

## Chapter 226

# Group-Sequential Non-Inferiority Tests for the Difference of Two Proportions (Simulation)

---

### Introduction

This procedure can be used to determine power, sample size and/or boundaries for group sequential non-inferiority tests comparing the difference between proportions from two groups. The tests that can be simulated in this procedure are the common two-sample Z-test with or without pooled standard error and with or without continuity correction, the T-test, and three score tests. 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 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 proportions, random samples of size  $N_1$  and  $N_2$  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  $N_1$  and  $N_2$  are 100 and there are 5 equally spaced looks, test statistics are generated from the random samples at  $N_1 = N_2 = 20$ ,  $N_1 = N_2 = 40$ ,  $N_1 = N_2 = 60$ ,  $N_1 = N_2 = 80$ , and  $N_1 = N_2 = 100$  for both null and alternative samples.
3. To generate the first significance boundary, the null distribution statistics of the first look (e.g., at  $N_1 = N_2 = 20$ ) 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  $N_1 = N_2 = 40$ ) 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 970<sup>th</sup> (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  $N_1 = N_2 = 60$  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.

## Small Sample Considerations

When the sample size is small, say 200 or fewer per group, the discrete nature of the number of possible differences in proportions in the sampling distribution comes into play. This has led to a large number of proposed tests for comparing two proportions (or testing the 2 by 2 table of counts). For example, Upton (1982) considers twenty-two alternative tests for comparing two proportions. Sweeping statements about the power of one test over another are impossible to make, because the size of the Type I error depends upon the proportions used. At some proportions, some tests are overly conservative while others are not, while at other proportions the reverse may be true.

This simulation procedure, however, is based primarily on the ordering of the sample statistics in the simulation. The boundaries are determined by the spending function alphas. Thus, if a test used happens to be conservative in the single-look traditional sense, the boundaries chosen in the simulation results of this procedure will generally remove the conservative nature of the test. This makes comparisons to the one-look case surprising in many cases.

## Definitions

Suppose you have two populations from which dichotomous (binary) responses will be recorded. The probability (or risk) of obtaining the event of interest in population 1 (the treatment group) is  $p_1$  and in population 2 (the control group) is  $p_2$ . The corresponding failure proportions are given by  $q_1 = 1 - p_1$  and  $q_2 = 1 - p_2$ .

The assumption is made that the responses from each group follow a binomial distribution. This means that the event probability,  $p_i$ , is the same for all subjects within the group and that the response from one subject is independent of that of any other subject.

Random samples of  $m$  and  $n$  individuals are obtained from these two populations. The data from these samples can be displayed in a 2-by-2 contingency table as follows

Group	Success	Failure	Total
Treatment	$a$	$c$	$m$
Control	$b$	$d$	$n$
Total	$s$	$f$	$N$

The following alternative notation is also used.

Group	Success	Failure	Total
Treatment	$x_{11}$	$x_{12}$	$n_1$
Control	$x_{21}$	$x_{22}$	$n_2$
Total	$m_1$	$m_2$	$N$

The binomial proportions  $p_1$  and  $p_2$  are estimated from these data using the formulae

$$\hat{p}_1 = \frac{a}{m} = \frac{x_{11}}{n_1} \quad \text{and} \quad \hat{p}_2 = \frac{b}{n} = \frac{x_{21}}{n_2}$$

## Comparing Two Proportions

Let  $p_{1.0}$  represent the group 1 proportion tested by the null hypothesis,  $H_0$ . The power of a test is computed at a specific value of the proportion which we will call  $p_{1.1}$ . Let  $\delta$  represent the smallest difference (margin of non-inferiority) between the two proportions that still results in the conclusion that the new treatment is not inferior to the current treatment. For a non-inferiority test,  $\delta < 0$ . The set of statistical hypotheses that are tested is

$$H_0: p_{1.0} - p_2 \leq \delta \text{ versus } H_1: p_{1.0} - p_2 > \delta$$

which can be rearranged to give

$$H_0: p_{1.0} \leq p_2 + \delta \text{ versus } H_1: p_{1.0} > p_2 + \delta$$

There are multiple methods of specifying the margin of non-inferiority. The most direct is to simply give values for  $p_2$  and  $p_{1.0}$ . However, it is often more meaningful to give  $p_2$  and then specify  $p_{1.0}$  implicitly by specifying the difference, ratio, or odds ratio. Mathematically, assuming higher proportions are better, the definitions of these parameterizations are

<u>Parameter</u>	<u>Computation</u>	<u>Hypotheses</u>
Difference	$\delta = p_{1.0} - p_2$	$H_0: p_{1.0} - p_2 \leq \delta_0$ vs. $H_1: p_{1.0} - p_2 > \delta_0$ , $\delta_0 < 0$
Ratio	$\phi = p_{1.0} / p_2$	$H_0: p_1 / p_2 \leq \phi_0$ vs. $H_1: p_1 / p_2 > \phi_0$ , $\phi_0 < 1$
Odds Ratio	$\psi = Odds_{1.0} / Odds_2$	$H_0: o_{1.0} / o_2 \leq \psi_0$ versus $H_1: o_{1.0} / o_2 > \psi_0$ , $\psi_0 < 1$

## Difference

The difference is perhaps the most direct method of comparison between two proportions. It is easy to interpret and communicate. It gives the absolute impact of the treatment. However, there are subtle difficulties that can arise with its interpretation.

One difficulty arises when the event of interest is rare. If a difference of 0.001 occurs when the baseline probability is 0.40, it would be dismissed as being trivial. However, if the baseline probability of a disease is 0.002, a 0.001 decrease would represent a reduction of 50%. Thus interpretation of the difference depends on the baseline probability of the event.

The following example is used to convey the concept of a *non-inferiority* test. Suppose 60% of patients respond to the current treatment method ( $p_2 = 0.60$ ). If the response rate of the new treatment is no less than 5 percentage points worse ( $\delta = -0.05$ ) than the existing treatment, it will be considered to be non-inferior. Substituting these figures into the statistical hypotheses gives

$$H_0: \delta \leq -0.05 \text{ versus } H_1: \delta > -0.05$$

In this example, when the null hypothesis is rejected, the concluded alternative is that the response rate is at least 55%, which means that the new treatment is not inferior to the current treatment.

## Test Statistics

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

### Z Test (Pooled and Unpooled)

This test statistic was first proposed by Karl Pearson in 1900. Although this test can be expressed as a Chi-Square statistic, it is expressed here as a  $z$  so that it can be used for one-sided hypothesis testing.

