

## Chapter 625

# Equivalence 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 equivalence 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 Lokhnygina (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  $z$

## Equivalence Tests for Two Survival Curves using Cox's Proportional Hazards Model in a Cluster-Randomized Design

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

$HR_0$  and  $1 / HR_0$  are the upper and lower equivalence limits of the hazard ratio

## 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 yield this conclusion when the null hypothesis is rejected is

$$H_0: |\log(HR)| \geq \log(HR_0) \quad \text{vs.} \quad H_1: |\log(HR)| < \log(HR_0), \quad \text{assuming } HR_0 > 1$$

## Test Statistic

It can be shown that the test of  $b$ , the regression coefficient representing the group, based on the partial likelihood method of Cox (1972) coincides with the common logrank test statistic. When testing equivalence, you can use  $b$  from a Cox regression or calculate the modified logrank statistic as follows

## Logrank Test

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

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

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

where  $K$  is the number of deaths,  $Y_{ij}$  is the number of subjects at risk just prior to the  $j^{\text{th}}$  observed event in the  $i^{\text{th}}$  group, and  $I_k$  is a binary variable indicating whether the  $k^{\text{th}}$  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  $(\log(HR) - \log(HR_0))\sqrt{P_1 P_2 dN}$  and unit variance and the distribution of  $L_L$  is approximately normal with mean  $(\log(HR) - \log(1/HR_0))\sqrt{P_1 P_2 dN}$  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,  $N_2 = N(P_2)$

## Equivalence Tests for Two Survival Curves using Cox's Proportional Hazards Model in a Cluster-Randomized Design

$Pev_1$  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,  $d = Pev_1P_1 + Pev_2P_2$

$HR$  is the observed hazard ratio

$HR_0$  is the upper equivalence 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  $\rho$  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 significance level of  $\alpha$  to detect a hazard ratio of  $HR (h_2 / h_1)$  is given by

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

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

$$\bar{M} = \frac{K_1M_1 + K_2M_2}{K_1 + K_2}$$

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

## Equivalence Tests for Two Survival Curves using Cox's Proportional Hazards Model in a Cluster-Randomized Design

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

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

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

## Example 1 – Finding Sample Size

A company is planning an equivalence trial using a cluster-randomized, two-group, equal sample allocation design to compare the survivability of a new treatment with that of the current treatment. The upper equivalence limit of the hazard ratio is 1.25. The desired power is 0.9 and significance level is 0.05. HR1 is set at 1. The probability of observing an event is 0.8 in the control group and 0.5 in the treatment group. The intraclass correlation coefficient is 0.05. The average number of subjects per cluster between 5 and 20 with a coefficient of variation of 0.65. The required sample size (number of clusters) is desired assuming that the number of clusters in each group is the same.

### 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**  
 Power..... **0.90**  
 Alpha..... **0.05**  
 M1 (Average Cluster Size)..... **5 10 15 20**  
 K2 (Number of Clusters) ..... **K1**  
 M2 (Average Cluster Size)..... **M1**  
 COV of Cluster Sizes ..... **0.65**  
 Pev1 (Probability of a Control Event)..... **0.8**  
 Pev2 (Probability of a Treatment Event) ..... **0.6**  
 HR0 (Upper Equivalence Hazard Ratio) ..... **1.25**  
 HR1 (Actual Hazard Ratio) ..... **1**  
 ρ (Intraclass Correlation) ..... **0.05**

#### Plots Tab

---

Open the X-Y Plot Format..... **Click on the 2D Plot Format button**  
 On the Titles Tab ..... **Change the Top 2 Titles font size to 7**

## 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  
 Hypotheses:  $H_0: HR \leq 1 / HR_0$  or  $HR \geq HR_0$  vs.  $H_1: 1 / HR_0 < HR < HR_0$

Group	Power	Sample Size Ni	Number of Events Ei	Number of Clusters Ki	Cluster Size		Probability of an Event Pevi	Hazard Ratio			Intracluster Correlation $\rho$	Alpha
					Average Mi	COV		Equivalence		Actual HR1		
								Lower 1/HR0	Upper HR0			
1: Control		815	851.3	163	5	0.65	0.8				0.05	
2: Treatment	0.90176	815	638.5	163	5	0.65	0.6	0.8	1.25	1	0.05	0.05
Total		1630	1489.7	326								
1: Control		1040	1382.2	104	10	0.65	0.8				0.05	
2: Treatment	0.90273	1040	1036.6	104	10	0.65	0.6	0.8	1.25	1	0.05	0.05
Total		2080	2418.8	208								
1: Control		1260	2033.0	84	15	0.65	0.8				0.05	
2: Treatment	0.90203	1260	1524.8	84	15	0.65	0.6	0.8	1.25	1	0.05	0.05
Total		2520	3557.8	168								
1: Control		1480	2809.0	74	20	0.65	0.8				0.05	
2: Treatment	0.90154	1480	2106.8	74	20	0.65	0.6	0.8	1.25	1	0.05	0.05
Total		2960	4915.8	148								

