Chapter 569

Two-Sample Equivalence Tests for Survival Data using Cox Regression

Introduction

This procedure uses Cox (proportional hazards) regression analysis, which models the relationship between a set of one or more covariates and the hazard rate, for making inference about the equivalence of a treatment hazard rate compared to a control hazard rate using survival data with independent groups. The question of interest is whether the treatment is equivalent to the control, that is, differ at most by a small margin.

Additional discrete or continuous covariates may be included for adjustment before comparison of the two groups. Inference about the equivalence of the treatment versus the control is based on the hazard ratio (or risk ratio), HR, where

$$HR = \frac{Hazard(\text{Treatment Group})}{Hazard(\text{Control Group})}$$

The details regarding Cox Regression are available in the Cox Regression chapter of the documentation and will not be repeated here. Only the details related specifically to equivalence testing will be covered in this chapter.

The Cox Regression Model

Survival analysis refers to the analysis of elapsed time. The response variable is the time between a *time origin* and an *end point.* The end point is either the occurrence of the event of interest, referred to as a *death* or *failure*, or the end of the subject's participation in the study. These elapsed times have two properties that invalidate standard statistical techniques, such as t-tests, analysis of variance, and multiple regression. First of all, the time values are often positively skewed. Standard statistical techniques require that the data be normally distributed. Although this skewness could be corrected with a transformation, it is easier to adopt a more realistic data distribution.

The second problem with survival data is that part of the data are *censored*. An observation is censored when the end point has not been reached when the subject is removed from study. This may be because the study ended before the subject's response occurred, or because the subject withdrew from active participation. This may be because the subject died for another reason, because the subject moved, or because the subject quit following the study protocol. All that is known is that the response of interest did not occur while the subject was being studied.

When analyzing survival data, two functions are of fundamental interest—the *survivor function* and the *hazard function*. Let *T* be the survival time. That is, *T* is the elapsed time from the beginning point, such as diagnosis of cancer, and death due to that disease. The values of *T* can be thought of as having a *probability distribution*. Suppose the *probability density function* of the random variable *T* is given by $f(T)$. The *probability distribution function* of *T* is then given by

$$F(T) = \Pr(t < T)$$

$$= \int_0^T f(t)\, dt$$

The *survivor function*, $S(T)$, is the probability that an individual survives past *T*. This leads to

$$S(T) = \Pr(T \geq t)$$

$$= 1 - F(T)$$

The *hazard function* is the probability that a subject experiences the event of interest (death, relapse, etc.) during a small time interval given that the individual has survived up to the beginning of that interval. The mathematical expression for the hazard function is

$$h(T) = \lim_{\Delta T \to 0} \frac{\Pr(T \leq t < (T + \Delta T) | T \leq t)}{\Delta T}$$

$$= \lim_{\Delta T \to 0} \frac{F(T + \Delta T) - F(T)}{\Delta T}$$

$$= \frac{f(T)}{S(T)}$$

The cumulative hazard function $H(T)$ is the sum of the individual hazard rates from time zero to time *T*. The formula for the cumulative hazard function is

$$H(T) = \int_0^T h(u) du$$

Thus, the hazard function is the derivative, or slope, of the cumulative hazard function. The cumulative hazard function is related to the cumulative survival function by the expression

$$S(T) = e^{-H(T)}$$

or

$$H(T) = -\ln\big(S(T)\big)$$

We see that the distribution function, the hazard function, and the survival function are mathematically related. As a matter of convenience and practicality, the hazard function is used in the basic Cox regression model.

