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*.

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, ..., 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)

$Z_k = \left(\hat{\delta}_k - \delta 0\right)\sqrt{\hat{I}_k}$ (the superiority z-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) 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 $\delta 0$, $\delta 1$, 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.

Converting from a T-Statistic to a Z-Statistic

A common problem is that the procedure requires a Z-statistic, but the results from an analysis usually provide a T-statistic with a given degrees of freedom. So, the T-statistic must be transformed to a Z-statistic. One way to do this is to use the associated p-value. This is accomplished using the following steps:

Step 1. **Find the p-value associated with the t-statistic**
For example, suppose we have a t-statistic of 2.33 with 30 degrees of freedom. Using the **PASS** Probability Calculator, the probability for Student's T distribution is Prob(t ≥ T) = 0.0133616.

Step 2. **Convert this p-value to a z-statistic**
Continuing the example, we can use the Normal distribution in the **PASS** Probability Calculator to determine that the z-score associated with a p-value of 0.0133616 is -2.215537403212. Since the t-statistic was positive, we use z = 2.215537403212.

Step 3. **Enter the z-statistic for Zk (Current Test Statistic)**
Continuing the example, we would enter 2.2155374 for Zk (Current Test Statistic).

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*.

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 δ0 = 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, δ1 = μ2 - μ1, of 1.5, 2, 2.5, 3, and 3.5.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For	**Conditional Power**
Higher Means Are	**Better (H1: δ > δ0)**
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**
δ0 (Superiority Difference)	**1**
δ1 (Actual Difference to Detect)	**1.5 2 2.5 3 3.5**
σ1 (Standard Deviation of Group 1)	**4**
σ2 (Standard Deviation of Group 2)	**σ1**
Test Statistic Input Type	**Tk**
Tk (Current Test Statistic)	**2.12**

Output

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

Numeric Reports

Numeric Results

Solve For: Conditional Power
Test Type: Two-Sample T-Test
Groups: 1 = Control, 2 = Treatment
Hypotheses: H0: δ ≤ δ0 vs. H1: δ > δ0

Power		Sample Size				Mean Difference (μ2 - μ1)		Standard Deviation		Test Statistic		
		Target		Look k		Superiority	Actual					
Conditional	Predictive	N1	N2	n1k	n2k	δ0	δ1	σ1	σ2	Tk	Alpha	Futility
0.41449	0.83393	60	60	30	30	1	1.5	4	4	2.12	0.025	0.58551
0.60569	0.83393	60	60	30	30	1	2.0	4	4	2.12	0.025	0.39431
0.77405	0.83393	60	60	30	30	1	2.5	4	4	2.12	0.025	0.22595
0.89184	0.83393	60	60	30	30	1	3.0	4	4	2.12	0.025	0.10816
0.95733	0.83393	60	60	30	30	1	3.5	4	4	2.12	0.025	0.04267

Conditional Power The probability of rejecting a false null hypothesis at the end of the study given the data that have emerged so far.
Predictive Power The result of averaging the conditional power over the posterior distribution of the effect size.
N1 and N2 The target sample sizes of groups 1 and 2, respectively.
n1k and n2k The actual sample sizes of groups 1 and 2, respectively, obtained through stage k.
δ The mean difference. δ = μ2 - μ1.
δ0 The superiority difference used to construct the hypotheses.
δ1 The actual difference to detect under the alternative hypothesis at which conditional power is calculated.
σ1 and σ2 The standard deviations of groups 1 and 2, respectively.
Tk The value of the test statistic from the observed data at stage k.
Alpha The probability of rejecting a true null hypothesis.
Futility Equal to 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

A parallel two-group design will be used to test whether the Group 2 (treatment) mean (μ2) is superior to the Group 1 (reference) mean (μ1) by a margin, with a superiority margin of 1 (H0: δ ≤ 1 versus H1: δ > 1, δ = μ2 - μ1). The comparison will be made using a one-sided, two-sample t-test, with a Type I error rate (α) of 0.025. The Group 1 standard deviation is assumed to be 4 and the Group 2 standard deviation is assumed to be 4. To detect a mean difference (δ = μ2 - μ1) of 1.5, with current sample sizes of n1k = 30 and n2k = 30 out of target sample sizes of 60 and 60, respectively, and with a current t-test statistic of 2.12, the conditional power is 0.41449. The predictive power is 0.83393, and the futility index is 0.58551.

