

# Survival Analysis: Non Proportional Hazards Model

*Nesma Magdi*

*October 21, 2019*

## Ovarian Cancer:

Survival in a randomised trial comparing two treatments for ovarian cancer. Format

Treatment of patients with advanced ovarian carcinoma (stages IIIB and IV) using either cyclophosphamide alone (1 g/m<sup>2</sup>) or cyclophosphamide (500 mg/m<sup>2</sup>) plus adriamycin (40 mg/m<sup>2</sup>) by iv injection every 3 weeks each produced partial regression in approximately one third of the patients.

```
data("ovarian")
attach(ovarian)

Status <- ifelse(fustat == 0, "Death", "Censored")
Progression <- ifelse(resid.ds == 1, "No Progression", "Progression")
Treatment <- ifelse(rx == 1, "Endoxan alone", "Endoxan + Adria")
Performance <- ifelse(ecog.ps == 1, "High", "Low")
dummy_Age <- ifelse(age >= 50, "Old", "Young")

ex <- data.frame(Status, Progression, Treatment, Performance, dummy_Age)

Ovarian <- cbind(ovarian, ex)

Ovarian$dummy_Age <- factor(Ovarian$dummy_Age)
Ovarian$Status <- factor(Ovarian$Status)
Ovarian$Progression <- factor(Ovarian$Progression)
Ovarian$Treatment <- factor(Ovarian$Treatment)
Ovarian$Performance <- factor(Ovarian$Performance)

write.csv(Ovarian, file="ovarian.csv")
ovarian <- read.csv(choose.files(), header = T)
attach(ovarian)
```

## Description:

```
##              Description
## ftime      Survival or Censoring time
## fustat      Censoring status
## age         Age in years
## resid.ds    Residual disease present
## rx          Treatment Group
## ecog.ps     ECOG Performance Status
```

```
summary(ovarian)
```

```
##      X      ftime      fustat      age
## Min.   : 1.00   Min.    : 59.0   Min.    :0.0000   Min.    :38.89
```

```
## 1st Qu.: 7.25 1st Qu.: 368.0 1st Qu.:0.0000 1st Qu.:50.17
## Median :13.50 Median : 476.0 Median :0.0000 Median :56.85
## Mean :13.50 Mean : 599.5 Mean :0.4615 Mean :56.17
## 3rd Qu.:19.75 3rd Qu.: 794.8 3rd Qu.:1.0000 3rd Qu.:62.38
## Max. :26.00 Max. :1227.0 Max. :1.0000 Max. :74.50
## resid.ds rx ecog.ps Status
## Min. :1.000 Min. :1.0 Min. :1.000 Censord:12
## 1st Qu.:1.000 1st Qu.:1.0 1st Qu.:1.000 Death :14
## Median :2.000 Median :1.5 Median :1.000
## Mean :1.577 Mean :1.5 Mean :1.462
## 3rd Qu.:2.000 3rd Qu.:2.0 3rd Qu.:2.000
## Max. :2.000 Max. :2.0 Max. :2.000
## Progression Treatment Performance dummy_Age
## No Progression:11 Endoxan + Adria:13 High:14 Old :20
## Progression :15 Endoxan alone :13 Low :12 Young: 6
##
##
##
##
```

```
Count <- table(Status, Treatment)
Percent <- prop.table(Count)*100

persent <- as.data.frame(cbind(Count, Percent))
persent
```

```
## Endoxan + Adria Endoxan alone Endoxan + Adria Endoxan alone
## Censord 5 7 19.23077 26.92308
## Death 8 6 30.76923 23.07692
```

```
Tret_Per <- as.data.frame(ovarian %>%
  summary_factorlist("Status", "Treatment",
    add_dependent_label = T, p=T, cont_cut = 1))

Tret_Per
```

```
## Dependent: Status Censord Death p
## 1 Treatment Endoxan + Adria 5 (38.5) 8 (61.5) 0.431
## 2 Endoxan alone 7 (53.8) 6 (46.2)
```

## Survival Analysis:

\*Kaplan-Meier Estimator:

- Pointwise CI S(t):

H0: No difference in Survival between two or more independent groups.

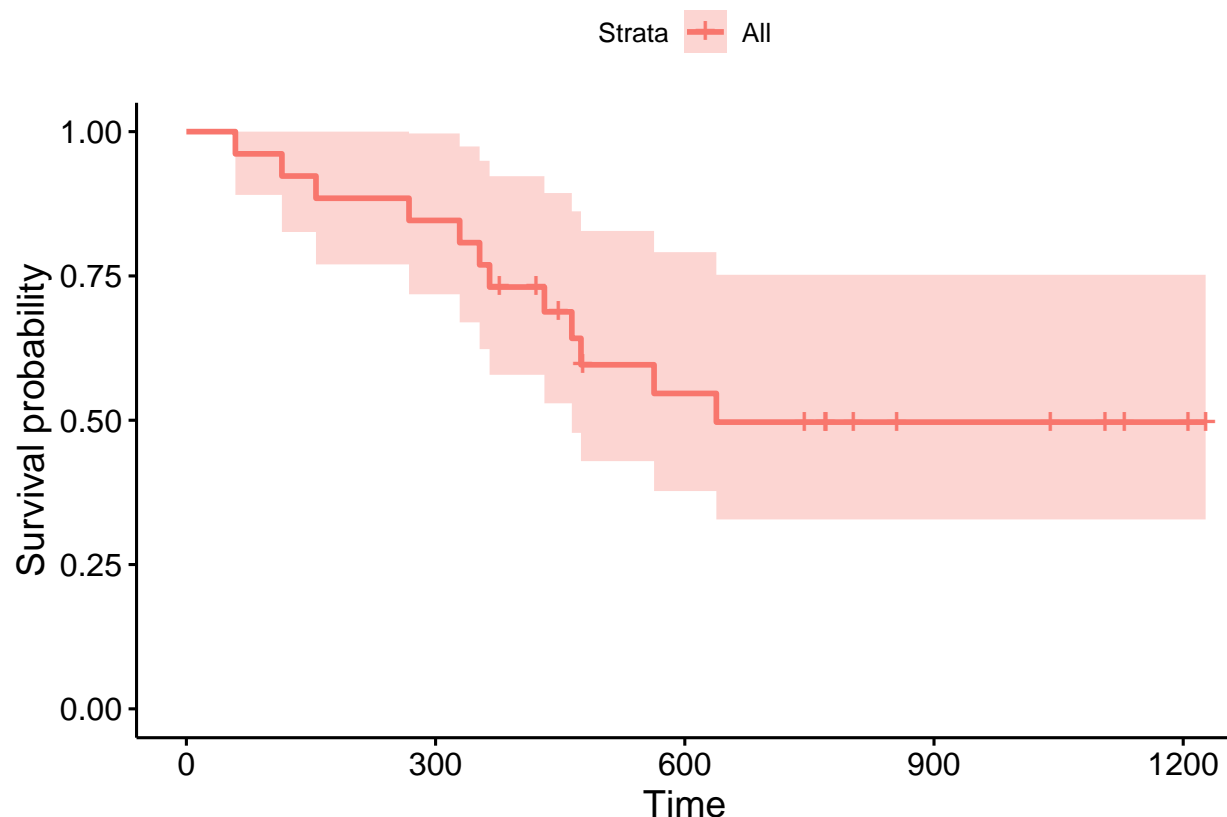
```
surviv <- survfit(Surv(futime, fustat) ~ 1, data = ovarian)
print(surviv, print.rmean = TRUE)
```

```
## Call: survfit(formula = Surv(futime, fustat) ~ 1, data = ovarian)
##
##           n      events      *rmean *se(rmean)      median      0.95LCL
##        26.0        12.0        794.0        91.5        638.0        464.0
##      0.95UCL
##          NA
##      * restricted mean with upper limit = 1227
```

```
summary(surviv)
```

```
## Call: survfit(formula = Surv(futime, fustat) ~ 1, data = ovarian)
##
##   time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    59     26      1    0.962  0.0377    0.890    1.000
##   115     25      1    0.923  0.0523    0.826    1.000
##   156     24      1    0.885  0.0627    0.770    1.000
##   268     23      1    0.846  0.0708    0.718    0.997
##   329     22      1    0.808  0.0773    0.670    0.974
##   353     21      1    0.769  0.0826    0.623    0.949
##   365     20      1    0.731  0.0870    0.579    0.923
##   431     17      1    0.688  0.0919    0.529    0.894
##   464     15      1    0.642  0.0965    0.478    0.862
##   475     14      1    0.596  0.0999    0.429    0.828
##   563     12      1    0.546  0.1032    0.377    0.791
##   638     11      1    0.497  0.1051    0.328    0.752
```

```
ggsurvplot(surviv, data = ovarian, pval = TRUE)
```