Cox (1972) expressed the relationship between the hazard rate and a set of covariates using the model

$$\ln[h(T)] = \ln[h_0(T)] + \sum_{i=1}^{p} x_i\beta_i$$

or

$$h(T) = h_0(T)e^{\sum_{i=1}^{p} x_i\beta_i}$$

where $x_1, x_2, \cdots, x_p$ are covariates, $\beta_1, \beta_2, \cdots, \beta_p$ are regression coefficients to be estimated, *T* is the elapsed time, and $h_0(T)$ is the baseline hazard rate when all covariates are equal to zero. Thus, the linear form of the regression model is

$$\ln\left[\frac{h(T)}{h_0(T)}\right] = \sum_{i=1}^{p} x_i\beta_i$$

Taking the exponential of both sides of the above equation, we see that this is the ratio between the actual hazard rate and the baseline hazard rate (i.e., the hazard ratio), and is sometimes called the *relative risk*. This can be rearranged to give the model

$$\frac{h(T)}{h_0(T)} = \exp\left(\sum_{i=1}^{p} x_i\beta_i\right)$$

$$= e^{x_1\beta_1}e^{x_2\beta_2} \ldots e^{x_p\beta_p}$$

The regression coefficients can thus be interpreted as the relative risk when the value of the covariate is increased by one unit.

Note that unlike most regression models, this model does not include an intercept term. This is because if an intercept term were included, it would become part of $h_0(T)$.

Also note that the above model does not include *T* on the right-hand side. That is, the relative risk is constant for all time values. This is why the method is called *proportional hazards*.

An interesting attribute of this model is that you only need to use the ranks of the failure times to estimate the regression coefficients. The actual failure times are not used except to generate the ranks. Thus, you will achieve the same regression coefficient estimates regardless of whether you enter the time values in days, months, or years.

TOST Equivalence Test (Two One-Sided Tests) using the Cox Proportional Hazard Regression Model

Schuirmann's (1987) two one-sided tests (TOST) approach is used to test equivalence. This equivalence test essentially reverses the roles of the usual null and alternative hypothesis. Using $HR = h_T(T)/h_C(T)$ (where $h_T(T)$ is the hazard rate of the treatment group at time *T* and $h_C(T)$ is the hazard rate of the reference or control group at time *T*) to represent the hazard ratio of the treatment and the control, *L* to represent the lower equivalence bound, and *U* to represent the upper equivalence bound, the null and alternative hypotheses are

$$\text{H0}: HR < L \text{ or } HR > U$$
$$\text{H1}: L < HR < U$$

The null hypothesis is made up of two simple one-sided hypotheses:

$$\text{H0}_1: HR < L$$
$$\text{H0}_2: HR > U$$

The null hypotheses can be tested at significance level α by constructing a 100(1 – 2*α*)% confidence interval for HR (not the usual 100(1 – *α*)% like you might suspect). If the upper bound of the 100(1 – 2*α*)% confidence interval is less than *U*, and the lower bound is greater than *L*, then you can conclude that the treatment and the control are equivalent. This is the same as conducting two one-sided tests of the null hypotheses using the Wald test. If both of these one-sided tests are rejected, we conclude H1 that the two groups are equivalent (their difference is confined within a small margin). Schuirmann showed that if we want the alpha level of the equivalence test to be *α*, then each of the one-sided tests should be *α* as well (not *α*/2 as you might expect). The probability level (p-value) of the equivalence test is equal to the maximum of the probability levels of the two one-sided tests.

Hazard Ratio (or Risk Ratio)

The hazard ratio for the group variable is computed from the Cox regression coefficient for the grouping variable as

$$HR = \exp\left(b_{Grp}\right)$$

100(1 – 2α)% Confidence Interval for HR

The 100(1 – 2*α*)% confidence interval for the hazard ratio computed by **NCSS** is based on the Wald statistic, which is valid for large samples. The formula for the limits of a 100(1 – 2*α*)% two-sided confidence interval for *HR* is

$$\exp\left(b_{Grp} \pm z_{1-\alpha} s_{b_{Grp}}\right),$$

where b_{Grp} is the Cox regression coefficient corresponding to the Group variable, and $s_{b_{Grp}}$ is the standard error of b_{Grp}.

Wald Test

The formula for the Wald test statistics are

$$Z_L = \frac{b_{Grp} - \ln(L)}{s_{b_{Grp}}} \quad \text{and} \quad Z_U = \frac{b_{Grp} - \ln(U)}{s_{b_{Grp}}}$$

where b_{Grp} is the Cox regression coefficient corresponding to the Group variable, $s_{b_{Grp}}$ is the standard error of b_{Grp}, and *L* and *U* is the are the lower and upper equivalence bounds, respectively. With large sample sizes, the distribution of *Z* is closely approximated by the normal distribution. With small and moderate sample sizes, the normal approximation is described as "adequate."

