Chapter 731

Non-Inferiority Tests for Two Survival Curves using Cox's Proportional Hazards Model

Introduction

A clinical trial may be employed to test the non-inferiority of a treatment over a control in regard to their survival distributions. Because survival times are not normally distributed and because some survival times are censored, Cox proportional-hazards regression is often used to analyze the data. 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.

This procedure is documented in Chow, Shao, and Wang (2008) which summarizes the work of Schoenfeld (1981, 1983). Note that there was an error in Chow, Shao, and Wang (2008) page 179 which caused the sample size to be doubled. This error has been corrected in this edition.

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, $h(t|1)/h(t|0)$

Non-Inferiority Hypothesis

Lower Hazards Better

Assuming that lower hazard rates are better, non-inferiority means that the treatment hazard rate is at most, only slightly higher than the control hazard rate. We find it more convenient to state the hypotheses in terms of the hazard ratio, *HR*, rather than the Cox regression coefficient, *b*. Remembering that $b = \log(HR)$ and assuming that $HR_0 > 1$, non-inferiority requires that $HR < HR_0$. Here, HR_0 is the boundary of clinical insignificance or the non-inferiority boundary.

The statistical hypotheses that results in the conclusion of non-inferiority when the null hypothesis is rejected is

$$H_0: \log(HR) \geq \log(HR_0) \quad \text{vs.} \quad H_a: \log(HR) < \log(HR_0)$$

Higher Hazards Better

Assuming that higher hazard rates are better, non-inferiority means that the treatment hazard rate is at most, only slightly lower than the control hazard rate. We find it more convenient to state the hypotheses in terms of the hazard ratio, *HR*, rather than the Cox regression coefficient, *b*. Remembering that $b = \log(HR)$ and assuming that $HR_0 < 1$, non-inferiority requires that $HR > HR_0$. Here, HR_0 is called the boundary of clinical insignificance or the non-inferiority boundary.

The statistical hypotheses that results in the conclusion of non-inferiority when the null hypothesis is rejected is

$$H_0: \log(HR) \leq \log(HR_0) \quad \text{vs.} \quad H_a: \log(HR) > \log(HR_0)$$

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. When testing non-inferiority, you can use *b* from a Cox regression or calculate the modified logrank statistic as follows

Logrank Test

The logrank test statistic is

$$L = \frac{\sum_{k=1}^{K}\left(I_k - \frac{Y_{1i}HR_0}{Y_{1i}HR_0 + Y_{2i}}\right)}{\left[\sum_{k=1}^{K}\left(\frac{Y_{1i}Y_{2i}HR_0}{(Y_{1i}HR_0 + Y_{2i})^2}\right)\right]^{-\frac{1}{2}}}$$

where *K* is the number of deaths, Y_{ij} is the number of subjects at risk just prior to the j^{th} observed event in the i^{th} group, and I_k is a binary variable indicating whether the k^{th} event is from group 1 or not.

The distribution of *L* is approximately normal with mean $\left(\log(HR) - \log(HR_0)\right)\sqrt{P_1 P_2 dN}$ 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, $d = Pev_1 P_1 + Pev_2 P_2$

HR is the observed hazard ratio

HR_0 is the non-inferiority boundary (limit) of the hazard ratio

Power Calculations

The power of this test is given by

$$\Phi\left(\left(\log(HR_1) - \log(HR_0)\right)\sqrt{P_1 P_2 dN} - z_{1-\alpha}\right)$$

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

Example 1 – Finding the Sample Size

A researcher is planning a non-inferiority trial using a parallel, two-group, equal sample allocation design to compare the survivability of a new treatment with that of the current treatment. The non-inferiority hazard ratio is 1.2. The desired power is 0.90 and significance level is 0.025. HR1 will be between 0.5 and 1. The probability of observing an event is 0.50 in the control group and 0.30 in the treatment group.