The 95% confidence interval of survival time is (464, NA); NA in this case means infinity. A 95% upper confidence limit of NA/infinity is common in survival analysis due to the fact that the data is skewed.

\*Comparing Survival Curves:

```
surviv1 <- survfit(Surv(futime, fustat) ~ Treatment, data = ovarian)
summary(surviv1)
```

```
## Call: survfit(formula = Surv(futime, fustat) ~ Treatment, data = ovarian)
```

```
##
```

```
## Treatment=Endoxan + Adria
```

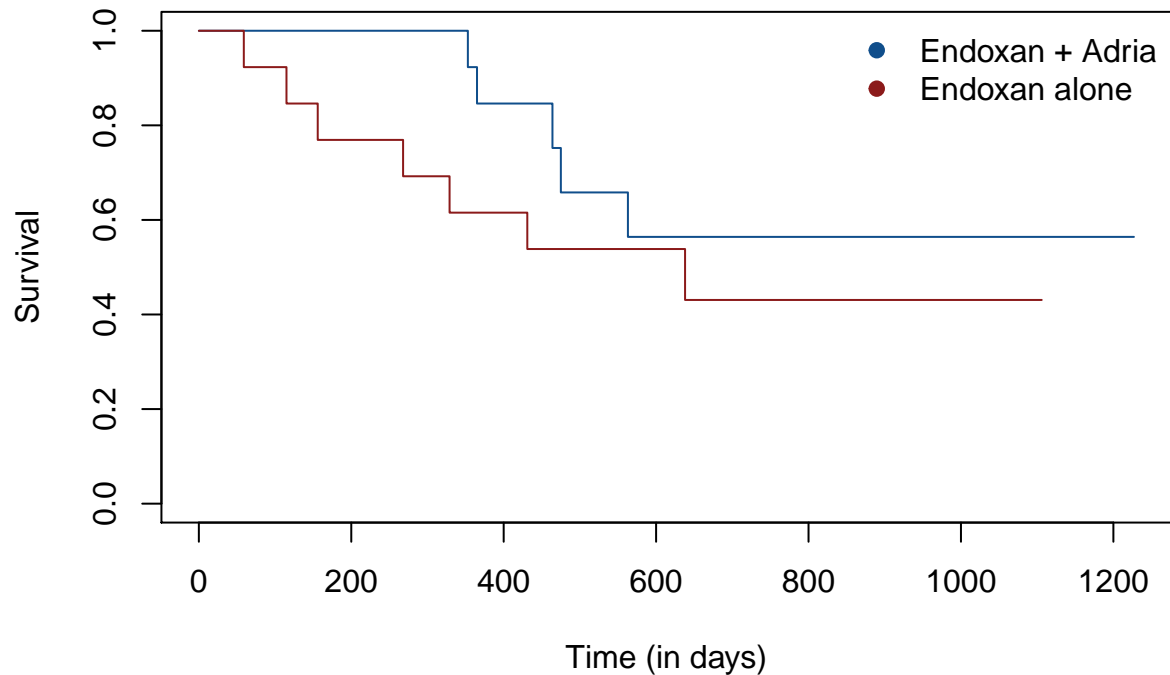
##	time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
##	353	13	1	0.923	0.0739	0.789	1.000
##	365	12	1	0.846	0.1001	0.671	1.000
##	464	9	1	0.752	0.1256	0.542	1.000
##	475	8	1	0.658	0.1407	0.433	1.000
##	563	7	1	0.564	0.1488	0.336	0.946

```
##
```

```
## Treatment=Endoxan alone
```

##	time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
##	59	13	1	0.923	0.0739	0.789	1.000
##	115	12	1	0.846	0.1001	0.671	1.000
##	156	11	1	0.769	0.1169	0.571	1.000
##	268	10	1	0.692	0.1280	0.482	0.995
##	329	9	1	0.615	0.1349	0.400	0.946
##	431	8	1	0.538	0.1383	0.326	0.891
##	638	5	1	0.431	0.1467	0.221	0.840

```
plot(surviv1, ylab = "Survival", xlab = "Time (in days)",
     col = c("dodgerblue4", "firebrick4"), marktime = TRUE, pval = TRUE)
legend("topright", col = c("dodgerblue4", "firebrick4"),
      legend = levels(ovarian$Treatment), bty = "n", pch = 19)
```

