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 $\theta$ at the end of the study, given the observed test statistic, $Z_k$, is computed as

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

and the general lower one-sided conditional power at stage $k$ is computed as

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

where

- $\theta = \text{the parameter being tested by the hypothesis}$
- $k = \text{an interim stage at which the conditional power is computed (} k = 1, ..., K - 1\text{)}$
- $K = \text{the stage at which the study is terminated, and the final test computed}$
- $Z_k = \text{the test statistic calculated from the observed data that has been collected up to stage } k$
- $I_k = \text{the information level at stage } k$
- $I_K = \text{the information level at the end of the study}$
- $z_{1-a} = \text{the standard normal value for the test with a type I error rate of } a.$

Let $\lambda_1$ and $\lambda_2$ be the population hazard rates in groups 1 and 2, respectively. If we define $HR = \lambda_2 / \lambda_1$, such that $HR_0$ is the superiority hazard ratio boundary and $HR_1$ is the true population difference under the alternative hypothesis, then the parameter $\theta$ to test the one-sided superiority by a margin alternative hypotheses of $H_1: HR < HR_0$ (higher hazards worse) or $H_1: HR > HR_0$ (higher hazards better) and other conditional power calculation components computed in Chang (2008) page 71 are

$$\theta = \log(HR_1) - \log(HR_0) \quad \text{(the expected log(hazard ratio) difference under the alternative hypothesis)}$$

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

$$I_k = E_k p_1 (1 - p_1) \quad \text{(the interim information level)}$$

$$I_K = E p_1 (1 - p_1) \quad \text{(the final information level)}$$

where

- $S_k$ is a logrank score statistic
- $I_k$ 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 $HR_0$ and $HR_1$. 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

$$P_{uk} = \Phi\left(\frac{Z_k \sqrt{I_k} - z_{1-\alpha} \sqrt{I_k}}{\sqrt{I_k - \hat{I}_k}}\right),$$

and the general lower one-sided predictive power at stage $k$ is given by

$$P_{lk} = \Phi\left(\frac{-Z_k \sqrt{I_k} - z_{1-\alpha} \sqrt{I_k}}{\sqrt{I_k - \hat{I}_k}}\right),$$

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 $HR_0 = 0.9$ 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.

Output

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

Numeric Reports

<table>
<thead>
<tr>
<th>Cond. Power</th>
<th>Pred. Power</th>
<th>Target Events E</th>
<th>Interim Prop Group 1 P1</th>
<th>Interim Prop Hazard Ratio HR0</th>
<th>Actual Hazard Ratio HR1</th>
<th>Test Statistic Zk</th>
<th>Alpha</th>
<th>Futility</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.87282</td>
<td>0.98878</td>
<td>200</td>
<td>100</td>
<td>0.5</td>
<td>0.75</td>
<td>-3.0</td>
<td>0.025</td>
<td>0.12718</td>
</tr>
<tr>
<td>0.73885</td>
<td>0.94244</td>
<td>200</td>
<td>100</td>
<td>0.5</td>
<td>0.75</td>
<td>-2.5</td>
<td>0.025</td>
<td>0.26115</td>
</tr>
<tr>
<td>0.55559</td>
<td>0.80743</td>
<td>200</td>
<td>100</td>
<td>0.5</td>
<td>0.75</td>
<td>-2.0</td>
<td>0.025</td>
<td>0.44441</td>
</tr>
<tr>
<td>0.35935</td>
<td>0.56409</td>
<td>200</td>
<td>100</td>
<td>0.5</td>
<td>0.75</td>
<td>-1.5</td>
<td>0.025</td>
<td>0.64065</td>
</tr>
<tr>
<td>0.19484</td>
<td>0.29262</td>
<td>200</td>
<td>100</td>
<td>0.5</td>
<td>0.75</td>
<td>-1.0</td>
<td>0.025</td>
<td>0.80516</td>
</tr>
</tbody>
</table>

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.
### Summary Statements

The first 100 of 200 total events in both groups achieves 87.282% conditional power to detect an actual hazard ratio of 0.75 using a logrank test for superiority by a margin with a hazard ratio bound of 0.9 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. The futility index is 0.12718.