Data Structure

Survival data sets require up to three components for the survival time: the ending survival time, the beginning survival time during which the subject was not observed, and an indicator of whether the observation was censored or failed.

Based on these three components, various types of data may be analyzed. *Right censored* data are specified using only the ending time variable and the censor variable. *Left truncated* and *Interval* data are entered using all three variables.

The table below shows survival data ready for analysis. These data are in the CoxRegSub dataset. The variables are

Time	Months of survival
Status	Censor indicator
Treatment	Treatment category (C = Control, T1 = Treatment)
Age	Age of patient in years
Gender	Gender of patient
Count	Frequency (or count) variable

CoxRegSub Dataset (Subset)

Time	Censor	Treatment	Age	Gender	Count
14.29	1	C	20	F	1
16.68	0	C	20	F	5
3.38	1	C	20	F	1
3.97	1	C	20	F	1
10.43	1	C	20	M	1
11.66	0	C	20	M	4
6.51	1	C	20	M	1
11.86	1	C	20	M	1
5.71	1	T1	20	F	1
6.35	0	T1	20	F	6
11.36	1	T1	20	F	1
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.

Example 1 – Two-Sample TOST Equivalence Test for Survival Data

This section presents an example of how to perform a two-sample equivalence test with survival data using Cox regression. Researchers want to determine if a new treatment (T1) is equivalent to the current standard treatment (C). The upper and lower hazard ratio bounds of equivalence are set at 1.25 and 0.8, respectively. The data for this study are contained in the **CoxRegSub** dataset.

Note: This example only shows the output directly related to the equivalence test. For description about the other reports available, see the Cox Regression chapter.

Setup

To run this example, complete the following steps:

1 Open the CoxRegSub example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **CoxRegSub** and click **OK**.

2 Specify the Two-Sample Equivalence Tests for Survival Data using Cox Regression procedure options

- Find and open the **Two-Sample Equivalence Tests for Survival Data using Cox Regression** procedure using the menus or the Procedure Navigator.
- The settings 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.

Variables, Model Tab

Time	**Time**
Group	**Treatment**
Reference Group	**C**
Censor	**Censor**
Frequencies	**Count**
Hazard Ratio Upper Equivalence Bound	**1.25**
Hazard Ratio Lower Equivalence Bound	**0.80**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Run Summary

Run Summary

Item	Value	Rows	Value
Time Variable	Time	Rows Processed	49
Censor Variable	Censor	Rows Used in Estimation	49
Frequency Variable	Count	Rows with X's Missing	0
Independent Variables Available	1	Rows with Y Missing	0
Number of X's in the Model	1	Rows Failed	36
Final Log-Likelihood	-142.3606	Rows Censored	13
Number of Likelihood Iterations	3 of 20	Sum of Frequencies	93
Convergence Criterion	1E-09	Sum of Censored Frequencies	57
Achieved Convergence	4.853289E-10	Sum of Failed Frequencies	36
Completion Status	Normal completion		
Starting B's	0		

This report summarizes the characteristics of the dataset and provides useful information about the analysis. It should be studied to make sure that the data were read in properly and that the estimation algorithm terminated normally. We will only discuss those parameters that need special explanation.

Final Log-Likelihood

This is the log-likelihood of the model.

Iterations

This is the number of iterations used by the maximum likelihood procedure. This value should be compared against the value of the Maximum Iterations option to see if the iterative procedure terminated early.

Achieved Convergence

This is the maximum of the relative changes in the regression coefficients on the last iteration. If this value is less than the Convergence Criterion, the procedure converged normally. Otherwise, the specified convergence precision was not achieved.

Rows Processed

This is the number of rows processed during the run. Check this count to make certain it agrees with what you anticipated.