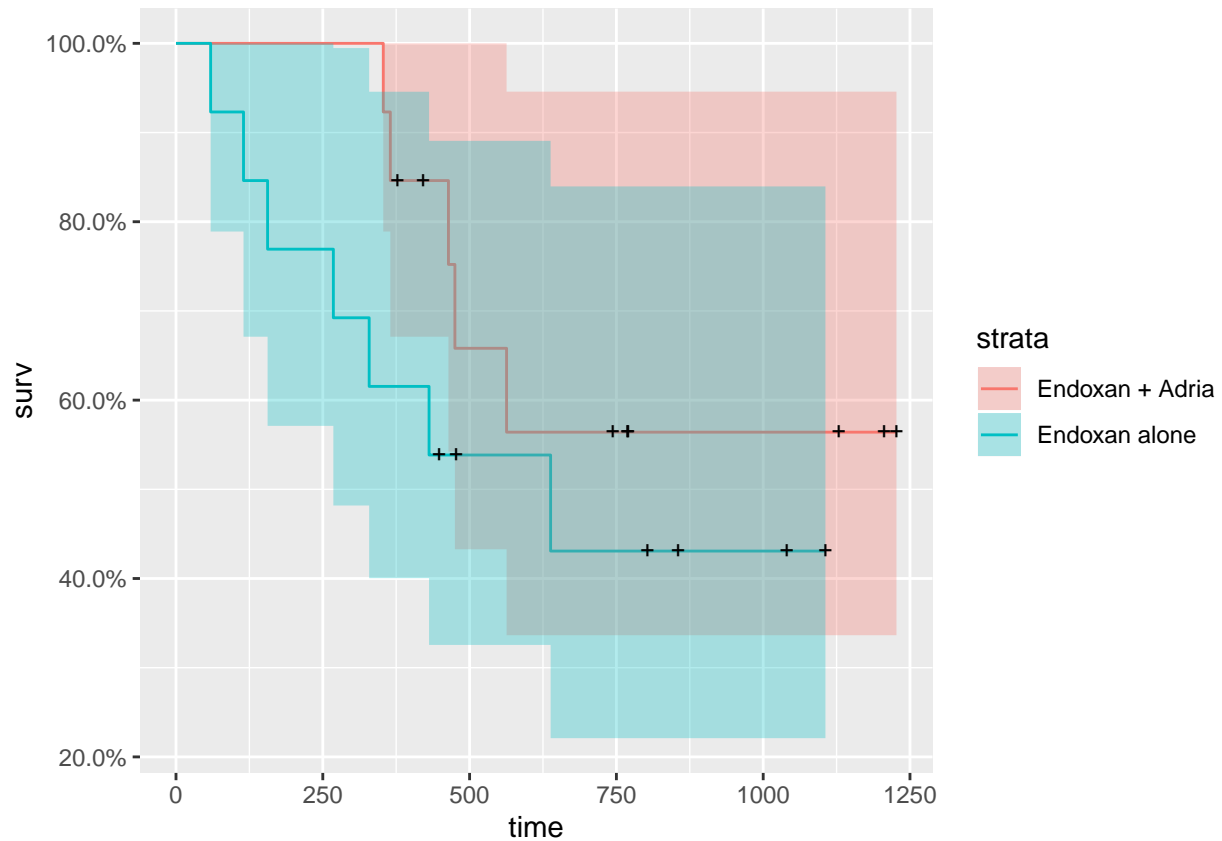


\*Comparing Survival:

```
surviv2 <- survdiff(Surv(futime, fustat) ~ Treatment, data = ovarian)
surviv2
```

```
## Call:
## survdiff(formula = Surv(futime, fustat) ~ Treatment, data = ovarian)
##
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## Treatment=Endoxan + Adria 13      5    6.77    0.461    1.06
## Treatment=Endoxan alone  13      7    5.23    0.596    1.06
##
## Chisq= 1.1  on 1 degrees of freedom, p= 0.3
```

```
autoplot(surviv1)
```



H0: rejected (p-value = 0.3).

\*Post-hoc analysis:

```
surviv_post <- pairwise_survdif(Surv(futime, fustat) ~ Treatment, data = ovarian)
```

```
surviv_post
```

```
##
## Pairwise comparisons using Log-Rank test
##
## data: ovarian and Treatment
##
##           Endoxan + Adria
## Endoxan alone 0.3
##
## P value adjustment method: BH
```

=====

## Cox Proportional Hazards:

-Hazard functions for each individual should be strictly parallel & Hazard ratio is constant over time.

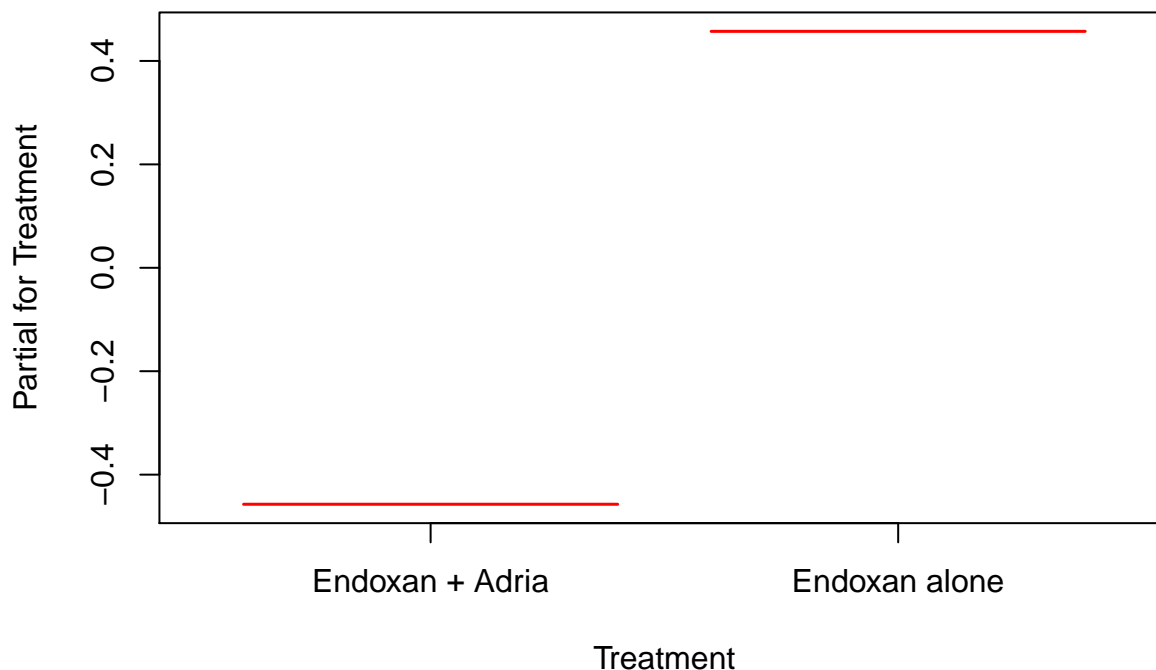
- HR = 1: No effect. HR > 1: Increase in Hazards. HR < 1: Reduced in Hazards.

```
Model2 <- coxph(Surv(futime, fustat) ~ Treatment+ age +
                Progression + Performance, data = ovarian)
```

```
summary(Model2)
```

```
## Call:
## coxph(formula = Surv(futime, fustat) ~ Treatment + age + Progression +
##       Performance, data = ovarian)
##
##      n= 26, number of events= 12
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## TreatmentEndoxan alone 0.91450   2.49553  0.65332  1.400  0.16158
## age                    0.12481   1.13294  0.04689  2.662  0.00777 **
## ProgressionProgression 0.82619   2.28459  0.78961  1.046  0.29541
## PerformanceLow         0.33621   1.39964  0.64392  0.522  0.60158
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## TreatmentEndoxan alone   2.496    0.4007    0.6935    8.980
## age                     1.133    0.8827    1.0335    1.242
## ProgressionProgression   2.285    0.4377    0.4861   10.738
## PerformanceLow          1.400    0.7145    0.3962    4.945
##
## Concordance= 0.807 (se = 0.068 )
## Likelihood ratio test= 17.04 on 4 df,  p=0.002
## Wald test              = 14.25 on 4 df,  p=0.007
## Score (logrank) test = 20.81 on 4 df,  p=3e-04
```

```
termplot(Model2, terms = "Treatment")
```



\*Interpretation:

1- Estimated Hazard Ratio for Young Ages ( $\leq 50$ ) Versus Old Ages ( $>50$ ) is 1.133 with confidence interval CI of (1.0335 - 1.242). Estimated Hazard Ratio for Old Ages Versus Young Ages is 0.8827 with confidence interval CI of (1.0335 - 1.242).

