Chapter 452

Conditional Power and Sample Size Reestimation of Superiority by a Margin Tests for Two Means in a 2x2 Cross-Over Design

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 predicted power for the case when a *t*-test, computed from data obtained from a 2x2 cross-over design, is used to test superiority by a margin for two population means. It also provides *sample size reestimation* to achieve a specified conditional power value.

Technical Details

The 2x2 cross-over design may be described as follows. Randomly assign the subjects to one of two sequence groups so that there are n_1 subjects in sequence one and n_2 subjects in sequence two. In order to achieve design balance, the sample sizes n_1 and n_2 are assumed to be equal so that $n_1 = n_2 = N/2$.

Sequence one is given treatment A followed by treatment B. Sequence two is given treatment B followed by treatment A.

The data may be analyzed by a *t*-test. See one of the cross-over chapters for further details.

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.

For a 2x2 cross-over design *t*-test of two means, with $\delta 0$ representing the hypothesized mean difference under the null hypothesis, and $\delta 1$ representing the actual mean difference under the alternative hypothesis, the parameter θ to test the superiority 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 71 are

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

$Z_k = \left(\frac{\overline{d}_{Y1k} + \overline{d}_{Y2k}}{2} - \delta 0\right)\sqrt{\hat{I}_k}$ (the *t*-statistic computed from the observed data)

$I_k = \frac{n_k}{\sigma_d^2}$ (the interim information level)

$I_K = \frac{N}{\sigma_d^2}$ (the final information level)

where

$\bar{d}_{Yik}$ the sample mean difference at stage k for those subjects receiving sequence i

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

n_k is the sample size at stage k

N is the total sample size

σ_d^2 is the variance of the subject differences

Computing conditional power requires you to set σ_d^2, $\delta 0$, and $\delta 1$. 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.

The Variance of the Pair Differences (σ_d^2)

If you have an estimate of σ_d^2 from the differences, you may enter the standard deviation of the pair differences (σ_d) directly. It is assumed that the variances of the two sequences are the same. A reasonable (but somewhat rough) estimate of σ_d^2 may be obtained using the range of pair differences as

$$\sigma_d^2 = \left(\frac{Range}{4}\right)^2$$

such that

$$\sigma_d = \frac{Range}{4}.$$

If you have estimates of the variances of the original pair variables (σ_1^2 and σ_2^2) and the Pearson correlation between the pair variables (ρ), the variance of the pair differences (σ_d^2) may be calculated using the equation

$$\sigma_d^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$$

such that

$$\sigma_d = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}.$$

If $\sigma_1^2 = \sigma_2^2 = \sigma_x^2$, then this formula reduces to

$$\sigma_d^2 = 2\sigma_x^2(1 - \rho)$$

such that

$$\sigma_d = \sqrt{2\sigma_x^2(1 - \rho)}.$$

If you have an estimate of the within-subject population variance (σ_w^2), then σ_d^2 may be calculated using the equation

$$\sigma_d^2 = 2\sigma_w^2$$

such that

$$\sigma_d = \sqrt{2\sigma_w^2}\,.$$

Often, cross-over designs are analyzed using repeated measures ANOVA. σ_w^2 is estimated by the within mean square error (WMSE) from the repeated measures ANOVA.

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 2x2 cross-over 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 total sample size is 52. The standard deviation of pair differences is expected to be about 1.8. 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 actual mean changes of 1.2, 1.4, 1.6, 1.8, and 2.

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**
N (Target Sample Size) **52**
nk (Sample Size at Look k)............................**26**
δ0 (Superiority Difference).............................**1**
δ1 (Actual Difference to Detect)....................**1.2 to 2 by 0.2**
Standard Deviation Input Type ....**Enter the SD of Pair Differences**
σd (SD of Pair Differences)............................**1.8**
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: 2x2 Cross-Over T-Test
Hypotheses: H0: δ ≤ δ0   vs.   H1: δ > δ0

| Power | | Sample Size | | Mean Difference | | | Standard Deviation | Test Statistic | | |
| | | Target | Look k | Superiority | Actual | | | | | |
Conditional	Predictive	N	nk	δ0	δ1	δ1 - δ0	σd	Tk	Alpha	Futility
0.42380	0.81244	52	26	1	1.2	0.2	1.8	2.12	0.025	0.57620
0.64594	0.81244	52	26	1	1.4	0.4	1.8	2.12	0.025	0.35406
0.82663	0.81244	52	26	1	1.6	0.6	1.8	2.12	0.025	0.17337
0.93416	0.81244	52	26	1	1.8	0.8	1.8	2.12	0.025	0.06584
0.98096	0.81244	52	26	1	2.0	1.0	1.8	2.12	0.025	0.01904

