Chapter 204

Conditional Power and Sample Size Reestimation of Superiority by a Margin Tests for the Difference Between Two Proportions

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 *z*-test is used to test superiority by a margin for two proportions. It also provides *sample size reestimation* to achieve a specified conditional power value.

Technical Details

All details and assumptions usually made when using a two-sample *z*-test to test the difference between two proportions 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 P_1 and P_2 be the population proportions in groups 1 and 2, respectively. If we define $\delta = P_2 - P_1$, such that $\delta 0 = P_{2.0} - P_1$ is the superiority difference boundary, $\delta 1 = P_{2.1} - P_1$ is the true population difference under the alternative hypothesis, and $\hat{\delta}_k = p_{2k} - p_{1k}$ is the estimated proportion 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 proportions better) or $H_1{:}\,\delta < \delta 0$ (higher proportions worse) and other conditional power calculation components as outlined in Chang (2008) pages 70 and 71 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)

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

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

p_{jk} is the sample proportion for group *j*, estimating P_j at stage *k*

$\hat{I}_k$ 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*

$\sigma^2 = \bar{p}(1-\bar{p})$ with $\bar{p} = \frac{P_1+P_{2.1}}{2}$

Computing conditional power requires you to set P_1, $P_{2.0}$ or $\delta 0$, and $P_{2.1}$ or $\delta 1$. 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.

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 *z*-test against a lower difference bound of δ0 = 0.05 at an alpha of 0.025. The reference group proportion (P1) is 0.6, so the difference bound of δ0 = 0.05 corresponds to a superiority proportion of P2.0 = P1 + δ0 = 0.6 + 0.05 = 0.65. The target sample size is 60 per group.

An interim analysis is planned after half the data have been collected. The data monitoring board would like to have the conditional power calculated for an actual difference of δ1 = 0.1 (P2.1 = P1 + δ1 = 0.6 + 0.1 = 0.7) and *z* values of 1, 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 Proportions 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**
Input Type	**Proportions**
P1 (Group 1 Proportion)	**0.6**
P2.0 (Superiority Proportion)	**0.65**
P2.1 (Actual Proportion)	**0.7**
Zk (Current Test Statistic)	**1 1.5 2 2.5 3 3.5**

Output

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

Numeric Reports

Numeric Results

Solve For: Conditional Power
Groups: 1 = Reference, 2 = Treatment
Test Type: Two-Sample Z-Test
Higher Proportions Are: Better
Hypotheses: H0: P2 - P1 ≤ δ0 vs. H1: P2 - P1 > δ0

Power		Sample Size						Proportion			Difference		Test	
		Target				Look k								
Conditional	Predictive	N1	N2	N	R	n1k	n2k	Ref. P1	Superiority P2.0	Actual P2.1	Superiority δ0	Actual δ1	Statistic Zk	Alpha
0.08600	0.29262	60	60	120	1	30	30	0.6	0.65	0.7	0.05	0.1	1.0	0.025
0.19330	0.56409	60	60	120	1	30	30	0.6	0.65	0.7	0.05	0.1	1.5	0.025
0.35725	0.80743	60	60	120	1	30	30	0.6	0.65	0.7	0.05	0.1	2.0	0.025
0.55337	0.94244	60	60	120	1	30	30	0.6	0.65	0.7	0.05	0.1	2.5	0.025
0.73702	0.98878	60	60	120	1	30	30	0.6	0.65	0.7	0.05	0.1	3.0	0.025
0.87164	0.99860	60	60	120	1	30	30	0.6	0.65	0.7	0.05	0.1	3.5	0.025

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, N2, and N	The target sample sizes at the end of the study of groups 1, 2, and both, respectively.
R	The sample allocation ratio which is used to calculate N2. R = N2 / N1.
n1k and n2k	The sample sizes of groups 1 and 2 through stage k, respectively.
P1	The response proportion for group 1.
P2.0	The superiority proportion for group 2 used to compute δ0.
P2.1	The actual proportion for group 2 to detect under the alternative hypothesis used to compute δ1.
δ	The difference in proportions. δ = P2 - P1.
δ0	The superiority difference used to construct the hypotheses. δ0 = P2.0 - P1.
δ1	The actual difference to detect under the alternative hypothesis at which conditional power is calculated. δ1 = P2.1 - P1.
Zk	The value of the test statistic from the observed data at stage k.
Alpha	The probability of rejecting a true null hypothesis.

