

Chapter 444

Conditional Power and Sample Size Reestimation of Two-Sample T-Tests for Superiority by a Margin

Introduction

In sequential designs, one or more intermediate analyses of the emerging data are conducted to evaluate whether the experiment should be continued. This may be done to conserve resources or to allow a data monitoring board to evaluate safety and efficacy when subjects are entered in a staggered fashion over a long period of time. *Conditional power* (a frequentist concept) is the probability that the final result will be significant, given the data obtained up to the time of the interim look. *Predictive power* (a Bayesian concept) is the result of averaging the conditional power over the posterior distribution of effect size. Both of these methods fall under the heading of *stochastic curtailment* techniques. Further reading about the theory of these methods can be found in Jennison and Turnbull (2000), Chow and Chang (2007), Chang (2008), Proschan et.al (2006), and Dmitrienko et.al (2005).

This program module computes conditional and predictive power for the case when a two-sample *t*-test is used to test superiority by a margin for two means. It provides *sample size reestimation* to achieve a specified conditional power value.

Technical Details

All details and assumptions usually made when using a two-sample non-inferiority *t*-test continue to be in force here.

Conditional Power

The power of an experiment indicates whether a study is likely to result in useful results, given the sample size. Low power means that the study is *futile*: little chance of statistical significance even though the alternative hypothesis is true. A study that is futile should not be started. However, futility may be determined only after the study has started. When this happens, the study is *curtailed*.

Conditional Power and Sample Size Reestimation of Two-Sample T-Tests for Superiority by a Margin

The futility of a study that is underway can be determined by calculating its *conditional power*: the probability of statistical significance at the completion of the study given the data obtained so far.

It is important to note that conditional power at the beginning of the study before any data are collected is equal to the unconditional power. So, conditional power will be high even if early results are negative. Hence, conditional power will seldom result in study curtailment very early in the study.

From Jennison and Turnbull (2000) pages 205 to 208, the general upper one-sided conditional power at stage k for rejecting a null hypothesis about a parameter θ at the end of the study, given the observed test statistic, Z_k , is computed as

$$P_{uk}(\theta) = \Phi\left(\frac{Z_k\sqrt{I_k} - z_{1-\alpha}\sqrt{I_K} + \theta(I_K - I_k)}{\sqrt{I_K - I_k}}\right),$$

and the general lower one-sided conditional power at stage k is computed as

$$P_{lk}(\theta) = \Phi\left(\frac{-Z_k\sqrt{I_k} - z_{1-\alpha}\sqrt{I_K} - \theta(I_K - I_k)}{\sqrt{I_K - I_k}}\right),$$

where

θ = the parameter being tested by the hypothesis

k = an interim stage at which the conditional power is computed ($k = 1, \dots, K - 1$)

K = the stage at which the study is terminated and the final test computed

Z_k = the test statistic calculated from the observed data that has been collected up to stage k

I_k = the information level at stage k

I_K = the information level at the end of the study

$z_{1-\alpha}$ = the standard normal value for the test with a type I error rate of α .

Let μ_1 and μ_2 be the population means in groups 1 and 2, respectively. If we define $\delta = \mu_2 - \mu_1$, such that $\delta 0$ is the superiority difference boundary, $\delta 1$ is the true population difference under the alternative hypothesis, and $\hat{\delta}_k = \bar{x}_{2k} - \bar{x}_{1k}$ is the estimated mean difference from the observed data at stage k , then the parameter θ to test the one-sided superiority by a margin alternative hypotheses of $H_1: \delta > \delta 0$ (higher means better) or $H_1: \delta < \delta 0$ (higher means worse) and other conditional power calculation components as outlined in Chang (2008) page 70 are

$\theta = \delta 1 - \delta 0$ (the expected difference under the alternative hypothesis)

$Z_k = (\hat{\delta}_k - \delta 0)\sqrt{\hat{I}_k}$ (the superiority t -statistic computed from the observed data)

$I_k = \left(\frac{\sigma_1^2}{n_{1k}} + \frac{\sigma_2^2}{n_{2k}}\right)^{-1}$ (the interim information level)

$I_K = \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)^{-1}$ (the final information level)

Conditional Power and Sample Size Reestimation of Two-Sample T-Tests for Superiority by a Margin

where

\hat{I}_k is the estimated information from the sample at stage k

n_{jk} is the sample size in group j at stage k

n_j is the final sample size in group j

σ_j^2 is the variance of group j

Computing conditional power requires you to set δ_0 , δ_1 , σ_1 , and σ_2 , in addition to the current test statistic value, Z_k . Their values can come from the values used during the planning of the study, from similar studies, or from estimates made from the data that has emerged.