### References


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.
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

\[ I_k = E_k P_1 (1 - P_1) \quad I_K = E_P (1 - P_1) \]
\[ = 100 \times 0.5(1 - 0.5) \quad = 200 \times 0.5(1 - 0.5) \]
\[ = 25 \quad = 50 \]

\[
P_{lk}(\theta) = \Phi \left( \frac{-Z_k \sqrt{I_k} - Z_{1-\alpha} \sqrt{I_K} - \theta (I_K - I_k)}{\sqrt{I_K - I_k}} \right)
\]
\[
= \Phi \left( \frac{-(-2)\sqrt{25} - 1.959964\sqrt{50} - (\log(0.75) - \log(0.9))(50 - 25)}{\sqrt{50 - 25}} \right)
\]
\[
= \Phi \left( \frac{0.699000}{5} \right)
\]
\[ = \Phi(0.139800) \]
\[ = 0.555591 \]

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.

<table>
<thead>
<tr>
<th>Design Tab</th>
<th>Conditional Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solve For</td>
<td><img src="https://via.placeholder.com/150" alt="Image" /></td>
</tr>
<tr>
<td>Higher Hazards Are</td>
<td><img src="https://via.placeholder.com/150" alt="Image" /></td>
</tr>
<tr>
<td>Alpha</td>
<td>0.025</td>
</tr>
<tr>
<td>E (Target Number of Events)</td>
<td>200</td>
</tr>
<tr>
<td>Ek (Events through Look k)</td>
<td>100</td>
</tr>
<tr>
<td>P1 (Proportion of Subjects in Group 1)</td>
<td>0.5</td>
</tr>
<tr>
<td>HR0 (Superiority Hazard Ratio)</td>
<td>0.9</td>
</tr>
<tr>
<td>HR1 (Actual Hazard Ratio to Detect)</td>
<td>0.75</td>
</tr>
<tr>
<td>Zk (Current Test Statistic)</td>
<td>-2</td>
</tr>
</tbody>
</table>
Output

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

<table>
<thead>
<tr>
<th>Cond. Power</th>
<th>Pred. Power</th>
<th>Target Events E</th>
<th>Interim Look k Events Ek</th>
<th>Prop. Group 1 P1</th>
<th>Super. Hazard Ratio HR0</th>
<th>Actual Hazard Ratio HR1</th>
<th>Test Statistic Zk</th>
<th>Alpha</th>
<th>Futility</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.55559</td>
<td>0.80743</td>
<td>200</td>
<td>100</td>
<td>0.5</td>
<td>0.9</td>
<td>0.75</td>
<td>-2</td>
<td>0.025</td>
<td>0.44441</td>
</tr>
</tbody>
</table>

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 $HR_0 = 0.9$ 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 ($H_1: HR < HR_0$)
Conditional Power: 0.8
Alpha: 0.025
$E_k$ (Events through Look k): 100
$P_1$ (Proportion of Subjects in Group 1): 0.5
$HR_0$ (Superiority Hazard Ratio): 0.9
$HR_1$ (Actual Hazard Ratio to Detect): 0.75
$Z_k$ (Current Test Statistic): -2.12

Output

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

<table>
<thead>
<tr>
<th>Cond. Power</th>
<th>Pred. Power</th>
<th>Target Events E</th>
<th>Interim Look k Events</th>
<th>Prop. Group 1 P1</th>
<th>Super. Hazard Ratio $HR_0$</th>
<th>Actual Hazard Ratio $HR_1$</th>
<th>Test Statistic $Z_k$</th>
<th>Alpha</th>
<th>Futility</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.80021</td>
<td>0.92602</td>
<td>602</td>
<td>100</td>
<td>0.5</td>
<td>0.9</td>
<td>0.75</td>
<td>-2.12</td>
<td>0.025</td>
<td>0.19979</td>
</tr>
</tbody>
</table>

The needed number of events has increased from 200 to 602.