

## Chapter 480

# Paired T-Tests for Non-Inferiority

---

### Introduction

This procedure computes power and sample size for non-inferiority tests in paired designs in which the paired differences are distributed as a normal random variable.

The details of sample size calculation for the paired design are presented in the Paired T-Tests chapter and they will not be duplicated here. This chapter only discusses those changes necessary for non-inferiority tests. Sample size formulas for non-inferiority tests of a single mean are presented in Chow et al. (2018) page 43.

The *one-sample t-test* is used to test whether a population mean is different from a specific value. When the data are differences between paired values, this test is known as the *paired t-test*.

---

### Paired Designs

Paired data may occur because two measurements are made on the same subject or because measurements are made on two subjects that have been matched according to other variables. Hypothesis tests on paired data can be analyzed by considering the difference between the paired items as the response. The distribution of differences is usually symmetric. In fact, the distribution must be symmetric if the individual distributions of the two items are identical. Hence, the paired *t-test* is appropriate for paired data even when the distributions of the individual items are not normal.

---

### The Statistical Hypotheses

Both non-inferiority and superiority tests are examples of directional (one-sided) tests and their power and sample size could be calculated using the One-Sample T-Tests procedure. However, at the urging of our users, we have developed this procedure which provides the input and output options that are convenient for non-inferiority tests for paired data. This section will review the specifics of non-inferiority testing.

Remember that in the usual *t-test* setting with  $\delta$  defined as the mean paired difference, the null ( $H_0$ ) and alternative ( $H_1$ ) hypotheses for one-sided upper-tail tests are defined as

$$H_0: \delta \leq \delta_0 \quad \text{versus} \quad H_1: \delta > \delta_0$$

Rejecting  $H_0$  implies that the mean is larger than the value  $\delta_0$ . This test is called an *upper-tail test* because  $H_0$  is rejected in samples in which the mean of paired differences is larger than  $\delta_0$ .

## Paired T-Tests for Non-Inferiority

The *lower-tail test* is

$$H_0: \delta \geq \delta_0 \quad \text{versus} \quad H_1: \delta < \delta_0$$

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

<u>Parameter</u>	<u>PASS Input/Output</u>	<u>Interpretation</u>
$\delta$	$\delta$	<i>Population mean paired difference.</i> This is the mean of paired differences in the population. This parameter will be estimated by the study.
$\delta_1$	$\delta 1$	<i>Actual paired difference at which power is calculated.</i> This is the value of the mean paired difference at which power is calculated.
$M_{NI}$	NIM	<i>Margin of non-inferiority.</i> This is a tolerance value that defines the magnitude of difference that is not of practical importance. This may be thought of as the largest difference that is considered to be trivial. The sign is determined by the specific design. This value is used to compute the bound, $\delta_0$ .

---

## Non-Inferiority Tests

A *non-inferiority test* tests that the mean difference is not less than (or greater than) zero by more than a small non-inferiority margin. The actual direction of the hypothesis depends on the whether higher values of the response are good or bad.

### Case 1: High Values Good

In this case, higher values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the mean difference no less than a small amount below zero. The value of  $\delta$  at which power is calculated is often set to zero. The null and alternative hypotheses with  $\delta_0 = -|M_{NI}|$  are

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

### Case 2: High Values Bad

In this case, lower values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the mean of the treatment group is no more than a small amount above the reference value. The value of  $\delta$  at which power is calculated is often set to zero. The null and alternative hypotheses with  $\delta_0 = |M_{NI}|$  are

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

## Paired T-Test Statistic

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

$$t_{n-1} = \frac{\bar{X} - \delta_0}{s/\sqrt{n}}$$

where

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n},$$

$$s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}},$$

and  $\delta_0$  is the value of the mean hypothesized by the null hypothesis which depends on the magnitude and sign of  $M_{NI}$ .

The significance of the test statistic is determined by computing the p-value. If this p-value is less than a specified level (usually 0.05), the hypothesis is rejected. Otherwise, no conclusion can be reached.

## Power Calculation for the Paired T-Test

When the standard deviation is unknown, the power is calculated as follows for a directional alternative (one-tailed test) in which  $\delta_1 > \delta_0$ .