## Test Significance of variance or interaction term in the Model:

\*Wald Test (Maximum Likelihood Estimates):

```
summary(Model2)$coef
```

```
##               coef exp(coef)  se(coef)      z
## TreatmentEndoxan alone 0.9144999  2.495527 0.65331563 1.3997827
## age                    0.1248131  1.132937 0.04689036 2.6618066
## ProgressionProgression 0.8261864  2.284590 0.78961106 1.0463208
## PerformanceLow         0.3362117  1.399635 0.64392306 0.5221302
##               Pr(>|z|)
## TreatmentEndoxan alone 0.161578399
## age                    0.007772252
## ProgressionProgression 0.295412966
## PerformanceLow         0.601579671
```



```
wald <- summary(Model2)$coef[4, 1]/summary(Model2)$coef[4, 3]
pvalue <- wald * pnorm(wald, lower.tail = FALSE)
pvalue
```

```
## [1] 0.1570515
```

\*p-value of Treatment, Progression, ECOG Performance status is > 0.05. Fail to Reject Null Hypothesis.

According to p-value of the test, Fail to Reject null hypothesis (for the Treatment, Progression, ECOG Performance status variable) Thus, the model must not include this variable.

\*Likelihood Ratio Test (LRT):

LRT Statistics has better statistical properties. First, The reduced. Second, The full.

```
Model3 <- coxph(Surv(futime, fustat) ~ age, data = ovarian)
anova(Model3, Model2)
```

```
## Analysis of Deviance Table
## Cox model: response is Surv(futime, fustat)
## Model 1: ~ age
## Model 2: ~ Treatment + age + Progression + Performance
##      loglik  Chisq Df P(>|Chi|)
## 1 -27.838
## 2 -26.463 2.7497 3 0.4318
```

```
Model2$loglik
```

```
## [1] -34.98494 -26.46329
```

The First is the log likelihood of a model that contains none of the predictors, So we need the second one.

```
chi <- 2 * Model2$loglik[2] - Model3$loglik[2]
pvalue2 <- 1 - pchisq(chi, df = 3)
pvalue2
```

```
## [1] 0
```

In this case, using  $\alpha=0.05$  & testing the significance of Performance variable. We must remove it from the model.

=====

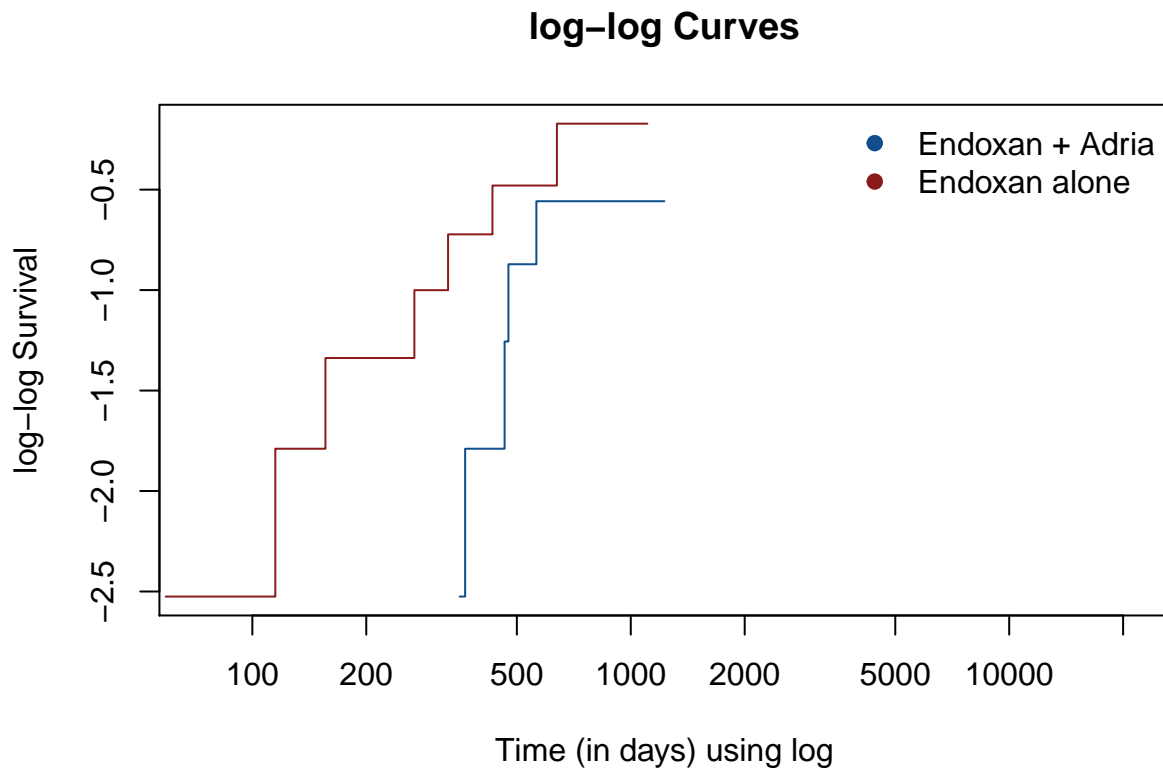
## Evaluate Proportional Hazard:

\*Note: If the graph of the hazards cross for two or more categories of the predictor of interest: PH assumption not met.

\*a- Graphical Approach:

```
Model4 <- survfit(Surv(futime, fustat) ~ Treatment, data = ovarian)

plot(Model4, fun = "cloglog", xlab = "Time (in days) using log", ylab = "log-log Survival",
      main = "log-log Curves", col = c("dodgerblue4", "firebrick4"))
legend("topright", col = c("dodgerblue4", "firebrick4"),
      legend = levels(ovarian$Treatment), bty = "n", pch = 19)
```



It seems that the proportional hazards assumption is violated as the log-log survival curves are not parallel.

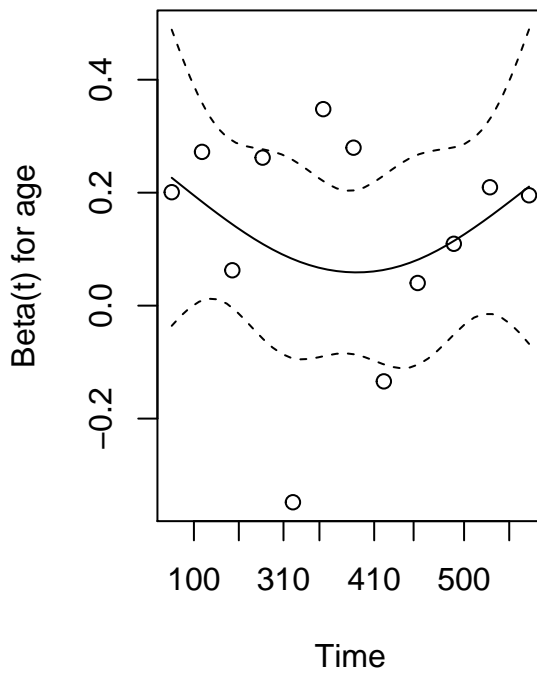
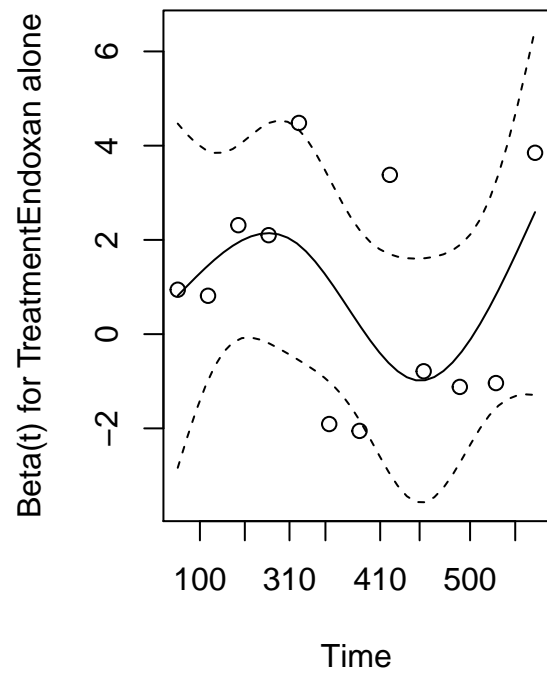
\*b- Schoenfeld Residuals:

Difference between the observed covariate and the expected given the risk set at that time.

