaries are being computed, null distribution samples with statistics outside the first futility boundary are not removed.

5. To generate the second significance boundary, the remaining null distribution statistics of the second look (e.g., at N1 = N2 = 40) are ordered and the percent of alpha to be spent at the second look is determined (again, using either the alpha-spending function or the input value). The percent of alpha to be spent at the second look is multiplied by the total number of simulations to determine the number of the statistic that is to be the second boundary statistic. The statistic for which that number of statistics is above it (or below, as the case may be) is the second boundary statistic. For example, suppose there are initially 1000 simulated samples, with 10 removed at the first look (from, say, alpha spent at Look 1 equal to 0.01), leaving 990 samples considered for the second look. Suppose further that the alpha to be spent at the second look is 0.02. This is multiplied by 1000 to give 20. The 990 still-considered statistics are ordered and the 970th (20 in from 990) statistic is the second boundary.

6. All null distribution samples that are outside the second significance boundary and the second futility boundary, if binding, at the second look are removed from consideration for the third look (e.g., leaving 970 statistics computed at N1 = N2 = 60 to be considered at the third look). Steps 4 and 5 are repeated until the final look is reached.

Futility boundaries are computed in a similar manner using the desired beta-spending function or custom beta-spending values and the alternative hypothesis simulated statistics at each look. For both binding and non-binding futility boundaries, samples for which alternative hypothesis statistics are outside either the significance or futility boundaries of the previous look are excluded from current and future looks.

Because the final futility and significance boundaries are required to be the same, futility boundaries are computed beginning at a small value of beta (e.g., 0.0001) and incrementing beta by that amount until the futility and significance boundaries meet.

When boundaries are entered directly, this procedure uses the null hypothesis and alternative hypothesis simulations to determine the number of test statistics that are outside the boundaries at each look. The cumulative proportion of alternative hypothesis statistics that are outside the significance boundaries is the overall power of the study.

Generating Random Distributions

Two methods are available in *PASS* to simulate random samples. The first method generates the random variates directly, one value at a time. The second method generates a large pool (over 10,000) of random values and then draws the random numbers from this pool. This second method can cut the running time of the simulation by 70%.

As mentioned above, the second method begins by generating a large pool of random numbers from the specified distributions. Each of these pools is evaluated to determine if its mean is within a small relative tolerance (0.0001) of the target mean. If the actual mean is not within the tolerance of the target mean, individual members of the population are replaced with new random numbers if the new random number moves the mean towards its target. Only a few hundred such swaps are required to bring the actual mean to within tolerance of the target mean. This population is then sampled with replacement using the uniform distribution. We have found that this method works well as long as the size of the pool is the maximum of twice the number of simulated samples desired and 10,000.

The Statistical Hypotheses

This section will review the specifics of non-inferiority testing.

Remember that in the usual t-test setting, the null (H0) and alternative (H1) hypotheses for one-sided tests are defined as

$$H_0: \mu_1 - \mu_2 \le D \text{ versus } H_1: \mu_1 - \mu_2 > D$$

Rejecting this test implies that the mean difference is larger than the value *D*. This test is called an *upper-tailed test* because it is rejected in samples in which the difference between the sample means is larger than *D*.

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

$$H_0: \mu_1 - \mu_2 \ge D \text{ versus } H_1: \mu_1 - \mu_2 < D$$

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

Parameter	PASS Input/Output	Interpretation
μ_1	Mean1	*Mean* of population 1. Population 1 is assumed to consist of those who have received the new treatment.
μ_2	Mean2	*Mean* of population 2. Population 2 is assumed to consist of those who have received the reference treatment.
M_{NI}	NIM	*Margin of non-inferiority.* 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	*True difference.* This is the value of $\mu_1 - \mu_2$, the difference between the means. This is the value at which the power is calculated.

Note that the actual values of μ_1 and μ_2 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.

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

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

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

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.

$$H_0 : \mu_1 \le \mu_2 - \left|M_{NI}\right| \quad \text{versus} \quad H_1 : \mu_1 > \mu_2 - \left|M_{NI}\right|$$

$$H_0 : \mu_1 - \mu_2 \le -\left|M_{NI}\right| \quad \text{versus} \quad H_1 : \mu_1 - \mu_2 > -\left|M_{NI}\right|$$

$$H_0 : \delta \le -\left|M_{NI}\right| \qquad \text{versus} \quad H_1 : \delta > -\left|M_{NI}\right|$$

Test Statistics

This section describes the test statistics that are available in this procedure.

Two-Sample T-Test

The two-sample t-test assumes that the data are a simple random sample from a population of normally-distributed values that all have the same mean and variance. This assumption implies that the data are continuous and their distribution is symmetric. The calculation of the t statistic is as follows

$$t_{df} = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{s_{\overline{X}_1 - \overline{X}_2}}$$

where

$$\overline{X}_k = \frac{\sum_{i=1}^{N_k} X_{ki}}{N_k}$$

$$s_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{\sum_{i=1}^{N_1}\left(X_{1i} - \overline{X}_1\right)^2 + \sum_{i=1}^{N_2}\left(X_{2i} - \overline{X}_2\right)^2}{N_1 + N_2 - 2}\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}$$

$$df = N_1 + N_2 - 2$$

The significance of the test statistic is determined by computing the p-value based on the t distribution with degrees of freedom *df*. If this p-value is less than a specified level (often 0.05), the null hypothesis is rejected. Otherwise, no conclusion can be reached.

Mann-Whitney U Test

This test is the nonparametric substitute for the equal-variance t-test. Two key assumptions for this test are that the distributions are at least ordinal and that they are identical under H0. This implies that ties (repeated values) are not acceptable. When ties are present, the approximation provided can be used, but know that the theoretic results no longer hold.

The Mann-Whitney test statistic is defined as follows in Gibbons (1985).

$$z = \frac{W_1 - \frac{N_1(N_1 + N_2 + 1)}{2} + C}{s_W}$$

where

$$W_1 = \sum_{k=1}^{N_1} Rank(X_{1k})$$

The ranks are determined after combining the two samples. The standard deviation is calculated as

$$s_W = \sqrt{\frac{N_1 N_2 (N_1 + N_2 + 1)}{12} - \frac{N_1 N_2 \sum_{i=1}^{} (t_i^3 - t_i)}{12(N_1 + N_2)(N_1 + N_2 - 1)}}$$

where t_i is the number of observations tied at value one, t_2 is the number of observations tied at some value two, and so forth.

The correction factor, C, is 0.5 if the rest of the numerator of z is negative or -0.5 otherwise. The value of z is then compared to the standard normal distribution.

Spending Functions

Spending functions can be used in this procedure to specify the proportion of alpha or beta that is spent at each look without having to specify the proportion directly.

Spending functions have the characteristics that they are increasing and that

$$\alpha(0) = 0$$
$$\alpha(1) = \alpha$$

The last characteristic guarantees a fixed α level when the trial is complete. This methodology is very flexible since neither the times nor the number of analyses must be specified in advance. Only the functional form of $\alpha(\tau)$ must be specified.