- 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 upper equivalence limit of the hazard ratio. The lower equivalence limit of the hazard ratio is 1 / HR0.
- HR1 The hazard ratio of the treatment group and the control group.  $HR_1 = h_2 / h_1$ . The subscript '1' indicates that this hazard ratio value is used for the alternative hypothesis, H1.
- $\rho$  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 will be used to test whether the Group 2 (treatment) hazard rate is equivalent to the Group 1 (control) hazard rate, with lower and upper equivalence limits for the hazard ratio of 0.8 and 1.25 ( $H_0: HR \leq 0.8$  or  $HR \geq 1.25$  versus  $H_1: 0.8 < HR < 1.25$ ,  $HR = h_2 / h_1$ ). The comparison will be made using two one-sided, two-sample Cox proportional hazards regression coefficient tests with an overall Type I error rate ( $\alpha$ ) of 0.05. It is anticipated that the probability of observing an event during the course of the study is 0.8 for each member of the control group and 0.6 for each member of the treatment group. The coefficient of variation of the cluster sizes within a group is assumed to be 0.65. The intracluster correlation coefficient of responses within a particular cluster is assumed to be 0.05. 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 1, with average cluster sizes of 5 subjects per cluster in Group 1 and 5 subjects per cluster in Group 2, with 90% power, the number of needed clusters is 163 for Group 1 (control) (totaling 815 subjects) and 163 for Group 2 (treatment) (totaling 815 subjects). The corresponding required number of events is 851.3 in Group 1 and 638.5 in Group 2 (totaling 1489.7 events).

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

**References**

---

Blackwelder, W.C. 1998. 'Equivalence Trials.' In Encyclopedia of Biostatistics, John Wiley and Sons. New York. Volume 2, 1367-1372.

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.

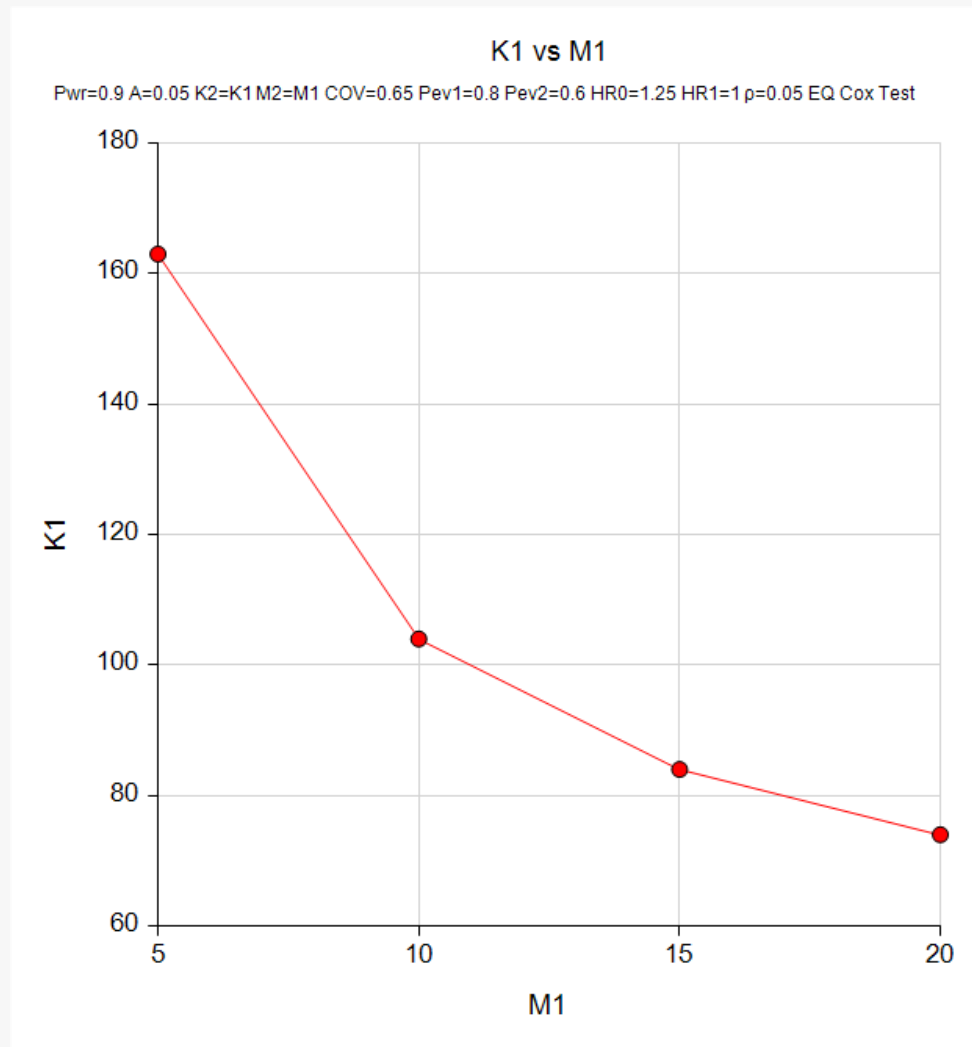
Schoenfeld, David A. 1983. 'Sample Size Formula for the Proportional-Hazards Regression Model', Biometrics, Volume 39, Pages 499-503.