Summary Statements

A parallel two-group design is used to test whether the Group 2 (treatment) proportion (P2) is superior to the Group 1 (reference) proportion (P1) by a margin, with a superiority margin of δ0 = P2.0 - P1 = 0.65 - 0.6 = 0.05 (H0: P2 - P1 ≤ 0.05 versus H1: P2 - P1 > 0.05). The comparison is made using a one-sided, two-sample Z-test, with a Type I error rate (α) of 0.025. The desired difference to detect is δ1 = P2.1 - P1 = 0.7 - 0.6 = 0.1. With current sample sizes of n1k = 30 and n2k = 30 out of target sample sizes of 60 and 60, respectively, and with a current z-value of 1, the conditional power is 0.086. The predictive power is 0.29262, and the futility index is 0.914.

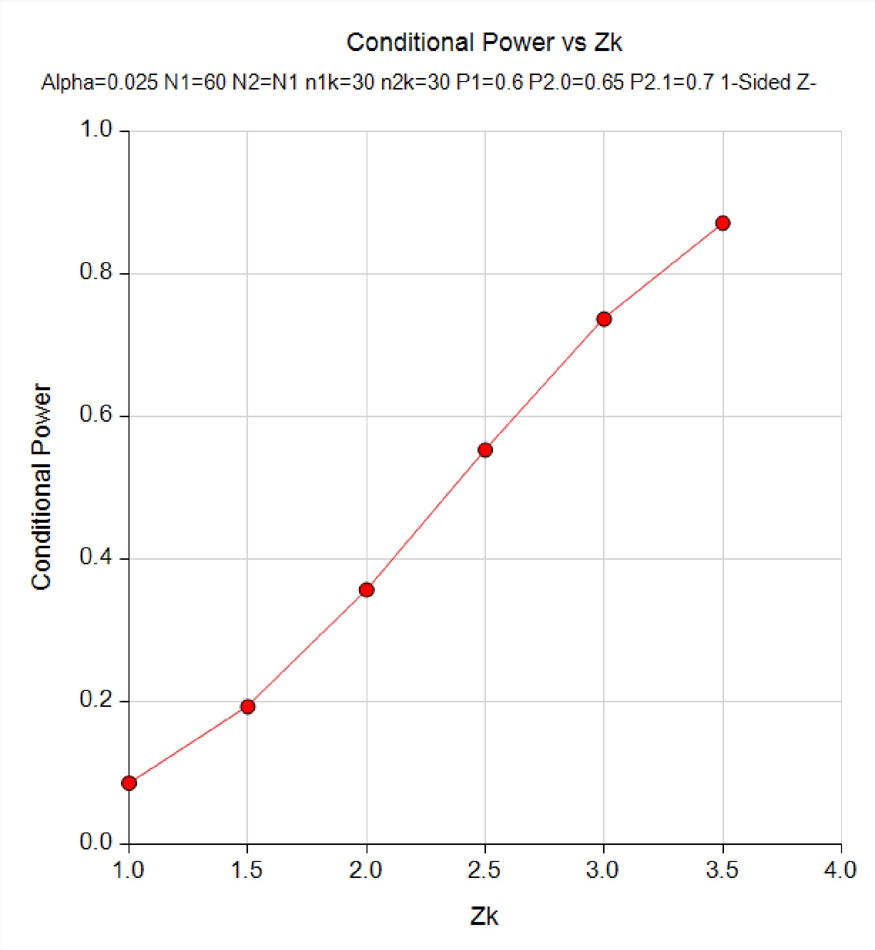
References

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 Z_k.