100(1 - 2α)% Confidence Interval Test for Equivalence (Two One-Sided Wald Tests)

100(1 - 2α)% Confidence Interval Test for Equivalence (Two One-Sided Wald Tests)

HR = Hazard(Treatment="T1") / Hazard(Treatment="C")

Null Hypothesis (H0): HR ≤ 0.8 or HR ≥ 1.25
Equivalence Hypothesis (H1): 0.8 < HR < 1.25

Test	Alternative Hypothesis	Hazard Ratio HR*	90% Confidence Interval Limits for HR		Wald Test		Reject H0 at α = 0.05?
			Lower	Upper	Z-Value	P-Value	
Lower Boundary	HR > 0.8	0.8590	0.4958	1.4884	0.2131	0.4156	No
Upper Boundary	HR < 1.25	0.8590	0.4958	1.4884	-1.1225	0.1308	No
Equivalence	0.8 < HR < 1.25	0.8590	0.4958	1.4884		0.4156	No

* In Cox Regression, the Hazard Ratio (HR) is commonly referred to as the Risk Ratio and is equal to Exp(B), where B is the estimated regression coefficient.

This report displays the results of the two one-sided tests for equivalence. The Wald P-value for the overall equivalence test is 0.4156, indicating that we cannon reject the null and cannot conclude equivalence at the 0.05 level. Notice that the upper limit of the 90% confidence interval for the hazard ratio (HR) is not less than the equivalence bound of 1.25 and the lower limit of the confidence interval is not greater than 0.80.

Analysis of Deviance Section

Analysis of Deviance

Term(s) Omitted	DF	-2 Log-Likelihood	Increase from Model Deviance (Chi²)	P-Value
All Terms	1	284.9279	0.2066	0.6495
Treatment	1	284.9279	0.2066	0.6495
None(Model)	1	284.7213		

The P-Value is for testing the significance of each term after adjusting for all other terms.

This report is the Cox regression analog of the analysis of variance table. It displays the results of a chi-square test used to test whether each of the individual terms are statistically significant after adjusting for all other terms in the model.

The DF (degrees of freedom) column indicates the number of binary variables needed to represent each term. The chi^2 test is used to test the significance of all binary variables associated with a particular term.

Log-Likelihood and R^2 Section

Log-Likelihood and R^2

Term(s) Omitted	DF	Log-Likelihood	R^2 of Remaining Term(s)	Reduction from Model R^2
All Terms	1	-142.4639	0.0000	0.0022
Treatment	1	-142.4639	0.0000	0.0022
None(Model)	1	-142.3606	0.0022	0.0000

This report displays the Log-Likelihood and R^2 that is achieved when each term is omitted from the regression model. The DF (degrees of freedom) column indicates the number of binary variables needed to represent each term. The chi^2 test is used to test the significance of all binary variables associated with a particular term.

Example 2 – Two-Sample TOST Equivalence Test for Survival Data with Adjustment for Additional Covariates

Continuing with Example 1, this example will demonstrate how to include additional covariates in the model in which the Group variable is tested for equivalence.

Setup

To run this example, complete the following steps:

1 Open the CoxRegSub example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **CoxRegSub** and click **OK**.

2 Specify the Two-Sample Equivalence Tests for Survival Data using Cox Regression procedure options

- Find and open the **Two-Sample Equivalence Tests for Survival Data using Cox Regression** procedure using the menus or the Procedure Navigator.
- The settings 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.

Variables, Model Tab	
Time	**Time**
Group	**Treatment**
Reference Group	**C**
Censor	**Censor**
Frequencies	**Count**
Hazard Ratio Upper Equivalence Bound	**1.25**
Hazard Ratio Lower Equivalence Bound	**0.80**
Adjust for Covariates	**Checked**
Categorical X's	**Age, Gender**

Reports Tab	
Regression Coefficients	**Checked**
Confidence Limits of Regression Coefficients	**Checked**

