Chapter 717

Conditional Power and Sample Size Reestimation of Non-Inferiority 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 non-inferiority 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 non-inferiority hazard ratio boundary and $HR1$ is the true population difference under the alternative hypothesis, then the parameter θ to test the one-sided non-inferiority 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 non-inferiority 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)

$I_K = EP_1(1 - P_1)$ (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*.

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 Hazards Are

Use this option to specify the direction of the non-inferiority test.

If Higher Hazards are "Worse", the alternative hypothesis is H1: HR < HR0, with HR0 > 1.

If Higher Hazards are "Better", the alternative hypothesis is H1: HR > HR0, with HR0 < 1.

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*.

Events

E (Target Number of Events)

Enter *E*, the target number of events (e.g. deaths) needed in the study. This is the sum of the number of events in both the treatment and control groups.

Ek (Events through Look k)

Enter the number of events (e.g. deaths) obtained through look *k* in both groups.

P1 (Proportion of Subjects in Group 1)

The proportion of all subjects (failed plus not failed) that are in the first (control) group. The rest are in the second (treatment) group.

Effect Size – Hazard Ratios

HR0 (Non-Inferiority Hazard Ratio)

This is the hazard ratio bound of the non-inferiority test hypothesis.

When Higher Hazards are "Worse", the alternative hypothesis is H1: HR < HR0, and you can enter values in the range HR0 > 1 and HR0 > HR1.

When Higher Hazards are "Better", the alternative hypothesis is H1: HR > HR0, and you can enter values in the range 0 < HR0 < 1 and HR0 < HR1.

HR1 (Actual Hazard Ratio to Detect)

This is the actual hazard ratio ($\lambda2/\lambda1$) at which conditional power is calculated.

For non-inferiority tests, this value is often set to one. When this value is not set to one, care should be taken that this value is consistent with the alternative hypothesis.

When Higher Hazards are "Worse", 0 < HR1 < HR0. When Higher Hazards are "Better", HR1 > HR0 > 0.

Effect Size – Current Test Statistic

Zk (Current Test Statistic)

Enter the value of the test statistic calculated from the data obtained so far. This *z*-value is equal to the non-inferiority logrank score statistic divided by the square root of the information level, *Ik*.

This value may be positive or negative. Usually, the *Zk* ranges between -5 and 5.

Example 1 – Computing Conditional Power

Suppose a study has been planned to detect a hazard ratio of 1.0 in a non-inferiority logrank test against an upper hazard ratio bound of $HR0 = 1.25$ 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

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 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 (Non-Inferiority Hazard Ratio) **1.25**

HR1 (Actual Hazard Ratio to Detect)...... **1**

Zk (Current Test Statistic)....................... **-1 -1.5 -2 -2.5 -3**

Annotated Output

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

Numeric Results

Numeric Results for Logrank Test ───
Hypotheses: H0: HR ≥ HR0   vs.   H1: HR < HR0

Cond. Power	Pred. Power	Target Events E	Interim Look k Events Ek	Prop. Group 1 P1	Non-Inf Hazard Ratio HR0	Actual Hazard Ratio HR1	Test Statistic Zk	Alpha	Futility
0.91051	0.98878	200	100	0.50	1.25	1.00	-3.000	0.025	0.08949
0.80064	0.94244	200	100	0.50	1.25	1.00	-2.500	0.025	0.19936
0.63454	0.80743	200	100	0.50	1.25	1.00	-2.000	0.025	0.36546
0.43798	0.56409	200	100	0.50	1.25	1.00	-1.500	0.025	0.56202
0.25588	0.29262	200	100	0.50	1.25	1.00	-1.000	0.025	0.74412

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.

E is the target number of events that are needed.

E_k is the number of events obtained through interim stage k.

P1 is the proportion of subjects that were assigned to group 1 (control group).

HR = λ_2/λ_1 is the hazard ratio.

HR0 is the non-inferiority hazard ratio used to construct the hypotheses.

HR1 is the actual hazard ratio to detect under the alternative hypothesis at which conditional power is calculated.

Z_k 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 100 of 200 total events in both groups achieves 91.051% conditional power to detect an actual hazard ratio of 1.00 using a logrank test for non-inferiority with a hazard ratio bound of 1.25 and a significance level of 0.025. The study was designed to have 50% of the subjects in group 1 (the control group). The z-value of the data that have emerged so far is -3.000. The futility index is 0.08949.

