

Chapter 624

Superiority by a Margin Tests for Two Survival Curves using Cox's Proportional Hazards Model in a Cluster-Randomized Design

Introduction

This procedure calculates power and sample size for testing the superiority by a margin of a treatment hazard rate compared to a control hazard rate using Cox's proportional hazards regression when the data are obtained in a cluster-randomized design. 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, Wang, and Lakhnygina (2018) and Machin, Campbell, Tan, and Tan (2018) which are based on the work of Schoenfeld (1981, 1983).

Cluster-randomized designs are those in which whole clusters of subjects (classes, hospitals, communities, etc.) are randomized into the treatment or control groups. In this design, the survival curves of the two groups, made up of K_i clusters of M_{ij} individuals each, are to be tested using a regression coefficient from a Cox regression.

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

The one-sided, statistical hypothesis testing survival superiority is a test of whether b is different from a constant value. This hypothesis is stated as

$$H_0: b \geq b_0 \quad \text{vs.} \quad H_1: b < b_0$$

Superiority by a Margin Hypothesis

Higher Hazards Worse

Assuming that lower hazard rates are better, *superiority by a margin* means that the treatment hazard rate is lower than the control hazard rate by a clinically important amount. 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 $HR_0 < 1$, superiority requires that $HR < HR_0$. Here, HR_0 is the superiority boundary.

The statistical hypotheses that results in the conclusion of superiority when the null hypothesis is rejected is

$$H_0: \log(HR) \geq \log(HR_0) \quad \text{vs.} \quad H_1: \log(HR) < \log(HR_0)$$

Higher Hazards Better

Assuming that higher hazard rates are better, *superiority by a margin* means that the treatment hazard rate is greater than the control hazard rate by a clinically important amount. 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 $HR_0 > 1$, superiority by a margin requires that $HR > HR_0$. Here, HR_0 is called the superiority boundary.

The statistical hypotheses that results in the conclusion of superiority by a margin when the null hypothesis is rejected is

$$H_0: \log(HR) \leq \log(HR_0) \quad \text{vs.} \quad H_1: \log(HR) > \log(HR_0)$$

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 shown next.

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 $(\log(HR) - \log(HR_0))\sqrt{P_1 P_2 d N}$ 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, $N_2 = N(P_2)$

Pe_{v1} is probability of the event of interest in the control group

Pe_{v2} is probability of the event of interest in the treatment group

d is the overall probability of an event, $d = Pe_{v1}P_1 + Pe_{v2}P_2$

HR is the observed hazard ratio

HR_0 is the superiority boundary (limit) of the hazard ratio

Cluster-Randomized Designs

Denote an observation by Y_{ijk} where $i = 1, 2$ gives the group, $j = 1, 2, \dots, K_i$ gives the cluster within group i , and $k = 1, 2, \dots, m_{ij}$ denotes an individual in cluster j of group i . In this chapter, we will assume that group 1 is the control group and group 2 is the treatment group.

Let ρ denote the intraclass correlation coefficient (ICC) among individuals from the same cluster. This correlation is the correlation of censor indicator variable. Also let COV denote the coefficient of variation of the cluster sizes. Machin *et al.* (2018) page 101 shows that the number of events, e , that are needed to obtain a given power of $1 - \beta$ and a one-sided significance level of α to detect a hazard ratio of HR (h_2 / h_1) is given by

$$e = \frac{(1 + r)^2}{r} \left[\frac{(z_{1-\alpha} + z_{1-\beta})^2}{(\log HR_1 - \log HR_0)^2} \right]$$

Superiority by a Margin Tests for Two Survival Curves using Cox's Prop. Hazards Model in a Cluster-Randomized Design

where $r = N_2 / N_1$ and $z_x = \Phi(x)$ is the standard normal distribution function.

The number of subjects in a regular design can be determined using

$$N = N_1 + N_2 = \frac{e}{F}$$

where

$$F = \frac{(Pev_1 + r(Pev_2))}{1 + r}$$

Design Effect

When using a cluster-randomized design Machin *et al.* (2018) page 202 show that the above formula is modified using a quantity that is known as the *design effect* (DE). The version of DE that is used in **PASS** is given as formula 12.7 on page 197 of Machin *et al.* (2018) which is

$$DE = 1 + \{[CV^2 + 1]\bar{M} - 1\}\rho$$

where \bar{M} is the average cluster size of all clusters given by

$$\bar{M} = \frac{K_1 M_1 + K_2 M_2}{K_1 + K_2}$$

CV is the coefficient of variation of cluster sizes of all clusters in the study and ρ is the ICC as defined above.