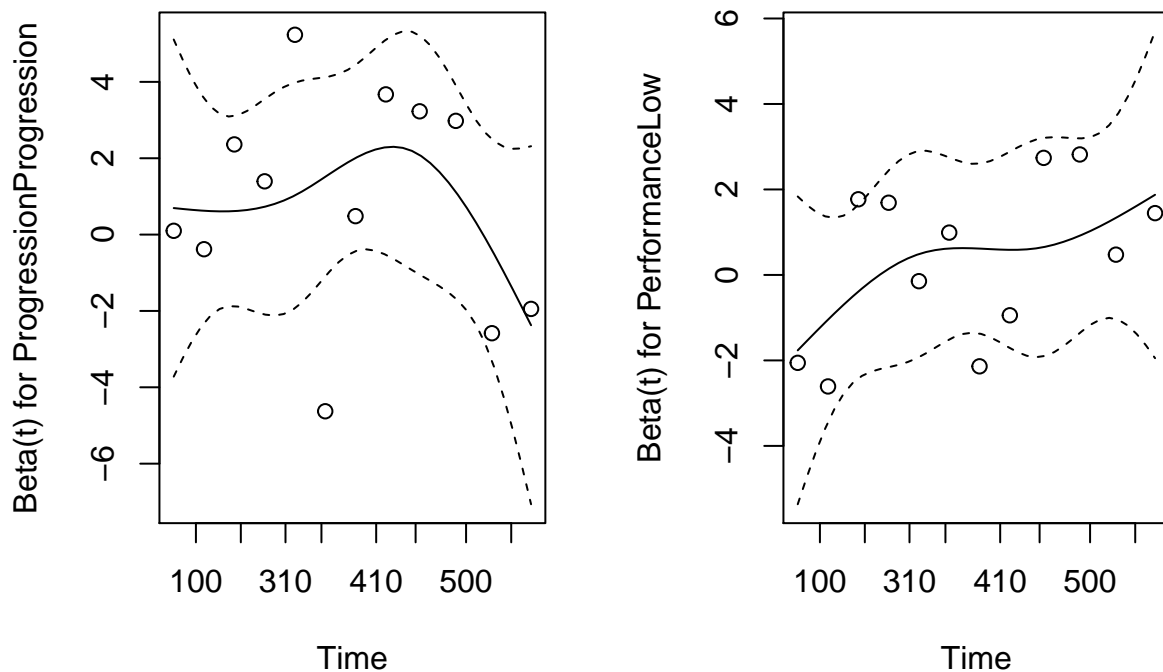
```
residuals_ov <- cox.zph(Model2)
residuals_ov
```

```
##               rho  chisq    p
## TreatmentEndoxan alone -0.1325 0.2001 0.655
## age                    -0.0399 0.0262 0.871
## ProgressionProgression -0.1417 0.2463 0.620
## PerformanceLow         0.4845 1.8819 0.170
## GLOBAL                 NA 3.3609 0.499
```

```
par(mfrow = c(1, 2))
plot(residuals_ov, var = 1)
plot(residuals_ov, var = 2)
```



```
plot(residuals_ov, var = 3)
plot(residuals_ov, var = 4)
```



It seems that the residuals not flat, not centered about zero..

## Non Proportional Hazard:

### Stratified Proportional Hazard Model (1- Treatment):

H0: The regression coefficient are assumed to be the same for each stratum.

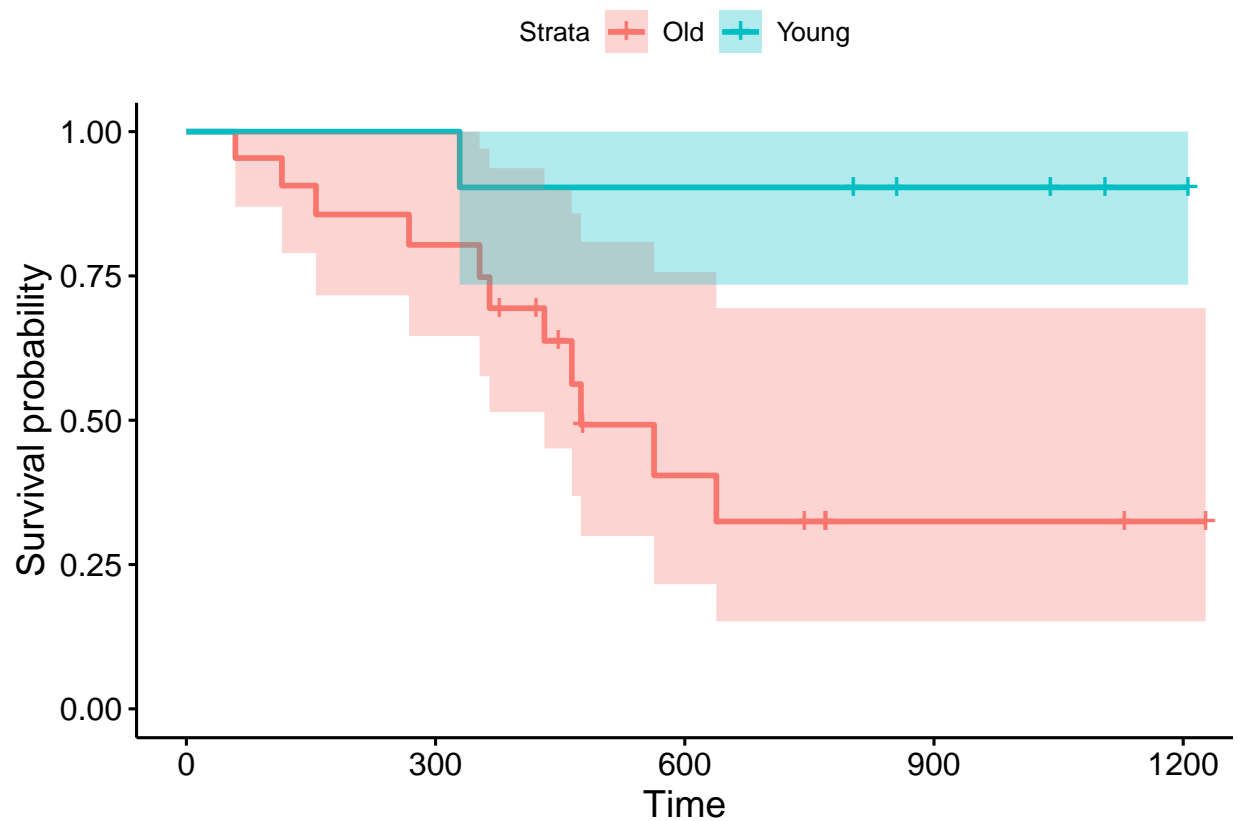
Ha: The baseline hazard functions may be different and completely unrelated.

```
Model15 <- coxph(Surv(futime, fustat) ~ Treatment + strata(dummy_Age), data = ovarian)
summary(Model15)
```

```
## Call:
## coxph(formula = Surv(futime, fustat) ~ Treatment + strata(dummy_Age),
##       data = ovarian)
##
##      n= 26, number of events= 12
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## TreatmentEndoxan alone 1.2421    3.4627  0.6134  2.025  0.0429 *
```

