

## Chapter 235

# Non-Inferiority Tests for the Difference of Two Proportions in a Cluster-Randomized Design

---

## Introduction

This module provides power analysis and sample size calculation for non-inferiority tests of the difference in two-sample, cluster-randomized designs in which the outcome is binary.

---

## Technical Details

Our formulation comes from Donner and Klar (2000). Denote a binary observation by  $Y_{gkm}$  where  $g = 1$  or  $2$  is the group,  $k = 1, 2, \dots, K_g$  is a cluster within group  $g$ , and  $m = 1, 2, \dots, M_g$  is an individual in cluster  $k$  of group  $g$ . The results that follow assume an equal number of individuals per cluster. When the number of subjects from cluster to cluster are about the same, the power and sample size values should be fairly accurate. In these cases, the average number of subjects per cluster can be used.

The statistical hypothesis that is tested concerns the difference between the two group proportions,  $p_1$  and  $p_2$ . When necessary, we assume that group 1 is the treatment group and group 2 is the control group. With a simple modification, all of the large-sample sample size formulas that are listed in the module for testing two proportions can be used here.

When the individual subjects are randomly assigned to one of the two groups, the variance of the sample proportion is

$$\sigma_{S,g}^2 = \frac{p_g(1-p_g)}{n_g}$$

When the randomization is by clusters of subjects, the variance of the sample proportion is

$$\begin{aligned}\sigma_{C,g}^2 &= \frac{p_g(1-p_g)(1+(m_g-1)\rho)}{k_g m_g} \\ &= \sigma_{S,g}^2 [1+(m_g-1)\rho] \\ &= F_{g,\rho} \sigma_{S,g}^2\end{aligned}$$

## Non-Inferiority Tests for the Difference of Two Proportions in a Cluster-Randomized Design

The factor  $\left[1 + (m_g - 1)\rho\right]$  is called the *inflation factor*. The Greek letter  $\rho$  is used to represent the *intracluster correlation coefficient (ICC)*. This correlation may be thought of as the simple correlation between any two subjects within the same cluster. If we stipulate that  $\rho$  is positive, it may also be interpreted as the proportion of total variability that is attributable to differences between clusters. This value is critical to the sample size calculation.

The asymptotic formulas that were used in comparing two proportions (see Chapter 210, “Non-Inferiority Tests for the Difference Between Two Proportions”) may be used with cluster-randomized designs as well, as long as an adjustment is made for the inflation factor.

---

## Power Calculations

A large sample approximation may be used that is most accurate when the values of  $n_1$  and  $n_2$  are large. The large approximation is made by replacing the values of  $\hat{p}_1$  and  $\hat{p}_2$  in the  $z$  statistic with the corresponding values of  $p_1$  and  $p_2$  under the alternative hypothesis, and then computing the results based on the normal distribution.

Note that in this case, exact calculations are not possible.

---

## Procedure Options

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

---

## Design Tab

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

---

### Solve For

#### Solve For

This option specifies the parameter to be solved for using the other parameters. The parameters that may be selected are *Power*, *Sample Size (K1)*, *Sample Size (M1)*, *Effect Size*, and *ICC*. Under most situations, you will select either *Power* or *Sample Size (K1)*.

Select *Sample Size (K1)* when you want to calculate the number of clusters per group needed to achieve a given power and alpha level.

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

---

### Test

#### Higher Proportions Are

This option specifies whether proportions represent successes (better) or failures (worse).

- **Better (Successes)**

If Higher Proportions are "Better", the alternative hypothesis is  $H_1: P_1 - P_2 > D_0$ .

- **Worse (Failures)**

If Higher Proportions are "Worse", the alternative hypothesis is  $H_1: P_1 - P_2 < D_0$ .

## Non-Inferiority Tests for the Difference of Two Proportions in a Cluster-Randomized Design

### Test Type

Specify which test statistic is used in searching and reporting. We recommend the likelihood score test.

---

### Power and Alpha

#### Power

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

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

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

#### Alpha

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

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

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

---

### Sample Size – Group 1 (Treatment)

#### K1 (Clusters in Group 1)

Enter a value (or range of values) for the number of clusters in group one. You may enter a range of values such as *10 to 20 by 2*. The sample size for this group is equal to the number of clusters times the number of subjects per cluster.