The resulting sample size formula in terms of cluster counts and size is

$$N = K_1 M_1 + K_2 M_2 = DE \left(\frac{e}{F} \right)$$

Power Calculations

The power of this test is given by

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

where HR_1 is the actual value of the hazard ratio under the alternative hypothesis and N is defined as above.

Example 1 – Showing Impact of ICC and COV on K1 and K2

This example will not only show how to find the sample size but will also show how ICC and COV impact the sample size.

Suppose that a cluster-randomized study is being planned in which higher hazards are worse; $HR0 = 0.8$; $HR1 = 0.6$; $Pev1 = 0.7$; $Pev2 = 0.5$; $\rho = 0.00, 0.01, 0.05$; $M1$ and $M2 = 20$; $alpha = 0.025$; $power = 0.90$; $COV = 0.0, 0.6$; and $K1 = K2$. Sample size (number of clusters) is desired.

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	Number of Clusters
Higher Hazards Are	Worse (H1: HR < HR0)
Power.....	0.90
Alpha.....	0.025
M1 (Average Cluster Size).....	20
K2 (Number of Clusters)	K1
M2 (Average Cluster Size).....	M1
COV of Cluster Sizes.....	0 0.6
Pev1 (Probability of a Control Event).....	0.7
Pev2 (Probability of a Treatment Event)	0.5
HR0 (Superiority Hazard Ratio)	0.8
HR1 (Actual Hazard Ratio)	0.6
ρ (Intracluster Correlation)	0 0.1 0.05

Output

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

Numeric Reports

Numeric Results

Solve For: Number of Clusters
 Groups: 1 = Control, 2 = Treatment
 Higher Hazards Are: Worse
 Hypotheses: H0: HR ≥ HR0 vs. H1: HR < HR0

Group	Power	Sample Size Ni	Number of Events Ei	Number of Clusters Ki	Cluster Size		Probability of an Event Pevi	Hazard Ratio		Intracluster Correlation ρ	Alpha
					Average Mi	COV		Superiority HR0	Actual HR1		
1: Control	0.91073	440	308.0	22	20	0.0	0.7	0.8	0.6	0.00	0.025
2: Treatment		440	220.0	22	20	0.0	0.5				
Total		880	528.0	44							
1: Control	0.90497	840	1146.6	42	20	0.0	0.7	0.8	0.6	0.05	0.025
2: Treatment		840	819.0	42	20	0.0	0.5				
Total		1680	1965.6	84							
1: Control	0.90291	1240	2517.2	62	20	0.0	0.7	0.8	0.6	0.10	0.025
2: Treatment		1240	1798.0	62	20	0.0	0.5				
Total		2480	4315.2	124							
1: Control	0.91073	440	308.0	22	20	0.6	0.7	0.8	0.6	0.00	0.025
2: Treatment		440	220.0	22	20	0.6	0.5				
Total		880	528.0	44							
1: Control	0.90070	980	1584.7	49	20	0.6	0.7	0.8	0.6	0.05	0.025
2: Treatment		980	1131.9	49	20	0.6	0.5				
Total		1960	2716.6	98							
1: Control	0.90148	1540	3902.4	77	20	0.6	0.7	0.8	0.6	0.10	0.025
2: Treatment		1540	2787.4	77	20	0.6	0.5				
Total		3080	6689.8	154							

- Group The group to which the results on this line belong. Group 1 is the control group and group 2 is the treatment group. The first (Total) line shows the total when appropriate.
- Power The probability of rejecting a false null hypothesis for this hazard ratio. This is the power calculated at HR1.
- Ni The number of items (subjects) in the ith group. The total sample size is shown as the last row of each three-row set.
- Ei The number of events in the ith group required to achieve the power indicated.
- Ki The number of clusters in the ith group. The total number of clusters is reported in the last row of each three-row set.
- Mi The average number of items per cluster (or average cluster size) in the ith group.
- COV The coefficient of variation of the cluster sizes of the clusters within the group. This value is used for both groups.
- Pevi The average probability that a subject in the ith group will have an event during the study. This probability includes the impact of various kinds of censoring.
- HR0 The superiority hazard ratio boundary used to declare whether a treatment is superior to the control.
- HR1 The hazard ratio of the treatment group and the control group. HR1 = h2 / h1. The subscript '1' indicates that this hazard ratio value is used for the alternative hypothesis, H1.
- ρ The intracluster correlation (ICC) between any two observations from within a cluster. This statistic has two common interpretations: (1) the proportion of variation related to the between-cluster variation or (2) the correlation between two subjects from the same cluster.
- Alpha The probability of rejecting a true null hypothesis.