Futility Index

The *futility index* is $1 - P_k(\theta)|H_1$. The study may be stopped if this index is above 0.8 or 0.9 (that is, if conditional power falls below 0.2 or 0.1).

Predictive Power

Predictive power (a Bayesian concept) is the result of averaging the conditional power over the posterior distribution of effect size. From Jennison and Turnbull (2000) pages 210 to 213, the general upper one-sided predictive power at stage k is given by

$$P_{uk} = \Phi \left(\frac{Z_k \sqrt{I_K} - z_{1-\alpha} \sqrt{I_k}}{\sqrt{I_K - I_k}} \right),$$

and the general lower one-sided predictive power at stage k is given by

$$P_{lk} = \Phi \left(\frac{-Z_k \sqrt{I_K} - z_{1-\alpha} \sqrt{I_k}}{\sqrt{I_K - I_k}} \right),$$

with all terms defined as in the equations for conditional power.

Sample Size Reestimation

As Chang (2014) points out, after an interim analysis, it is often desirable to recalculate the target sample size using updated values for various nuisance parameters such as the variance. This process is known as *sample size reestimation*.

One method of calculating an adjusted sample size estimate is to search for the sample size that results in a predetermined value of conditional power. PASS conducts a binary search using the conditional power as the criterion. The result is called the *target sample size*.

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 you will be concerned with.

Solve For

Solve For

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are *Conditional Power* or *Sample Size Reestimation*.

Note that the value selected here always appears as the vertical axis on the charts.

The program is set up to calculate conditional power directly. To find appropriate values of sample size, a binary search is made using an iterative procedure until an appropriate value is found.

Test

Higher Means Are

Use this option to specify the direction of the superiority by a margin test. If Higher Means are “Better”, the alternative hypothesis is $H_1: \delta > \delta_0$, with $\delta_0 > 0$. If Higher Means are “Worse”, the alternative hypothesis is $H_1: \delta < \delta_0$, with $\delta_0 < 0$.

Conditional Power and Alpha

Conditional Power

Enter one or more values for the conditional power that is to be attained by the sample size search.

Conditional power is the probability of rejecting a false null hypothesis given the data that has been obtained so far. It is a function of the effect size and the data that have been collected.

The valid range is between 0 to 1. Typically, values over 0.70 are considered high enough to continue a study, while values below 0.10 considered grounds for curtailing the study.

You can enter a single value such as 0.80 or a series of values such as 0.80 0.85 0.90 0.95 or 0.7 to 0.97 by 0.05.

Alpha

This option specifies one or more values for the probability of a type-I error at the end of the study. 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 two-sided tests and 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

N1 (Group 1 Target Sample Size)

Enter one or more values of the target sample size of group 1, the total number of subjects planned for this group. This value must be an integer greater than one. Note that you may enter a list of values using the syntax *50,100,150,200,250* or *50 to 250 by 50*.

N2 (Group 2 Target Sample Size)

Enter a value (or range of values) for the target sample size of group 2 or enter *Use R* to calculate *N2* from *N1*. You may enter a range of values such as *10 to 100 by 10*.

- **Use R**

If *Use R* is entered here, *N2* is calculated using the formula

$$N2 = [R(N1)]$$

where *R* is the Sample Allocation Ratio and the operator $[Y]$ is the first integer greater than or equal to *Y*. For example, if you want $N1 = N2$, select *Use R* and set $R = 1$.

R (Sample Allocation Ratio)

Enter a value (or range of values) for *R*, the allocation ratio between samples. This value is only used when *N2* is set to *Use R*.

When used, *N2* is calculated from *N1* using the formula: $N2 = [R(N1)]$ where $[Y]$ is the next integer greater than or equal to *Y*. Note that setting $R = 1.0$ forces $N2 = N1$.

n1k (Group 1 Sample Size at Look k)

Enter the group 1 sample size obtained through look *k*. If this value is greater than *N1*, the value of *N1* is increased to this amount.

n2k (Group 2 Sample Size at Look k)

Enter the group 2 sample size obtained through look *k*. If this value is greater than *N2*, the value of *N2* is increased to this amount.

Equal Sample Sizes

If *n1k* is entered here, *n2k* is set equal to *n1k*.