#### M1 (Items per Cluster in Group 1)

This is the average number of items (subjects) per cluster in group one. This value must be a positive number that is at least 1. You can use a list of values such as *100 150 200*.

---

### Sample Size – Group 2 (Reference)

#### K2 (Clusters in Group 2)

This is the number of clusters in group two. The sample size for this group is equal to the number of clusters times the number of subjects per cluster. This value must be a positive number.

If you simply want a multiple of the value for group one, you would enter the multiple followed by *K1*, with no blanks. If you want to use *K1* directly, you do not have to pre-multiply by 1. For example, all of the following are valid entries: *10 K1 2K1 0.5K1*.

You can use a list of values such as *10 20 30* or *K1 2K1 3K1*.

#### M2 (Items per Cluster in Group 2)

This is the number of items (subjects) per cluster in group two. This value must be a positive number.

If you simply want a multiple of the value for group one, you would enter the multiple followed by *M1*, with no blanks. If you want to use *M1* directly, you do not have to pre-multiply by 1. For example, all of the following are valid entries: *10 M1 2M1 0.5M1*.

You can use a list of values such as *10 20 30* or *M1 2M1 3M1*.

## Non-Inferiority Tests for the Difference of Two Proportions in a Cluster-Randomized Design

---

### Effect Size

#### Input Type

Indicate what type of values to enter to specify the differences. Regardless of the entry type chosen, the test statistics used in the power and sample size calculations are the same. This option is simply given for convenience in specifying the differences.

---

#### Effect Size – Differences

*Only displayed 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.

##### 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.

---

#### Effect Size – Group 1 (Treatment)

*Only displayed 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.

Proportions must be between 0 and 1. They cannot take on the values 0 or 1. This value should not be set to exactly the value of P2.

You may enter a range of values such as *0.03 0.05 0.10* or *0.01 to 0.05 by 0.01*.

## Non-Inferiority Tests for the Difference of Two Proportions in a Cluster-Randomized Design

### P1.1 (Actual Proportion)

This option specifies the value of P1.1, the value of the treatment proportion at which the power is to be calculated. It is only used for *Proportions*. Proportions must be between 0 and 1. They cannot take on the values 0 or 1.

You may enter a range of values such as *0.03 0.05 0.10* or *0.01 to 0.05 by 0.01*.

---

### Effect Size – Group 2 (Reference)

#### P2 (Group 2 Proportion)

Specify the value of  $p_2$ , the control, baseline, or standard group's proportion. The null hypothesis is that the two proportions differ by a specified amount (See *Specify Group 1 Proportion using* below).

Since  $p_2$  is a proportion, these values must be between 0 and 1.

You may enter a range of values such as *0.1 0.2 0.3* or *0.1 to 0.9 by 0.1*.

---

### Effect Size – Intraclass Correlation

#### ICC (Intraclass Correlation)

Enter a value (or range of values) for the intraclass correlation. This correlation may be thought of as the simple correlation between any two observations in the same cluster. It may also be thought of as the proportion of total variance in the observations that can be attributed to difference between clusters.

Although the actual range for this value is between 0 to 1, typical values range from 0.002 to 0.05.

## Example 1 – Finding Power

A study is being designed to study the effectiveness of a new treatment. Historically, the standard treatment has enjoyed a 60% cure rate. The new treatment reduces the seriousness of certain side effects that occur with the standard treatment. Thus, the new treatment will be adopted even if it is slightly less effective than the standard treatment. The researchers will recommend adoption of the new treatment if it has a cure rate of at least 55%. That is, the margin of inferiority is -5%.

The researchers will recruit patients from various hospitals. All patients at a particular hospital will receive the same treatment. They anticipate an average of 100 patients per hospital. Based on similar studies, they estimate the intraclass correlation to be 0.002.

The researchers plan to use the Farrington and Manning likelihood score test statistic to analyze the data. They want to study the power of the one-sided Farrington and Manning test at group cluster sizes ranging from 2 to 10 for detecting a difference of -0.05 when the actual cure rate of the new treatment ranges from 60% to 66%. The significance level will be 0.05.

## Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Non-Inferiority Tests for the Difference of Two Proportions in a Cluster-Randomized Design** procedure window by expanding **Proportions**, then **Two Proportions – Cluster Randomized**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for the Difference of Two Proportions in a Cluster-Randomized Design**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
<b>Design Tab</b>	
Solve For .....	<b>Power</b>
Higher Proportions Are .....	<b>Better</b>
Test Type .....	<b>Likelihood Score (Farr. &amp; Mann.)</b>
Alpha .....	<b>0.05</b>
K1 (Clusters in Group 1) .....	<b>2 4 6 8 10</b>
M1 (Items per Cluster in Group 1) .....	<b>100</b>
K2 (Clusters in Group 2) .....	<b>K1</b>
M2 (Items per Cluster in Group 2) .....	<b>M1</b>
Input Type .....	<b>Differences</b>
D0 (Non-Inferiority Difference) .....	<b>-0.05</b>
D1 (Actual Difference) .....	<b>0 0.02 0.04 0.06</b>
P2 (Group 2 Proportion) .....	<b>0.6</b>
ICC (Intraclass Correlation) .....	<b>0.002</b>

## Non-Inferiority Tests for the Difference of Two Proportions in a Cluster-Randomized Design

### Output

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

### Numeric Results

#### Numeric Results for Non-Inferiority Tests for the Difference of Two Proportions (Cluster-Randomized)

Test Statistic: Likelihood Score Test (Farrington & Manning)

H0:  $P_1 - P_2 \leq D_0$ . H1:  $P_1 - P_2 = D_1 > D_0$ .

Power	Group 1 Clusters/ Items K1/M1	Group 2 Clusters/ Items K2/M2	Group 2 Prop P2	Group 1 Non-Inf. Prop P1.0	Group 1 Actual Prop P1.1	Non-Inf. Diff D0	Actual Diff D1	Intra- Cluster Corr. ICC	Alpha
0.23867	2/100	2/100	0.6000	0.5500	0.6000	-0.0500	0.0000	0.0020	0.050
0.37284	4/100	4/100	0.6000	0.5500	0.6000	-0.0500	0.0000	0.0020	0.050
0.48884	6/100	6/100	0.6000	0.5500	0.6000	-0.0500	0.0000	0.0020	0.050
0.58780	8/100	8/100	0.6000	0.5500	0.6000	-0.0500	0.0000	0.0020	0.050
0.67076	10/100	10/100	0.6000	0.5500	0.6000	-0.0500	0.0000	0.0020	0.050
0.36988	2/100	2/100	0.6000	0.5500	0.6200	-0.0500	0.0200	0.0020	0.050
0.58359	4/100	4/100	0.6000	0.5500	0.6200	-0.0500	0.0200	0.0020	0.050
0.73497	6/100	6/100	0.6000	0.5500	0.6200	-0.0500	0.0200	0.0020	0.050
0.83629	8/100	8/100	0.6000	0.5500	0.6200	-0.0500	0.0200	0.0020	0.050
0.90129	10/100	10/100	0.6000	0.5500	0.6200	-0.0500	0.0200	0.0020	0.050
0.52013	2/100	2/100	0.6000	0.5500	0.6400	-0.0500	0.0400	0.0020	0.050
0.77421	4/100	4/100	0.6000	0.5500	0.6400	-0.0500	0.0400	0.0020	0.050
0.90176	6/100	6/100	0.6000	0.5500	0.6400	-0.0500	0.0400	0.0020	0.050
0.95959	8/100	8/100	0.6000	0.5500	0.6400	-0.0500	0.0400	0.0020	0.050
0.98407	10/100	10/100	0.6000	0.5500	0.6400	-0.0500	0.0400	0.0020	0.050
0.66945	2/100	2/100	0.6000	0.5500	0.6600	-0.0500	0.0600	0.0020	0.050
0.90359	4/100	4/100	0.6000	0.5500	0.6600	-0.0500	0.0600	0.0020	0.050
0.97530	6/100	6/100	0.6000	0.5500	0.6600	-0.0500	0.0600	0.0020	0.050
0.99420	8/100	8/100	0.6000	0.5500	0.6600	-0.0500	0.0600	0.0020	0.050
0.99872	10/100	10/100	0.6000	0.5500	0.6600	-0.0500	0.0600	0.0020	0.050

#### Report Definitions

Power is the probability of rejecting a false null hypothesis. It should be close to one.