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Output

Run Summary

Item	Value	Rows	Value
Time Variable	Time	Rows Processed	49
Censor Variable	Censor	Rows Used in Estimation	49
Frequency Variable	Count	Rows with X's Missing	0
Independent Variables Available	3	Rows with Y Missing	0
Number of X's in the Model	4	Rows Failed	36
Final Log-Likelihood	-141.0197	Rows Censored	13
Number of Likelihood Iterations	5 of 20	Sum of Frequencies	93
Convergence Criterion	1E-09	Sum of Censored Frequencies	57
Achieved Convergence	1.996924E-16	Sum of Failed Frequencies	36
Completion Status	Normal completion		
Starting B's	0		

100(1 - 2α)% Confidence Interval Test for Equivalence (Two One-Sided Wald Tests)

HR = Hazard(Treatment="T1") / Hazard(Treatment="C")

Additional Categorical X's: Age, Gender
Null Hypothesis (H0): HR ≤ 0.8 or HR ≥ 1.25
Equivalence Hypothesis (H1): 0.8 < HR < 1.25

Test	Alternative Hypothesis	Hazard Ratio HR*	90% Confidence Interval Limits for HR		Wald Test		Reject H0 at α = 0.05?
			Lower	Upper	Z-Value	P-Value	
Lower Boundary	HR > 0.8	0.8108	0.4599	1.4296	0.0390	0.4844	No
Upper Boundary	HR < 1.25	0.8108	0.4599	1.4296	-1.2555	0.1046	No
Equivalence	0.8 < HR < 1.25	0.8108	0.4599	1.4296		0.4844	No

* In Cox Regression, the Hazard Ratio (HR) is commonly referred to as the Risk Ratio and is equal to Exp(B), where B is the estimated regression coefficient.

Regression Coefficients

Independent Variable	Regression Coefficient (B)	Standard Error of B	Risk Ratio Exp(B)	Mean	Wald Z-Value	P-Value	Pseudo R^2
B1: (Treatment="T1")	-0.209688	0.344742	0.8108	0.4946237	-0.6082	0.5430	0.0105
B2: (Age=40)	0.232735	0.413718	1.2620	0.3225806	0.5625	0.5737	0.0090
B3: (Age=60)	0.695165	0.425375	2.0040	0.3333333	1.6342	0.1022	0.0712
B4: (Gender="M")	0.084476	0.344212	1.0881	0.5268817	0.2454	0.8061	0.0017

Confidence Interval Limits for Regression Coefficients

Independent Variable	Regression Coefficient (B)	95% Confidence Interval Limits for B		Risk Ratio Exp(B)	95% Confidence Interval Limits for Exp(B)	
		Lower	Upper		Lower	Upper
B1: (Treatment="T1")	-0.209688	-0.885371	0.465994	0.8108	0.4126	1.5936
B2: (Age=40)	0.232735	-0.578138	1.043608	1.2620	0.5609	2.8394
B3: (Age=60)	0.695165	-0.138555	1.528885	2.0040	0.8706	4.6130
B4: (Gender="M")	0.084476	-0.590167	0.759119	1.0881	0.5542	2.1364

Analysis of Deviance

Term(s) Omitted	DF	-2 Log-Likelihood	Increase from Model Deviance (Chi²)	P-Value
All Terms	4	284.9279	2.8884	0.5767
Treatment	1	282.4091	0.3697	0.5432
Age	2	284.7057	2.6662	0.2637
Gender	1	282.0997	0.0603	0.8061
None(Model)	4	282.0394		

The P-Value is for testing the significance of each term after adjusting for all other terms.

Log-Likelihood and R^2

Term(s) Omitted	DF	Log-Likelihood	R^2 of Remaining Term(s)	Reduction from Model R^2
All Terms	4	-142.4639	0.0000	0.0306
Treatment	1	-141.2046	0.0267	0.0039
Age	2	-142.3528	0.0024	0.0282
Gender	1	-141.0498	0.0300	0.0006
None(Model)	4	-141.0197	0.0306	0.0000

Adjusting for other covariates, it is still not possible to reject the null hypothesis and conclude equivalence.

The Regression Coefficients and Confidence Limits reports give the Cox regression parameter estimates for both the group variable and the other variables included in the model. The hypotheses tested by these reports are that the regression coefficients are equal to zero.