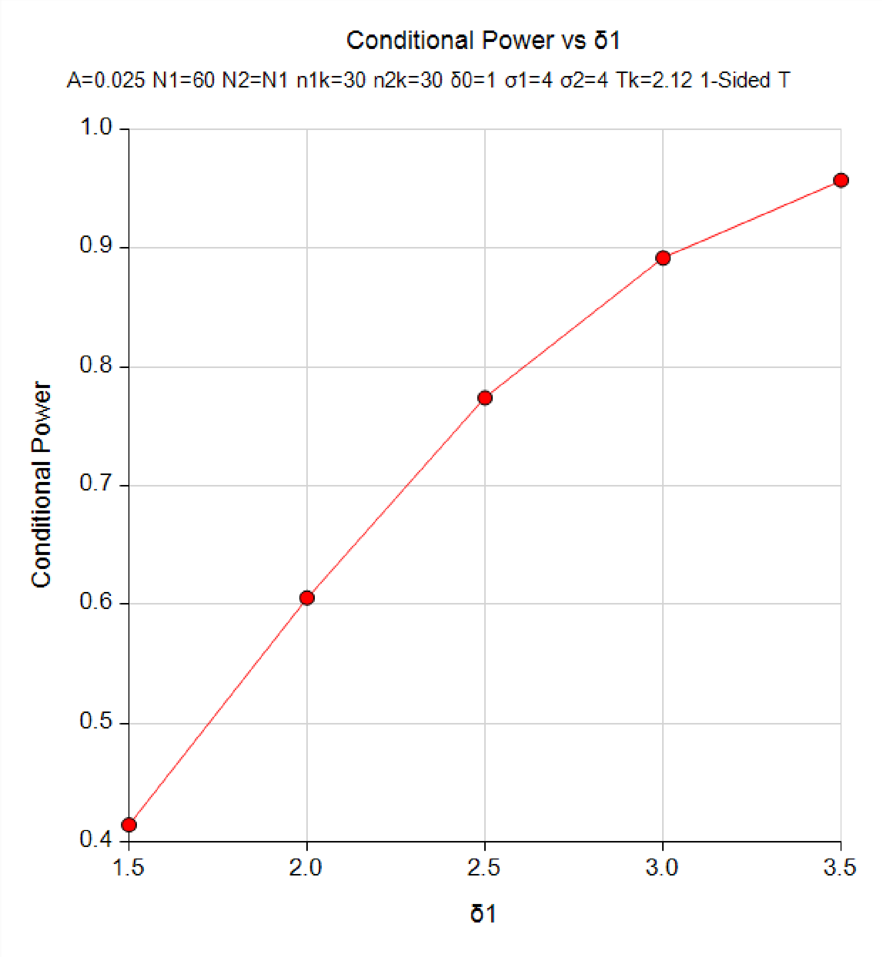
References

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.

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

Plots



This plot shows the relationship between conditional power and δ1.

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 (θ = δ1 – δ0 = 0.5) of Example 1 by hand, except that we use *Zk* instead of *Tk*.

In this case

$$I_k = \left(\frac{\sigma_1^2}{n_{1k}} + \frac{\sigma_2^2}{n_{2k}}\right)^{-1} \qquad I_K = \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)^{-1}$$

$$= \left(\frac{16}{30} + \frac{16}{30}\right)^{-1} \qquad = \left(\frac{16}{60} + \frac{16}{60}\right)^{-1}$$

$$= 0.9375 \qquad = 1.875$$

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

$$= \Phi\left(\frac{2.12\sqrt{0.9375} - 1.9599640\sqrt{1.875} + 0.5(1.875 - 0.9375)}{\sqrt{1.875 - 0.9375}}\right)$$

$$= \Phi(-0.1676848)$$

$$= 0.4334156$$

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For ...... **Conditional Power**
Higher Means Are ...... **Better (H1: δ > δ0)**
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**
δ0 (Superiority Difference)...... **1**
δ1 (Actual Difference to Detect)...... **1.5**
σ1 (Standard Deviation of Group 1)...... **4**
σ2 (Standard Deviation of Group 2)...... **σ1**
Test Statistic Input Type ...... **Zk**
Zk (Current Test Statistic) **2.12**

Output

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

Numeric Results

Solve For: Conditional Power
Test Type: Two-Sample T-Test
Groups: 1 = Control, 2 = Treatment
Hypotheses: H0: δ ≤ δ0 vs. H1: δ > δ0

Power		Sample Size				Mean Difference (μ2 - μ1)		Standard Deviation		Test Statistic		
		Target		Look k								
Conditional	Predictive	N1	N2	n1k	n2k	Superiority δ0	Actual δ1	σ1	σ2	Zk	Alpha	Futility
0.43342	0.8504	60	60	30	30	1	1.5	4	4	2.12	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 δ0 = 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

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 3** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For .... **Sample Size Reestimation**
Higher Means Are .... **Better (H1: δ > δ0)**
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**
δ0 (Superiority Difference) **1**
δ1 (Actual Difference to Detect)..................... **2**
σ1 (Standard Deviation of Group 1).............. **6.7**
σ2 (Standard Deviation of Group 2).............. **σ1**
Test Statistic Input Type .... **Zk**
Zk (Current Test Statistic) **2.12**

Output

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

Numeric Results

Solve For: Sample Size Reestimation
Test Type: Two-Sample T-Test
Groups: 1 = Control, 2 = Treatment
Hypotheses: H0: δ ≤ δ0 vs. H1: δ > δ0

Power		Sample Size				Mean Difference (μ2 - μ1)		Standard Deviation				
		Target		Look k						Test Statistic		
Conditional	Predictive	N1	N2	n1k	n2k	Superiority δ0	Actual δ1	σ1	σ2	Zk	Alpha	Futility
0.8	0.95534	520	520	30	30	1	2	6.7	6.7	2.12	0.025	0.2

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