PASS provides several popular spending functions plus the ability to enter and analyze your own percents of alpha or beta spent. These are calculated as follows (beta may be substituted for alpha for beta-spending functions):

1. **Hwang-Shih-DeCani (gamma family)** $\alpha\left[\frac{1 - e^{-\gamma t}}{1 - e^{-\gamma}}\right], \gamma \neq 0 ; \quad \alpha t, \gamma = 0$

O'Brien-Fleming Boundaries with Alpha = 0.05
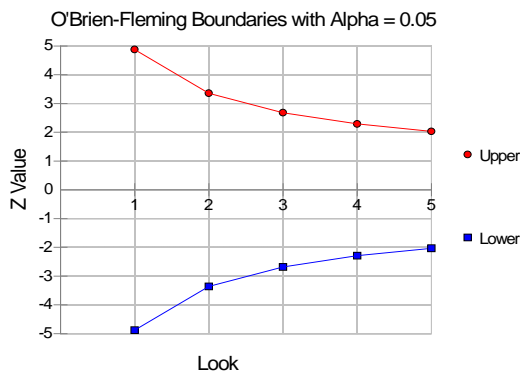
2. O'Brien-Fleming Analog $\quad 2 - 2\Phi\left(\dfrac{Z_{\alpha/2}}{\sqrt{t}}\right)$

O'Brien-Fleming Boundaries with Alpha = 0.05

3. Pocock Analog $\quad \alpha \cdot \ln\left(1 + (e - 1)t\right)$
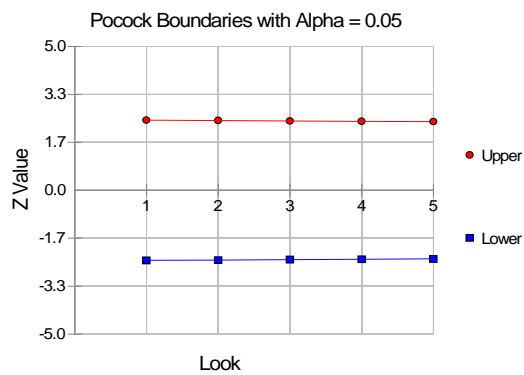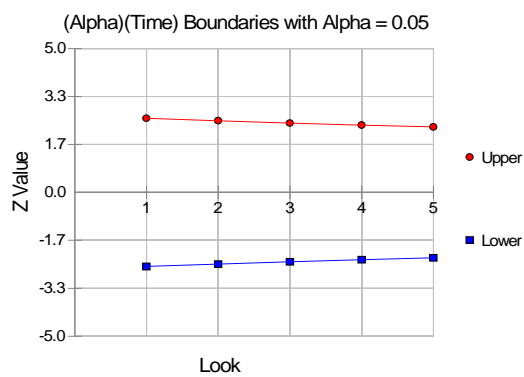
Pocock Boundaries with Alpha = 0.05

4. Alpha * time $\quad \alpha \cdot t$

(Alpha)(Time) Boundaries with Alpha = 0.05

5. Alpha * time^1.5 $\alpha \cdot t^{3/2}$

(Alpha)(Time^1.5) Boundaries with Alpha = 0.05

6. Alpha * time^2 $\alpha \cdot t^{2}$

(Alpha)(Time^2.0) Boundaries with Alpha = 0.05

7. Alpha * time^C $\alpha \cdot t^{C}$

(Alpha)(Time^2.0) Boundaries with Alpha = 0.05
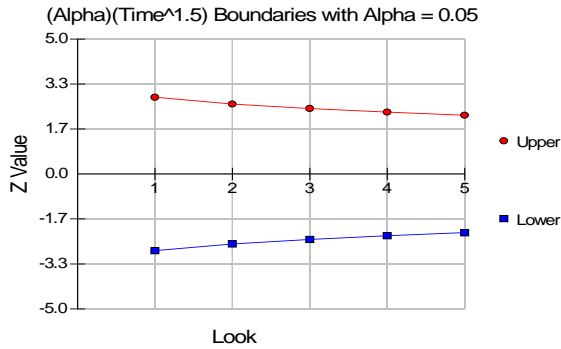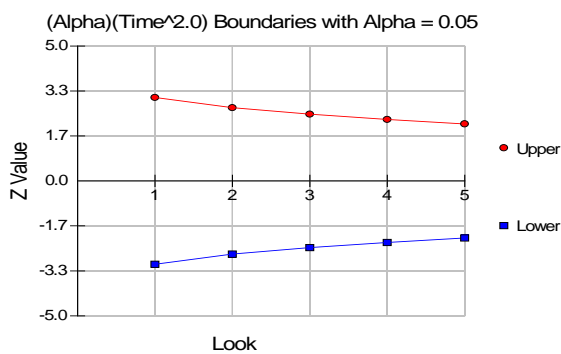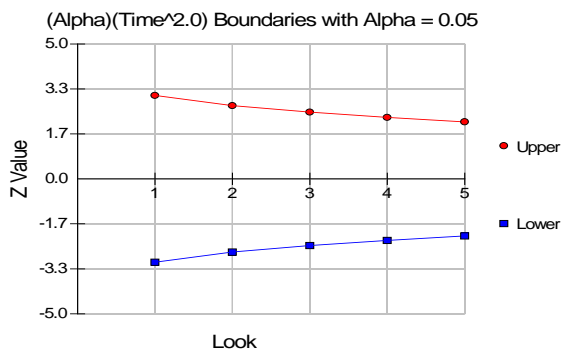
8. User Supplied Percents

A custom set of percents of alpha to be spent at each look may be input directly.

The O'Brien-Fleming Analog spends very little alpha or beta at the beginning and much more at the final looks. The Pocock Analog and (Alpha or Beta)(Time) spending functions spend alpha or beta more evenly across the looks. The Hwang-Shih-DeCani (C) (gamma family) spending functions and (Alpha or Beta)(Time^C) spending functions are flexible spending functions that can be used to spend more alpha or beta early or late or evenly, depending on the choice of C.

Procedure Options

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

Design Tab

The Design tab contains most of the parameters and options for the general setup of the procedure.

Solve For

Solve For

Solve for either power, sample size, or enter the boundaries directly and solve for power and alpha.

When solving for power or N1, the early-stopping boundaries are also calculated. High accuracy for early-stopping boundaries requires a very large number of simulations (Recommended 100,000 to 10,000,000).

The parameter selected here is the parameter displayed on the vertical axis of the plot.

Because this is a simulation based procedure, the search for the sample size may take several minutes or hours to complete. You may find it quicker and more informative to solve for Power for a range of sample sizes.

Test and Simulations

Test Type

Specify which test statistic is to be reported on.

The T-Test is the standard T-test based on the pooled variance and SS1 + SS2 - 2 degrees of freedom, where SS1 and SS2 are the sample sizes of groups 1 and 2 at the corresponding look.

The Mann-Whitney Test is the common non-parametric alternative and is based on the ranks of the observations. It is sometimes called the Mann-Whitney U, Mann-Whitney-Wilcoxon, Wilcoxon Rank-Sum, or Wilcoxon-Mann-Whitney test.

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 Mean1 - Mean2 $\geq$ NIM and the alternative hypothesis is Mean1 - Mean2 < NIM.

If Higher Means Are Worse the null hypothesis is Mean1 - Mean2 $\leq$ -NIM and the alternative hypothesis is Mean1 - Mean2 > -NIM.

Simulations

This option specifies the number of iterations, M, used in the simulation. As the number of iterations is increased, the accuracy and running time of the simulation will be increased also.

The precision of the simulated power estimates are calculated from the binomial distribution. Thus, confidence intervals may be constructed for various power values. The following table gives an estimate of the precision that is achieved for various simulation sizes when the power is either 0.50 or 0.95. The table values are interpreted as follows: a 95% confidence interval of the true power is given by the power reported by the simulation plus and minus the 'Precision' amount given in the table.

Group-Sequential Non-Inferiority Tests for Two Means (Simulation)

Simulation Size M	Precision when Power = 0.50	Precision when Power = 0.95
100	0.100	0.044
500	0.045	0.019
1000	0.032	0.014
2000	0.022	0.010
5000	0.014	0.006
10000	0.010	0.004
50000	0.004	0.002
100000	0.003	0.001

Notice that a simulation size of 1000 gives a precision of plus or minus 0.01 when the true power is 0.95. Also note that as the simulation size is increased beyond 5000, there is only a small amount of additional accuracy achieved.