K1 and K2 are the number of clusters in groups 1 and 2, respectively.

M1 and M2 are the average number of items (subjects) per cluster in groups 1 and 2, respectively.

P2 is the proportion for group 2. This is the standard, reference, baseline, or control group.

P1.0 is the proportion for group 1 (treatment group) assuming the null hypothesis (H0).

P1.1 is the proportion for group 1 (treatment group) assuming the alternative hypothesis (H1).

D0 = P1.0 - P2 is the non-inferiority margin. It is the difference assuming H0.

D1 = P1.1 - P2 is the actual difference at which the power is calculated.

ICC is the intracluster correlation.

Alpha is the probability of rejecting a true null hypothesis.

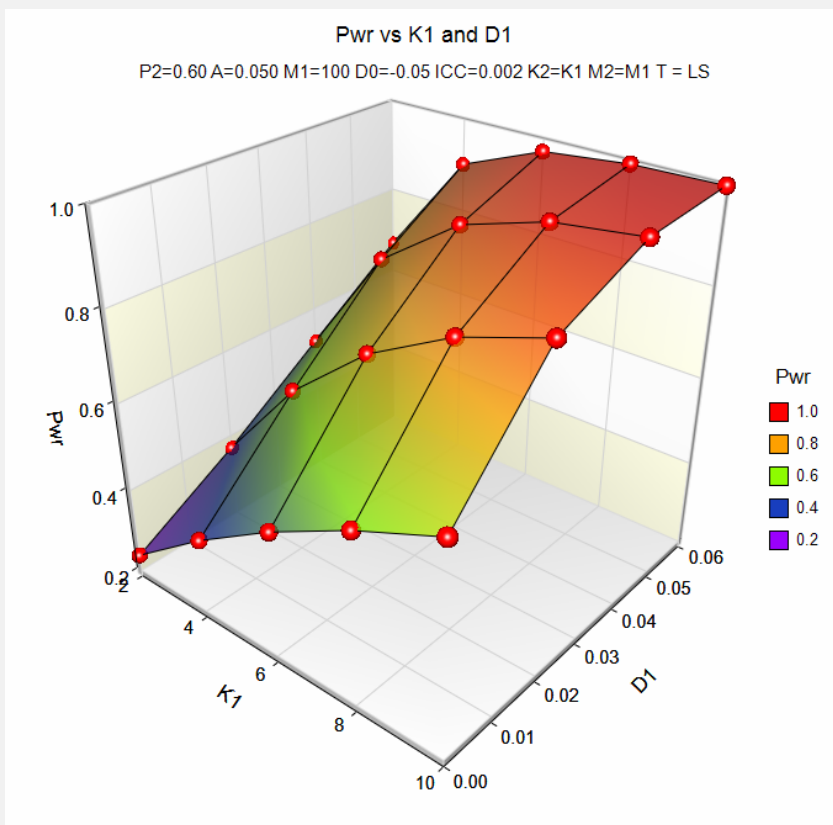
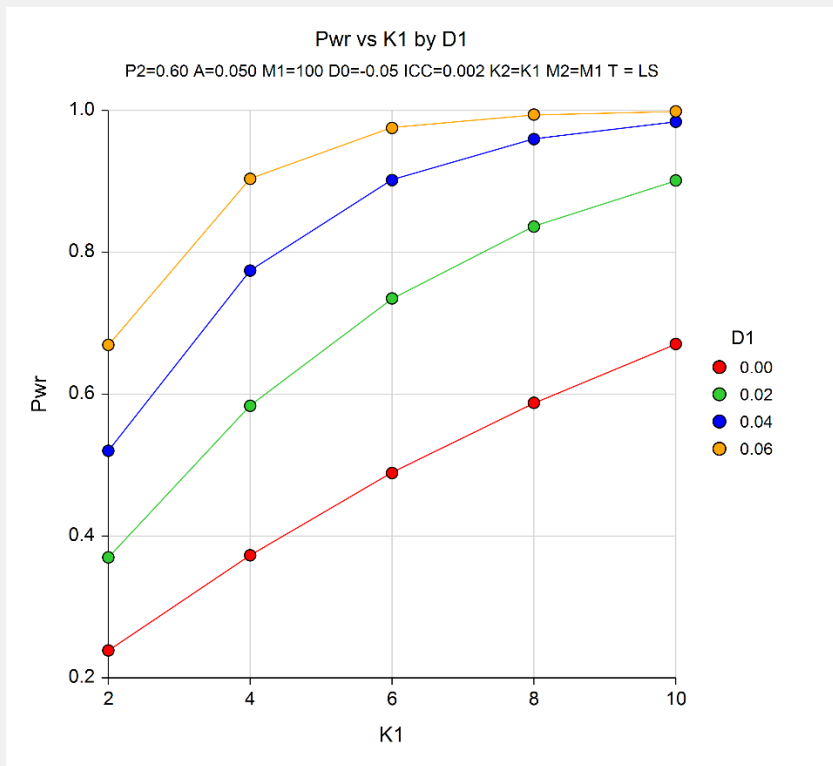
#### Summary Statements

