Time to heart failure survival analysis

Christophe Mpaga, Ahmed Oulad Amara, Adrien Parruitte

Contents

Context

Heart failure is a chronic condition where the heart is unable to effectively pump enough blood to meet the body's needs. It occurs when the heart muscle becomes weakened or damaged, leading to symptoms like shortness of breath, fatigue, and fluid retention. Heart failures happen from a variety of reason such as coronary disease, diabetes, obesity etc. In this study we try to determine the importance of various parameter on the survival of patient having heart failure. The event we analyse is then the death of the patient.

Introduction

The individuals of the data were patients admitted to Institute of Cardiology and Allied hospital Faisalabad-Pakistan during April-December (2015). From the 299 patients of the dataset, 105 are women and are 194 men. They are between 40 and 95 years. All have left ventricular systolic dysfunction, belonging to New York Heart Association (NYHA) class III and IV. Class III means patients have marked limitations of physical activity. They are comfortable at rest but experience symptoms with less than ordinary physical activity. Class IV means patients are unable to carry out any physical activity without discomfort. They may have symptoms even at rest and are often bedridden. From the 299 patients of the dataset, 105 are women and are 194 men. They are between 40 and 95 years.

Methods

data acquisition preparation and investigation.

data dictionary

Data originate from the study of Ahmad et al. 2017 [1]. The dataset has 13 features: Age, Anemia, High Blood Pressure, Creatinine phosphokinase, Diabetes, Ejection Fraction, Sex, Platelets, Serum Creatinine, Serum Sodium, Smoking, Time and Death Event. We explain some of the non-evident features:
- anemia : lower than normal haemoglobin concentration in blood - Creatinine phosphokinase : an enzyme notably found in the heart. Can leak in the blood in case of heart damage. - serum creatinine : a waste formed by the functioning of muscle. It is present in the blood and eliminated by the kidney through urine. As it is a serum creatinine the amount is not error induced by the taking of supplement creatine
- ejection fraction : percentage of blood pump out of the left ventricle with each contraction. If the EF is less than 40%, it indicates an heart failure or cardiomyopathy. - platelets : the normal amount of platelets ranges between 150,000 to 450,000 per μL of blood.
- serum sodium : sodium amount in blood is a well known indicator of heart failure. Is normal value ranges between 135-145 milli equivalents per liter. The presence of time and death event make this dataset perfectly adapted for a survival analysis. he unit of time is day. As all the patient didn't die, the dataset has right censored data.

```
## Rows: 299 Columns: 13
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (13): age, anaemia, creatinine_phosphokinase, diabetes, ejection_fractio...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
## # A tibble: 3 x 13
##     age anaemia creatini~1 diabe~2 eject~3 high_~4 plate~5 serum~6 serum~7   sex
##   <dbl>   <dbl>      <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl> <dbl>
## 1    75       0        582       0      20       1 265000      1.9     130     1
## 2    55       0       7861       0      38       0 263358.     1.1     136     1
## 3    65       0        146       0      20       0 162000      1.3     129     1
## # ... with 3 more variables: smoking <dbl>, time <dbl>, DEATH_EVENT <dbl>, and
## #   abbreviated variable names 1: creatinine_phosphokinase, 2: diabetes,
## #   3: ejection_fraction, 4: high_blood_pressure, 5: platelets,
## #   6: serum_creatinine, 7: serum_sodium
```