However, when solving for Power or N1, the simulations are also used to calculate the look boundaries. To obtain precise boundary estimates, the number of simulations needs to be high. This consideration competes with the length of time to complete the simulation. When solving for power, a large number of simulations (100,000 or 1,000,000) will finish in several minutes. When solving for N1, perhaps 10,000 simulations can be run for each iteration. Then, a final run with the resulting N1 solving for power can be run with more simulations.

Power and Alpha

Power

Power is the probability of rejecting the null hypothesis when it is false. Power is equal to 1-Beta, so specifying power implicitly specifies beta.

Beta is the probability obtaining a false negative on the statistical test. That is, it is the probability of accepting a false null hypothesis.

In the context of simulated group sequential trials, the power is the proportion of the alternative hypothesis simulations that cross any one of the significance (efficacy) boundaries.

The valid range is between 0 to 1.

Different disciplines and protocols have different standards for setting power. A common choice is 0.90, but 0.80 is also popular.

You can enter a range of values such as *0.70 0.80 0.90* or *0.70 to 0.90 by 0.1*.

Alpha

Alpha is the probability of obtaining a false positive on the statistical test. That is, it is the probability of rejecting a true null hypothesis.

The null hypothesis is usually that the parameters (the means, proportions, etc.) are all equal.

In the context of simulated group sequential trials, alpha is the proportion of the null hypothesis simulations that cross any one of the significance (efficacy) boundaries.

Since Alpha is a probability, it is bounded by 0 and 1. Commonly, it is between 0.001 and 0.250.

Alpha is often set to 0.05 for two-sided tests and to 0.025 for one-sided tests.

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 (When Solving for Sample Size)

Group Allocation

Select the option that describes the constraints on *N1* or *N2* or both.

The options are

- **Equal (N1 = N2)**

    This selection is used when you wish to have equal sample sizes in each group. Since you are solving for both sample sizes at once, no additional sample size parameters need to be entered.

- **Enter N2, solve for N1**

    Select this option when you wish to fix *N2* at some value (or values), and then solve only for *N1*. Please note that for some values of *N2*, there may not be a value of *N1* that is large enough to obtain the desired power.

- **Enter R = N2/N1, solve for N1 and N2**

    For this choice, you set a value for the ratio of *N2* to *N1*, and then PASS determines the needed *N1* and *N2*, with this ratio, to obtain the desired power. An equivalent representation of the ratio, *R*, is

$$N2 = R * N1.$$

- **Enter percentage in Group 1, solve for N1 and N2**

    For this choice, you set a value for the percentage of the total sample size that is in Group 1, and then PASS determines the needed *N1* and *N2* with this percentage to obtain the desired power.

N2 (Sample Size, Group 2)

This option is displayed if Group Allocation = "Enter N2, solve for N1"

N2 is the number of items or individuals sampled from the Group 2 population.

N2 must be ≥ 2. You can enter a single value or a series of values.

R (Group Sample Size Ratio)

This option is displayed only if Group Allocation = "Enter R = N2/N1, solve for N1 and N2."

R is the ratio of *N2* to *N1*. That is,

$$R = N2 / N1.$$

Use this value to fix the ratio of *N2* to *N1* while solving for *N1* and *N2*. Only sample size combinations with this ratio are considered.

N2 is related to *N1* by the formula:

$$N2 = [R \times N1],$$

where the value *[Y]* is the next integer $\geq Y$.

For example, setting *R* = 2.0 results in a Group 2 sample size that is double the sample size in Group 1 (e.g., *N1* = 10 and *N2* = 20, or *N1* = 50 and *N2* = 100).

R must be greater than 0. If *R* < 1, then *N2* will be less than *N1*; if *R* > 1, then *N2* will be greater than *N1*. You can enter a single or a series of values.

Percent in Group 1

This option is displayed only if Group Allocation = "Enter percentage in Group 1, solve for N1 and N2."

Use this value to fix the percentage of the total sample size allocated to Group 1 while solving for *N1* and *N2*. Only sample size combinations with this Group 1 percentage are considered. Small variations from the specified percentage may occur due to the discrete nature of sample sizes.

The Percent in Group 1 must be greater than 0 and less than 100. You can enter a single or a series of values.

Sample Size (When <u>Not</u> Solving for Sample Size)

Group Allocation

Select the option that describes how individuals in the study will be allocated to Group 1 and to Group 2.

The options are

- **Equal (N1 = N2)**

  This selection is used when you wish to have equal sample sizes in each group. A single per group sample size will be entered.

- **Enter N1 and N2 individually**

  This choice permits you to enter different values for *N1* and *N2*.

- **Enter N1 and R, where N2 = R * N1**

  Choose this option to specify a value (or values) for *N1*, and obtain *N2* as a ratio (multiple) of *N1*.

- **Enter total sample size and percentage in Group 1**

  Choose this option to specify a value (or values) for the total sample size (*N*), obtain *N1* as a percentage of *N*, and then *N2* as *N - N1*.

Sample Size Per Group

This option is displayed only if Group Allocation = "Equal (N1 = N2)."

The Sample Size Per Group is the number of items or individuals sampled from each of the Group 1 and Group 2 populations. Since the sample sizes are the same in each group, this value is the value for *N1*, and also the value for *N2*.

The Sample Size Per Group must be ≥ 2. You can enter a single value or a series of values.

N1 (Sample Size, Group 1)

*This option is displayed if Group Allocation = "Enter N1 and N2 individually" or "Enter N1 and R, where N2 = R * N1."*

N1 is the number of items or individuals sampled from the Group 1 population.

N1 must be ≥ 2. You can enter a single value or a series of values.

N2 (Sample Size, Group 2)

This option is displayed only if Group Allocation = "Enter N1 and N2 individually."

N2 is the number of items or individuals sampled from the Group 2 population.

N2 must be ≥ 2. You can enter a single value or a series of values.

R (Group Sample Size Ratio)

*This option is displayed only if Group Allocation = "Enter N1 and R, where N2 = R * N1."*

R is the ratio of *N2* to *N1*. That is,

$$R = N2/N1$$

Use this value to obtain *N2* as a multiple (or proportion) of *N1*.

N2 is calculated from *N1* using the formula:

$$N2=[R \ x \ N1],$$

where the value *[Y]* is the next integer $\geq$ *Y*.

For example, setting *R* = 2.0 results in a Group 2 sample size that is double the sample size in Group 1.

R must be greater than 0. If *R* < 1, then *N2* will be less than N1; if *R* > 1, then *N2* will be greater than *N1*. You can enter a single value or a series of values.

Total Sample Size (N)

This option is displayed only if Group Allocation = "Enter total sample size and percentage in Group 1."

This is the total sample size, or the sum of the two group sample sizes. This value, along with the percentage of the total sample size in Group 1, implicitly defines *N1* and *N2*.

The total sample size must be greater than one, but practically, must be greater than 3, since each group sample size needs to be at least 2.

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

Percent in Group 1

This option is displayed only if Group Allocation = "Enter total sample size and percentage in Group 1."

This value fixes the percentage of the total sample size allocated to Group 1. Small variations from the specified percentage may occur due to the discrete nature of sample sizes.

The Percent in Group 1 must be greater than 0 and less than 100. You can enter a single value or a series of values.

Effect Size

Mean1 (Mean of Group 1, Control)

Enter a value for the hypothesized mean of this group. The difference between the two means and the direction of the difference relative to the alternative hypothesis is of primary importance here.

For the simulations under the null and alternative hypothesis, the test statistics are calculated as mean of group 1 minus mean of group 2. For the purposes of simulation, if a two-sided alternative is selected, Mean1 should be larger than Mean2.

If you want to analyze the difference, enter the value of the difference for one of the means and zero for the other.

You may enter a range of values such as *10 20 30* or *0 to 100 by 25*.

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.

The % box may be checked to specify NIM as a percentage of Mean1.

% (NIM as a Percent of Mean1)

If this box is checked, the non-inferiority margin will be generated as a percent of Mean1.