Summary Statements

A parallel, two-group cluster-randomized design (where higher hazard rates are considered worse) will be used to test whether the Group 2 (treatment) hazard rate is superior to the Group 1 (control) hazard rate by a margin, with a superiority hazard ratio of 0.8 ($H_0: HR \geq 0.8$ versus $H_1: HR < 0.8$, $HR = h_2 / h_1$). The comparison will be made using a one-sided, two-sample Cox proportional hazards regression coefficient 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 0.7 for each member of the control group and 0.5 for each member of the treatment group. The coefficient of variation of the cluster sizes within a group is assumed to be 0. The intracluster correlation coefficient of responses within a particular cluster is assumed to be 0. The calculations are based on the assumption that the hazard ratio is constant throughout the study. To detect a hazard ratio (h_2 / h_1) of 0.6, with average cluster sizes of 20 subjects per cluster in Group 1 and 20 subjects per cluster in Group 2, with 90% power, the number of needed clusters is 22 for Group 1 (control) (totaling 440 subjects) and 22 for Group 2 (treatment) (totaling 440 subjects). The corresponding required number of events is 308 in Group 1 and 220 in Group 2 (totaling 528 events).

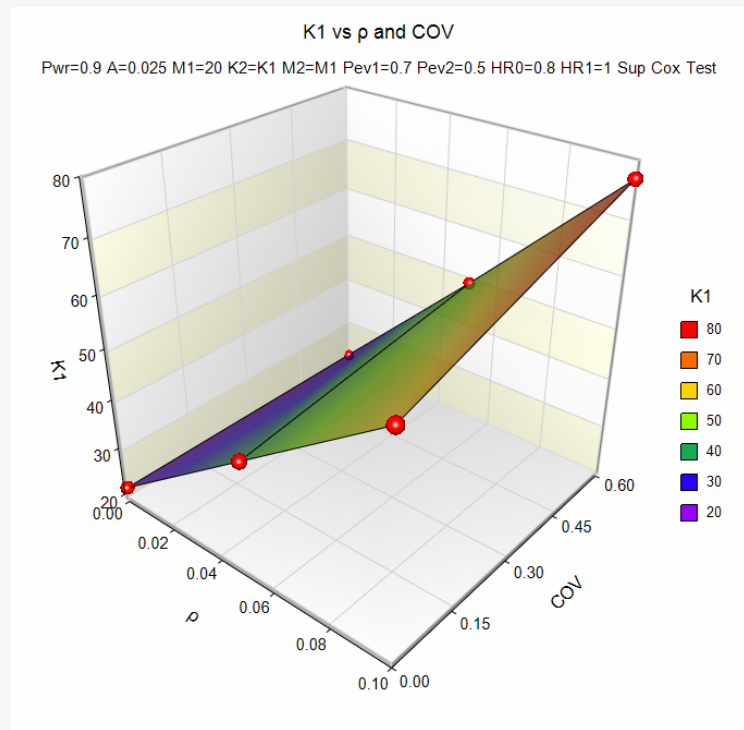
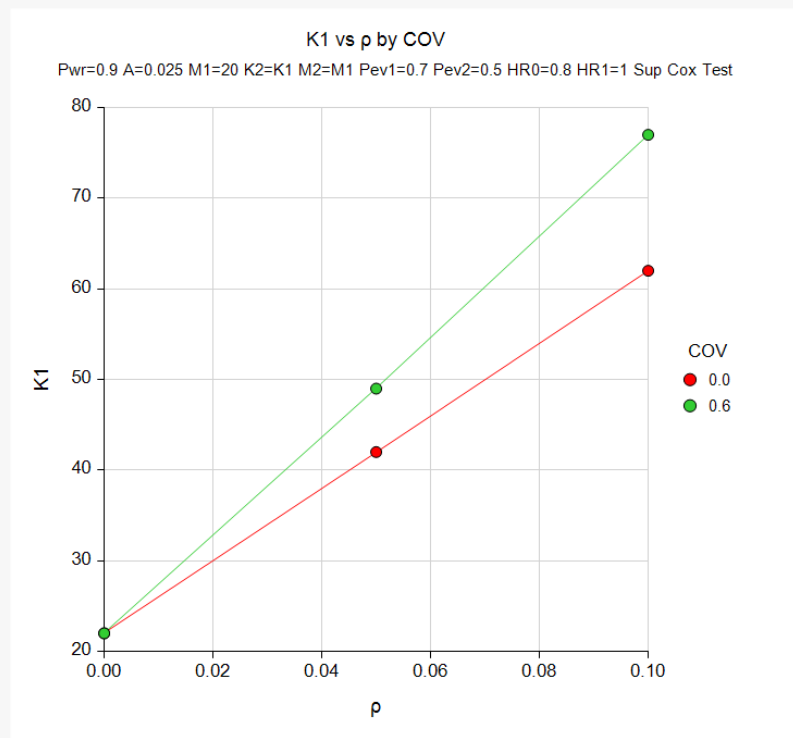
References