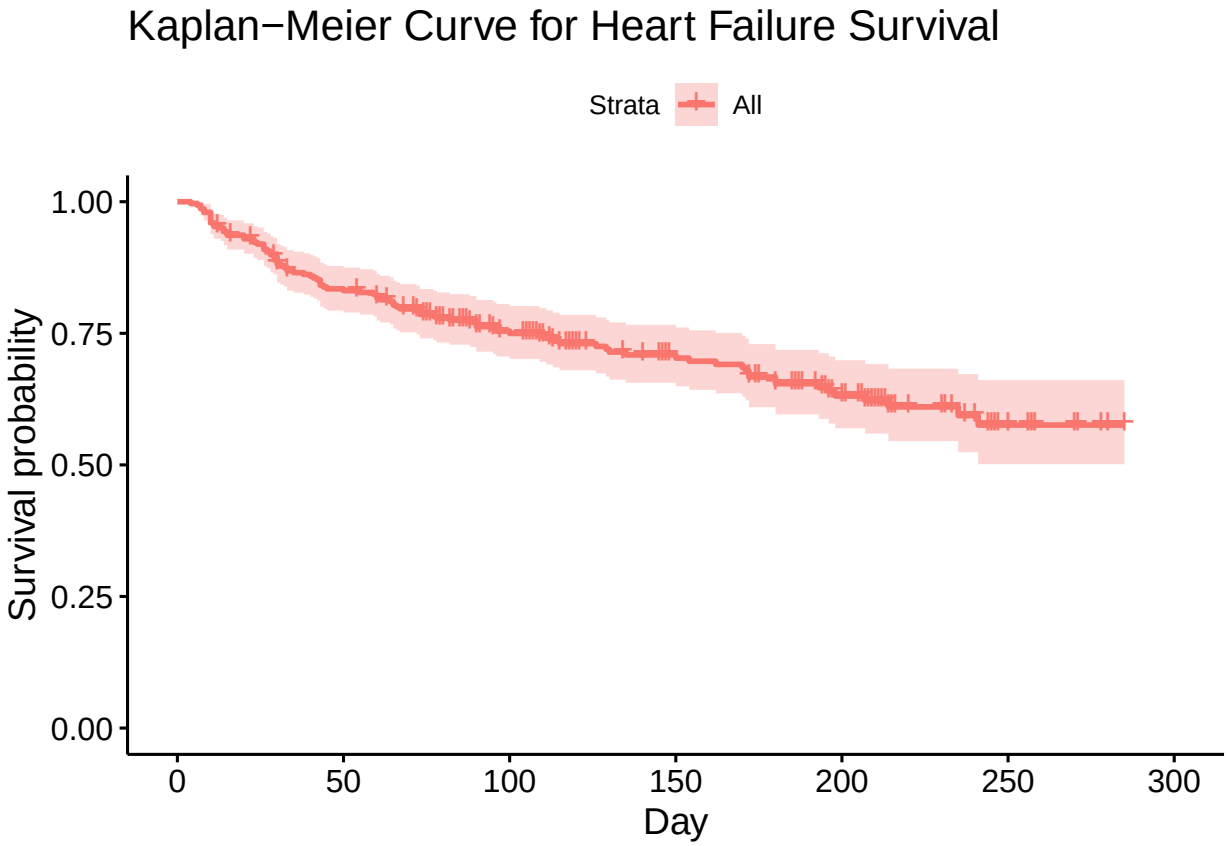
[1] Ahmad, Tanvir; Munir, Assia; Bhatti, Sajjad Haider; Aftab, Muhammad; Ali Raza, Muhammad (2017). DATA_MINIMAL.. PLOS ONE. Dataset. https://doi.org/10.1371/journal.pone.0181001.s001

overall Kaplan-Meyer estimator

We estimate the survival probability with the Kaplan-meier estimator.

```
## Call: survfit(formula = Surv(time, DEATH_EVENT) ~ 1, data = data)
##
##         n events median 0.95LCL 0.95UCL
## [1,] 299     96     NA      NA      NA
```