Effect Size – Mean Difference

δ_0 (Superiority Difference)

This is the difference bound of the superiority by a margin test hypothesis. When Higher Means are “Better”, the alternative hypothesis is $H1: \delta > \delta_0$ and you can enter values in the range $0 < \delta_0 < \delta_1$. When Higher Means are “Worse”, the alternative hypothesis is $H1: \delta < \delta_0$ and you can enter values in the range $0 > \delta_0 > \delta_1$.

δ_1 (Actual Difference to Detect)

This is the actual difference between the treatment mean and the reference mean at which power is calculated. Care should be taken that this value is consistent with the alternative hypothesis.

When Higher Means are “Better”, $\delta_1 > \delta_0 > 0$. When Higher Means are “Worse”, $\delta_1 < \delta_0 < 0$.

Effect Size – Standard Deviations

σ_1 (Standard Deviation of Group 1)

Enter a value for σ_1 , the standard deviation of group 1. This value is controversial. Some think you should enter the value used in planning the study. Others think you should use the value estimated from the data obtained so far. Still others think you should use a confidence limit for σ_1 created from the current sample.

You can enter a range of values such as *1 2 3* or *1 to 5 by 1*.

Press the σ button to the right to load the Standard Deviation Estimator window.

σ_2 (Standard Deviation of Group 2)

Enter a value for σ_2 , the standard deviation of group 2. This value is controversial. Some think you should enter the value used in planning the study. Others think you should use the value estimated from the data obtained so far. Still others think you should use a confidence limit for σ_1 created from the current sample.

You can enter a range of values such as *1 2 3* or *1 to 5 by 1*.

Equal Variance

Enter 'σ1' to set $\sigma_2 = \sigma_1$.

Effect Size – Current Test Statistic

Z_k (Current Test Statistic)

Enter the value of the T -statistic calculated from the data that have been obtained so far. The test statistic is calculated as

$$Z_k = (\bar{x}_{2k} - \bar{x}_{1k} - \delta_0) \times \sqrt{\hat{I}_k}$$

where \bar{x}_{1k} and \bar{x}_{2k} are the group 1 and group 2 sample means observed at stage k , respectively, and \hat{I}_k is the information estimate at stage k .

Z_k may be positive or negative. Usually, Z_k ranges between -5 and 5.

The conditional power calculations used here are for large samples which are at least 20.

Example 1 – Computing Conditional Power

Suppose a study has been planned and is to be analyzed using a one-sided superiority *t*-test against a lower difference bound of $\delta_0 = 1$ at an alpha of 0.025. The sample size is 60. The standard deviations are expected to be about 4. An interim analysis is run after half the data have been collected. This analysis yields a *t*-test value of 2.12.

The data monitoring board would like to have the conditional power calculated for mean changes, $\delta_1 = \mu_2 - \mu_1$, of 1.5, 2, 2.5, 3, and 3.5.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the procedure window. You may make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	Conditional Power
Higher Means Are	Better (H1: $\delta > \delta_0$)
Alpha	0.025
N1 (Group 1 Target Sample Size)	60
N2 (Group 2 Target Sample Size)	Use R
R (Sample Allocation Ratio)	1.0
n1k (Group 1 Sample Size at Look k).....	30
n2k (Group 2 Sample Size at Look k).....	n1k
δ_0 (Superiority Difference).....	1
δ_1 (Actual Difference to Detect)	1.5 2 2.5 3 3.5
σ_1 (Standard Deviation of Group 1)	4
σ_2 (Standard Deviation of Group 2)	σ_1
Zk (Current Test Statistic).....	2.12

Annotated Output

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

Numeric Results

Numeric Results for Two-Sample T-Test									
Hypotheses: H0: $\delta \leq \delta_0$ vs. H1: $\delta > \delta_0$									
Cond	Pred	Target	Current	Superiority	Actual	Std	Test		
Power	Power	Sample	Sample	Diff	Diff	Dev	Stat	Alpha	Futility
		Sizes	Sizes	δ_0	δ_1	$\sigma_1 \sigma_2$	Zk		
		N1 N2	n1k n2k						
0.43342	0.85040	60 60	30 30	1.0	1.5	4.0 4.0	2.120	0.025	0.56658
0.62417	0.85040	60 60	30 30	1.0	2.0	4.0 4.0	2.120	0.025	0.37583
0.78831	0.85040	60 60	30 30	1.0	2.5	4.0 4.0	2.120	0.025	0.21169
0.90055	0.85040	60 60	30 30	1.0	3.0	4.0 4.0	2.120	0.025	0.09945
0.96154	0.85040	60 60	30 30	1.0	3.5	4.0 4.0	2.120	0.025	0.03846

Conditional Power and Sample Size Reestimation of Two-Sample T-Tests for Superiority by a Margin

References

- Jennison, C., and Turnbull, B.W. 2000. Group Sequential Methods with Applications to Clinical Trials. Chapman & Hall/CRC. New York.
- Proschan, M., Lan, K.K.G., Wittes, J.T. 2006. Statistical Monitoring of Clinical Trials. Springer. New York.
- Chang, Mark. 2008. Classical and Adaptive Clinical Trial Designs. John Wiley & Sons. Hoboken, New Jersey.
- Chang, Mark. 2014. Adaptive Design Theory and Implementation Using SAS and R. CRC Press. New York.