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.
N	The target sample size from both sequences combined.
nk	The total sample size obtained through stage k from both sequences combined.
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.
δ1 - δ0	The difference of the mean differences. This is the minimum value that is to be detected by the completed study.
σd	The standard deviation of the paired differences of subjects within the same sequence. It is assumed that the standard deviations of the two sequences are the same.
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 2×2 cross-over design will be used to test whether the Group 1 mean is superior to the Group 2 mean by a margin, with a superiority margin of $\delta0 = 1$ (H0: δ ≤ 1 versus H1: δ > 1). The comparison will be made using a one-sided, two-sample t-test, with a Type I error rate (α) of 0.025. The standard deviation of the paired differences of subjects within the same sequence is assumed to be 1.8 for both sequences. The desired difference to detect is $\delta1 = 1.2$ (or $\delta1 - \delta0 = 0.2$). With a current total sample size of nk = 26 out of a target sample size of 52 (assuming half in Sequence 1 and half in Sequence 2), and with a current t-test statistic of 2.12, the conditional power is 0.4238. The predictive power is 0.81244, and the futility index is 0.5762.

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
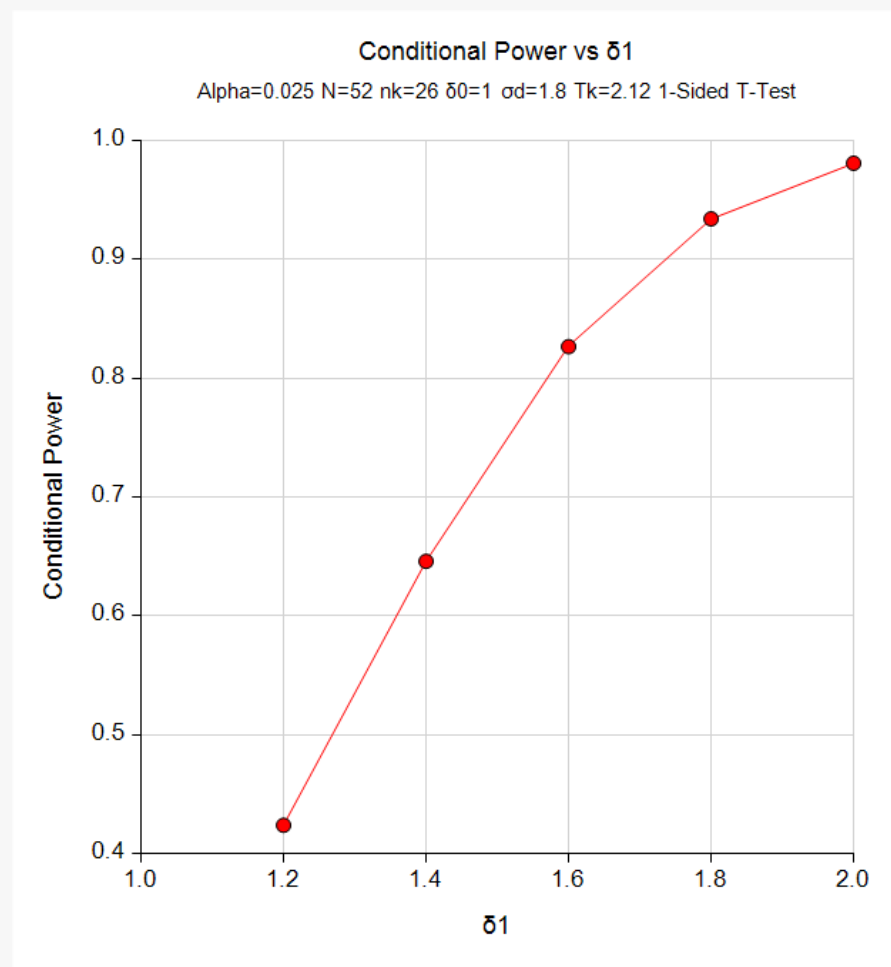
This plot shows the relationship between conditional power and δ1.