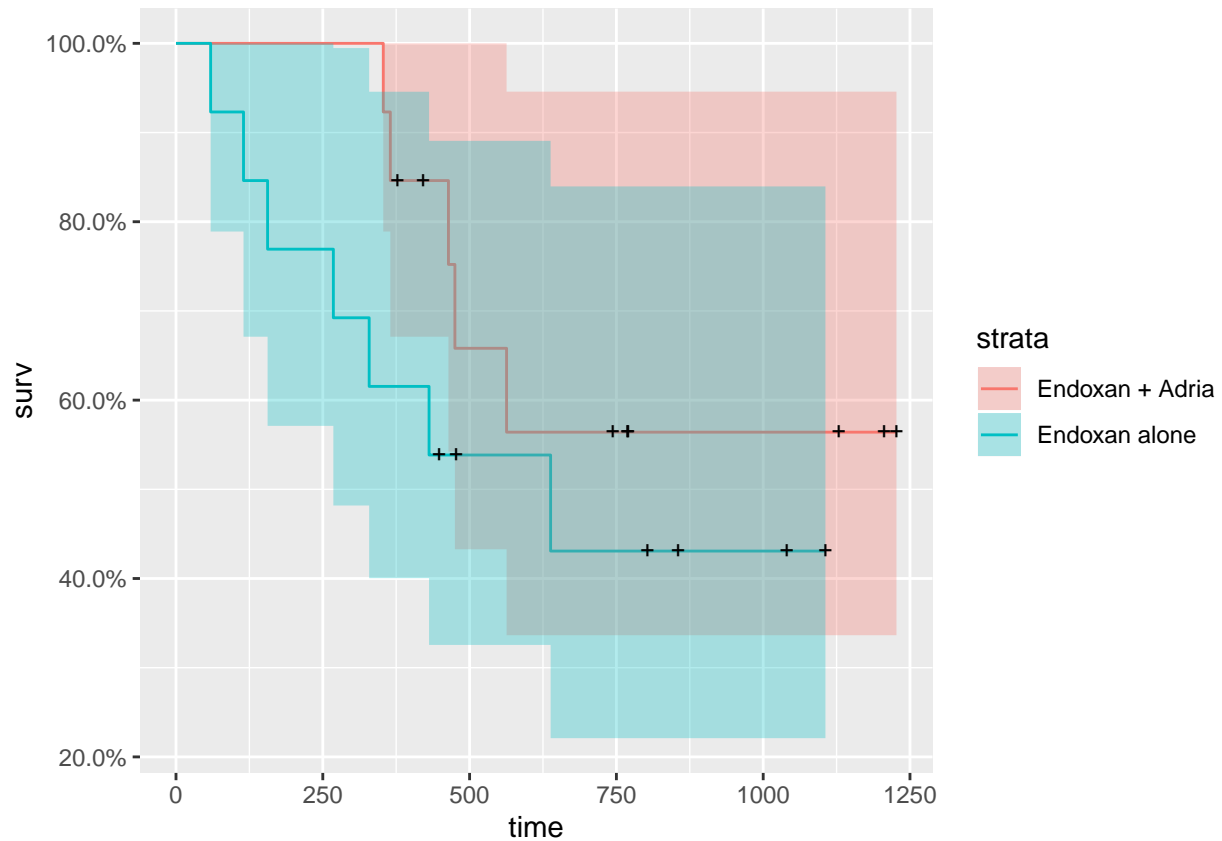
```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## TreatmentEndoxan alone    3.463    0.2888    1.041    11.52
##
## Concordance= 0.671  (se = 0.065 )
## Likelihood ratio test= 4.07  on 1 df,  p=0.04
## Wald test            = 4.1   on 1 df,  p=0.04
## Score (logrank) test = 4.54  on 1 df,  p=0.03
```

```
ggsurvplot(survfit(Model5), data = ovarian, conf.int = T)
```



p-value of Treatment Now (p-value = 0.0429) is Significant, not similar to previous model without stratification (p-value = 0.16158).

```
Fit2 <- survfit(Surv(futime, fustat) ~ factor(Treatment), data = ovarian)
autoplot(Fit2)
```



## Stratified Proportional Hazard Model (2- Progression):

H0: The regression coefficient are assumed to be the same for each stratum.

Ha: The baseline hazard functions may be different and completely unrelated.

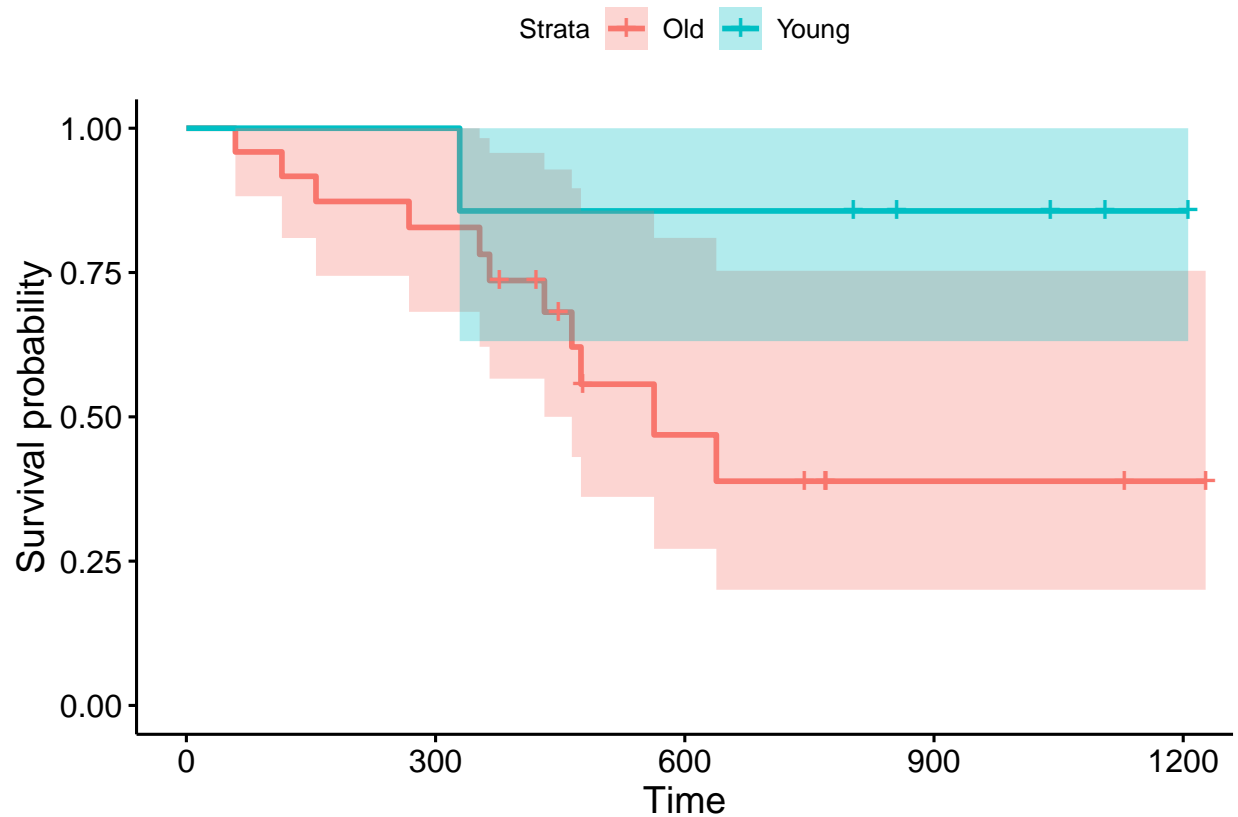
```
Model15_1 <- coxph(Surv(futime, fustat) ~ Progression + strata(dummy_Age), data = ovarian)
```

```
summary(Model15_1)
```

```
## Call:
## coxph(formula = Surv(futime, fustat) ~ Progression + strata(dummy_Age),
##       data = ovarian)
##
##      n= 26, number of events= 12
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## Progression 1.1842   3.2682  0.6771 1.749  0.0803 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## Progression 3.268      0.306    0.8668    12.32
```