Mean2 (Mean of Group 2, Treatment | H1)

Enter a value for the hypothesized mean of this group. The difference between the two means and the direction of the difference relative to the alternative hypothesis is of primary importance here.

For the simulations under the null and alternative hypothesis, the test statistics are calculated as mean of group 1 minus mean of group 2. For the purposes of simulation, if a two-sided alternative is selected, Mean1 should be larger than Mean2.

If you want to analyze the difference, enter the value of the difference for one of the means and zero for the other.

You may enter a range of values such as *10 20 30* or *0 to 100 by 25*.

Standard Deviation

Enter the desired value for the standard deviation of the population from which simulations will be generated. The standard deviation must be a positive number.

This standard deviation is the standard deviation of each group.

Press the "SD" button to obtain help on estimating the standard deviation.

You can enter a range of values such as *1 2 3* or *1 to 10 by 1*.

Looks & Boundaries Tab when Solving for Power or Sample Size

The Looks & Boundaries tab contains settings for the looks and significance boundaries.

Looks and Boundaries

Specification of Looks and Boundaries

Choose whether spending functions will be used to divide alpha and beta for each look (Simple Specification), or whether the percents of alpha and beta to be spent at each look will be specified directly (Custom Specification).

Under Simple Specification, the looks are automatically considered to be equally spaced. Under Custom Specification, the looks may be equally spaced or custom defined based on the percent of accumulated information.

Looks and Boundaries – Simple Specification

Number of Equally Spaced Looks

Select the total number of looks that will be used if the study is not stopped early for the crossing of a boundary.

Alpha Spending Function

Specify the type of alpha spending function to use.

The O'Brien-Fleming Analog spends very little alpha at the beginning and much more at the final looks. The Pocock Analog and (Alpha)(Time) spending functions spend alpha more evenly across the looks. The Hwang-Shih-DeCani (C) (sometimes called the gamma family) spending functions and (Alpha)(Time^C) spending functions are flexible spending functions that can be used to spend more alpha early or late or evenly, depending on the choice of C.

C (Alpha Spending)

C is used to define the Hwang-Shih-DeCani (C) or (Alpha)(Time^C) spending functions.

For the Hwang-Shih-DeCani (C) spending function, negative values of C spend more of alpha at later looks, values near 0 spend alpha evenly, and positive values of C spend more of alpha at earlier looks.

For the (Alpha)(Time^C) spending function, only positive values for C are permitted. Values of C near zero spend more of alpha at earlier looks, values near 1 spend alpha evenly, and larger values of C spend more of alpha at later looks.

Type of Futility Boundary

This option determines whether or not futility boundaries will be created, and if so, whether they are binding or non-binding.

Futility boundaries are boundaries such that, if crossed at a given look, stop the study in favor of H0.

Binding futility boundaries are computed in concert with significance boundaries. They are called binding because they require the stopping of a trial if they are crossed. If the trial is not stopped, the probability of a false positive will exceed alpha.
When Non-binding futility boundaries are computed, the significance boundaries are first computed, ignoring the futility boundaries. The futility boundaries are then computed. These futility boundaries are non-binding because continuing the trial after they are crossed will not affect the overall probability of a false positive declaration.

Number of Skipped Futility Looks

In some trials it may be desirable to wait a number of looks before examining the trial for futility. This option allows the beta to begin being spent after a specified number of looks.

The Number of Skipped Futility Looks should be less than the number of looks.

Beta Spending Function

Specify the type of beta spending function to use.

The O'Brien-Fleming Analog spends very little beta at the beginning and much more at the final looks. The Pocock Analog and (Beta)(Time) spending functions spend beta more evenly across the looks. The Hwang-Shih-DeCani (C) (sometimes called the gamma family) spending functions and (Beta)(Time^C) spending functions are flexible spending functions that can be used to spend more beta early or late or evenly, depending on the choice of C.

C (Beta Spending)

C is used to define the Hwang-Shih-DeCani (C) or (Beta)(Time^C) spending functions.

For the Hwang-Shih-DeCani (C) spending function, negative values of C spend more of beta at later looks, values near 0 spend beta evenly, and positive values of C spend more of beta at earlier looks.

For the (Beta)(Time^C) spending function, only positive values for C are permitted. Values of C near zero spend more of beta at earlier looks, values near 1 spend beta evenly, and larger values of C spend more of beta at later looks.

Looks and Boundaries – Custom Specification

Number of Looks

This is the total number of looks of either type (significance or futility or both).

Equally Spaced

If this box is checked, the Accumulated Information boxes are ignored and the accumulated information is evenly spaced.

Type of Futility Boundary

This option determines whether or not futility boundaries will be created, and if so, whether they are binding or non-binding.

Futility boundaries are boundaries such that, if crossed at a given look, stop the study in favor of H0.

Binding futility boundaries are computed in concert with significance boundaries. They are called binding because they require the stopping of a trial if they are crossed. If the trial is not stopped, the probability of a false positive will exceed alpha.

When Non-binding futility boundaries are computed, the significance boundaries are first computed, ignoring the futility boundaries. The futility boundaries are then computed. These futility boundaries are non-binding because continuing the trial after they are crossed will not affect the overall probability of a false positive declaration.

Accumulated Information

The accumulated information at each look defines the proportion or percent of the sample size that is used at that look.

These values are accumulated information values so they must be increasing.

Proportions, percents, or sample sizes may be entered. All proportions, percents, or sample sizes will be divided by the value at the final look to create an accumulated information proportion for each look.

Percent of Alpha Spent

This is the percent of the total alpha that is spent at the corresponding look. It is not the cumulative value.

Percents, proportions, or alphas may be entered here. Each of the values is divided by the sum of the values to obtain the proportion of alpha that is used at the corresponding look.

Percent of Beta Spent

This is the percent of the total beta (1-power) that is spent at the corresponding look. It is not the cumulative value.

Percents, proportions, or betas may be entered here. Each of the values is divided by the sum of the values to obtain the proportion of beta that is used at the corresponding look.

Looks & Boundaries Tab when Solving for Alpha and Power

The Looks & Boundaries tab contains settings for the looks and significance boundaries.

Looks and Boundaries

Number of Looks

This is the total number of looks of either type (significance or futility or both).

Equally Spaced

If this box is checked, the Accumulated Information boxes are ignored and the accumulated information is evenly spaced.

Types of Boundaries

This option determines whether or not futility boundaries will be entered.

Futility boundaries are boundaries such that, if crossed at a given look, stop the study in favor of H0.

Accumulated Information

The accumulated information at each look defines the proportion or percent of the sample size that is used at that look.

These values are accumulated information values so they must be increasing.

Proportions, percents, or sample sizes may be entered. All proportions, percents, or sample sizes will be divided by the value at the final look to create an accumulated information proportion for each look.

Significance Boundary

Enter the value of the significance boundary corresponding to the chosen test statistic. These are sometimes called efficacy boundaries.

Futility Boundary

Enter the value of the futility boundary corresponding to the chosen test statistic.

Options Tab

The Options tab contains limits on the number of iterations and various options about individual tests.

Maximum Sample Size Searched

Maximum N1 Before Search Termination

Specify the maximum N1 before the search for N1 is aborted.

Since simulations for large sample sizes are very computationally intensive and hence time-consuming, this value can be used to stop searches when N1 is larger than reasonable sample sizes for the study.

This applies only when "Solve For" is set to Sample Size (N1).

The procedure uses a binary search when searching for N1. If a value for N1 is tried that exceeds this value, and the power is not reached, a warning message will be shown on the output indicating the desired power was not reached.

We recommend a value of at least 20000.

Random Numbers

Random Number Pool Size

This is the size of the pool of random values from which the random samples will be drawn. Pools should be at least the maximum of 10,000 and twice the number of simulations. You can enter *Automatic* and an appropriate value will be calculated.