Example 2 – Validation

We could not find an example of a conditional power calculation for a one-sided superiority by a margin proportions test in the literature. Since the calculations are relatively simple, we will validate the calculation of the third scenario (Z_k = 2) of Example 1 by hand.

In this case

$$\bar{p} = \frac{P_1 + P_{2.1}}{2} = \frac{0.6 + 0.7}{2} = 0.65 \qquad \sigma^2 = \bar{p}(1-\bar{p}) = 0.65(0.35) = 0.2275$$

$$I_k = \frac{1}{\sigma^2}\left(\frac{1}{n_{1k}} + \frac{1}{n_{2k}}\right)^{-1} = \frac{1}{0.2275}\left(\frac{1}{30} + \frac{1}{30}\right)^{-1} = 65.9341 \qquad I_K = \frac{1}{\sigma^2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1} = \frac{1}{0.2275}\left(\frac{1}{60} + \frac{1}{60}\right)^{-1} = 131.8681$$

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

$$= \Phi\left(\frac{2\sqrt{65.9341} - 1.9599640\sqrt{131.8681} + 0.05(131.8681 - 65.9341)}{\sqrt{131.8681 - 65.9341}}\right)$$

$$= \Phi(-0.36581)$$

$$= 0.35725$$

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 Proportions 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**
Input Type.. **Proportions**
P1 (Group 1 Proportion)................................. **0.6**
P2.0 (Superiority Proportion) **0.65**
P2.1 (Actual Proportion)................................. **0.7**
Zk (Current Test Statistic).............................. **2**

Output

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

Numeric Results

Solve For:	Conditional Power
Groups:	1 = Reference, 2 = Treatment
Test Type:	Two-Sample Z-Test
Higher Proportions Are:	Better
Hypotheses:	H0: P2 - P1 ≤ δ0 vs. H1: P2 - P1 > δ0

Power		Sample Size						Proportion			Difference		Test	
		Target				Look k								
Conditional	Predictive	N1	N2	N	R	n1k	n2k	Ref. P1	Superiority P2.0	Actual P2.1	Superiority δ0	Actual δ1	Statistic Zk	Alpha
0.35725	0.80743	60	60	120	1	30	30	0.6	0.65	0.7	0.05	0.1	2	0.025

The conditional power of 0.35725 matches the value calculated by hand.

Example 3 – Sample Size Reestimation

Suppose a study has started and is to be analyzed using a one-sided superiority *z*-test against a lower difference bound of δ0 = 0.05 at an alpha of 0.025. The reference group proportion (P1) was set at 0.6 and the difference bound (δ0) is set at 0.05. The target sample size was 60 per group.

An interim analysis was run after half the data had been collected. This analysis yielded a z-test value of 2.12. The value of P1 (the proportion for the reference group) was found to be 0.643. The data monitoring board would like to have the sample size recalculated for an actual difference of δ1 = 0.1 and 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 Proportions 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**
Input Type	**Difference**
P1 (Group 1 Proportion)	**0.643**
δ0 (Superiority Difference)	**0.05**
δ1 (Actual Difference to Detect)	**0.1**
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
Groups: 1 = Reference, 2 = Treatment
Test Type: Two-Sample Z-Test
Higher Proportions Are: Better
Hypotheses: H0: δ ≤ δ0 vs. H1: δ > δ0

Power		Sample Size						Proportion			Difference		Test	
		Target				Look k								
Conditional	Predictive	N1	N2	N	R	n1k	n2k	Ref. P1	Superiority P2.0	Actual P2.1	Superiority δ0	Actual δ1	Statistic Zk	Alpha
0.80007	0.96541	1068	1068	2136	1	30	30	0.64	0.69	0.74	0.05	0.1	2.12	0.025

Notice that the target sample size has increased from 60 per group (N = 120), to 1068 per group (N = 2136).