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
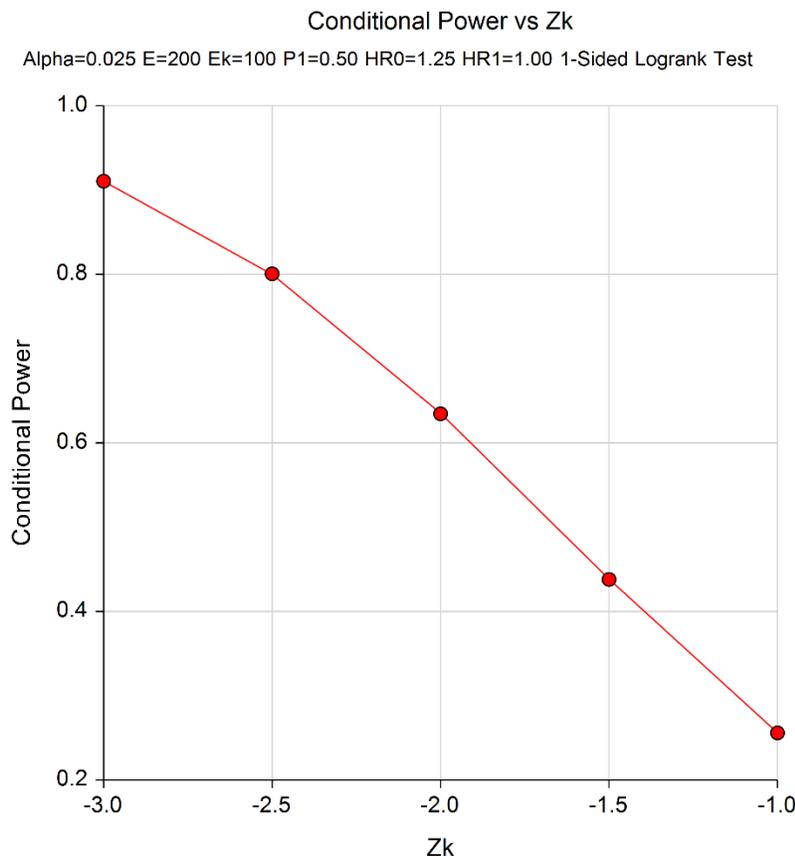
This plot shows the relationship between conditional power and Z_k.

Example 2 – Validation

We could not find an example of a conditional power calculation for a non-inferiority 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

$$
\begin{aligned}
I_k &= E_k P_1 (1 - P_1) & I_K &= E P_1 (1 - P_1) \\
&= 100 \times 0.5(1 - 0.5) & &= 200 \times 0.5(1 - 0.5) \\
&= 25 & &= 50
\end{aligned}
$$

$$
\begin{aligned}
P_{lk}(\theta) &= \Phi\left(\frac{-Z_k\sqrt{I_k} - Z_{1-\alpha}\sqrt{I_K} - \theta(I_K - I_k)}{\sqrt{I_K - I_k}} \right) \\
&= \Phi\left(\frac{-(-2)\sqrt{25} - 1.959964\sqrt{50} - (\log(1) - \log(1.25))(50 - 25)}{\sqrt{50 - 25}} \right) \\
&= \Phi\left(\frac{1.719551}{5} \right) \\
&= \Phi(0.343910) \\
&= 0.634543
\end{aligned}
$$

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**.

Option	**Value**
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 (Non-Inferiority Hazard Ratio)	**1.25**
HR1 (Actual Hazard Ratio to Detect)......	**1**
Zk (Current Test Statistic).......................	**-2**

Output

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

Numeric Results

Numeric Results for Logrank Test ——
Hypotheses: H0: HR ≥ HR0   vs.   H1: HR < HR0

Cond. Power	Pred. Power	Target Events E	Interim Look k Events Ek	Prop. Group 1 P1	Non-Inf Hazard Ratio HR0	Actual Hazard Ratio HR1	Test Statistic Zk	Alpha	Futility
0.63454	0.80743	200	100	0.50	1.25	1.00	-2.000	0.025	0.36546

The conditional power of 0.63454 matches the value calculated by hand.

Example 3 – Sample Size Reestimation

Suppose a study has been planned to detect a hazard ratio of 1.0 in a non-inferiority logrank test against an upper hazard ratio bound of *HR0* = 1.25 at an alpha of 0.025. The target number of events needed 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

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**.

Option	**Value**
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 (Non-Inferiority Hazard Ratio)	**1.25**
HR1 (Actual Hazard Ratio to Detect)......	**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 Logrank Test —————————————————————————————————
Hypotheses: H0: HR ≥ HR0   vs.   H1: HR < HR0

Cond. Power	Pred. Power	Target Events E	Interim Look k Events Ek	Prop. Group 1 P1	Non-Inf Hazard Ratio HR0	Actual Hazard Ratio HR1	Test Statistic Zk	Alpha	Futility
0.80011	0.90228	376	100	0.50	1.25	1.00	-2.120	0.025	0.19989

The needed number of events has increased from 200 to 376.