Both *pooled* and *unpooled* versions of this test have been discussed in the statistical literature. The pooling refers to the way in which the standard error is estimated. In the pooled version, the two proportions are averaged, and only one proportion is used to estimate the standard error. In the unpooled version, the two proportions are used separately.

The formula for the test statistic is

$$z_t = \frac{\hat{p}_1 - \hat{p}_2}{\hat{\sigma}_D}$$

#### Pooled Version

$$\hat{\sigma}_D = \sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

#### Unpooled Version

$$\hat{\sigma}_D = \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

### Continuity Correction

Frank Yates is credited with proposing a correction to the Pearson Chi-Square test for the lack of continuity in the binomial distribution. However, the correction was in common use when he proposed it in 1922.

The continuity corrected  $z$ -test is

$$z = \frac{(\hat{p}_1 - \hat{p}_2) + \frac{F}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}{\hat{\sigma}_D}$$

where  $F$  is -1 for upper-tailed, 1 for lower-tailed, and either -1 or 1 for two-sided hypotheses, depending on whether the numerator difference is positive or negative.

### T-Test

Based on a study of the behavior of several tests, D'Agostino (1988) and Upton (1982) proposed using the usual two-sample  $t$ -test for testing whether two proportions are equal. One substitutes a '1' for a success and a '0' for a failure in the usual, two-sample  $t$ -test formula. The test statistic is computed as

$$t_{N-2} = (ad - bc) \left( \frac{N - 2}{N(nac + mbd)} \right)^{\frac{1}{2}}$$

which can be compared to the  $t$  distribution with  $N-2$  degrees of freedom.

## Group-Sequential Non-Inferiority Tests for the Difference of Two Proportions (Simulation)

### Miettinen and Nurminen's Likelihood Score Test

Miettinen and Nurminen (1985) proposed a test statistic for testing whether the difference is equal to a specified, non-zero, value,  $\delta_0$ . The regular MLE's,  $\hat{p}_1$  and  $\hat{p}_2$ , are used in the numerator of the score statistic while MLE's  $\tilde{p}_1$  and  $\tilde{p}_2$ , constrained so that  $\tilde{p}_1 - \tilde{p}_2 = \delta_0$ , are used in the denominator. A correction factor of  $N/(N-1)$  is applied to make the variance estimate less biased. The significance level of the test statistic is based on the asymptotic normality of the score statistic. The formula for computing this test statistic is

$$z_{MND} = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\hat{\sigma}_{MND}}$$

where

$$\hat{\sigma}_{MND} = \sqrt{\left( \frac{\tilde{p}_1 \tilde{q}_1}{n_1} + \frac{\tilde{p}_2 \tilde{q}_2}{n_2} \right) \left( \frac{N}{N-1} \right)}$$

$$\tilde{p}_1 = \tilde{p}_2 + \delta_0$$

$$\tilde{p}_1 = 2B \cos(A) - \frac{L_2}{3L_3}$$

$$A = \frac{1}{3} \left[ \pi + \cos^{-1} \left( \frac{C}{B^3} \right) \right]$$

$$B = \text{sign}(C) \sqrt{\frac{L_2^2}{9L_3} - \frac{L_1}{3L_3}}$$

$$C = \frac{L_2^3}{27L_3^3} - \frac{L_1 L_2}{6L_3^2} + \frac{L_0}{2L_3}$$

$$L_0 = x_{21} \delta_0 (1 - \delta_0)$$

$$L_1 = [n_2 \delta_0 - N - 2x_{21}] \delta_0 + m_1$$

$$L_2 = (N + n_2) \delta_0 - N - m_1$$

$$L_3 = N$$

### Farrington and Manning's Likelihood Score Test

Farrington and Manning (1990) proposed a test statistic for testing whether the difference is equal to a specified value  $\delta_0$ . The regular MLE's,  $\hat{p}_1$  and  $\hat{p}_2$ , are used in the numerator of the score statistic while MLE's  $\tilde{p}_1$  and  $\tilde{p}_2$ , constrained so that  $\tilde{p}_1 - \tilde{p}_2 = \delta_0$ , are used in the denominator. The significance level of the test statistic is based on the asymptotic normality of the score statistic.

The formula for computing the test statistic is

$$z_{FMD} = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\sqrt{\left( \frac{\tilde{p}_1 \tilde{q}_1}{n_1} + \frac{\tilde{p}_2 \tilde{q}_2}{n_2} \right)}}$$

## Group-Sequential Non-Inferiority Tests for the Difference of Two Proportions (Simulation)

where the estimates  $\tilde{p}_1$  and  $\tilde{p}_2$  are computed as in the corresponding test of Miettinen and Nurminen (1985) given above.

### Gart and Nam's Likelihood Score Test

Gart and Nam (1990), page 638, proposed a modification to the Farrington and Manning (1988) difference test that corrects for skewness. Let  $z_{FMD}(\delta)$  stand for the Farrington and Manning difference test statistic described above. The skewness corrected test statistic,  $z_{GND}$ , is the appropriate solution to the quadratic equation

$$(-\tilde{\gamma})z_{GND}^2 + (-1)z_{GND} + (z_{FMD}(\delta) + \tilde{\gamma}) = 0$$

where

$$\tilde{\gamma} = \frac{\tilde{V}^{3/2}(\delta)}{6} \left( \frac{\tilde{p}_1 \tilde{q}_1 (\tilde{q}_1 - \tilde{p}_1)}{n_1^2} - \frac{\tilde{p}_2 \tilde{q}_2 (\tilde{q}_2 - \tilde{p}_2)}{n_2^2} \right)$$

---

## 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

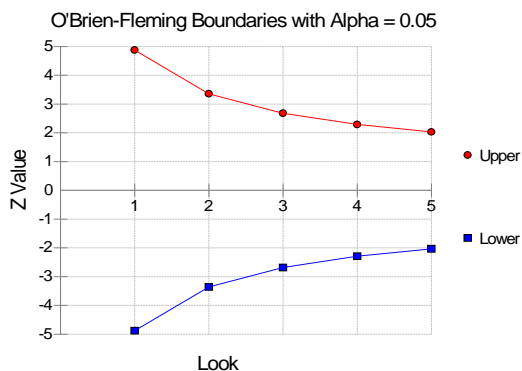
$$\alpha(0) = 0$$

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

The last characteristic guarantees a fixed  $\alpha$  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)  $\alpha \left[ \frac{1 - e^{-\gamma t}}{1 - e^{-\gamma}} \right], \gamma \neq 0; \alpha t, \gamma = 0$