Example 2 – Validation

We could not find an example of a conditional power calculation for a 2x2 cross-over superiority by a margin *t*-test in the literature. Since the calculations are relatively simple, as a validation we will perform the calculation of the third scenario ($\delta 1 = 1.6$) of Example 1 by hand, except that we will use $Z_k = 2.12$ instead of $t_k = 2.12$.

In this case

$$I_k = \frac{n_k}{\sigma_d^2} \qquad\qquad I_K = \frac{N}{\sigma_d^2}$$

$$= \frac{26}{1.8^2} \qquad\qquad = \frac{52}{1.8^2}$$

$$= 8.0246914 \qquad\qquad = 16.0493827$$

$$P_{uk}(\theta) = \Phi\left(\frac{Z_k\sqrt{I_k} - Z_{1-\alpha}\sqrt{I_K} + (\delta_1 - \delta_0)(I_K - I_k)}{\sqrt{I_K - I_k}}\right)$$

$$= \Phi\left(\frac{2.12\sqrt{8.0246914} - 1.959964\sqrt{16.0493827} + (1.6 - 1)(16.0493827 - 8.0246914)}{\sqrt{16.0493827 - 8.0246914}}\right)$$

$$= \Phi\left(\frac{2.9682814}{2.8327886}\right)$$

$$= \Phi(1.0478655)$$

$$= 0.8526497$$

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**
N (Target Sample Size) **52**
nk (Sample Size at Look k)...........................**26**
δ0 (Superiority Difference)............................**1**
δ1 (Actual Difference to Detect)....................**1.6**
Standard Deviation Input Type ....**Enter the SD of Pair Differences**
σd (SD of Pair Differences)...........................**1.8**
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: 2x2 Cross-Over T-Test
Hypotheses: H0: δ ≤ δ0   vs.   H1: δ > δ0
_____

| Power | | Sample Size | | Mean Difference | | | Standard Deviation | Test Statistic | | |
Conditional	Predictive	Target N	Look k nk	Superiority δ0	Actual δ1	δ1 - δ0	σd	Zk	Alpha	Futility
0.85265	0.8504	52	26	1	1.6	0.6	1.8	2.12	0.025	0.14735

The conditional power of 0.85265 matches the value calculated by hand.

Example 3 – Sample Size Reestimation

Suppose a 2x2 cross-over study has started. It is being conducted to detect a mean change of 1.6 using a one-sided superiority *t*-test to compare the result against a lower difference bound of $\delta 0 = 1.0$ at an alpha of 0.025. The target sample size is 52. The standard deviation of pair differences is expected to be about 1.8.

An interim analysis was run after half the data were collected. This analysis yields a z-test value of 2.12. This test value was the result of transforming the p-value of the t-test with the inverse, standard normal distribution. The sample standard deviation of pair differences was 3.28.

The data monitoring board would like to have sample size reestimated to obtain a conditional power of 0.8.

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**
nk (Sample Size at Look k) **26**
δ0 (Superiority Difference)..............................**1**
δ1 (Actual Difference to Detect).....................**1.6**
Standard Deviation Input Type ....**Enter the SD of Pair Differences**
σd (SD of Pair Differences)............................**3.28**
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: 2x2 Cross-Over T-Test
Hypotheses: H0: δ ≤ δ0   vs.   H1: δ > δ0
_____

Power		Sample Size		Mean Difference			Standard Deviation	Test Statistic		
Conditional	Predictive	Target N	Look k nk	Superiority δ0	Actual δ1	δ1 - δ0	σd	Zk	Alpha	Futility
0.8013	0.92396	149	26	1	1.6	0.6	3.28	2.12	0.025	0.1987

Notice that the target sample size has increased from 52 to 149.