Sample sizes of 200 in group 1 and 200 in group 2, which were obtained by sampling 2 clusters with 100 subjects each in group 1 and 2 clusters with 100 subjects each in group 2, achieve 23.867% power to detect a non-inferiority margin difference between the group proportions of -0.0500. The proportion in group 1 (the treatment group) is assumed to be 0.5500 under the null hypothesis and 0.6000 under the alternative hypothesis. The proportion in group 2 (the control group) is 0.6000. The test statistic used is the one-sided Likelihood Score Test (Farrington & Manning). The intracluster correlation is 0.0020, and the significance level of the test is 0.050.

This report shows the values of each of the parameters, one scenario per row. The total number of items sampled in group 1 is  $N_1 = K_1 \times M_1$ . The total number of items sampled in group 2 is  $N_2 = K_2 \times M_2$ .

Non-Inferiority Tests for the Difference of Two Proportions in a Cluster-Randomized Design

Plots Section



The values from the table are displayed on the above charts. These charts give a quick look at the sample sizes that will be required for various values of D1.



## Example 2 – Finding the Sample Size (Number of Clusters)

Continuing with the scenario given in Example 1, the researchers want to determine the number of clusters necessary for each value of D1 when the target power is set to 0.80.

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Non-Inferiority Tests for the Difference of Two Proportions in a Cluster-Randomized Design** procedure window by expanding **Proportions**, then **Two Proportions – Cluster Randomized**, then clicking on **Non-Inferiority**, and then clicking on **Non-Inferiority Tests for the Difference of Two Proportions in a Cluster-Randomized Design**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
<b>Design Tab</b>	
Solve For .....	<b>Sample Size (K1)</b>
Higher Proportions Are .....	<b>Better</b>
Test Type .....	<b>Likelihood Score (Farr. &amp; Mann.)</b>
Power .....	<b>0.80</b>
Alpha .....	<b>0.05</b>
M1 (Items per Cluster in Group 1) .....	<b>100</b>
K2 (Clusters in Group 2) .....	<b>K1</b>
M2 (Items per Cluster in Group 2) .....	<b>M1</b>
Input Type .....	<b>Differences</b>
D0 (Non-Inferiority Difference) .....	<b>-0.05</b>
D1 (Actual Difference) .....	<b>0 0.02 0.04 0.06</b>
P2 (Group 2 Proportion) .....	<b>0.6</b>
ICC (Intracluster Correlation) .....	<b>0.002</b>

### Output

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

### Numeric Results

#### Numeric Results for Non-Inferiority Tests for the Difference of Two Proportions (Cluster-Randomized)

Test Statistic: Likelihood Score Test (Farrington & Manning)

H0:  $P_1 - P_2 \leq D_0$ . H1:  $P_1 - P_2 = D_1 > D_0$ .