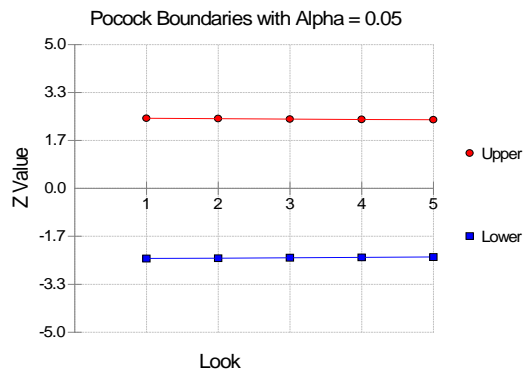


Group-Sequential Non-Inferiority Tests for the Difference of Two Proportions (Simulation)

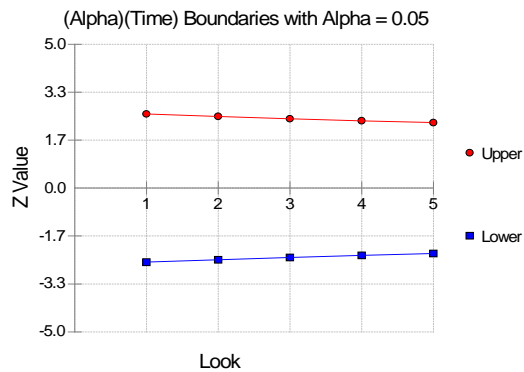
2. O'Brien-Fleming Analog  $2 - 2\Phi\left(\frac{Z_{\alpha/2}}{\sqrt{t}}\right)$



3. Pocock Analog  $\alpha \cdot \ln(1 + (e - 1)t)$

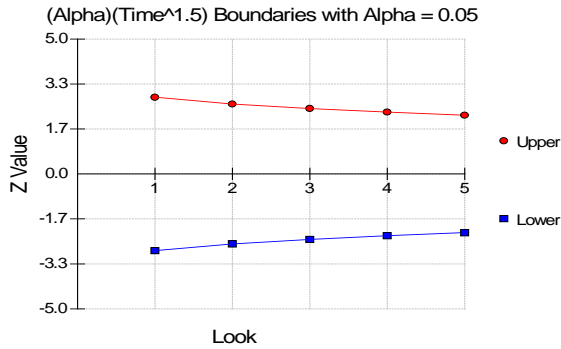
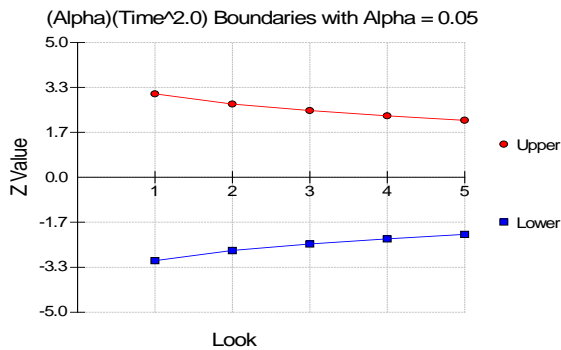
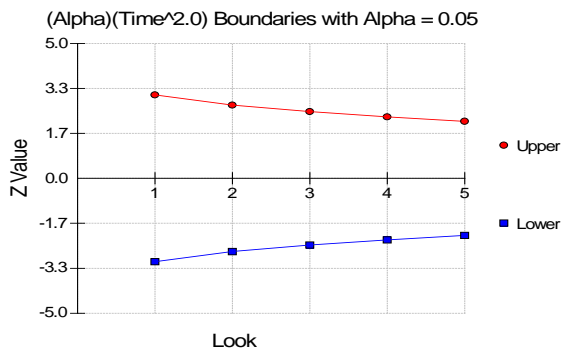


4. Alpha \* time  $\alpha \cdot t$





## Group-Sequential Non-Inferiority Tests for the Difference of Two Proportions (Simulation)

5.  $\text{Alpha} * \text{time}^{1.5}$   $\alpha \cdot t^{3/2}$ 6.  $\text{Alpha} * \text{time}^2$   $\alpha \cdot t^2$ 7.  $\text{Alpha} * \text{time}^C$   $\alpha \cdot t^C$ 

## 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 sample size, 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 simulated and reported on.

For details and formulation of the tests, see the section in this manual above.

##### Higher Proportions Are

Use this option to specify the direction of the test.

If Higher Proportions are "Better", the alternative hypothesis is  $H_1: P_1 - P_2 > NIM$ .

If Higher Proportions are "Worse", the alternative hypothesis is  $H_1: P_1 - P_2 < NIM$ .

##### Simulations

Specify the number of Monte Carlo iterations,  $M$ .

The following table gives an estimate of the precision that is achieved for various simulation sizes when the power is near 0.50 and 0.95. The values in the table are the "Precision" amounts that are added and subtracted to form a 95% confidence interval.

Simulation Size	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

## Group-Sequential Non-Inferiority Tests for the Difference of Two Proportions (Simulation)

However, when solving for power or sample size, the simulations are used to calculate the look boundaries. To obtain precise boundary estimates, the number of simulations needs to be high. However, 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 sample size, perhaps 10,000 simulations can be run for each iteration. Then, a final run with the resulting sample size 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.

## Group-Sequential Non-Inferiority Tests for the Difference of Two Proportions (Simulation)

- **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  $\geq 2$ . 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 Not 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.

## Group-Sequential Non-Inferiority Tests for the Difference of Two Proportions (Simulation)

- **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  $\geq 2$ . 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  $\geq 2$ . 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  $\geq 2$ . 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:

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

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.

$R$  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  $N2$ .

## Group-Sequential Non-Inferiority Tests for the Difference of Two Proportions (Simulation)

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

### Input Type

Indicate what type of values to enter to specify the non-inferiority and actual differences. Regardless of the entry type chosen, the calculations are the same. This option is simply given for convenience in specifying the differences.

---

## Effect Size – Differences (P1 – P2)

*These options are displayed only if Input Type = "Differences"*

### D0 (Non-Inferiority Difference)

Specify the difference between P1.0 and P2 that will be considered non-inferior.

