Chapter 346

Multi-Arm Equivalence Tests for Treatment and Control Survival Curves using Cox's Proportional Hazards Model

Introduction

This module computes power and sample size for multiple equivalence comparisons of treatment survival curves versus a control survival curve based on the results in Machin, Campbell, Tan, and Tan (2018). In this design, there are k treatment groups and one control group. A survival curve is measured in each group. A total of k equivalence tests are anticipated, each comparing a treatment group with the common control group using standard TOST z-tests based on a Cox proportional hazards regression coefficient.

The formulation for testing the significance of a Cox regression coefficient is identical to the standard logrank test. Thus, the power and sample size formulas for one analysis also work for the other. The Cox Regression model has the added benefit over the exponential model that it does not assume that the hazard rates are constant, but only that they are proportional. That is, that the hazard ratio remains constant throughout the experiment, even if the hazard rates vary.

A Bonferroni adjustment of the type I error rate may be optionally made because several comparisons are being tested using the same data. Making a multiplicity adjustment is usually recommended, but not always. In fact, Saville (1990) advocates not applying it and Machin, Campbell, Tan, and Tan (2018) include omitting it as a possibility.

Background

Whether you want to test several doses of a single treatment or several types of treatments, good research practice requires that each treatment be compared with a control. For example, a popular three-arm design consists of three groups: control, treatment A, and treatment B. Two tests are run: treatment A versus control and treatment B versus the same control. This avoids having to obtain a second control group for treatment B. Besides the obvious efficiency in subjects, it may be easier to recruit subjects if their chances of receiving the new treatment are better than 50-50.

Technical Details

Cox's Proportional Hazards Regression

Cox's proportional hazards regression is widely used for survival data. The regression model is

$$h(t|z) = h(t|0) \exp(bz)$$

where

b is the regression coefficient which is equal to $\log[h(t|1)/h(t|0)] = \log(HR)$

z is a binary indicator variable of treatment group

t is elapsed time

$h(t|z)$ is the hazard rate at time t, given covariate z

HR is the hazard ratio, $h(t|1)/h(t|0)$

Equivalence Hypothesis

The equivalence of two hazard rates is established by concluding that their ratio is within a clinically insignificant margin from one. The statistical hypotheses that yields this conclusion when the null hypothesis is rejected is

$$H_0: |\log(HR)| \geq \log(HR_U) \quad \text{vs.} \quad H_1: |\log(HR)| < \log(HR_U), \quad \text{assuming } HR_U > 1$$

Test Statistic

It can be shown that the test of b based on the partial likelihood method of Cox (1972) coincides with the common logrank test statistic shown next.

Logrank Test

In this case, the test statistic uses two one-sided logrank tests. Define

$$L_U = \frac{\sum_{k=1}^{K}\left(I_k - \frac{Y_{1i}HR_U}{Y_{1i}HR_U + Y_{2i}}\right)}{\left[\sum_{k=1}^{K}\left(\frac{Y_{1i}Y_{2i}HR_U}{(Y_{1i}HR_U + Y_{2i})^2}\right)\right]^{-\frac{1}{2}}}$$

$$L_L = \frac{\sum_{k=1}^{K}\left(I_k - \frac{Y_{1i}HR_L}{Y_{1i}HR_L + Y_{2i}}\right)}{\left[\sum_{k=1}^{K}\left(\frac{Y_{1i}Y_{2i}HR_L}{(Y_{1i}HR_L + Y_{2i})^2}\right)\right]^{-\frac{1}{2}}}$$

where K is the number of deaths, Y_{ij} is the number of subjects at risk just prior to the jth observed event in the ith group, and I_k is a binary variable indicating whether the kth event is from group 1 or not.

The null hypothesis is rejected if $L_U < z_\alpha$ and $L_L > z_{1-\alpha}$.