1. Find  $t_\alpha$  such that  $1 - T_{df}(t_\alpha) = \alpha$ , where  $T_{df}(t_\alpha)$  is the area under a central- $t$  curve to the left of  $x$  and  $df = n - 1$ .
2. Calculate:  $X_1 = \delta_0 + t_\alpha \frac{\sigma}{\sqrt{n}}$ .
3. Calculate the noncentrality parameter:  $\lambda = \frac{\delta_1 - \delta_0}{\frac{\sigma}{\sqrt{n}}}$ .
4. Calculate:  $t_1 = \frac{X_1 - \delta_1}{\frac{\sigma}{\sqrt{n}}} + \lambda$ .
5. Power =  $1 - T'_{df,\lambda}(t_1)$ , where  $T'_{df,\lambda}(x)$  is the area to the left of  $x$  under a noncentral- $t$  curve with degrees of freedom  $df$  and noncentrality parameter  $\lambda$ .

## Procedure Options

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

### Design Tab

The Design tab contains most of the parameters and options that will be of interest.

#### Solve For

##### Solve For

This option specifies the parameter to be calculated from the values of the other parameters. Under most conditions, you would select either *Power* or *Sample Size*.

Select *Sample Size* when you want to determine the sample size needed to achieve a given power and alpha error level.

Select *Power* when you want to calculate the power.

#### Test

##### Higher Means Are

This option defines whether higher values of the response variable are to be considered better or worse. The choice here determines the direction of the non-inferiority test. Note that  $\delta$  is the population mean of paired differences.

- **Better (H1:  $\delta > -NIM$ )**

If higher means are Better the null hypothesis is  $H_0: \delta \leq -NIM$ , and the alternative hypothesis is  $H_1: \delta > -NIM$ .

- **Worse (H1:  $\delta < NIM$ )**

If higher means are Worse the null hypothesis is  $H_0: \delta \geq NIM$ , and the alternative hypothesis is  $H_1: \delta < NIM$ .

##### Population Size

This is the number of subjects in the population. Usually, you assume that samples are drawn from a very large (infinite) population. Occasionally, however, situations arise in which the population of interest is of limited size. In these cases, appropriate adjustments must be made.

When a finite population size is specified, the standard deviation is reduced according to the formula:

$$\sigma_1^2 = \left(1 - \frac{n}{N}\right) \sigma^2$$

where  $n$  is the sample size,  $N$  is the population size,  $\sigma$  is the original standard deviation, and  $\sigma_1$  is the new standard deviation.

The quantity  $n/N$  is often called the sampling fraction. The quantity  $\left(1 - \frac{n}{N}\right)$  is called the *finite population correction factor*.

## Paired T-Tests for Non-Inferiority

---

### Power and Alpha

#### Power

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

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

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

#### Alpha

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

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

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

---

### Sample Size

#### N (Sample Size)

This option specifies one or more values of the sample size, the number of individuals in the study. This value must be an integer greater than one. You may enter a list of values using the syntax *50 100 150 200 250* or *50 to 250 by 50*.

---

### Effect Size – Means

#### NIM (Non-Inferiority Margin)

This is the magnitude of the margin of non-inferiority for the mean of paired differences. It should be entered as a positive number. If a negative value is entered, the absolute value is used.

When higher means are better, this value is the distance below zero for which the new treatment will still be considered non-inferior to the standard treatment.

When higher means are worse, this value is the distance above zero for which the new treatment will still be considered non-inferior to the standard treatment.

#### $\delta_1$ (Mean of Paired Differences)

Enter a value (or range of values) for the mean of paired differences at which power and sample size are calculated. This value indicates the minimum detectible paired difference.

For non-inferiority tests, this value is often set to 0. When higher means are better, this value should be greater than -NIM. When higher means are worse, this value should be less than NIM.

---

### Effect Size – Standard Deviation

#### $\sigma$ (Std Dev of Paired Differences)

Enter an estimate of the standard deviation of the paired differences (must be positive).

When this value is not known, you must supply an estimate of it. **PASS** includes a special tool for estimating the standard deviation. This tool may be loaded by pressing the *SD* button. Refer to the Standard Deviation Estimator chapter for further details.

## Paired T-Tests for Non-Inferiority

## Example 1 – Power Analysis

Suppose that a test is to be conducted to determine if a new cancer treatment adversely affects the mean bone density. The adjusted mean bone density (AMBD) in the population of interest is 0.002300 gm/cm. Clinicians decide that if the treatment reduces AMBD by more than 5% (0.000115 gm/cm), it poses a significant health threat. They also want to consider what would happen if the margin of non-inferiority is set to 2.5% (0.0000575 gm/cm). The standard deviation of paired differences is 0.000300 gm/cm.