Report Definitions

Conditional Power is the probability of rejecting a false null hypothesis at the end of the study given the data that have emerged so far.

Predictive Power is the result of averaging the conditional power over the posterior distribution of the effect size.

$N1|N2$ are the target sample sizes of groups 1 and 2, respectively.

$n1k|n2k$ are the sample sizes of groups 1 and 2, respectively, obtained through stage k.

$\delta = \mu2 - \mu1$ is the mean difference.

$\delta0$ is the superiority difference used to construct the hypotheses.

$\delta1$ is the actual difference to detect under the alternative hypothesis at which conditional power is calculated.

$\sigma1|\sigma2$ are the standard deviations of groups 1 and 2, respectively.

Zk is the value of the test statistic from the observed data at stage k.

Alpha is the probability of rejecting a true null hypothesis.

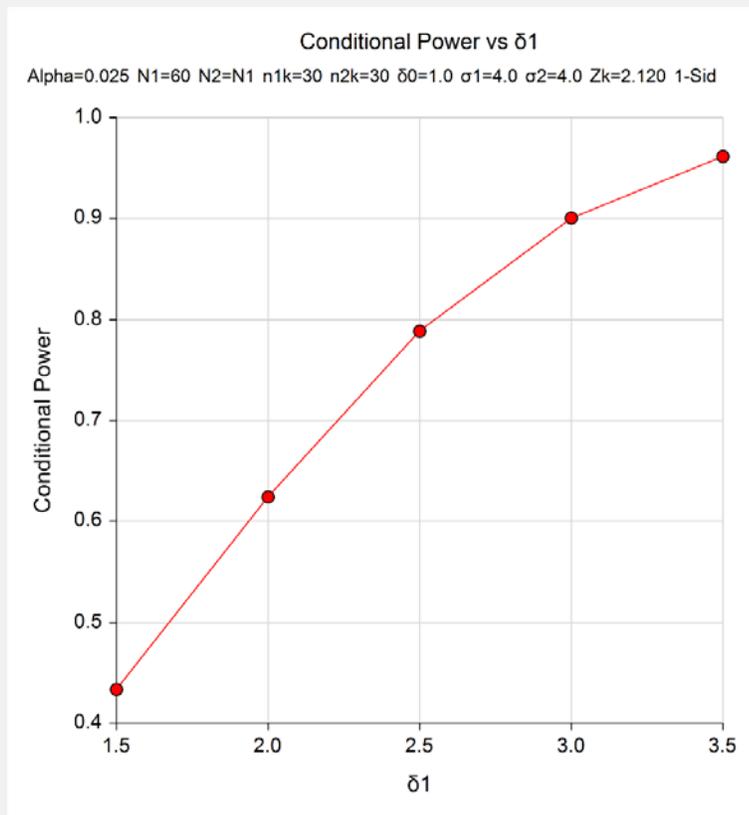
Futility is one minus the conditional power. A value greater than 0.9 or 0.8 indicates the study should be stopped because there is little chance of achieving statistical significance.

Summary Statements

The first 30 of 60 subjects in group 1 and 30 of 60 subjects in group 2 achieve 43.342% conditional power to detect a difference of $\delta1 = 1.5$ using a T-test for superiority by a margin with a difference bound of $\delta0 = 1.0$, group 1 standard deviation of $\sigma1 = 4.0$, group 2 standard deviation of $\sigma2 = 4.0$, and a significance level of $\alpha = 0.025$. The value of the test statistic from data that have emerged so far is $Zk = 2.120$. The futility index is 0.56658.

This report shows the values of each of the parameters, one scenario per row. The definitions of each column are given in the Report Definitions section.

Plots Section



This plot shows the relationship between conditional power and $\delta1$.

Example 2 – Validation

We could not find an example of a conditional power calculation for a one-sided superiority by a margin t -test in the literature. Since the calculations are relatively simple, we will validate the calculation of the first scenario ($\theta = \delta_1 - \delta_0 = 0.5$) of Example 1 by hand.