When higher proportions are "Better", the non-inferiority difference is that amount that P1 can be less than P2 and still have the treatment group declared non-inferior to the reference group. When higher proportions are "Better", D0 should be negative.

When higher proportions are "Worse", the non-inferiority difference is that amount that P1 can be greater than P2 and still have the treatment group declared non-inferior to the reference group. When higher proportions are "Worse", D0 should be positive.

The power calculations assume that P1.0 is the value of the P1 under the null hypothesis. This value is used with P2 to calculate the value of P1.0 using the formula:  $P1.0 = D0 + P2$

You may enter a range of values such as  $-.03$   $-.05$   $-.10$  or  $-.05$  to  $-.01$  by  $.01$ .

Differences must be between -1 and 1. D0 cannot take on the values -1, 0, or 1. D0 should be selected such that P1.0 is between zero and one.

### D1 (Actual Difference)

Specify the actual difference between P1.1 (the actual value of P1) and P2. This is the value at which the power is calculated.

In non-inferiority and equivalence trials, this difference is often set to zero.

The power calculations assume that P1.1 is the actual value of the proportion in group 1 (experimental or treatment group). This difference is used with P2 to calculate the value of P1 using the formula:  $P1.1 = D1 + P2$

You may enter a range of values such as  $-.05$   $0.5$  or  $-.05$  to  $.05$  by  $.02$ .

Actual differences must be between -1 and 1. They cannot take on the values -1 or 1.

## Group-Sequential Non-Inferiority Tests for the Difference of Two Proportions (Simulation)

---

### Effect Size – Group 1 (Treatment)

*These options are displayed only if Input Type = "Proportions"*

#### P1.0 (Non-Inferiority Proportion)

Specify the value P1.0 directly.

When higher proportions are "Better", P1.0 is the treatment group proportion such that proportions greater than P1.0 are considered non-inferior to the reference group proportion. When higher proportions are "Better", P1.0 should be less than P2.

When higher proportions are "Worse", P1.0 is the treatment group proportion such that proportions less than P1.0 are considered non-inferior to the reference group proportion. When higher proportions are "Worse", P1.0 should be greater than P2.

When P1.0 is close to zero or one, a Zero Count Adjustment may be needed.

You can enter a list of values such as *0.4 0.5 0.6* or *0.3 to 0.7 by 0.05*.

#### P1.1 (Actual Proportion)

This is the value of the treatment proportion (P1) at which the power calculations are made.

When P1.1 is close to zero or one, a Zero Count Adjustment may be needed.

You can enter a list of values such as *0.4 0.5 0.6* or *0.3 to 0.7 by 0.05*.

---

### Effect Size – Group 2 (Reference)

#### P2 (Group 2 Proportion)

Enter a value for the proportion in the control (baseline, standard, or reference) group, P2.

When P2 is close to zero or one, a Zero Count Adjustment may be needed.

Values must be between 0 and 1.

You can enter a list of values such as *0.4 0.5 0.6* or *0.3 to 0.7 by 0.05*.

---

## 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.

## Group-Sequential Non-Inferiority Tests for the Difference of Two Proportions (Simulation)

### 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<sup>C</sup>) 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<sup>C</sup>) 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<sup>C</sup>) 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 H<sub>0</sub>.

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<sup>C</sup>) 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<sup>C</sup>) 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<sup>C</sup>) 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  $H_0$ .

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 H<sub>0</sub>.

#### 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 Iterations

#### 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 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.

---

### 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.

---

### Zero Count Adjustment

#### Zero Count Adjustment Method

Zero cell counts cause many calculation problems. To compensate for this, a small value (called the Zero Count Adjustment Value) may be added either to all cells or to all cells with zero counts. This option specifies whether you want to use the adjustment and which type of adjustment you want to use.

Adding a small value is controversial, but may be necessary. Some statisticians recommend adding 0.5 while others recommend 0.25. We have found that adding values as small as 0.0001 seems to work well.

#### Zero Count Adjustment Value

Zero cell counts cause many calculation problems. To compensate for this, a small value may be added either to all cells or to all zero cells. This is the amount that is added.

Some statisticians recommend that the value of 0.5 be added to all cells (both zero and non-zero). Others recommend 0.25. We have found that even a value as small as 0.0001 works well.

## Example 1 – Power and Output

A clinical trial is to be conducted over a two-year period to compare the proportion response of a new treatment to that of the current treatment. The current response proportion is 0.58. The researchers would like to show that the new treatment is no worse than 0.05 worse than the standard treatment. Although the researchers do not know the true proportion of patients that will survive with the new treatment, they would like to examine the power that is achieved if the proportion under the new treatment is also 0.58. The sample size at the final look is to be 1000 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 the Difference of Two Proportions (Simulation)** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Group-Sequential**, and then clicking on the **Group-Sequential Non-Inferiority Tests for the Difference of Two Proportions (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**.

<u>Option</u>	<u>Value</u>
<b>Design Tab</b>	
Solve For .....	<b>Power</b>
Test Type .....	<b>Z-Test (Pooled)</b>
Higher Proportions Are .....	<b>Better</b>
Simulations .....	<b>100000</b>
Alpha .....	<b>0.05</b>
Group Allocation .....	<b>Equal (N1 = N2)</b>
Sample Size Per Group .....	<b>1000</b>
Input Type .....	<b>Proportions</b>
P1.0 (Non-Inferiority Proportion) .....	<b>0.53</b>
P1.1 (Actual Proportion) .....	<b>0.58</b>
P2 (Group 2 Proportion) .....	<b>0.58</b>
<b>Looks &amp; Boundaries Tab</b>	
Specification of Looks and Boundaries ..	<b>Simple</b>
Number of Equally Spaced Looks .....	<b>5</b>
Alpha Spending Function .....	<b>O'Brien-Fleming Analog</b>
Type of Futility Boundary .....	<b>None</b>

## Group-Sequential Non-Inferiority Tests for the Difference of Two Proportions (Simulation)

**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 Testing Proportion Difference = NIM.**

Hypotheses: H0: Proportion 1 - Proportion 2 = NIM; H1: Proportion 1 - Proportion 2 > NIM

Test Statistic: Z-Test (Pooled)

Zero Adjustment Method: None

Alpha-Spending Function: O'Brien-Fleming Analog

Beta-Spending Function: None

Futility Boundary Type: None

Number of Looks: 5

Simulations: 100000

**Numeric Summary for Scenario 1**

Power				Alpha				
Value	95% LCL	95% UCL	Target	Actual	95% LCL	95% UCL	Beta	
0.718	0.715	0.721	0.050	0.050	0.049	0.051	0.282	

Average Sample Size											
-- Given H0 --											
-- Given H1 --											
N1	N2	Grp1	Grp2	Grp1	Grp2	D0	D1	P1.0	P1.1	P2	
1000	1000	992	992	819	819	-0.05	0.00	0.53	0.58	0.58	

**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.

Alpha or 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.