If you do not want to draw numbers from a pool, enter 0 here.

Matching Boundaries at Final Look

Beta Search Increment

For each simulation, when futility bounds are computed, the appropriate beta is found by searching from 0 to 1 by this increment. Smaller increments are more refined, but the search takes longer.

We recommend 0.001 or 0.0001.

Example 1 – Power and Output

A clinical trial is to be conducted over a two-year period to compare the mean response of a new treatment to that of the current treatment. The current response mean is 108. The researchers would like to determine if the new treatment is not inferior to the current treatment, as it has fewer side effects and is less expensive. The new treatment is considered non-inferior if it has a mean greater than 103. Although the researchers do not know the true mean of the new treatment, they would like to examine the power that is achieved if the mean of the new treatment is also 108. The standard deviation for both groups is assumed to be 20. The sample size at the final look is to be 200 per group. Testing will be done at the 0.05 significance level. A total of five tests are going to be performed on the data as they are obtained. The O'Brien-Fleming (Analog) boundaries will be used

Find the power and test boundaries assuming equal sample sizes per arm.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Group-Sequential Non-Inferiority Tests for Two Means (Simulation)** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Group-Sequential**, and then clicking on **Group-Sequential Non-Inferiority Tests for Two Means (Simulation)**. 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**.

Option	Value
Design Tab	
Solve For	**Power**
Test Type	**T-Test**
Higher Means Are	**Better**
Simulations	**20000**
Alpha	**0.05**
Group Allocation	**Equal (N1 = N2)**
Sample Size Per Group	**200**
Mean1 (Control)	**108**
NIM	**5**
Mean2 (Treatment | H1)	**108**
S	**20**
Looks & Boundaries Tab	
Specification of Looks and Boundaries ..	**Simple**
Number of Equally Spaced Looks	**5**
Alpha Spending Function	**O'Brien-Fleming Analog**
Type of Futility Boundary	**None**

Output

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

Numeric Results and Plots

Scenario 1 Numeric Results for Group Sequential Test of Non-Inferiority
Higher Means are Better
Hypotheses: H0: Mean1 - Mean2 = NIM; H1: Mean1 - Mean2 < NIM
Test Statistic: T-Test
Alpha-Spending Function: O'Brien-Fleming Analog
Beta-Spending Function: None
Futility Boundary Type: None
Number of Looks: 5
Simulations: 20000
Pool Size: 40000

Numeric Summary for Scenario 1

------------------ Power ------------------			------------------------- Alpha -------------------------				
Value	95% LCL	95% UCL	Target	Actual	95% LCL	95% UCL	Beta
0.7938	0.7882	0.7994	0.0500	0.0500	0.0470	0.0530	0.2062

		----- Average Sample Size ----							
		-- Given H0 --		-- Given H1 --		Non-Inf.			Std
N1	N2	Grp1	Grp2	Grp1	Grp2	Margin	Mean1	Mean2	Dev
200	200	198	198	158	158	5.0	108.0	108.0	20.0

Report Definitions
Power is the probability of rejecting a false null hypothesis at one of the looks. It is the total proportion
  of alternative hypothesis simulations that are outside the significance boundaries.
Power 95% LCL and UCL are the lower and upper confidence limits for the power estimate. The width of the
  interval is based on the number of simulations.
Target Alpha is the user-specified probability of rejecting a true null hypothesis. It is the total alpha
  spent.
Actual Alpha is the alpha level that was actually achieved by the experiment. It is the total proportion of
  the null hypothesis simulations that are outside the significance boundaries.
Alpha 95% LCL and UCL are the lower and upper confidence limits for the actual alpha estimate. The width of
  the interval is based on the number of simulations.
Beta is the probability of accepting a false null hypothesis. It is the total proportion of alternative
  hypothesis simulations that do not cross the significance boundaries.
N1 and N2 are the sample sizes of each group if the study reaches the final look.
Average Sample Size Given H0 Grp1 and Grp2 are the average or expected sample sizes of each group if H0 is
  true. These are based on the proportion of null hypothesis simulations that cross the significance or
  futility boundaries at each look.
Average Sample Size Given H1 Grp1 and Grp2 are the average or expected sample sizes of each group if H1 is
  true. These are based on the proportion of alternative hypothesis simulations that cross the significance
  or futility boundaries at each look.
Non-inferiority margin is the distance from the control mean that is still considered non-inferior.
Mean1, Mean2, and Std Dev are the parameters that were set by the user to define the null and alternative
  simulation distributions.

Summary Statements
Group sequential trials with group sample sizes of 200 and 200 at the final look achieve 78%
power to detect a difference of 5.0 at the 0.0500 significance level (alpha) using a one-sided
T-Test.

Group-Sequential Non-Inferiority Tests for Two Means (Simulation)

Accumulated Information Details for Scenario 1

Look	Accumulated Information Percent	-------- Accumulated Sample Size --------		
		Group 1	Group 2	Total
1	20.0	40	40	80
2	40.0	80	80	160
3	60.0	120	120	240
4	80.0	160	160	320
5	100.0	200	200	400

Accumulated Information Details Definitions
Look is the number of the look.
Accumulated Information Percent is the percent of the sample size accumulated up to the corresponding look.
Accumulated Sample Size Group 1 is total number of individuals in group 1 at the corresponding look.
Accumulated Sample Size Group 2 is total number of individuals in group 2 at the corresponding look.
Accumulated Sample Size Total is total number of individuals in the study (group 1 + group 2) at the
  corresponding look.

Boundaries for Scenario 1

Look	-- Significance Boundary --	
	T-Value Scale	P-Value Scale
1	-4.6117	0.0000
2	-2.9561	0.0018
3	-2.3362	0.0102
4	-1.9639	0.0252
5	-1.7773	0.0381

Boundaries Definitions
Look is the number of the look.
Significance Boundary T-Value Scale is the value such that statistics outside this boundary at the
  corresponding look indicate termination of the study and rejection of the null hypothesis. They are
  sometimes called efficacy boundaries.
Significance Boundary P-Value Scale is the value such that P-Values outside this boundary at the corresponding
  look indicate termination of the study and rejection of the null hypothesis. This P-Value corresponds to
  the T-Value Boundary and is sometimes called the nominal alpha.

Boundary Plot
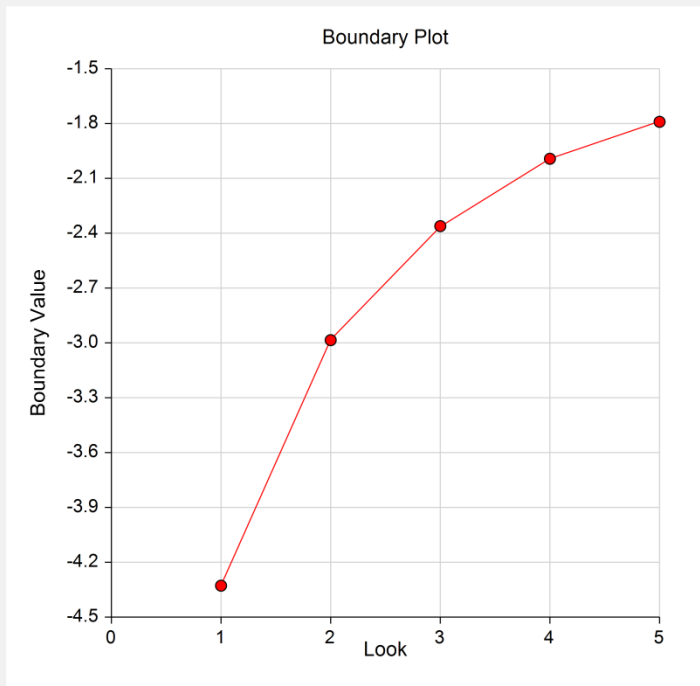
Boundary Plot

Group-Sequential Non-Inferiority Tests for Two Means (Simulation)