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**
Higher Hazards Are	**Worse (Ha: HR < HR0)**
Power	**0.90**
Alpha	**0.025**
Group Allocation	**Equal (N1 = N2)**
Pev1 (Event Probability in Group 1)	**0.5**
Pev2 (Event Probability in Group 2)	**0.3**
HR1 (Actual Hazard Ratio)	**0.5 to 1 by 0.1**
HR0 (Non-Inferiority Hazard Ratio)	**1.2**

Output

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

Numeric Reports

Numeric Results

Solve For: Sample Size
Groups: 1 = Control, 2 = Treatment
Hypotheses: H0: HR ≥ HR0 vs. Ha: HR < HR0

	Sample Size				Number of Events			Hazard Ratio		Event Probability		
Power	N	N1	N2	Percent Group 1 %N1	E	E1	E2	Non-Inferiority HR0	Actual HR1	Pev1	Pev2	Alpha
0.9019	138	69	69	50	55.2	34.5	20.7	1.2	0.5	0.5	0.3	0.025
0.9001	219	109	110	50	87.5	54.5	33.0	1.2	0.6	0.5	0.3	0.025
0.9003	362	181	181	50	144.8	90.5	54.3	1.2	0.7	0.5	0.3	0.025
0.9004	640	320	320	50	256.0	160.0	96.0	1.2	0.8	0.5	0.3	0.025
0.9001	1270	635	635	50	508.0	317.5	190.5	1.2	0.9	0.5	0.3	0.025
0.9001	3162	1581	1581	50	1264.8	790.5	474.3	1.2	1.0	0.5	0.3	0.025

Power	The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
N	The total sample size.
N1 and N2	The sample sizes of groups 1 and 2, respectively.
%N1	The percent of the total sample that is in group 1, the control group.
E, E1, and E2	The number of events required in both groups and groups 1 and 2, respectively.
HR	The hazard ratio. HR = h2/h1
HR0	The non-inferiority hazard ratio.
HR1	The actual hazard ratio at which power is calculated.
Pev1 and Pev2	The probabilities of an event in the control and the treatment groups, respectively.
Alpha	The probability of rejecting a true null hypothesis.

Summary Statements

A parallel, two-group design (where higher hazard rates are considered worse) will be used to test whether the Group 2 (treatment) hazard rate is non-inferior to the Group 1 (control) hazard rate, with a non-inferiority hazard ratio of 1.2 (H0: HR ≥ 1.2 versus Ha: HR < 1.2, HR = h2 / h1). The comparison will be made using a one-sided, two-sample Cox proportional hazards regression test (or equivalent non-inferiority logrank test) with a Type I error rate (α) of 0.025. It is anticipated that the probability of observing an event during the course of the study is 0.5 for each member of the control group and 0.3 for each member of the treatment group. The calculations are based on the assumption that the hazard ratio is constant throughout the study. To detect a hazard ratio (h2 / h1) of 0.5 with 90% power, the number of needed subjects will be 69 in Group 1 and 69 in Group 2 (totaling 138 subjects). The corresponding required number of events is 34.5 in Group 1 and 20.7 in Group 2 (totaling 55.2 events).

