Chapter 718

Conditional Power and Sample Size Reestimation of Superiority by a Margin Logrank Tests

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 superiority by a margin logrank test is used to compare the hazard rate of a treatment group to that of a control group. It provides *sample size reestimation* to achieve a specified conditional power value.

Technical Details

All details and assumptions usually made when using a logrank test are used 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 λ_1 and λ_2 be the population hazard rates in groups 1 and 2, respectively. If we define $HR = \lambda_2/\lambda_1$, such that $HR0$ is the superiority hazard ratio boundary and $HR1$ is the true population difference under the alternative hypothesis, then the parameter θ to test the one-sided superiority by a margin alternative hypotheses of $H_1: HR < HR0$ (higher hazards worse) or $H_1: HR > HR0$ (higher hazards better) and other conditional power calculation components computed in Chang (2008) page 71 are

$\theta = \log(HR1) - \log(HR0)$ (the expected log(hazard ratio) difference under the alternative hypothesis)

$Z_k = S_k/\sqrt{\hat{I}_k}$ (the superiority by a margin logrank test statistic computed from the observed data, see an alternate representation for Z_k in Jennison & Turnbull (2000) on page 216)

$I_k = E_k P_1(1 - P_1)$ (the interim information level)

$I_K = E P_1(1 - P_1)$ (the final information level) where

S_k is a logrank score statistic

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

E_k is the number of events at stage *k*

E is the total number of events

P_1 is the proportion of the subjects assigned to group 1, the control group

Computing conditional power requires you to set *HR0* and *HR1*. These 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 to detect a hazard ratio of 0.75 in a superiority by a margin logrank test against an upper hazard ratio bound of *HR0* = 0.9 at an alpha of 0.025. The total number of events needed is 200. The design is to have an equal number of subjects in both groups. An interim analysis will be run after half the data have been collected. The data monitoring board would like to have the conditional power calculated for z-values of -1.0, -1.5, -2.0, -2.5, and -3.0.

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 Hazards Are .... **Worse (H1: HR < HR0)**
Alpha... **0.025**
E (Target Number of Events)......................... **200**
Ek (Events through Look k)........................... **100**
P1 (Proportion of Subjects in Group 1) **0.5**
HR0 (Superiority Hazard Ratio) **0.9**
HR1 (Actual Hazard Ratio to Detect) **0.75**
Zk (Current Test Statistic) .... **-1 -1.5 -2 -2.5 -3**

Output

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

Numeric Reports

Numeric Results

Solve For: Conditional Power
Groups: 1 = Control, 2 = Treatment
Test Type: Logrank
Higher Hazards Are: Worse
Hypotheses: H0: HR ≥ HR0 vs. H1: HR < HR0

Power		Number of Events		Proportion of Subjects	Hazard Ratio		Test		
Conditional	Predictive	Target E	Look k Ek	in Group 1 P1	Superiority HR0	Actual HR1	Statistic Zk	Alpha	Futility
0.87282	0.98878	200	100	0.5	0.9	0.75	-3.0	0.025	0.12718
0.73885	0.94244	200	100	0.5	0.9	0.75	-2.5	0.025	0.26115
0.55559	0.80743	200	100	0.5	0.9	0.75	-2.0	0.025	0.44441
0.35935	0.56409	200	100	0.5	0.9	0.75	-1.5	0.025	0.64065
0.19484	0.29262	200	100	0.5	0.9	0.75	-1.0	0.025	0.80516

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.
E	The target number of events that are needed.
Ek	The number of events obtained through interim stage k.
P1	The proportion of subjects that were assigned to group 1 (control group).
HR	The hazard ratio. HR = λ2/λ1.
HR0	The superiority hazard ratio used to construct the hypotheses.
HR1	The actual hazard ratio to detect under the alternative hypothesis at which conditional power is calculated.
Zk	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 is used to test whether the Group 2 (treatment) hazard rate (λ2) is superior to the Group 1 (control) hazard rate (λ1) by a margin, with a superiority hazard ratio of 0.9 (H0: HR ≥ 0.9 versus H1: HR < 0.9, HR = λ2 / λ1). The comparison is made using a one-sided, two-sample logrank test, with a Type I error rate (α) of 0.025. The anticipated percentage of subjects in Group 1 (the control group) is 50%. To detect a hazard ratio of 0.75, with 100 events realized at the current look out of a target of 200 total events (in both groups combined), and with a current z-value of -3, the conditional power is 0.87282. The predictive power is 0.98878, and the futility index is 0.12718.