Following accepted procedure, the analysis will be a non-inferiority test using the paired  $t$ -test at the 0.025 significance level. Power is to be calculated assuming that the new treatment has no effect on AMBD. Several sample sizes between 20 and 300 will be analyzed. The researchers want to achieve a power of at least 90%. All numbers have been multiplied by 10000 to make the reports and plots easier to read.

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Paired T-Tests for Non-Inferiority** procedure window by expanding **Means**, then **Paired Means**, then clicking on **Non-Inferiority**, and then clicking on **Paired T-Tests for Non-Inferiority**. 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 Means Are.....	<b>Better (H1: <math>\delta &gt; -NIM</math>)</b>
Population Size.....	<b>Infinite</b>
Alpha.....	<b>0.025</b>
N (Sample Size).....	<b>20 40 60 80 100 150 200 300</b>
NIM (Non-Inferiority Margin) .....	<b>0.575 1.15</b>
$\delta_1$ (Mean of Paired Differences).....	<b>0</b>
$\sigma$ (Std Dev of Paired Differences).....	<b>3</b>

### Annotated Output

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

#### Numeric Results

Higher Means are Better

Hypotheses:  $H_0: \delta \leq -NIM$  vs.  $H_1: \delta > -NIM$

	Non-Inferiority Margin -NIM	Mean of Paired Differences $\delta_1$	Standard Deviation $\sigma$	Alpha	Beta	
Power	N					
0.12601	20	-0.575	0.000	3.0	0.025	0.87399
0.21844	40	-0.575	0.000	3.0	0.025	0.78156
0.30873	60	-0.575	0.000	3.0	0.025	0.69127
0.39493	80	-0.575	0.000	3.0	0.025	0.60507
0.47532	100	-0.575	0.000	3.0	0.025	0.52468
0.64517	150	-0.575	0.000	3.0	0.025	0.35483
0.76959	200	-0.575	0.000	3.0	0.025	0.23041
0.91135	300	-0.575	0.000	3.0	0.025	0.08865
0.36990	20	-1.150	0.000	3.0	0.025	0.63010
0.65705	40	-1.150	0.000	3.0	0.025	0.34295
0.83164	60	-1.150	0.000	3.0	0.025	0.16836
0.92317	80	-1.150	0.000	3.0	0.025	0.07683
0.96682	100	-1.150	0.000	3.0	0.025	0.03318

(report continues)

### Paired T-Tests for Non-Inferiority

#### Report Definitions

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

N is the sample size, the number of subjects (or pairs) in the study.

-NIM is the magnitude and direction of the margin of non-inferiority. Since higher means are better, this value is negative and is the distance below the reference value that is still considered non-inferior.

$\delta$  is the mean of paired differences.

$\delta_1$  is the value of the mean of paired differences at which power and sample size are calculated.

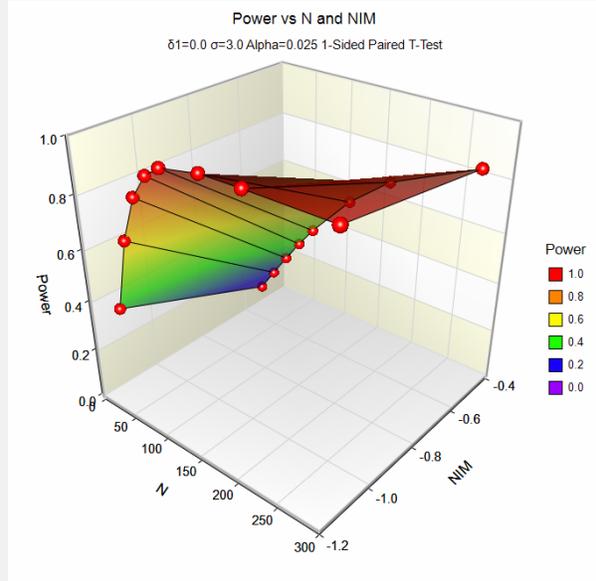
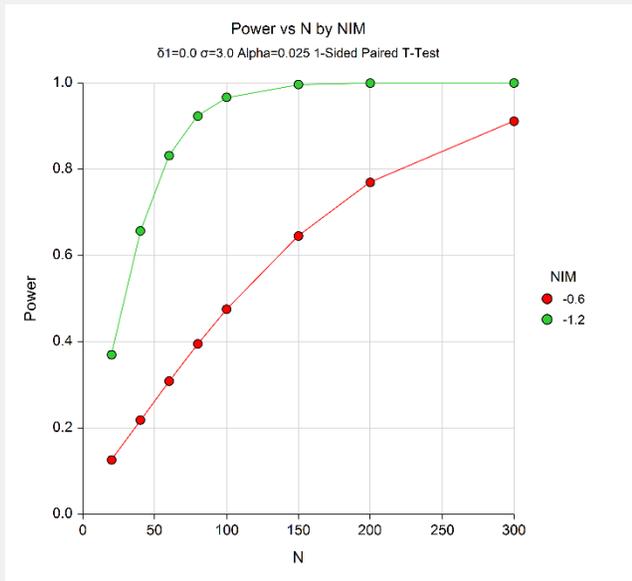
$\sigma$  is the standard deviation of paired differences for the population.