---

This report shows the sample size for each of the scenarios. Note each scenario requires three report lines.

**Plots Section**

**Plots**



This plot shows the relationship between the number of clusters and the average cluster size.

## 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.

The parameter values will be set as in Example 1 except that  $M1 = M2 = 4$  and  $COV = 0.6$ .

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

Design Tab	
Solve For .....	<b>Sample Size</b>
Power.....	<b>0.90</b>
Alpha.....	<b>0.05</b>
Group Allocation .....	<b>Equal (N1 = N2)</b>
Pev1 (Event Probability in Group 1).....	<b>0.8</b>
Pev2 (Event Probability in Group 2).....	<b>0.6</b>
HR1 (Actual Hazard Ratio) .....	<b>1.0</b>
HR0 (Equivalence Hazard Ratio).....	<b>1.25</b>

This set of options generates the following report.

Numeric Results													
Solve For: <b>Sample Size</b>													
Groups: 1 = Control, 2 = Treatment													
Hypotheses: H0: HR ≤ 1/HR0 or HR ≥ HR0 vs. Ha: 1/HR0 < HR < HR0													
Power	Sample Size			Percent Group 1 %N1	Number of Events			Hazard Ratio		Event Probability			Alpha
	N	N1	N2		E	E1	E2	Equivalence HR0	Actual HR1	Pev1	Pev2		
0.90001	1242	621	621	50	869.4	496.8	372.6	1.25	1	0.8	0.6	0.05	

The result is  $N1 = 621$  and  $N2 = 621$ .

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  $\rho$

$$DE = 1 + \{[0.36 + 1]4 - 1\}(0.05) = 1.222$$

Multiplying 1.222 times 1242 gives 1517.724 which we round up to 1540 since the average cluster size was set at 4 and 1540 is the next integer after 1517.7 that is exactly divisible by 4. Finally, we divide 1520 by 8 to obtain the number of clusters per group as 190.

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 .....	<b>Number of Clusters</b>
Power.....	<b>0.90</b>
Alpha.....	<b>0.05</b>
M1 (Average Cluster Size).....	<b>4</b>
K2 (Number of Clusters) .....	<b>K1</b>
M2 (Average Cluster Size).....	<b>M1</b>
COV of Cluster Sizes.....	<b>0.6</b>
Pev1 (Probability of a Control Event).....	<b>0.8</b>
Pev2 (Probability of a Treatment Event) .....	<b>0.6</b>
HR0 (Upper Equivalence Hazard Ratio) .....	<b>1.25</b>
HR1 (Actual Hazard Ratio) .....	<b>1</b>
$\rho$ (Intraclass Correlation) .....	<b>0.05</b>

## Output

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

Numeric Results												
Solve For: <b>Number of Clusters</b>												
Groups: 1 = Control, 2 = Treatment												
Hypotheses: $H_0: HR \leq 1 / HR_0$ or $HR \geq HR_0$ vs. $H_1: 1 / HR_0 < HR < HR_0$												
Group	Power	Sample Size Ni	Number of Events Ei	Number of Clusters Ki	Cluster Size		Probability of an Event Pevi	Hazard Ratio			Intraclass Correlation $\rho$	Alpha
					Average Mi	COV		Equivalence		Actual HR1		
								Lower 1/HR0	Upper HR0			
1: Control		760	743.0	190	4	0.6	0.8				0.05	
2: Treatment	0.90052	760	557.2	190	4	0.6	0.6	0.8	1.25	1	0.05	0.05
Total		1520	1300.2	380								

**PASS** also calculates  $K1 = K2 = 190$  which validates the procedure.