Boundary Plot - P-Value
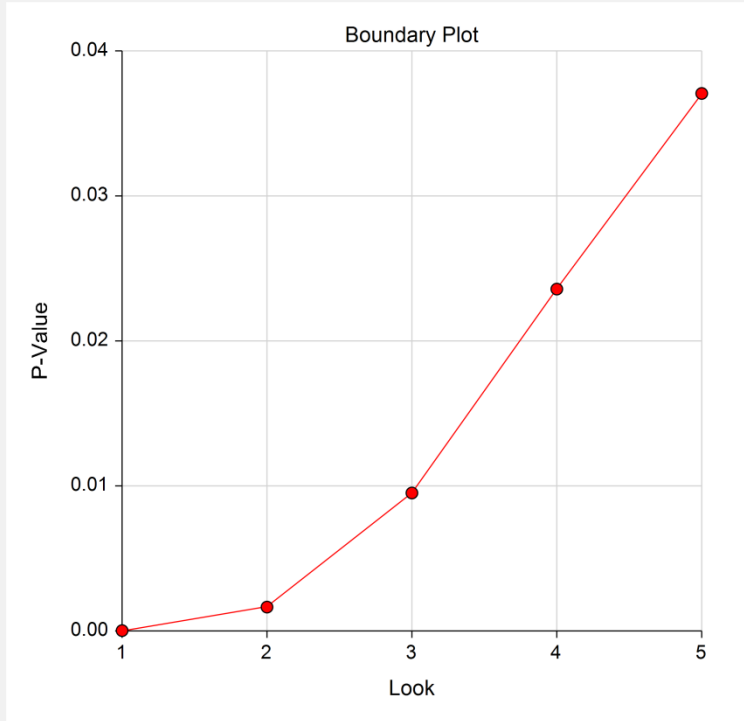
Significance Boundaries with 95% Simulation Confidence Intervals for Scenario 1

Look	--------- T-Value Boundary ---------			--------- P-Value Boundary ---------		
	Value	95% LCL	95% UCL	Value	95% LCL	95% UCL
1	-4.6117			0.0000		
2	-2.9561	-3.1336	-2.8627	0.0018	0.0010	0.0024
3	-2.3362	-2.4246	-2.2891	0.0102	0.0080	0.0115
4	-1.9639	-1.9969	-1.9248	0.0252	0.0233	0.0276
5	-1.7773	-1.8062	-1.7478	0.0381	0.0358	0.0406

Significance Boundary Confidence Limit Definitions
Look is the number of the look.
Look is the number of the look.
T-Value Boundary Value is the value such that statistics outside this boundary at the corresponding look
   indicate termination of the study and rejection of the null hypothesis. They are sometimes called efficacy
   boundaries.
P-Value Boundary Value is the value such that P-Values outside this boundary at the corresponding look
   indicate termination of the study and rejection of the null hypothesis. This P-Value corresponds to the
   T-Value Boundary and is sometimes called the nominal alpha.
95% LCL and UCL are the lower and upper confidence limits for the boundary at the given look. The width of the
   interval is based on the number of simulations.

Group-Sequential Non-Inferiority Tests for Two Means (Simulation)

Alpha-Spending and Null Hypothesis Simulation Details for Scenario 1

	--- Signif. Boundary---		--------- Target ---------		--------- Actual ----------		Proportion	Cum.
				Cum.			H1 Sims	H1 Sims
			Spending	Spending		Cum.	Outside	Outside
	T-Value	P-Value	Function	Function	Alpha	Alpha	Signif.	Signif.
Look	Scale	Scale	Alpha	Alpha	Spent	Spent	Boundary	Boundary
1	-4.6117	0.0000	0.0000	0.0000	0.0000	0.0000	0.0007	0.0007
2	-2.9561	0.0018	0.0019	0.0019	0.0020	0.0020	0.0856	0.0862
3	-2.3362	0.0102	0.0095	0.0114	0.0095	0.0114	0.2593	0.3455
4	-1.9639	0.0252	0.0170	0.0284	0.0171	0.0285	0.2699	0.6154
5	-1.7773	0.0381	0.0216	0.0500	0.0216	0.0500	0.1630	0.7783

**Alpha-Spending Details Definitions**

Look is the number of the look.

Significance Boundary T-Value Scale is the value such that statistics outside this boundary at the
   corresponding look indicate termination of the study and rejection of the null hypothesis. They are
   sometimes called efficacy boundaries.

Significance Boundary P-Value Scale is the value such that P-Values outside this boundary at the corresponding
   look indicate termination of the study and rejection of the null hypothesis. This P-Value corresponds to
   the Significance T-Value Boundary and is sometimes called the nominal alpha.

Spending Function Alpha is the intended portion of alpha allocated to the particular look based on the
   alpha-spending function.

Cumulative Spending Function Alpha is the intended accumulated alpha allocated to the particular look. It is
   the sum of the Spending Function Alpha up to the corresponding look.

Alpha Spent is the proportion of the null hypothesis simulations resulting in statistics outside the
   Significance Boundary at this look.

Cumulative Alpha Spent is the proportion of the null hypothesis simulations resulting in Significance Boundary
   termination up to and including this look. It is the sum of the Alpha Spent up to the corresponding look.

Proportion H1 Sims Outside Significance Boundary is the proportion of the alternative hypothesis simulations
   resulting in statistics outside the Significance Boundary at this look. It may be thought of as the
   incremental power.

Cumulative H1 Sims Outside Significance Boundary is the proportion of the alternative hypothesis simulations
   resulting in Significance Boundary termination up to and including this look. It is the sum of the
   Proportion H1 Sims Outside Significance Boundary up to the corresponding look.

The values obtained from any given run of this example will vary slightly due to the variation in simulations.

Example 2 – Power for One-Sided Test with Futility Boundaries

Suppose researchers would like to compare two treatments with a non-inferiority test at each look. Further, suppose they would like to terminate the study early when it can be deemed highly unlikely that the new treatment is non-inferior to the standard. Suppose the control group mean is 108. The researchers wish to know the power of the test if the treatment group mean is also 108. The sample size at the final look is to be 200 per group. Testing will be done at the 0.05 significance level. A total of five tests are going to be performed on the data as they are obtained. The O'Brien-Fleming (Analog) boundaries will be used for both significance and futility boundaries.

Find the power and test boundaries assuming equal sample sizes per arm.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Group-Sequential Non-Inferiority Tests for Two Means (Simulation)** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Group-Sequential**, and then clicking on **Group-Sequential Non-Inferiority Tests for Two Means (Simulation)**. 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**.

Option	Value
Design Tab	
Solve For	**Power**
Test Type...	**T-Test**
Higher Means Are..................................	**Better**
Simulations	**20000**
Alpha..	**0.05**
Group Allocation	**Equal (N1 = N2)**
Sample Size Per Group	**200**
Mean1 (Control).....................................	**108**
NIM	**5**
Mean2 (Treatment | H1)	**108**
S...	**20**
Looks & Boundaries Tab	
Specification of Looks and Boundaries ..	**Simple**
Number of Equally Spaced Looks	**5**
Alpha Spending Function	**O'Brien-Fleming Analog**
Type of Futility Boundary.......................	**Non-Binding**
Number of Skipped Futility Looks...........	**0**
Beta Spending Function	**O'Brien-Fleming Analog**

Output

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

Numeric Results and Plots

Scenario 1 Numeric Results for Group Sequential Test of Non-Inferiority
Higher Means are Better
Hypotheses: H0: Mean1 - Mean2 = NIM; H1: Mean1 - Mean2 < NIM
Test Statistic: T-Test
Alpha-Spending Function: O'Brien-Fleming Analog
Beta-Spending Function: O'Brien-Fleming Analog
Futility Boundary Type: Non-binding
Number of Looks: 5
Simulations: 20000
Pool Size: 40000