The distribution of L_U is approximately normal with mean $\left(\log(HR) - \log(HR_U)\right)\sqrt{P_1 P_2 dN}$ and unit variance. The distribution of L_L is approximately normal with mean $\left(\log(HR) - \log(HR_L)\right)\sqrt{P_1 P_2 dN}$ and unit variance, where

> P_1 is the proportion of N that is in the control group
>
> P_2 is the proportion of N that is in the treatment group
>
> N is the total sample size
>
> N_1 is the sample size from the control group, $N_1 = N(P_1)$
>
> N_2 is the sample size from the treatment group, $N_2 = N(P_2)$
>
> Pev_1 is probability of the event of interest in the control group
>
> Pev_2 is probability of the event of interest in the treatment group
>
> d is the overall probability of an event, $d = Pev_1 P_1 + Pev_2 P_2$
>
> b is the Cox regression coefficient, $b = \log(HR)$
>
> HR_0 is the equivalence limit value entered by the user
>
> HR_L is the lower equivalence boundary (limit) of the hazard ratio, and $HR_L = \text{Min}(HR_0, 1/HR_0)$
>
> HR_U is the upper equivalence boundary (limit) of the hazard ratio, and $HR_U = \text{Max}(HR_0, 1/HR_0)$

Power Calculations

The power of this test is given by

$$\Phi\left(\left(\log(HR_U) - \log(HR_1)\right)\sqrt{P_1 P_2 dN} - z_{1-\alpha}\right) + \Phi\left(\left(\log(HR_U) + \log(HR_1)\right)\sqrt{P_1 P_2 dN} - z_{1-\alpha}\right) - 1$$

where HR_1 is the actual value of the hazard ratio under the alternative hypothesis.

Testing Multiple Treatment Groups versus a Single Control Group

Suppose you have k treatment groups with samples of size N_i and one control group with a sample of size N_C. The total sample size is $N = N_1 + N_2 + … + N_k + N_C$. The response for each subject is their survival time until they either exhibit the event of interest or they are censored from the study.

A Cox proportional hazards regression model is fit to the data in which one of the independent variables is a binary variable that is zero if the subject is from the control group or one if they are from the ith treatment group. Suppose that the regression coefficient associated with this independent variable is called b_i. As pointed out above, it turns out that

$$b_i = \log(HR_i)$$

where HR_i is the hazard ratio comparing the treatment and control groups. If HR_i = 1, there is no difference between the groups.

The data may be analyzed using k separate regressions each producing an equivalence test of the hazard ratio comparing a treatment group to the common control group.

The power for each of the k tests can be computed using the formula given above.

Multiplicity Adjustment

Because k z-tests between treatment groups and the control group are run when analyzing the results of this study, many statisticians recommend that the Bonferroni adjustment be applied. This adjustment is easy to apply: the value of alpha that is used in the test is found by dividing the original alpha by the number of tests. For example, if the original alpha is set at 0.05 and the number of treatment (not including the control) groups is five, the individual tests will be conducted using an alpha of 0.01.

The main criticism of this procedure is that if there are many tests, the value of alpha becomes very small. To mitigate against this complaint, some statisticians recommend separating the treatment groups into those that are of primary interest and those that are of secondary interest. The Bonferroni adjustment is made by the using the number of primary treatments rather than the total number of treatments.

There are some who advocate ignoring the adjustment entirely in the case of randomized clinical trials. See for example Saville (1990) and the discussion in chapter 14 of Machin, Campbell, Tan, and Tan (2018).

Size of the Control Group

Because the control group is used over and over, some advocate increasing the number of subjects in this group. The standard adjustment is to include $\sqrt{k}$ subjects in the control group for each subject in one of the treatment groups. See Machin, Campbell, Tan, and Tan (2018, pages 231-232). Note that often, the treatment groups all have the same size.

Example 1 – Finding the Sample Size

A parallel-group, equivalence trial is being designed to compare the survivability associated with three doses of a test compound against the standard (control) therapy in patients with a specific type of disease.

The proportion surviving one-year after the current treatment is 0.50 ($h_c = 0.693$). The researchers want to determine the sample size necessary to show that the proportion surviving after the new treatment is the same. Hence, they want to compute the power when

$$HR = 1$$

The equivalence hazard ratio limits are 0.8 and 1.25. The researchers would like to study the influence of *HR* on the sample size, so they would like to look at a range of possible values: 1.0 to 1.06

For planning purposes, they decide that the probability of observing an event is 0.60 for all groups. The researchers decide to use a 0.05 significance level and a power of 0.8.

Following standard procedure, the control group multiplier will be set to $\sqrt{k} = \sqrt{3} = 1.732$ since the control group is used for three comparisons in this design.