Alpha is the probability of rejecting a true null hypothesis. It should be small.

Beta is the probability of accepting a false null hypothesis. It should be small.

#### Summary Statements

A sample size of 20 achieves 13% power to detect non-inferiority using a one-sided paired t-test when the non-inferiority margin is -0.575 and the actual mean of paired differences is 0.000. The an estimated standard deviation of paired differences is 3.0. The significance level (alpha) of the test is 0.025.



The above report shows that for NIM = 1.15, the sample size necessary to obtain 90% power is just under 80. However, if NIM = 0.575, the required sample size is about 300.

## Paired T-Tests for Non-Inferiority

## Example 2 – Finding the Sample Size

Continuing with Example 1, the researchers want to know the exact sample size for each value of NIM.

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Paired T-Tests for Non-Inferiority** procedure window by expanding **Means**, then **Paired Means**, then clicking on **Non-Inferiority**, and then clicking on **Paired T-Tests for Non-Inferiority**. 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</b>
Higher Means Are .....	<b>Better (H1: <math>\delta &gt; -NIM</math>)</b>
Population Size .....	<b>Infinite</b>
Power .....	<b>0.90</b>
Alpha .....	<b>0.025</b>
NIM (Non-Inferiority Margin) .....	<b>0.575 1.15</b>
$\delta_1$ (Mean of Paired Differences) .....	<b>0</b>
$\sigma$ (Std Dev of Paired Differences) .....	<b>3</b>

### Output

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

#### Numeric Results

Numeric Results						
Higher Means are Better						
Hypotheses: $H_0: \delta \leq -NIM$ vs. $H_1: \delta > -NIM$						
	Non-Inferiority Margin -NIM	Mean of Paired Differences $\delta_1$	Standard Deviation $\sigma$		Alpha	Beta
Power	N					
0.90005	288	-0.575	0.000	3.0	0.025	0.09995
0.90215	74	-1.150	0.000	3.0	0.025	0.09785

This report shows the exact sample size requirement for each value of NIM.

## Paired T-Tests for Non-Inferiority

## Example 3 – Validation using Chow, Shao, Wang, and Lokhnygina (2018)

Chow, Shao, Wang, and Lokhnygina (2018) page 46 has an example of a sample size calculation for a paired non-inferiority  $t$ -test. Their example obtains a sample size of 8 when  $\delta_1 = 0.5$ , NIM = 0.5,  $\sigma = 1$ , Alpha = 0.05, and Power = 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 **Paired T-Tests for Non-Inferiority** procedure window by expanding **Means**, then **Paired Means**, then clicking on **Non-Inferiority**, and then clicking on **Paired T-Tests for Non-Inferiority**. 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>Sample Size</b>
Higher Means Are .....	<b>Better (H1: <math>\delta &gt; -\text{NIM}</math>)</b>
Population Size .....	<b>Infinite</b>
Power .....	<b>0.80</b>
Alpha .....	<b>0.05</b>
NIM (Non-Inferiority Margin) .....	<b>0.5</b>
$\delta_1$ (Mean of Paired Differences) .....	<b>0.5</b>
$\sigma$ (Std Dev of Paired Differences) .....	<b>1</b>

### Output

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

<b>Numeric Results</b>						
Higher Means are Better						
Hypotheses: $H_0: \delta \leq -\text{NIM}$ vs. $H_1: \delta > -\text{NIM}$						
	<b>Non-Inferiority Margin -NIM</b>	<b>Mean of Paired Differences <math>\delta_1</math></b>	<b>Standard Deviation <math>\sigma</math></b>	<b>Alpha</b>	<b>Beta</b>	
<b>Power</b>	<b>N</b>					
0.81502	8	-0.5	0.5	1.0	0.050	0.18498

PASS also obtains a sample size of 8.