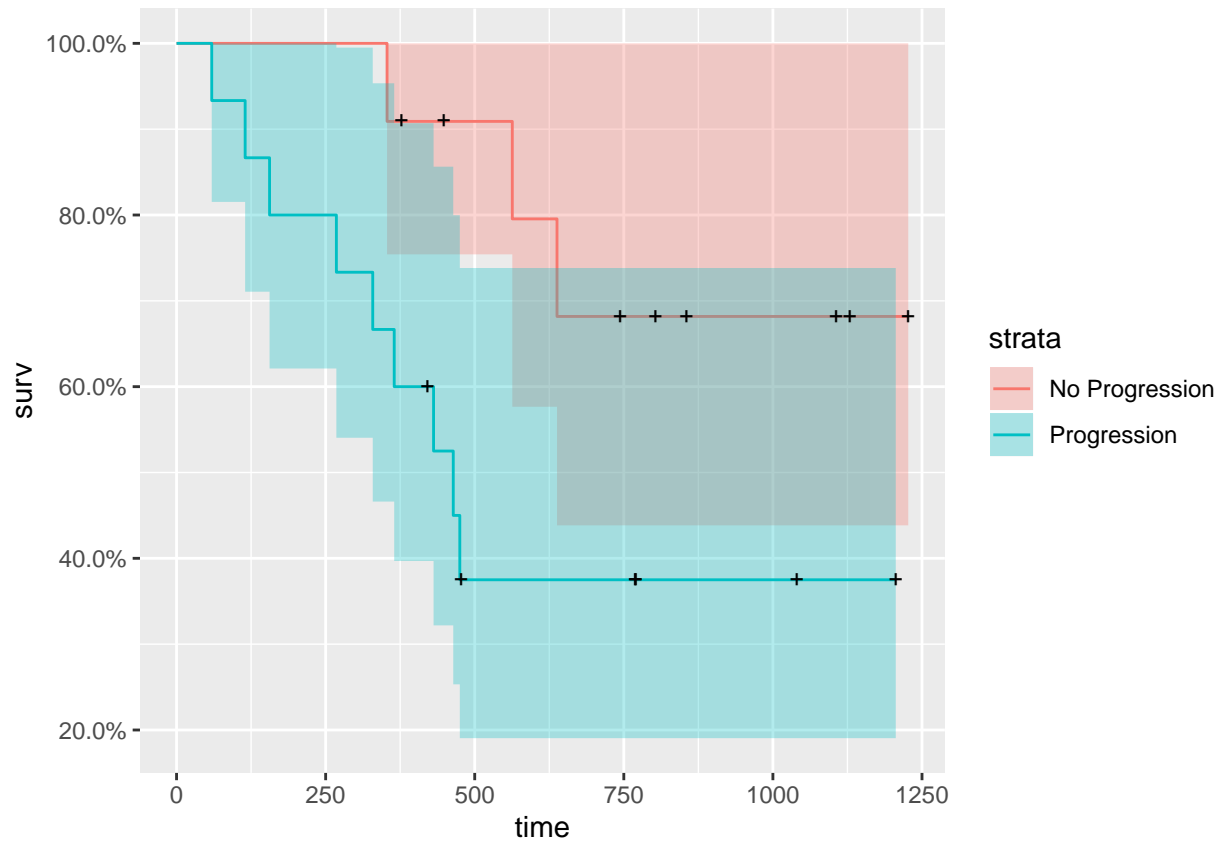
```
##
## Concordance= 0.661 (se = 0.062 )
## Likelihood ratio test= 3.54 on 1 df, p=0.06
## Wald test = 3.06 on 1 df, p=0.08
## Score (logrank) test = 3.4 on 1 df, p=0.07
```

```
ggsurvplot(survfit(Model5_1), data = ovarian, conf.int = T)
```



p-value of Treatment Now (p-value = 0.0803) is not Significant, but also not similar to previous model without stratification (p-value = 0.29541).

```
Fit3 <- survfit(Surv(futime, fustat) ~ factor(Progression), data = ovarian)
autoplot(Fit3)
```



## Stratified Proportional Hazard Model (3- Ecog Performance Status):

H0: The regression coefficient are assumed to be the same for each stratum.

Ha: The baseline hazard functions may be different and completely unrelated.