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 ....**Sample Size**
Power of Each Test **0.8**
Overall Alpha **0.05**
Bonferroni Adjustment ....**Standard Bonferroni**
Group Allocation ....**Enter Group Allocation Pattern, solve for group sample sizes**
Pev (Default Probability of an Event) **0.6**
HR0 (Equivalence Limit Value) **1.25**
Control Probability of an Event ....**Pev**
Control Sample Size Allocation......................**1.732**
Set A Number of Groups...............................**3**
Set A Hazard Ratio **1 1.02 1.04 1.06**
Set A Probability of an Event ....**Pev**
Set A Sample Size Allocation **1**
Set B Number of Groups...............................**0**
Set C Number of Groups **0**
Set D Number of Groups **0**
More...**Unchecked**

Output

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

Numeric Reports

Numeric Results
―――

Solve For: Sample Size
Group Allocation: Enter Group Allocation Pattern, solve for group sample sizes
Test Type: Two, One-Sided Z-Tests Based on the Log Hazard Ratio
Hypotheses: H0: HR ≤ HR$_L$ or HR ≥ HR$_U$ vs. H1: HR$_L$ < HR < HR$_U$
Number of Groups: 4
Bonferroni Adjustment: Standard Bonferroni (Divisor = 3)
―――

	Power		Sample Size		Events Ei	Hazard Ratio HRi	Equivalence Limits			Probability of an Event Pevi	Alpha	
Comparison	Target	Actual	Ni	Allocation			Lower HR$_L$	Upper HR$_U$	Value HR0		Overall	Bonferroni-Adjusted
Control			1063	1.732	637.8					0.6		
vs A1	0.8	0.80011	614	1.000	368.4	1.00	0.8	1.25	1.25	0.6	0.05	0.01667
vs A2	0.8	0.80011	614	1.000	368.4	1.00	0.8	1.25	1.25	0.6	0.05	0.01667
vs A3	0.8	0.80011	614	1.000	368.4	1.00	0.8	1.25	1.25	0.6	0.05	0.01667
Total			2905		1743.0							
Control			1102	1.732	661.2					0.6		
vs A1	0.8	0.80028	636	1.000	381.6	1.02	0.8	1.25	1.25	0.6	0.05	0.01667
vs A2	0.8	0.80028	636	1.000	381.6	1.02	0.8	1.25	1.25	0.6	0.05	0.01667
vs A3	0.8	0.80028	636	1.000	381.6	1.02	0.8	1.25	1.25	0.6	0.05	0.01667
Total			3010		1806.0							
Control			1230	1.732	738.0					0.6		
vs A1	0.8	0.80002	710	1.000	426.0	1.04	0.8	1.25	1.25	0.6	0.05	0.01667
vs A2	0.8	0.80002	710	1.000	426.0	1.04	0.8	1.25	1.25	0.6	0.05	0.01667
vs A3	0.8	0.80002	710	1.000	426.0	1.04	0.8	1.25	1.25	0.6	0.05	0.01667
Total			3360		2016.0							
Control			1484	1.732	890.4					0.6		
vs A1	0.8	0.80039	857	1.000	514.2	1.06	0.8	1.25	1.25	0.6	0.05	0.01667
vs A2	0.8	0.80039	857	1.000	514.2	1.06	0.8	1.25	1.25	0.6	0.05	0.01667
vs A3	0.8	0.80039	857	1.000	514.2	1.06	0.8	1.25	1.25	0.6	0.05	0.01667
Total			4055		2433.0							

Comparison	The group that is involved in the comparison between the treatment and control displayed on this report line. The comparison is made using the hazard ratio.
Target Power	The power desired. Power is probability of rejecting a false null hypothesis for this comparison. This power is of the comparison shown on this line only.
Actual Power	The power actually achieved.
Ni	The number of subjects in the ith group. The total sample size shown below the groups is equal to the sum of all individual group sample sizes.
Allocation	The group sample size allocation ratio of the ith group. The value on each row represents the relative number of subjects assigned to the group.
Ei	The number of events in the ith group required to achieve the power indicated. Ei = Pevi × Ni.
HRi	The hazard ratio of the ith treatment group. HR = hi / hc.
HR$_L$	The lower hazard ratio equivalence limit.
HR$_U$	The upper hazard ratio equivalence limit.
HR0	The equivalence limit value from which the lower and upper equivalence limits were created.
Pevi	The average probability that a subject the ith group will have an event during the study. Pevi also represents the proportion of individuals in the ith group that are expected to have an event during the study. This probability includes the impact of various kinds of censoring.
Overall Alpha	The probability of rejecting at least one of the comparisons in this experiment when each null hypothesis is true.
Bonferroni Alpha	The adjusted significance level at which each individual comparison is made.