	Group 1 Clusters/ Items	Group 2 Clusters/ Items	Group 2 Prop P2	Group 1 Non-Inf. Prop P1.0	Group 1 Actual Prop P1.1	Non-Inf. Diff D0	Actual Diff D1	Intra- Cluster Corr. ICC	Alpha
Power	K1/M1	K2/M2							
0.81875	15/100	15/100	0.6000	0.5500	0.6000	-0.0500	0.0000	0.0020	0.050
0.83629	8/100	8/100	0.6000	0.5500	0.6200	-0.0500	0.0200	0.0020	0.050
0.84985	5/100	5/100	0.6000	0.5500	0.6400	-0.0500	0.0400	0.0020	0.050
0.81783	3/100	3/100	0.6000	0.5500	0.6600	-0.0500	0.0600	0.0020	0.050

The required sample size depends a great deal on the value of D1. The researchers should spend time determining the most appropriate value for D1.

## Example 3 – Finding Power after an Experiment

A group of researchers want to show that a new, less expensive treatment works at least as well as the current treatment. They believe, in fact, that the new treatment is about 0.10 higher in proportion of success. One hundred patients at each of 10 randomly chosen hospitals were given the current treatment. One hundred patients at each of 10 randomly chosen hospitals were given the new treatment. It was agreed before the experiment that the new treatment needed to be no less than 0.05 in proportion of success below the current treatment to be considered non-inferior. The proportion of patients responding to the current treatment was  $821/1000 = 0.821$ . The proportion of patients responding to the new treatment was  $819/1000 = 0.819$ . This result did not show significant non-inferiority at the 0.05 level. The researchers want to know the power of their non-inferiority test. They decide to use the intracluster correlation coefficient estimated from the data, which was 0.0068. Although the observed difference in proportions is  $0.819 - 0.821 = -0.002$ , the trivial difference is still -0.05. This value is used in the power calculation.

### Setup

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

<u>Option</u>	<u>Value</u>
<b>Design Tab</b>	
Solve For .....	<b>Power</b>
Higher Proportions Are .....	<b>Better</b>
Test Type .....	<b>Likelihood Score (Farr. &amp; Mann.)</b>
Alpha .....	<b>0.05</b>
K1 (Clusters in Group 1) .....	<b>10</b>
M1 (Items per Cluster in Group 1) .....	<b>100</b>
K2 (Clusters in Group 2) .....	<b>K1</b>
M2 (Items per Cluster in Group 2) .....	<b>M1</b>
Input Type .....	<b>Differences</b>
D0 (Non-Inferiority Difference) .....	<b>-0.05</b>
D1 (Actual Difference) .....	<b>0.0 0.10</b>
P2 (Group 2 Proportion) .....	<b>0.821</b>
ICC (Intracluster Correlation) .....	<b>0.0068</b>

## Non-Inferiority Tests for the Difference of Two Proportions in a Cluster-Randomized Design

### Output

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

### Numeric Results

#### Numeric Results for Non-Inferiority Tests for the Difference of Two Proportions (Cluster-Randomized)

Test Statistic: Likelihood Score Test (Farrington & Manning)

H0:  $P_1 - P_2 \leq D_0$ . H1:  $P_1 - P_2 = D_1 > D_0$ .

	Group 1 Clusters/ Items	Group 2 Clusters/ Items	Group 2 Prop P2	Group 1 Non-Inf. Prop P1.0	Group 1 Actual Prop P1.1	Non-Inf. Diff D0	Actual Diff D1	Intra- Cluster Corr. ICC	Alpha
Power	K1/M1	K2/M2							
0.72698	10/100	10/100	0.8210	0.7710	0.8210	-0.0500	0.0000	0.0068	0.050
1.00000	10/100	10/100	0.8210	0.7710	0.9210	-0.0500	0.1000	0.0068	0.050

If indeed the new treatment were 0.10 higher in proportion of success, the power for showing non-inferiority would be 1.0000. If the true proportions are the same, the power would be 0.72698.

### Example 4 – Validation

This procedure uses the same mechanics as the Tests for Two Proportions in a Cluster-Randomized Design procedure. We refer the user to Example 4 of Chapter 230 for the validation.