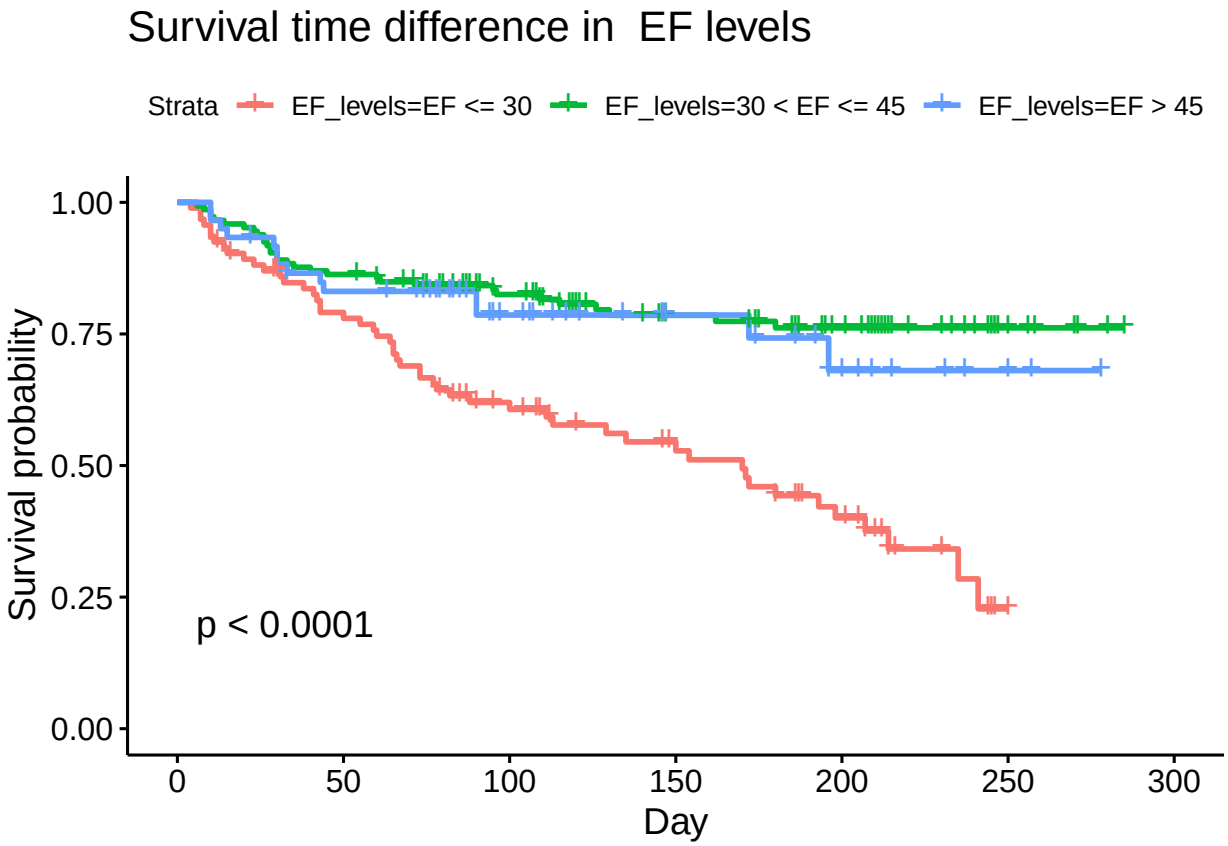
96 (32%) patients died due to the Cardiovascular Heart Disease (CHD). The median, 0.95LCL and 0.95UCl are NA because too many data are right censored. We need to go deeper in the analysis.



As the EF level and the high tension are the directly heart related covariates, we use them for the Kaplan Meier survival estimate