- 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.
- Donner, A. and Klar, N. 2000. Design and Analysis of Cluster Randomization Trials in Health Research. Arnold. London.
- 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 sample size for each of the scenarios. Note each scenario requires three report lines.

Plots Section

Plots



These plots show the relationship between the number of clusters, COV, and ICC.

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 (**Superiority by a Margin Tests for Two Survival Curves using Cox's Proportional Hazards Model**) to produce the results for the following example.

The parameters for this procedure are as follows: $M = 2$, $COV = 0.6$, $\rho = 0.05$, $alpha = 0.025$, $power = 0.8$, $HR0 = 0.8$, $HR1 = 0.6$, $Pev1 = 0.8$; and $Pev2 = 0.8$. Higher hazards are worse. Solve for sample size using equal allocation. Note that the parameters COV , M , and ρ relate to cluster-randomization and so they are not used in this procedure.

The **Superiority by a Margin Tests for Two Survival Curves using Cox's Proportional Hazards Model** procedure is set up as follows.

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).....	0.8
Pev2 (Event Probability in Group 2).....	Pev1
HR1 (Actual Hazard Ratio)	0.6
HR0 (Superiority Hazard Ratio)	0.8

This set of options generates the following report.

Numeric Results												
Solve For: Sample Size												
Groups: 1 = Control, 2 = Treatment												
Hypotheses: H0: HR ≥ HR0 vs. Ha: HR < HR0												
Power	Sample Size			Percent Group 1 %N1	Number of Events			Hazard Ratio		Event Probability		Alpha
	N	N1	N2		E	E1	E2	Superiority HR0	Actual HR1	Pev1	Pev2	
0.90009	635	317	318	50	508	253.6	254.4	0.8	0.6	0.8	0.8	0.025

The result is $N1 = 317$ and $N2 = 318$. Since equal group sizes is desired, we will increase $N1$ to 318.

Next, we calculate DE by hand. The formula for DE is

$$DE = 1 + \{[CV^2 + 1]\bar{M} - 1\}\rho$$

Plugging in we will now set $COV = 0.6$ and $\rho = 0.05$. the above values for COV and ρ

$$DE = 1 + \{[0.36 + 1]2 - 1\}(0.05) = 1.086$$

Multiplying 1.086 times 636 gives 690.696 which we round up to 692 since average cluster size was set at 2. Finally, we divide 692 by 4 to obtain the number of clusters per group as 173.

The next step is to run the CR routine and determine if the number of clusters match.

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	Number of Clusters
Higher Hazards Are	Worse (H1: HR < HR0)
Power.....	0.90
Alpha.....	0.025
M1 (Average Cluster Size).....	2
K2 (Number of Clusters)	K1
M2 (Average Cluster Size).....	M1
COV of Cluster Sizes	0.6
Pev1 (Probability of a Control Event).....	0.8
Pev2 (Probability of a Treatment Event)	Pev1
HR0 (Superiority Hazard Ratio)	0.8
HR1 (Actual Hazard Ratio)	0.6
ρ (Intraclass Correlation)	0.05

Output

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

Numeric Results											
Solve For:		Number of Clusters									
Groups:		1 = Control, 2 = Treatment									
Higher Hazards Are:		Worse									
Hypotheses:		H0: HR \geq HR0 vs. H1: HR < HR0									
Group	Power	Sample Size Ni	Number of Events Ei	Number of Clusters Ki	Cluster Size		Probability of an Event Pevi	Hazard Ratio		Intraclass Correlation ρ	Alpha
					Average Mi	COV		Superiority HR0	Actual HR1		
1: Control		346	300.6	173	2	0.6	0.8			0.05	
2: Treatment	0.90107	346	300.6	173	2	0.6	0.8	0.8	0.6	0.05	0.025
Total		692	601.2	346							

PASS also calculates $K1 = K2 = 173$ which validates the procedure.