In this case

$$\begin{aligned} I_k &= \left(\sigma_1^2 / n_{1k} + \sigma_2^2 / n_{2k} \right)^{-1} & I_K &= \left(\sigma_1^2 / n_1 + \sigma_2^2 / n_2 \right)^{-1} \\ &= \left(\frac{16}{30} + \frac{16}{30} \right)^{-1} & &= \left(\frac{16}{60} + \frac{16}{60} \right)^{-1} \\ &= 0.9375 & &= 1.875 \end{aligned}$$

$$\begin{aligned} P_{uk}(\theta) &= \Phi \left(\frac{Z_k \sqrt{I_k} - Z_{1-\alpha} \sqrt{I_K} + \theta(I_K - I_k)}{\sqrt{I_K - I_k}} \right) \\ &= \Phi \left(\frac{2.12 \sqrt{0.9375} - 1.9599640 \sqrt{1.875} + 0.5(1.875 - 0.9375)}{\sqrt{1.875 - 0.9375}} \right) \\ &= \Phi(-0.1676848) \\ &= 0.4334156 \end{aligned}$$

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the procedure window. You may make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Conditional Power
Higher Means Are.....	Better (H1: $\delta > \delta_0$)
Alpha.....	0.025
N1 (Group 1 Target Sample Size).....	60
N2 (Group 2 Target Sample Size).....	Use R
R (Sample Allocation Ratio)	1.0
n1k (Group 1 Sample Size at Look k).....	30
n2k (Group 2 Sample Size at Look k).....	n1k
δ_0 (Superiority Difference).....	1
δ_1 (Actual Difference to Detect)	1.5
σ_1 (Standard Deviation of Group 1)	4
σ_2 (Standard Deviation of Group 2)	σ_1
Zk (Current Test Statistic).....	2.12

Output

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

Numeric Results

Numeric Results for Two-Sample T-Test

Hypotheses: $H_0: \delta \leq \delta_0$ vs. $H_1: \delta > \delta_0$

Cond Power	Pred Power	Target Sample Sizes N1 N2	Current Sample Sizes n1k n2k	Superiority Diff δ_0	Actual Diff δ_1	Std Dev $\sigma_1 \sigma_2$	Test Stat Zk	Alpha	Futility
0.43342	0.85040	60 60	30 30	1.0	1.5	4.0 4.0	2.120	0.025	0.56658

The conditional power of 0.43342 matches the value calculated by hand.

Example 3 – Sample Size Reestimation

Suppose a study has been planned and is to be analyzed using a one-sided superiority *t*-test against a lower difference bound of $\delta_0 = 1$ at an alpha of 0.025. The original sample size was set at 60 per group. The standard deviations were expected to be about 4. An interim analysis was run after half the data were collected. This analysis yields a z-test value of 2.12. This value was obtained by transforming the t-test p-value using the inverse normal.

The data monitoring board would like to recalculate the sample size for a mean change of 2, a conditional power of 0.8, and a standard deviation of 6.7.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the procedure window. You may make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

<u>Option</u>	<u>Value</u>
Design Tab	
Solve For	Sample Size Reestimation
Higher Means Are.....	Better (H1: $\delta > \delta_0$)
Conditional Power.....	0.8
Alpha.....	0.025
N2 (Group 2 Target Sample Size)	Use R
R (Sample Allocation Ratio)	1.0
n1k (Group 1 Sample Size at Look k).....	30
n2k (Group 2 Sample Size at Look k).....	n1k
δ_0 (Superiority Difference).....	1
δ_1 (Actual Difference to Detect)	2
σ_1 (Standard Deviation of Group 1)	6.7
σ_2 (Standard Deviation of Group 2)	σ_1
Zk (Current Test Statistic).....	2.12

Output

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

Numeric Results

Numeric Results for Two-Sample T-Test									
Hypotheses: H0: $\delta \leq \delta_0$ vs. H1: $\delta > \delta_0$									
Cond	Pred Power	Target Sample Sizes N1 N2	Current Sample Sizes n1k n2k	Superiority Diff δ_0	Actual Diff δ_1	Std Dev $\sigma_1 \sigma_2$	Test Stat Zk	Alpha	Futility
0.80000	0.95534	520 520	30 30	1.0	2.0	6.7 6.7	2.120	0.025	0.20000

Notice that the target sample size has increased from 60 per group (N = 120), to 520 per group (N = 1040).