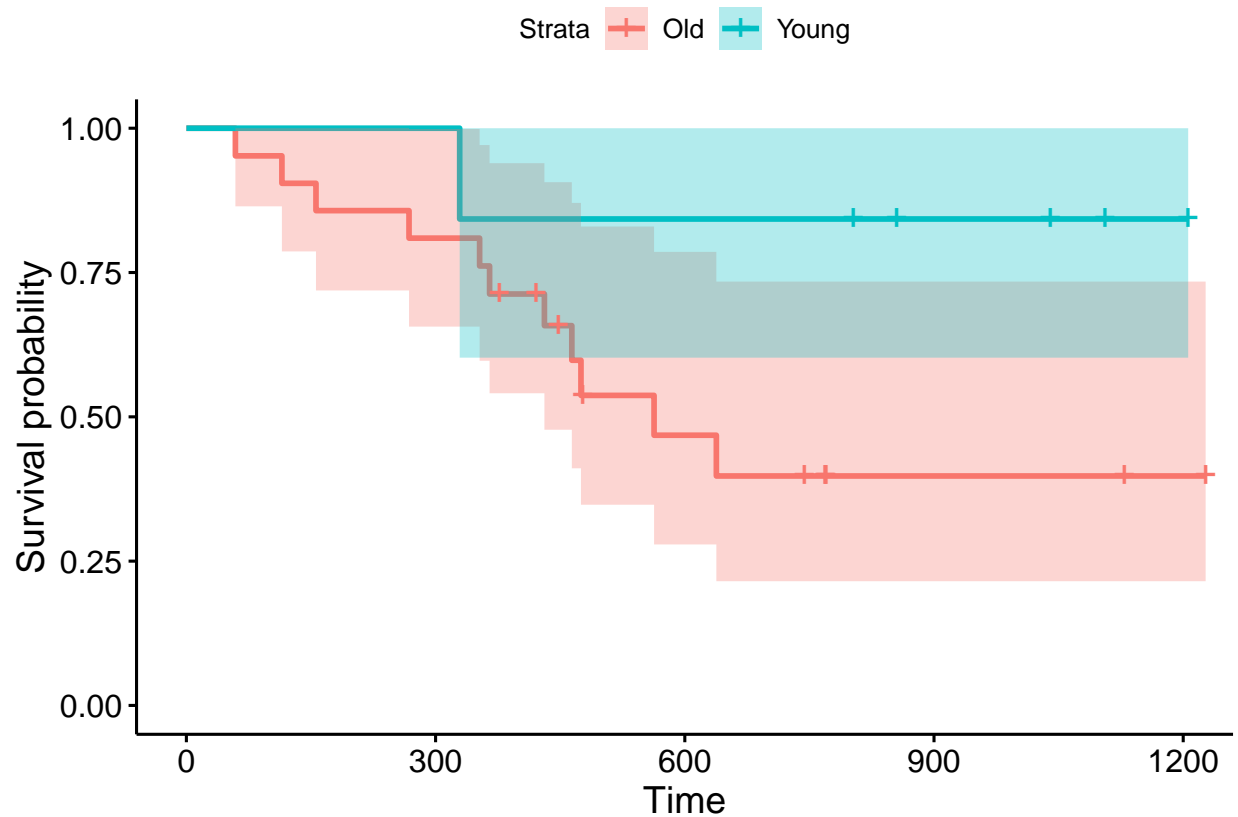
```
Model15_2 <- coxph(Surv(futime, fustat) ~ Performance + strata(dummy_Age), data = ovarian)
summary(Model15_2)
```

```
## Call:
## coxph(formula = Surv(futime, fustat) ~ Performance + strata(dummy_Age),
##       data = ovarian)
##
## n= 26, number of events= 12
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## PerformanceLow 0.2675    1.3067   0.5911 0.453   0.651
##
##               exp(coef) exp(-coef) lower .95 upper .95
## PerformanceLow    1.307    0.7653   0.4102    4.162
```



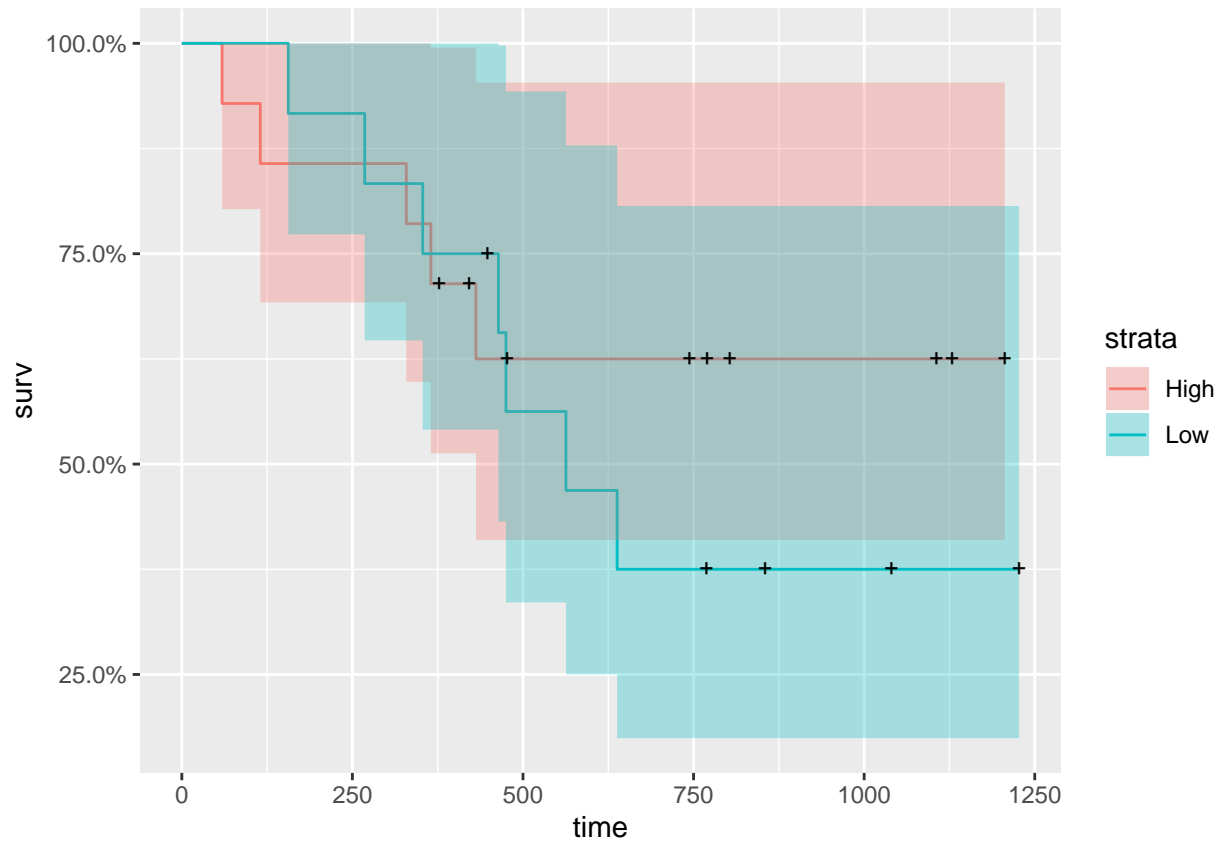
```
##
## Concordance= 0.507 (se = 0.085 )
## Likelihood ratio test= 0.21 on 1 df, p=0.6
## Wald test = 0.2 on 1 df, p=0.7
## Score (logrank) test = 0.21 on 1 df, p=0.6
```

```
ggsurvplot(survfit(Model5_2), data = ovarian, conf.int = T)
```



p-value of Treatment Now (p-value = 0.651) is not Significant, but also not similar to previous model without stratification (p-value = 0.60158).

```
Fit4 <- survfit(Surv(futime, fustat) ~ factor(Performance), data = ovarian)
autoplot(Fit4)
```



## Conclusion:

1- Estimated Hazard Ratio for Old Patients ( $\geq 50$ ) Receiving Cyclophosphamide alone Versus Old Patients ( $\geq 50$ ) Receiving Cyclophosphamide & Adriamycin is 3.463 (Increased Hazards) with confidence interval CI of (1.041 - 11.52).

Estimated Hazard Ratio for Old Patients ( $\geq 50$ ) Receiving Cyclophosphamide & Adriamycin Versus Old Patients ( $\geq 50$ ) Receiving Cyclophosphamide alone is 0.2888 (Reduced Hazards) with confidence interval CI of (1.0335 - 1.242).

2- This means, Receiving Cyclophosphamide (500 mg/m<sup>2</sup>) plus Adriamycin (40 mg/m<sup>2</sup>) in old Female Patients ( $\geq 50$ ) Reduce hazard of death by 28.9%, than Receiving Cyclophosphamide alone (1 g/m<sup>2</sup>) in old Female Patients ( $\geq 50$ ).