D0, or non-inferiority difference, is the proportion difference between groups (Grp1 - Grp2) assuming the null hypothesis, H0.

D1 is the proportion difference between groups (Grp1 - Grp2) assuming the alternative hypothesis, H1.

P1.0 is the proportion used in the simulations for Group 1 under H0.

P1.1 is the proportion used in the simulations for Group 1 under H1.

P2 is the proportion used in the simulations for Group 2 under H0 and H1.

**Summary Statements**

Group sequential trials with sample sizes of 1000 and 1000 at the final look achieve 72% power to detect a difference of 0.00 between a treatment group proportion of 0.58 and a control group proportion of 0.58 with a non-inferiority margin of -0.05 at the 0.050 significance level (alpha) using a one-sided Z-Test (Pooled).

## Group-Sequential Non-Inferiority Tests for the Difference of Two Proportions (Simulation)

### Accumulated Information Details for Scenario 1

Look	Accumulated Information Percent	----- Accumulated Sample Size -----		
		Group 1	Group 2	Total
1	20.0	200	200	400
2	40.0	400	400	800
3	60.0	600	600	1200
4	80.0	800	800	1600
5	100.0	1000	1000	2000

### 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.

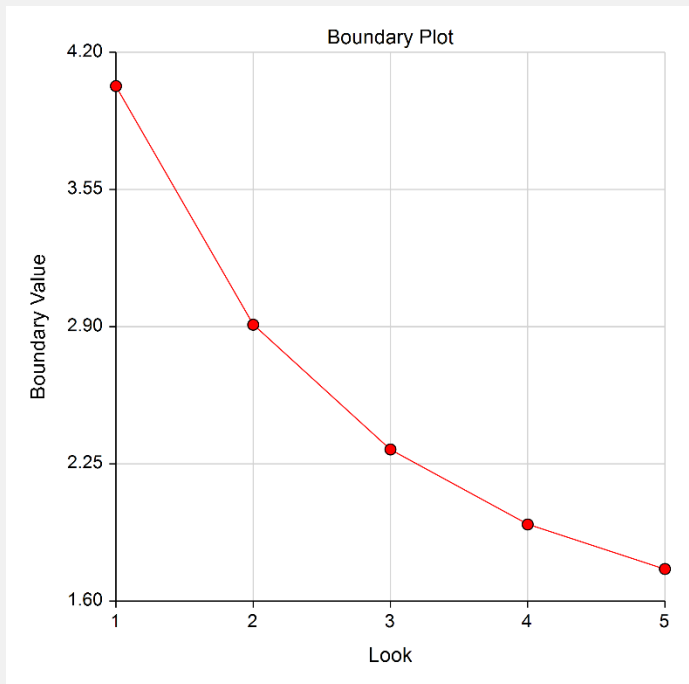
### Boundaries for Scenario 1

Look	-- Significance Boundary --	
	Z-Value Scale	P-Value Scale
1	4.03979	0.00003
2	2.90905	0.00181
3	2.31812	0.01022
4	1.96274	0.02484
5	1.75172	0.03991

### Boundaries Definitions

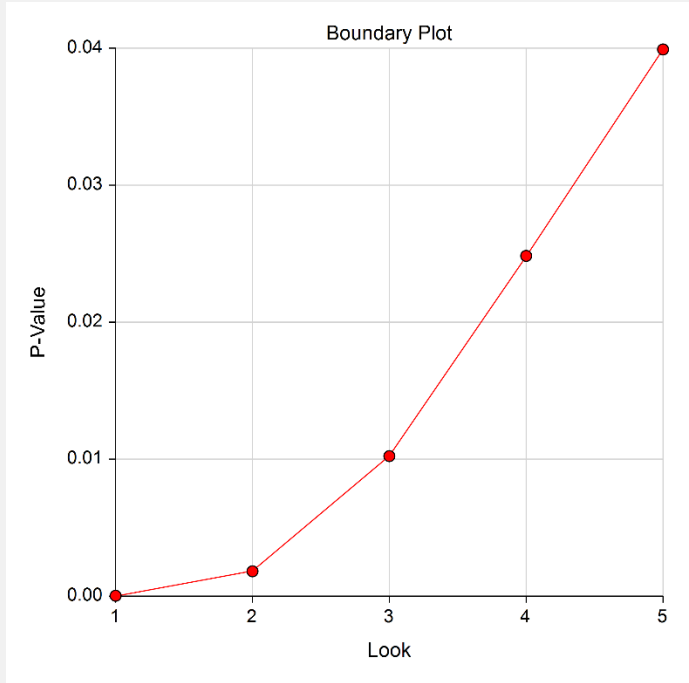
Look is the number of the look.  
 Significance Boundary Z-Value Scale is the value such that statistics outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. They are sometimes called efficacy boundaries.  
 Significance Boundary P-Value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

### Boundary Plot



**Group-Sequential Non-Inferiority Tests for the Difference of Two Proportions (Simulation)**

**Boundary Plot - P-Value**



**Significance Boundaries with 95% Simulation Confidence Intervals for Scenario 1**

Look	----- Z-Value Boundary -----			----- P-Value Boundary -----		
	Value	95% LCL	95% UCL	Value	95% LCL	95% UCL
1	4.03979			0.00003		
2	2.90905	2.84901	2.93345	0.00181	0.00168	0.00219
3	2.31812	2.27572	2.32621	0.01022	0.01000	0.01143
4	1.96274	1.95729	1.96975	0.02484	0.02443	0.02516
5	1.75172	1.71681	1.75538	0.03991	0.03960	0.04301

**Significance Boundary Confidence Limit Definitions**

Look is the number of the look.