Summary Statements
──

A parallel, 4-group design (with one control group and 3 treatment groups) will be used to test whether the hazard rate for each treatment group is equivalent to the control group hazard rate, with equivalence ratio limits of 0.8 and 1.25 (H0: HR ≤ 0.8 or HR ≥ 1.25 versus H1: 0.8 < HR < 1.25, HR = treatment hazard rate i / control hazard rate). Each of the 3 equivalence comparisons will be made using two one-sided, two-sample, Bonferroni-adjusted, Cox's proportional hazards regression term Z-tests. The overall (experiment-wise) Type I error rate (α) is 0.05. It is anticipated that the proportions of subjects in each group that will have an event during the course of the study (beginning with the control group) will be 0.6, 0.6, 0.6, and 0.6. To detect the treatment to control hazard ratios 1, 1, and 1 with at least 80% power for each test, the control group sample size needed will be 1063 and the number of needed subjects for the treatment groups will be 614, 614, and 614 (totaling 2905 subjects overall). The corresponding total number of events is 1743. These results assume that the hazard ratios are constant throughout the study.
──

References
──

Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. 2018. Sample Size Calculations in Clinical Research, 3rd
    Edition. Chapman & Hall/CRC. Boca Raton, FL. Pages 86-88.
Machin, D., Campbell, M.J., Tan, S.B, and Tan, S.H. 2018. Sample Sizes for Clinical, Laboratory, and
    Epidemiology Studies, 4th Edition. Wiley Blackwell.
Schoenfeld, David A. 1983. 'Sample Size Formula for the Proportional-Hazards Regression Model', Biometrics,
    Volume 39, Pages 499-503.
──

This report shows the numeric results of this power study. Notice that the results are shown in blocks of three rows at a time. Each block represents a single design.

Plots Section

Plots
_____
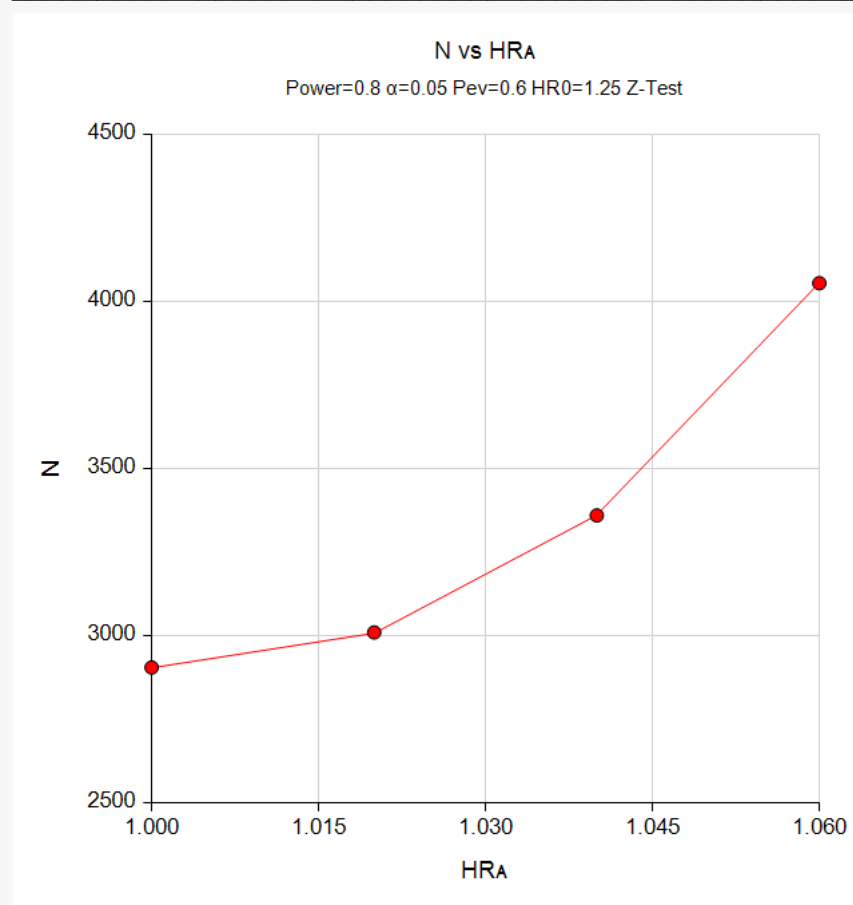
This plot gives a visual presentation to the results in the Numeric Report. We can quickly see the impact on the sample size of changing the hazard ratio from 1.0 to 1.06.