## Paired T-Tests for Non-Inferiority

## Example 4 – Validation of a Cross-Over Design given in Julious (2004)

Julious (2004) page 1953 gives an example of a sample size calculation for a cross-over design. His example obtains a sample size of 87 when  $\delta_1 = 0$ ,  $NIM = 10$ ,  $\sigma = 28.28427$ ,  $\text{Alpha} = 0.025$ , and  $\text{Beta} = 0.10$ . When  $\delta_1$  is changed to 2, the resulting sample size is 61.

Note that in Julius's example, the population standard deviation is given as 20. Assuming that the correlation between items in a pair is 0, the standard deviation of the difference is calculated to be  $\sigma = \sqrt{20^2 + 20^2 - 0(20)(20)} = 28.284271$ . Actually, the value of  $\sigma$  probably should be less because the correlation is usually greater than 0 (at least 0.2).

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Paired T-Tests for Non-Inferiority** procedure window by expanding **Means**, then **Paired Means**, then clicking on **Non-Inferiority**, and then clicking on **Paired T-Tests for Non-Inferiority**. 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 .....	<b>Sample Size</b>
Higher Means Are .....	<b>Better (H1: <math>\delta &gt; -NIM</math>)</b>
Population Size .....	<b>Infinite</b>
Power .....	<b>0.90</b>
Alpha .....	<b>0.025</b>
NIM (Non-Inferiority Margin) .....	<b>10</b>
$\delta_1$ (Mean of Paired Differences) .....	<b>0 2</b>
$\sigma$ (Std Dev of Paired Differences) .....	<b>28.284271</b>

### Output

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

<b>Numeric Results</b>						
Higher Means are Better						
Hypotheses: $H_0: \delta \leq -NIM$ vs. $H_1: \delta > -NIM$						
	<b>Non-Inferiority Margin</b>	<b>Mean of Paired Differences</b>	<b>Standard Deviation</b>		<b>Alpha</b>	<b>Beta</b>
<b>Power</b>	<b>N</b>	<b>-NIM</b>	<b><math>\delta_1</math></b>	<b><math>\sigma</math></b>		
0.90332	87	-10.0	0.0	28.3	0.025	0.09668
0.90323	61	-10.0	2.0	28.3	0.025	0.09677

PASS also obtains sample sizes of 87 and 61.

## Paired T-Tests for Non-Inferiority

## Example 5 – Validation of a Cross-Over Design given in Chow, Shao, Wang, and Lokhnygina (2018)

Chow, Shao, Wang, and Lokhnygina (2018) page 59 gives an example of a sample size calculation for a cross-over design. Their example calculates sample sizes of 13 and 14 (13 by normal approximation formula and 14 from their table for the  $t$ -test) in each sequence (26 or 28 total) when  $\delta_1 = -0.1$ , NIM = 0.2,  $\sigma = 0.2$ , Alpha = 0.05, and Beta = 0.20.

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Paired T-Tests for Non-Inferiority** procedure window by expanding **Means**, then **Paired Means**, then clicking on **Non-Inferiority**, and then clicking on **Paired T-Tests for Non-Inferiority**. You may then make the appropriate entries as listed below, or open **Example 5** 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</b>
Higher Means Are.....	<b>Better (H1: <math>\delta &gt; -\text{NIM}</math>)</b>
Population Size.....	<b>Infinite</b>
Power.....	<b>0.80</b>
Alpha.....	<b>0.05</b>
NIM (Non-Inferiority Margin) .....	<b>0.2</b>
$\delta_1$ (Mean of Paired Differences).....	<b>-0.1</b>
$\sigma$ (Std Dev of Paired Differences).....	<b>0.2</b>

### Output

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

<b>Numeric Results</b>						
Higher Means are Better						
Hypotheses: $H_0: \delta \leq -\text{NIM}$ vs. $H_1: \delta > -\text{NIM}$						
	<b>Non-Inferiority Margin</b>	<b>Mean of Paired Differences</b>	<b>Standard Deviation</b>		<b>Alpha</b>	<b>Beta</b>
<b>Power</b>	<b>N</b>	<b>-NIM</b>	<b><math>\delta_1</math></b>	<b><math>\sigma</math></b>		
0.81183	27	-0.2	-0.1	0.2	0.050	0.18817

PASS obtains a sample size of 27, which is between the values of 26 and 28 that were obtained by Chow et al. (2018)