Z-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 Z-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 the Difference of Two Proportions (Simulation)**

**Alpha-Spending and Null Hypothesis Simulation Details for Scenario 1**

Look	--- Signif. Boundary---		----- Target -----		----- Actual -----		Proportion	Cum.
	Z-Value Scale	P-Value Scale	Spending Function Alpha	Spending Function Alpha	Alpha Spent	Cum. Alpha Spent	H1 Sims Outside Signif. Boundary	H1 Sims Outside Signif. Boundary
1	4.03979	0.00003	0.000	0.000	0.000	0.000	0.001	0.001
2	2.90905	0.00181	0.002	0.002	0.002	0.002	0.070	0.071
3	2.31812	0.01022	0.009	0.011	0.009	0.011	0.219	0.290
4	1.96274	0.02484	0.017	0.028	0.017	0.028	0.251	0.541
5	1.75172	0.03991	0.022	0.050	0.022	0.050	0.177	0.718

**Alpha-Spending Details Definitions**

Look is the number of the look.

Significance Boundary Z-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 Z-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 with Futility Boundaries

Continuing with Example 1, suppose that the researchers would also like to terminate the study early if there is indication that the treatment is not non-inferior.

### 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 the Difference of Two Proportions (Simulation)** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Group-Sequential**, and then clicking on the **Group-Sequential Non-Inferiority Tests for the Difference of Two Proportions (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**.

<u>Option</u>	<u>Value</u>
<b>Design Tab</b>	
Solve For .....	<b>Power</b>
Test Type .....	<b>Z-Test (Pooled)</b>
Higher Proportions Are .....	<b>Better</b>
Simulations .....	<b>100000</b>
Alpha .....	<b>0.05</b>
Group Allocation .....	<b>Equal (N1 = N2)</b>
Sample Size Per Group .....	<b>1000</b>
Input Type .....	<b>Proportions</b>
P1.0 (Non-Inferiority Proportion) .....	<b>0.53</b>
P1.1 (Actual Proportion) .....	<b>0.58</b>
P2 (Group 2 Proportion) .....	<b>0.58</b>

### Looks & Boundaries Tab

Specification of Looks and Boundaries ..	<b>Simple</b>
Number of Equally Spaced Looks .....	<b>5</b>
Alpha Spending Function .....	<b>O'Brien-Fleming Analog</b>
Type of Futility Boundary .....	<b>Non-Binding</b>
Number of Skipped Futility Looks .....	<b>0</b>
Beta Spending Function .....	<b>O'Brien-Fleming Analog</b>

### 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 Testing Proportion Difference = NIM.

Hypotheses: H0: Proportion 1 - Proportion 2 = NIM; H1: Proportion 1 - Proportion 2 > NIM

Test Statistic: Z-Test (Pooled)

Zero Adjustment Method: None

Alpha-Spending Function: O'Brien-Fleming Analog

Beta-Spending Function: O'Brien-Fleming Analog

Futility Boundary Type: Non-binding

Number of Looks: 5

Simulations: 100000

### Group-Sequential Non-Inferiority Tests for the Difference of Two Proportions (Simulation)

#### Numeric Summary for Scenario 1

Power			Alpha		Beta	
Value	95% LCL	95% UCL	Target	Actual	95% LCL	95% UCL
0.624	0.621	0.627	0.050	0.039	0.038	0.040

Average Sample Size					
Given H0			Given H1		
N1	N2	Grp1	Grp2	Grp1	Grp2
1000	1000	451	451	665	665

D0	D1	P1.0	P1.1	P2
-0.05	0.00	0.53	0.58	0.58

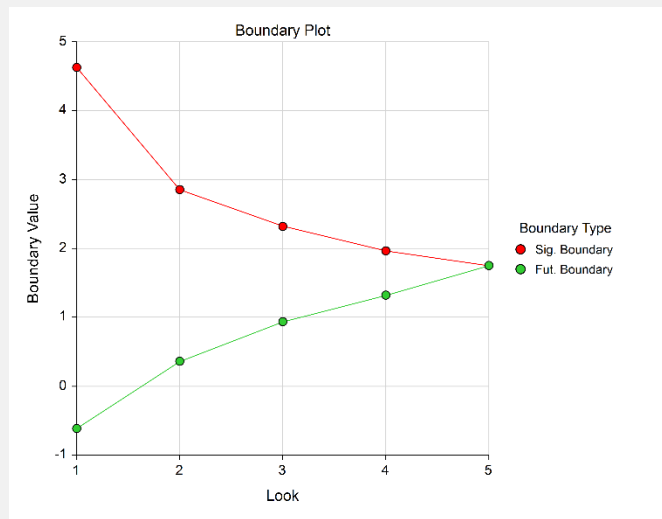
#### Accumulated Information Details for Scenario 1

Look	Accumulated Information	Accumulated Sample Size		
	Percent	Group 1	Group 2	Total
1	20.0	200	200	400
2	40.0	400	400	800
3	60.0	600	600	1200
4	80.0	800	800	1600
5	100.0	1000	1000	2000

#### Boundaries for Scenario 1

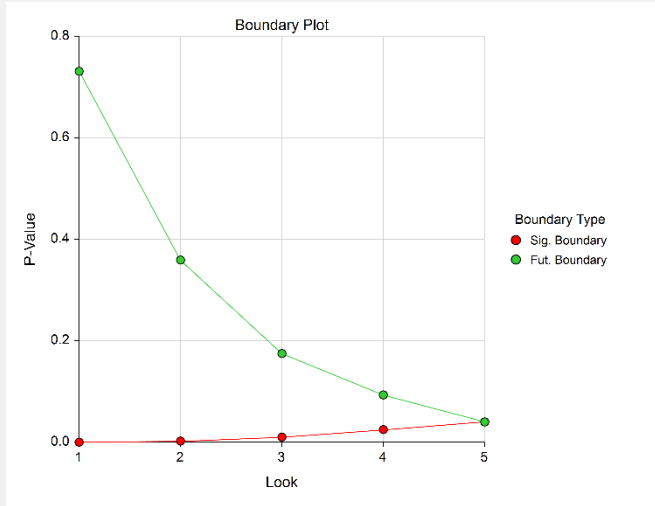
Look	Significance Boundary		Futility Boundary	
	Z-Value Scale	P-Value Scale	Z-Value Scale	P-Value Scale
1	4.62809	0.00000	-0.61507	0.73075
2	2.85456	0.00215	0.36142	0.35889
3	2.31953	0.01018	0.93633	0.17455
4	1.96510	0.02470	1.32159	0.09315
5	1.75206	0.03988	1.75206	0.03988