Numeric Summary for Scenario 1

------------------ Power -------------------			-------------------------- Alpha --------------------------				
Value	95% LCL	95% UCL	Target	Actual	95% LCL	95% UCL	Beta
0.7305	0.7244	0.7366	0.0500	0.0406	0.0379	0.0433	0.2695

		----- Average Sample Size ----				Non-Inf.			Std
		-- Given H0 --		-- Given H1 --					
N1	N2	Grp1	Grp2	Grp1	Grp2	Margin	Mean1	Mean2	Dev
200	200	101	101	139	139	5.0	108.0	108.0	20.0

Accumulated Information Details for Scenario 1

	Accumulated Information	-------- Accumulated Sample Size --------		
Look	Percent	Group 1	Group 2	Total
1	20.0	40	40	80
2	40.0	80	80	160
3	60.0	120	120	240
4	80.0	160	160	320
5	100.0	200	200	400

Boundaries for Scenario 1

	-- Significance Boundary --		----- Futility Boundary -----	
	T-Value	P-Value	T-Value	P-Value
Look	Scale	Scale	Scale	Scale
1	-3.7891	0.0001	1.0859	0.8596
2	-2.9619	0.0018	-0.1501	0.4404
3	-2.3396	0.0101	-0.8176	0.2072
4	-1.9647	0.0252	-1.2866	0.0996
5	-1.7319	0.0420	-1.7319	0.0420

Group-Sequential Non-Inferiority Tests for Two Means (Simulation)
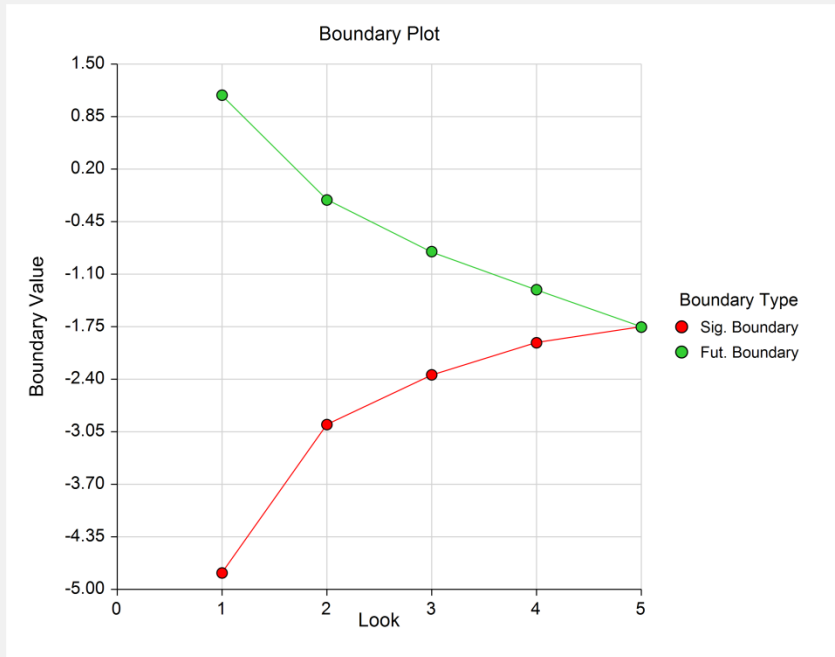
Boundary Plot

Boundary Plot - P-Value
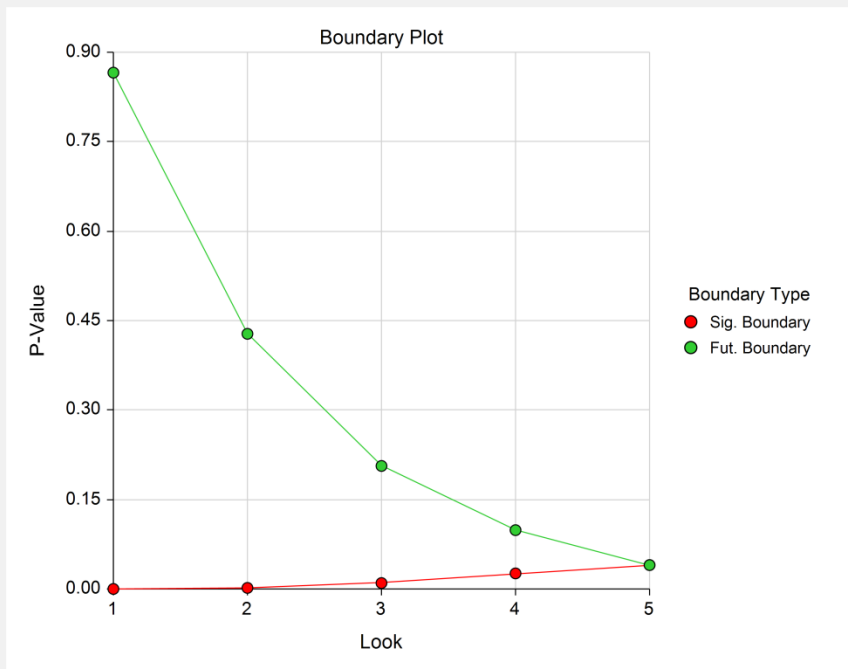
Significance Boundaries with 95% Simulation Confidence Intervals for Scenario 1

	--------- T-Value Boundary ---------			--------- P-Value Boundary ---------		
Look	**Value**	**95% LCL**	**95% UCL**	**Value**	**95% LCL**	**95% UCL**
1	-3.7891			0.0001		
2	-2.9619	-3.1283	-2.8905	0.0018	0.0010	0.0022
3	-2.3396	-2.3970	-2.2897	0.0101	0.0087	0.0115
4	-1.9647	-2.0082	-1.9309	0.0252	0.0227	0.0272
5	-1.7319	-1.7635	-1.7055	0.0420	0.0393	0.0444

Group-Sequential Non-Inferiority Tests for Two Means (Simulation)

Futility Boundaries with 95% Simulation Confidence Intervals for Scenario 1

Look	--------- T-Value Boundary --------- Value	95% LCL	95% UCL	--------- P-Value Boundary --------- Value	95% LCL	95% UCL
1	1.0859	1.0468	1.1278	0.8596		
2	-0.1501	-0.1696	-0.1269	0.4404	0.4328	0.4496
3	-0.8176	-0.8403	-0.7962	0.2072	0.2008	0.2134
4	-1.2866	-1.3064	-1.2633	0.0996	0.0962	0.1037
5	-1.7319	-1.7582	-1.7175	0.0420	0.0397	0.0433

Alpha-Spending and Null Hypothesis Simulation Details for Scenario 1

Look	--- Signif. Boundary--- T-Value Scale	P-Value Scale	-------- Target -------- Spending Function Alpha	Cum. Spending Function Alpha	-------- Actual --------- Alpha Spent	Cum. Alpha Spent	Proportion H0 Sims Outside Futility Boundary	Cum. H0 Sims Outside Futility Boundary
1	-3.7891	0.0001	0.0000	0.0000	0.0000	0.0000	0.1432	0.1432
2	-2.9619	0.0018	0.0019	0.0019	0.0020	0.0020	0.4281	0.5713
3	-2.3396	0.0101	0.0095	0.0114	0.0095	0.0114	0.2424	0.8137
4	-1.9647	0.0252	0.0170	0.0284	0.0165	0.0279	0.1016	0.9153
5	-1.7319	0.0420	0.0216	0.0500	0.0128	0.0406	0.0442	0.9594

Beta-Spending and Alternative Hypothesis Simulation Details for Scenario 1

Look	-- Futility Boundary-- T-Value Scale	P-Value Scale	-------- Target -------- Spending Function Beta	Cum. Spending Function Beta	-------- Actual --------- Beta Spent	Cum. Beta Spent	Proportion H1 Sims Outside Signif. Boundary	Cum. H1 Sims Outside Signif. Boundary
1	1.0859	0.8596	0.0136	0.0136	0.0137	0.0137	0.0052	0.0052
2	-0.1501	0.4404	0.0675	0.0811	0.0675	0.0812	0.0821	0.0873
3	-0.8176	0.2072	0.0733	0.1544	0.0733	0.1545	0.2567	0.3439
4	-1.2866	0.0996	0.0630	0.2175	0.0631	0.2175	0.2674	0.6113
5	-1.7319	0.0420	0.0525	0.2700	0.0520	0.2695	0.1193	0.7305

The values obtained from any given run of this example will vary slightly due to the variation in simulations.