References

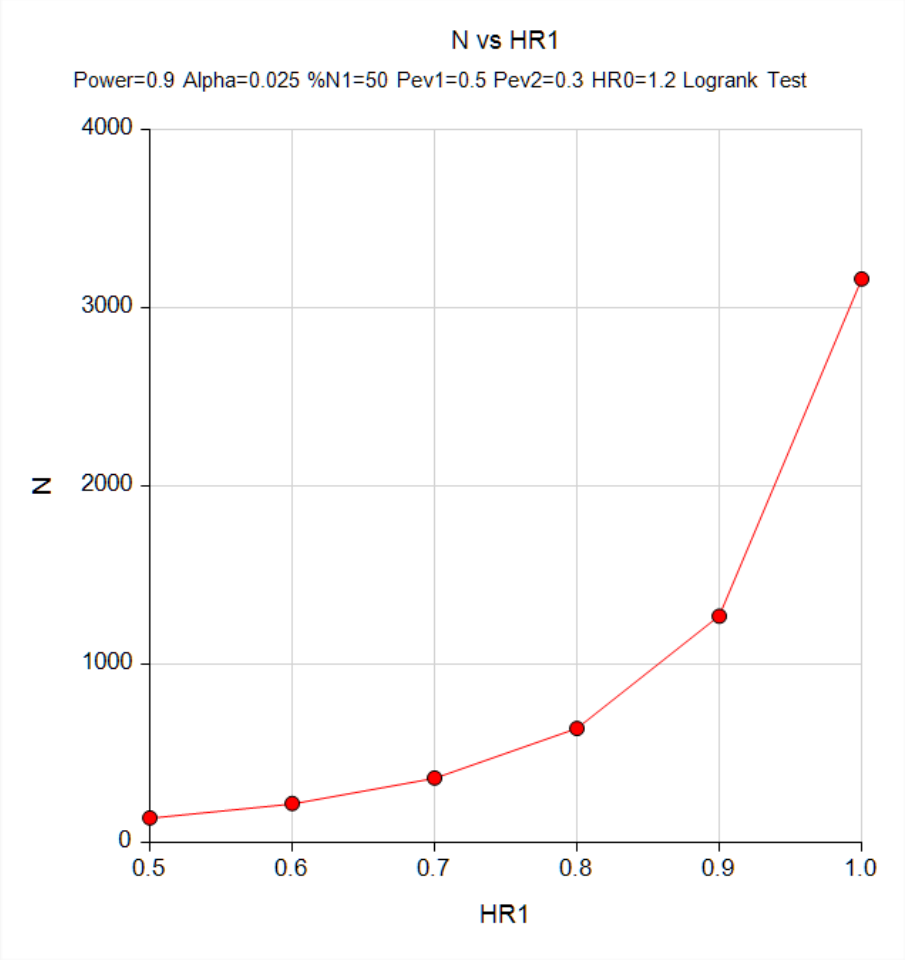
Chow, S.C., Shao, J., Wang, H. 2008. Sample Size Calculations in Clinical Research, 2nd Edition. Chapman & Hall/CRC.

Schoenfeld, David A. 1983. 'Sample Size Formula for the Proportional-Hazards Regression Model', Biometrics, Volume 39, Pages 499-503.

These reports show the values of each of the parameters, one scenario per row.

Plots Section

Plots



This plot shows the relationship between HR1 and N.

Example 2 – Validation using Chow et al. (2008)

Chow et al. (2008) page 179 presents an example that we will use for validation. In their example, HR1 = 2, log(HR0) = 0.3, Pev1 = Pev2 = 0.8, P1 = 0.5, alpha = 0.05, and power = 0.8. They obtained a value of about 100 per group (this is a correction from the 200 that they originally printed in error).

Actually, this example is for a superiority test, but if HR1 and HR0 are switched, it can be used to validate the non-inferiority test.

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**
Higher Hazards Are	**Worse (Ha: HR < HR0)**
Power	**0.80**
Alpha	**0.05**
Group Allocation	**Equal (N1 = N2)**
Pev1 (Event Probability in Group 1)	**0.8**
Pev2 (Event Probability in Group 2)	**Pev1**
HR1 (Actual Hazard Ratio)	**1.35 (which is exp(0.3))**
HR0 (Non-Inferiority Hazard Ratio)	**2**

Output

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

Numeric Results

Solve For: Sample Size
Groups: 1 = Control, 2 = Treatment
Hypotheses: H0: HR ≥ HR0 vs. Ha: HR < HR0

	Sample Size			Percent	Number of Events			Hazard Ratio		Event Probability		
Power	N	N1	N2	Group 1 %N1	E	E1	E2	Non-Inferiority HR0	Actual HR1	Pev1	Pev2	Alpha
0.8015	201	100	101	50	160.8	80	80.8	2	1.35	0.8	0.8	0.05

PASS also calculates the value of N = 201 which is within rounding of the 200 that Chow calculated.