#### Boundary Plot



Group-Sequential Non-Inferiority Tests for the Difference of Two Proportions (Simulation)

Boundary Plot - P-Value



Significance Boundaries with 95% Simulation Confidence Intervals for Scenario 1

Look	----- Z-Value Boundary -----			----- P-Value Boundary -----		
	Value	95% LCL	95% UCL	Value	95% LCL	95% UCL
1	4.62809			0.00000		
2	2.85456	2.83752	2.91449	0.00215	0.00178	0.00227
3	2.31953	2.27685	2.32815	0.01018	0.00995	0.01140
4	1.96510	1.96007	1.97837	0.02470	0.02394	0.02499
5	1.75206	1.71993	1.75538	0.03988	0.03960	0.04272

Futility Boundaries with 95% Simulation Confidence Intervals for Scenario 1

Look	----- Z-Value Boundary -----			----- P-Value Boundary -----		
	Value	95% LCL	95% UCL	Value	95% LCL	95% UCL
1	-0.61507	-0.70148	-0.61237	0.73075		
2	0.36142	0.35994	0.42662	0.35889	0.33483	0.35945
3	0.93633	0.93433	0.93796	0.17455	0.17413	0.17507
4	1.32159	1.31863	1.32821	0.09315	0.09205	0.09365
5	1.75206	1.72959	1.76382	0.03988	0.03888	0.04185

Alpha-Spending and Null Hypothesis Simulation Details for Scenario 1

Look	--- Signif. Boundary ---		----- Target -----		----- Actual -----		Proportion H0 Sims Outside Futility Boundary	Cum. H0 Sims Outside Futility Boundary
	Z-Value Scale	P-Value Scale	Spending Function Alpha	Spending Function Alpha	Alpha Spent	Cum. Alpha Spent		
1	4.62809	0.00000	0.000	0.000	0.000	0.000	0.256	0.256
2	2.85456	0.00215	0.002	0.002	0.002	0.002	0.409	0.665
3	2.31953	0.01018	0.009	0.011	0.009	0.011	0.188	0.853
4	1.96510	0.02470	0.017	0.028	0.015	0.027	0.076	0.929
5	1.75206	0.03988	0.022	0.050	0.012	0.039	0.032	0.961

Beta-Spending and Alternative Hypothesis Simulation Details for Scenario 1

Look	-- Futility Boundary --		----- Target -----		----- Actual -----		Proportion H1 Sims Outside Signif. Boundary	Cum. H1 Sims Outside Signif. Boundary
	Z-Value Scale	P-Value Scale	Spending Function Beta	Spending Function Beta	Beta Spent	Cum. Beta Spent		
1	-0.61507	0.73075	0.048	0.048	0.048	0.048	0.000	0.000
2	0.36142	0.35889	0.114	0.162	0.114	0.162	0.079	0.079
3	0.93633	0.17455	0.092	0.254	0.092	0.254	0.212	0.291
4	1.32159	0.09315	0.069	0.323	0.069	0.323	0.228	0.519
5	1.75206	0.03988	0.054	0.377	0.053	0.376	0.105	0.624

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 the Difference of Two Proportions (Simulation)** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Group-Sequential**, and then clicking on the **Group-Sequential Non-Inferiority Tests for the Difference of Two Proportions (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**.

<u>Option</u>	<u>Value</u>
<b>Design Tab</b>	
Solve For .....	<b>Alpha and Power (Enter Boundaries)</b>
Test Type .....	<b>Z-Test (Pooled)</b>
Higher Proportions Are .....	<b>Better</b>
Simulations .....	<b>100000</b>
Group Allocation .....	<b>Equal (N1 = N2)</b>
Sample Size Per Group .....	<b>1000</b>
Input Type .....	<b>Proportions</b>
P1.0 (Non-Inferiority Proportion) .....	<b>0.53</b>
P1.1 (Actual Proportion) .....	<b>0.58</b>
P2 (Group 2 Proportion) .....	<b>0.58</b>
<b>Looks &amp; Boundaries Tab</b>	
Number of Looks .....	<b>5</b>
Equally Spaced .....	<b>Checked</b>
Types of Boundaries .....	<b>Significance and Futility Boundaries</b>
Significance Boundary .....	<b>3 3 3 2 1 (for looks 1 through 5)</b>
Futility Boundary .....	<b>-2 -1 0 0 1 (for looks 1 through 5)</b>

### 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 Testing Proportion Difference = 0.

Hypotheses: H0: Proportion 1 = Proportion 2; H1: Proportion 1 > Proportion 2

Test Statistic: Z-Test (Pooled)

Zero Adjustment Method: Add 0 to each zero

Type of Boundaries: Significance and Futility Boundaries

Number of Looks: 5

Simulations: 100000

### Group-Sequential Non-Inferiority Tests for the Difference of Two Proportions (Simulation)

#### Numeric Summary for Scenario 1

Power				Alpha				Beta	
Value	95% LCL	95% UCL		Value	95% LCL	95% UCL			
0.885	0.883	0.887		0.146	0.143	0.148		0.115	

Average Sample Size					
Given H0			Given H1		
N1	N2	Grp1	Grp2	Grp1	Grp2
1000	1000	737	737	832	832

D0	D1	P1.0	P1.1	P2
-0.05	0.00	0.53	0.58	0.58

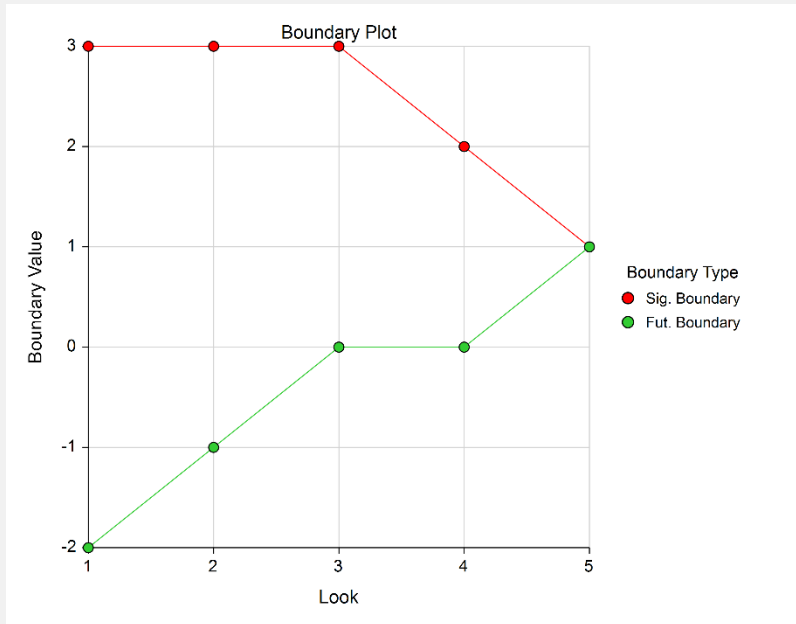
#### Accumulated Information Details for Scenario 1