Example 2 – Validation using a Previously Validated Procedure

We could not find a validation result in the statistical literature, so we will use a previously validated **PASS** procedure (**Equivalence Tests for Two Survival Curves Using Cox's Proportional Hazards Model**) to produce the results for the following example.

A parallel-group clinical trial is being designed to compare the survivability induced by three doses of a test compound against the standard (control) therapy in patients with a specific type of disease.

The hazard ratio for the power calculation is set to 1. The equivalence hazard ratio limits are 0.8 and 1.25.

For planning purposes, they decide that the probability of an event is 0.60 for all groups. The researchers decide on a 0.05 significance level and a power of 0.8. Since a Bonferroni adjustment is made, the significance level is reduced to 0.05 / 3 = 0.01667.

The sample sizes of all groups will be equal.

The **Equivalence Tests for Two Survival Curves Using Cox's Proportional Hazards Model** procedure is set up as follows.

Design Tab

Solve For ....**Sample Size**
Power...**0.8**
Alpha...**0.01667** (which is Alpha / k)
Group Allocation ....**Equal (N1 = N2)**
Pev1 (Event Probability in Group 1)...............**0.6**
Pev2 (Event Probability in Group 2)..............**Pev1**
HR1 (Actual Hazard Ratio) **1.0**
HR0 (Equivalence Hazard Ratio)...................**1.25**

This set of options generates the following report.

Numeric Results
───
Solve For: Sample Size
Groups: 1 = Control, 2 = Treatment
Hypotheses: H0: HR ≤ 1/HR0 or HR ≥ HR0   vs.   Ha: 1/HR0 < HR < HR0
───

	Sample Size			Percent Group 1 %N1	Number of Events			Hazard Ratio		Event Probability		
Power	N	N1	N2		E	E1	E2	Equivalence HR0	Actual HR1	Pev1	Pev2	Alpha
0.80022	1557	778	779	50	934.2	466.8	467.4	1.25	1	0.6	0.6	0.01667

In order to maintain a power of 80% for all three groups, it is apparent that the groups will all need to have a sample size of 779 per group if an equal group size is required. This table contains the validation values. We will now run these values through the current procedure and compare the results with these values.

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 ....**Sample Size**

Power of Each Test **0.8**

Overall Alpha **0.05**

Bonferroni Adjustment ....**Standard Bonferroni**

Group Allocation ....**Equal (Nc = N1 = N2 = ...)**

Pev (Default Probability of an Event) **0.6**

HR0 (Equivalence Limit Value) **1.25**

Control Probability of an Event ....**Pev**

Set A Number of Groups...............................**3**

Set A Hazard Ratio **1**

Set A Probability of an Event ....**Pev**

Set B Number of Groups...............................**0**

Set C Number of Groups **0**

Set D Number of Groups **0**

More...**Unchecked**

Output

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

Numeric Results
───
Solve For: Sample Size
Group Allocation: Equal (Nc = N1 = N2 = ...)
Test Type: Two, One-Sided Z-Tests Based on the Log Hazard Ratio
Hypotheses: H0: HR ≤ HRʟ or HR ≥ HRᵤ vs. H1: HRʟ < HR < HRᵤ
Number of Groups: 4
Bonferroni Adjustment: Standard Bonferroni (Divisor = 3)
───

	Power		Sample Size Ni	Events Ei	Hazard Ratio HRi	Equivalence Limits			Probability of an Event Pevi	Alpha	
Comparison	Target	Actual				Lower HRʟ	Upper HRᵤ	Value HR0		Overall	Bonferroni-Adjusted
Control			779	467.4					0.6		
vs A1	0.8	0.80058	779	467.4	1	0.8	1.25	1.25	0.6	0.05	0.01667
vs A2	0.8	0.80058	779	467.4	1	0.8	1.25	1.25	0.6	0.05	0.01667
vs A3	0.8	0.80058	779	467.4	1	0.8	1.25	1.25	0.6	0.05	0.01667
Total			3116	1869.6							

As you can see, the sample sizes are all 779. This matches the sample size found in the validation run above. The procedure is validated.