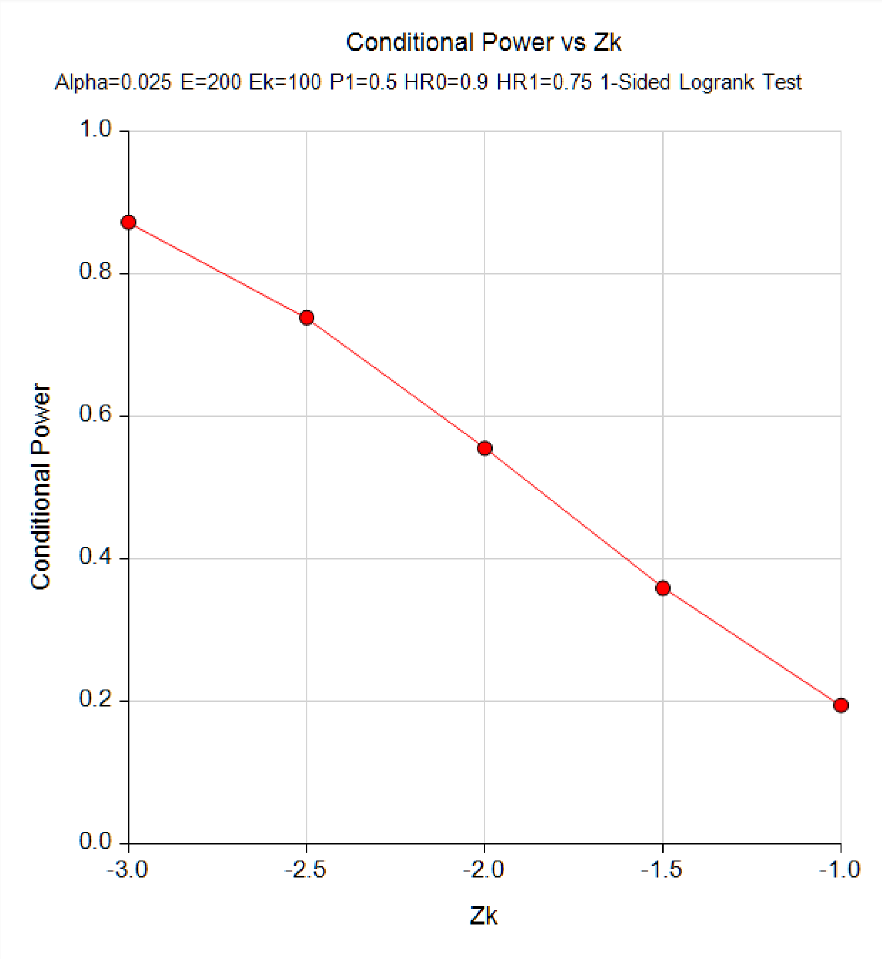
References

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 superiority by a margin logrank 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

$$I_k = E_k P_1(1-P_1) = 100 \times 0.5(1-0.5) = 25 \qquad I_K = E P_1(1-P_1) = 200 \times 0.5(1-0.5) = 50$$

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

$$= \Phi\left(\frac{-(-2)\sqrt{25} - 1.959964\sqrt{50} - \big(\log(0.75) - \log(0.9)\big)(50-25)}{\sqrt{50-25}}\right)$$

$$= \Phi\left(\frac{0.699000}{5}\right)$$

$$= \Phi(0.139800)$$

$$= 0.555591$$

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 Hazards Are .... **Worse (H1: HR < HR0)**
Alpha.. **0.025**
E (Target Number of Events)......................... **200**
Ek (Events through Look k)........................... **100**
P1 (Proportion of Subjects in Group 1) **0.5**
HR0 (Superiority Hazard Ratio) **0.9**
HR1 (Actual Hazard Ratio to Detect) **0.75**
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 = Control, 2 = Treatment
Test Type: Logrank
Higher Hazards Are: Worse
Hypotheses: H0: HR ≥ HR0 vs. H1: HR < HR0

Power		Number of Events		Proportion	Hazard Ratio		Test		
		Target	Look k	of Subjects in Group 1	Superiority	Actual	Statistic		
Conditional	Predictive	E	Ek	P1	HR0	HR1	Zk	Alpha	Futility
0.55559	0.80743	200	100	0.5	0.9	0.75	-2	0.025	0.44441

The conditional power of 0.55559 matches the value calculated by hand.

Example 3 – Sample Size Reestimation

Suppose a study has been planned to detect a hazard ratio of 0.75 in a superiority by a margin logrank test against an upper hazard ratio bound of *HR0* = 0.9 at an alpha of 0.025. At the beginning of the study, the target number of events is 200. The design is to have an equal number of subjects in both groups. An interim analysis is run after half the data have been collected. This analysis yields a z-test value of -2.12.

The data monitoring board would like to recalculate the sample size for 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 ....**E (Target Number of Events)**
Higher Hazards Are ....**Worse (H1: HR < HR0)**
Conditional Power..**0.8**
Alpha..**0.025**
Ek (Events through Look k)...........................**100**
P1 (Proportion of Subjects in Group 1) **0.5**
HR0 (Superiority Hazard Ratio) **0.9**
HR1 (Actual Hazard Ratio to Detect) **0.75**
Zk (Current Test Statistic) ....**-2.12**

Output

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

Numeric Results

Solve For:	E (Target Number of Events)
Groups:	1 = Control, 2 = Treatment
Test Type:	Logrank
Higher Hazards Are:	Worse
Hypotheses:	H0: HR ≥ HR0 vs. H1: HR < HR0

Power		Number of Events		Proportion of Subjects	Hazard Ratio		Test Statistic		
Conditional	Predictive	Target E	Look k Ek	in Group 1 P1	Superiority HR0	Actual HR1	Zk	Alpha	Futility
0.80021	0.92602	602	100	0.5	0.9	0.75	-2.12	0.025	0.19979

The needed number of events increased from 200 to 602.