Look	Accumulated Information	Accumulated Sample Size		
	Percent	Group 1	Group 2	Total
1	20.0	200	200	400
2	40.0	400	400	800
3	60.0	600	600	1200
4	80.0	800	800	1600
5	100.0	1000	1000	2000

#### Boundaries for Scenario 1

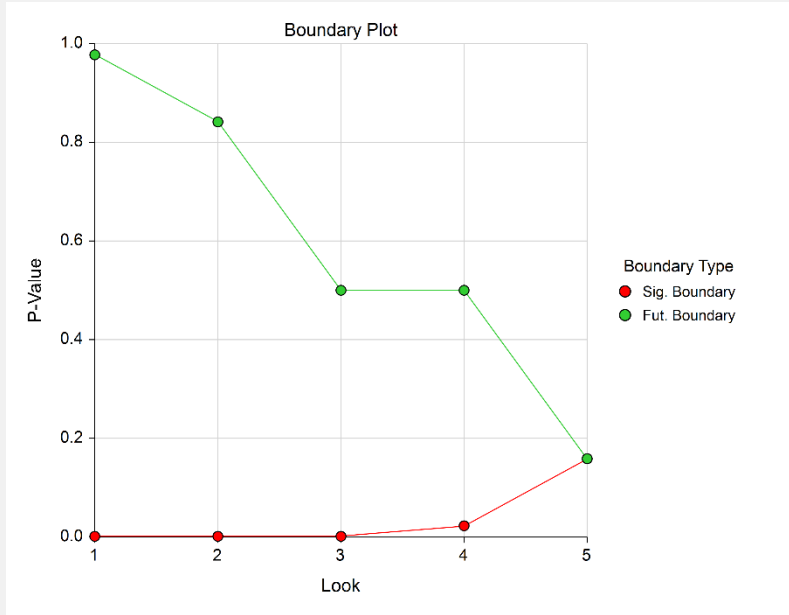
Look	Significance Boundary		Futility Boundary	
	Z-Value Scale	P-Value Scale	Z-Value Scale	P-Value Scale
1	3.00000	0.00135	-2.00000	0.97725
2	3.00000	0.00135	-1.00000	0.84134
3	3.00000	0.00135	0.00000	0.50000
4	2.00000	0.02275	0.00000	0.50000
5	1.00000	0.15866	1.00000	0.15866

#### Boundary Plot



**Group-Sequential Non-Inferiority Tests for the Difference of Two Proportions (Simulation)**

**Boundary Plot - P-Value**



**Alpha-Spending and Null Hypothesis Simulation Details for Scenario 1**

Look	--- Signif. Boundary---		Alpha Spent	Cum. Alpha Spent	Proportion H0 Sims Outside Futility Boundary	Cum. H0 Sims Outside Futility Boundary
	Z-Value Scale	P-Value Scale				
1	3.00000	0.00135	0.001	0.001	0.024	0.024
2	3.00000	0.00135	0.001	0.002	0.134	0.158
3	3.00000	0.00135	0.001	0.004	0.350	0.508
4	2.00000	0.02275	0.022	0.025	0.082	0.591
5	1.00000	0.15866	0.120	0.146	0.264	0.854

**Beta-Spending and Alternative Hypothesis Simulation Details for Scenario 1**

Look	-- Futility Boundary--		Beta Spent	Cum. Beta Spent	Proportion H1 Sims Outside Signif. Boundary	Cum. H1 Sims Outside Signif. Boundary
	Z-Value Scale	P-Value Scale				
1	-2.00000	0.97725	0.001	0.001	0.024	0.024
2	-1.00000	0.84134	0.007	0.009	0.046	0.070
3	0.00000	0.50000	0.033	0.042	0.061	0.131
4	0.00000	0.50000	0.006	0.047	0.384	0.515
5	1.00000	0.15866	0.067	0.115	0.371	0.885

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 2, we examine the power and alpha generated by the set of significance (4.62809, 2.85456, 2.31953, 1.96510, 1.75206) and futility (-0.61507, 0.36142, 0.93633, 1.32159, 1.75206) 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 the Difference of Two Proportions (Simulation)** procedure window by expanding **Proportions**, then **Two Independent Proportions**, then clicking on **Group-Sequential**, and then clicking on the **Group-Sequential Non-Inferiority Tests for the Difference of Two Proportions (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**.

<u>Option</u>	<u>Value</u>
<b>Design Tab</b>	
Solve For .....	Alpha and Power (Enter Boundaries)
Test Type .....	Z-Test (Pooled)
Higher Proportions Are .....	Better
Simulations .....	100000
Group Allocation .....	Equal (N1 = N2)
Sample Size Per Group .....	1000
Input Type .....	Proportions
P1.0 (Non-Inferiority Proportion) .....	0.53
P1.1 (Actual Proportion) .....	0.58
P2 (Group 2 Proportion) .....	0.58
<b>Looks &amp; Boundaries Tab</b>	
Number of Looks .....	5
Equally Spaced .....	Checked
Types of Boundaries .....	Significance and Futility Boundaries
Significance Boundary .....	4.62809, 2.85456, 2.31953, 1.96510, 1.75206
Futility Boundary .....	-0.61507, 0.36142, 0.93633, 1.32159, 1.75206

### Output

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

#### Numeric Results and Plots

Numeric Summary for Scenario 1											
Power						Alpha				Beta	
Value	95% LCL	95% UCL	Value	95% LCL	95% UCL						
0.623	0.620	0.626	0.038	0.037	0.039					0.377	
---- Average Sample Size ----											
-- Given H0 --						-- Given H1 --					
N1	N2	Grp1	Grp2	Grp1	Grp2	D0	D1	P1.0	P1.1	P2	
1000	1000	451	451	665	665	-0.05	0.00	0.53	0.58	0.58	

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 2.