Example 3 – Enter Boundaries

With a set-up similar to Example 2, suppose we wish to investigate the properties of a set of significance (-3, -3, -3, -2, -1) and futility (2, 1, 0, 0, -1) boundaries.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Group-Sequential Non-Inferiority Tests for Two Means (Simulation)** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Group-Sequential**, and then clicking on **Group-Sequential Non-Inferiority Tests for Two Means (Simulation)**. 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**.

Option   **Value**

Design Tab
Solve For .... **Alpha and Power (Enter Boundaries)**
Test Type... **T-Test**
Higher Means Are................................... **Better**
Simulations **20000**
Group Allocation .... **Equal (N1 = N2)**
Sample Size Per Group.......................... **200**
Mean1 (Control)..................................... **108**
NIM **5**
Mean2 (Treatment | H1) **108**
S... **20**

Looks & Boundaries Tab
Number of Looks **5**
Equally Spaced...................................... **Checked**
Types of Boundaries.............................. **Significance and Futility Boundaries**
Significance Boundary............................ **-3 -3 -3 -2 -1 (for looks 1 through 5)**
Futility Boundary .... **2 1 0 0 -1 (for looks 1 through 5)**

Output

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

Numeric Results and Plots

Scenario 1 Numeric Results for Group Sequential Test of Non-Inferiority
Higher Means are Better
Hypotheses: H0: Mean1 - Mean2 = NIM; H1: Mean1 - Mean2 < NIM
Test Statistic: T-Test
Type of Boundaries: Significance and Futility Boundaries
Number of Looks: 5
Simulations: 20000
Pool Size: 40000

Numeric Summary for Scenario 1

------------------- Power -------------------			------------------------- Alpha -------------------------			
Value	**95% LCL**	**95% UCL**	**Value**	**95% LCL**	**95% UCL**	**Beta**
0.9244	0.9207	0.9281	0.1544	0.1493	0.1594	0.0756

		----- Average Sample Size ----							
		-- Given H0 --		-- Given H1 --		Non-Inf.			Std
N1	**N2**	**Grp1**	**Grp2**	**Grp1**	**Grp2**	**Margin**	**Mean1**	**Mean2**	**Dev**
200	200	147	147	162	162	5.0	108.0	108.0	20.0

Accumulated Information Details for Scenario 1

	Accumulated Information	-------- Accumulated Sample Size --------		
Look	**Percent**	**Group 1**	**Group 2**	**Total**
1	20.0	40	40	80
2	40.0	80	80	160
3	60.0	120	120	240
4	80.0	160	160	320
5	100.0	200	200	400

Boundaries for Scenario 1

	-- Significance Boundary --		----- Futility Boundary -----	
	T-Value	**P-Value**	**T-Value**	**P-Value**
Look	**Scale**	**Scale**	**Scale**	**Scale**
1	-3.0000	0.0018	2.0000	0.9755
2	-3.0000	0.0016	1.0000	0.8406
3	-3.0000	0.0015	0.0000	0.5000
4	-2.0000	0.0232	0.0000	0.5000
5	-1.0000	0.1590	-1.0000	0.1590

Group-Sequential Non-Inferiority Tests for Two Means (Simulation)

Boundary Plot

Boundary Plot - P-Value

Alpha-Spending and Null Hypothesis Simulation Details for Scenario 1

| | --- Signif. Boundary--- | | | | Proportion H0 Sims Outside | Cum. H0 Sims Outside |
Look	T-Value Scale	P-Value Scale	Alpha Spent	Cum. Alpha Spent	Futility Boundary	Futility Boundary
1	-3.0000	0.0018	0.0020	0.0020	0.0226	0.0226
2	-3.0000	0.0016	0.0011	0.0030	0.1412	0.1638
3	-3.0000	0.0015	0.0008	0.0038	0.3429	0.5067
4	-2.0000	0.0232	0.0241	0.0279	0.0796	0.5862
5	-1.0000	0.1590	0.1265	0.1544	0.2595	0.8457

Beta-Spending and Alternative Hypothesis Simulation Details for Scenario 1

| | -- Futility Boundary-- | | | | Proportion H1 Sims Outside | Cum. H1 Sims Outside |
Look	T-Value Scale	P-Value Scale	Beta Spent	Cum. Beta Spent	Signif. Boundary	Signif. Boundary
1	2.0000	0.9755	0.0011	0.0011	0.0334	0.0334
2	1.0000	0.8406	0.0040	0.0051	0.0618	0.0951
3	0.0000	0.5000	0.0226	0.0277	0.0779	0.1730
4	0.0000	0.5000	0.0033	0.0310	0.4180	0.5910
5	-1.0000	0.1590	0.0446	0.0756	0.3334	0.9244

The values obtained from any given run of this example will vary slightly due to the variation in simulations.

Example 4 – Validation Using Simulation

With a set-up similar to Example 1, we examine the power and alpha generated by the set of two-sided significance boundaries (-4.6117, -2.9561, -2.3362, -1.9639, -1.7773).

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Group-Sequential Non-Inferiority Tests for Two Means (Simulation)** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Group-Sequential**, and then clicking on **Group-Sequential Non-Inferiority Tests for Two Means (Simulation)**. You may then make the appropriate entries as listed below, or open **Example 4** by going to the **File** menu and choosing **Open Example Template**.

Option   **Value**

Design Tab
Solve For .... **Alpha and Power (Enter Boundaries)**
Test Type... **T-Test**
Higher Means Are.................................... **Better**
Simulations **20000**
Group Allocation .... **Equal (N1 = N2)**
Sample Size Per Group **200**
Mean1 (Control)....................................... **108**
NIM **5**
Mean2 (Treatment | H1) **108**
S.. **20**

Group-Sequential Non-Inferiority Tests for Two Means (Simulation)

Looks & Boundaries Tab

Number of Looks **5**
Equally Spaced .... **Checked**
Types of Boundaries .... **Significance Boundaries**
Significance Boundary .... **-4.6117, -2.9561, -2.3362, -1.9639, -1.7773**

Output

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

Numeric Results and Plots

Scenario 1 Numeric Results for Group Sequential Test of Non-Inferiority
Higher Means are Better
Hypotheses: H0: Mean1 - Mean2 = NIM; H1: Mean1 - Mean2 < NIM
Test Statistic: T-Test
Type of Boundaries: Significance Boundaries Only
Number of Looks: 5
Simulations: 20000
Pool Size: 40000

Numeric Summary for Scenario 1

------------------ Power ------------------			------------------------- Alpha -------------------------			
Value	95% LCL	95% UCL	Value	95% LCL	95% UCL	Beta
0.7851	0.7794	0.7907	0.0458	0.0429	0.0487	0.2150

		----- Average Sample Size ----							
		-- Given H0 --		-- Given H1 --		Non-Inf.			Std
N1	N2	Grp1	Grp2	Grp1	Grp2	Margin	Mean1	Mean2	Dev
200	200	198	198	157	157	5.0	108.0	108.0	20.0

The values obtained from any given run of this example will vary slightly due to the variation in simulations. The power and alpha generated with these boundaries are very close to the values of Example 1.