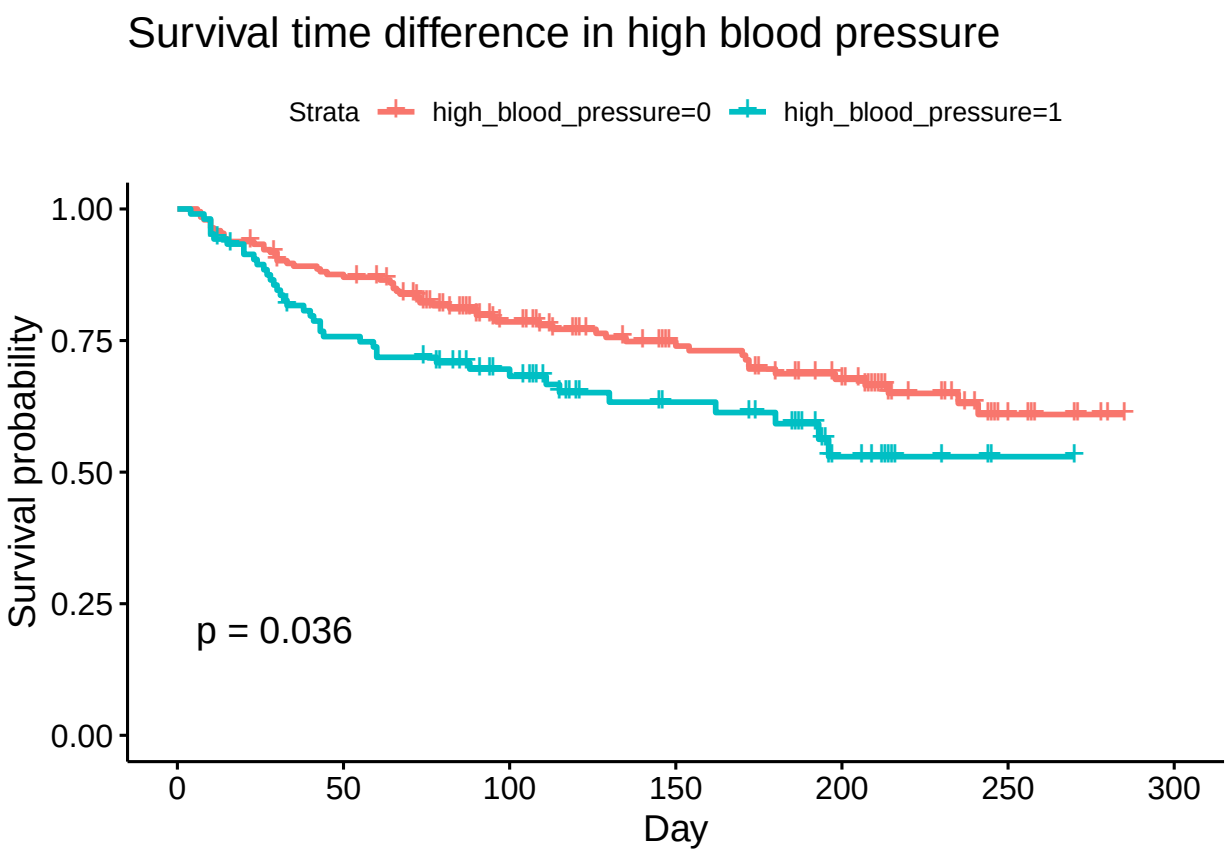
univariate analysis : group comparison

EF



The EF levels are indeed heavily correlated to the death for patient with heart failure as shown bby the plot and the p-value. The high pressure is less correlated.

high blood pressure



0.0358075, there is an effect of hbp in time to heart failure.

Sex

```
## p-value =  0.9497523 . No statistical difference, i.e. no difference of time to heart failure in sex
```

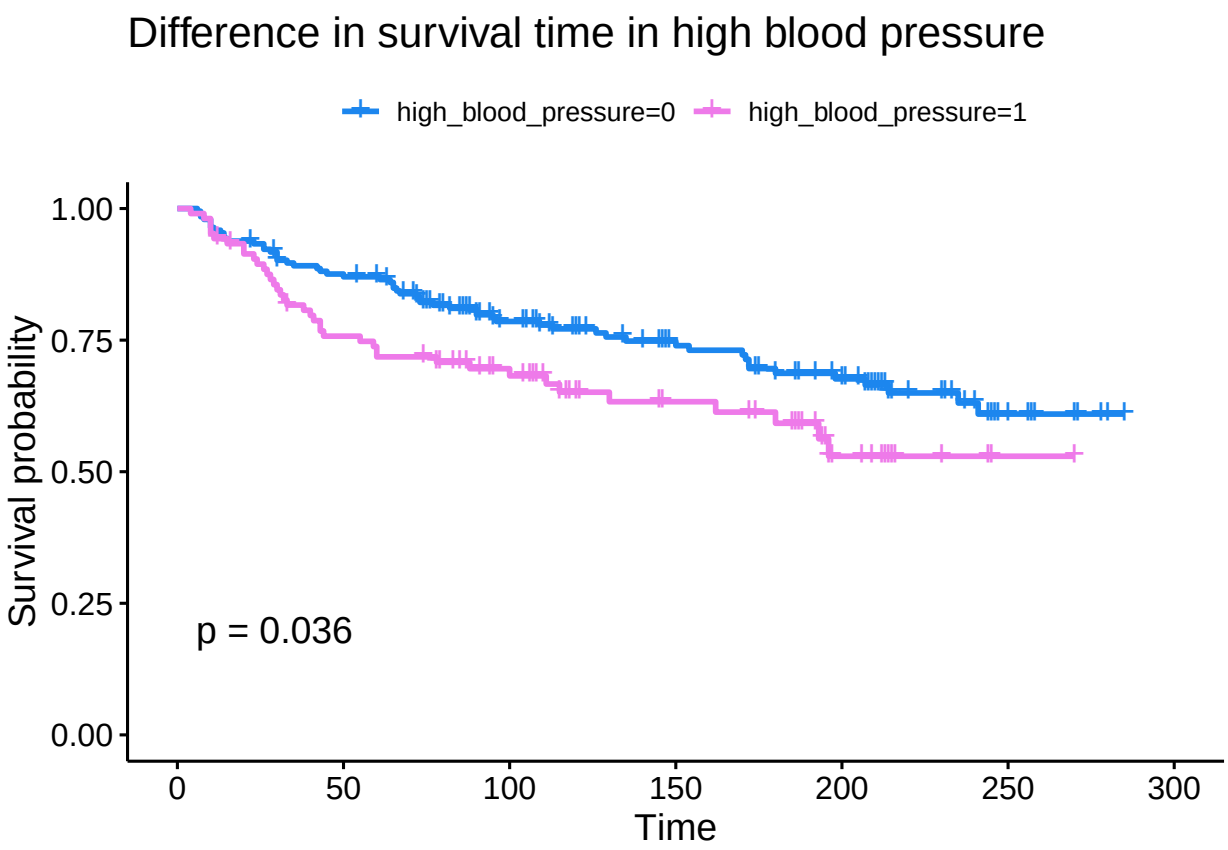
CDF:

From survival curves of male and female and using the long rank test we can concluded that the sex has no significant impact

Smoking.

```
## p-value =  0.9639597 There no difference in survival time in Smoking.
```

blood pressure



The p-value is 0.04, which is less than 0.05. Therefore, you can conclude that there is evidence of a statistically significant difference in survival between the two groups based on the presence or absence of high blood pressure

diabetes

```
## p-value =  0.840452 There is no difference in survival time in diabetes
```

anaemia

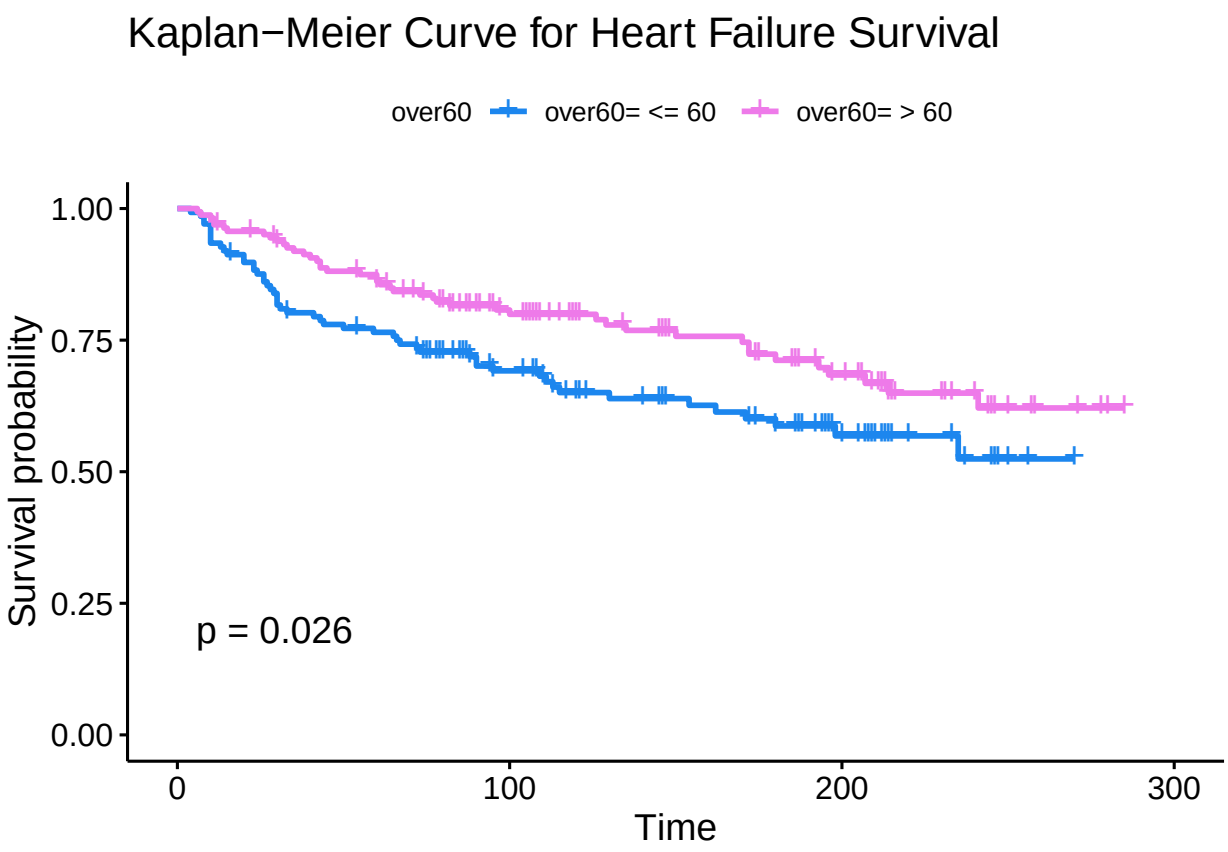
```
## p-value =  0.1 there is no difference in survival time
```

age

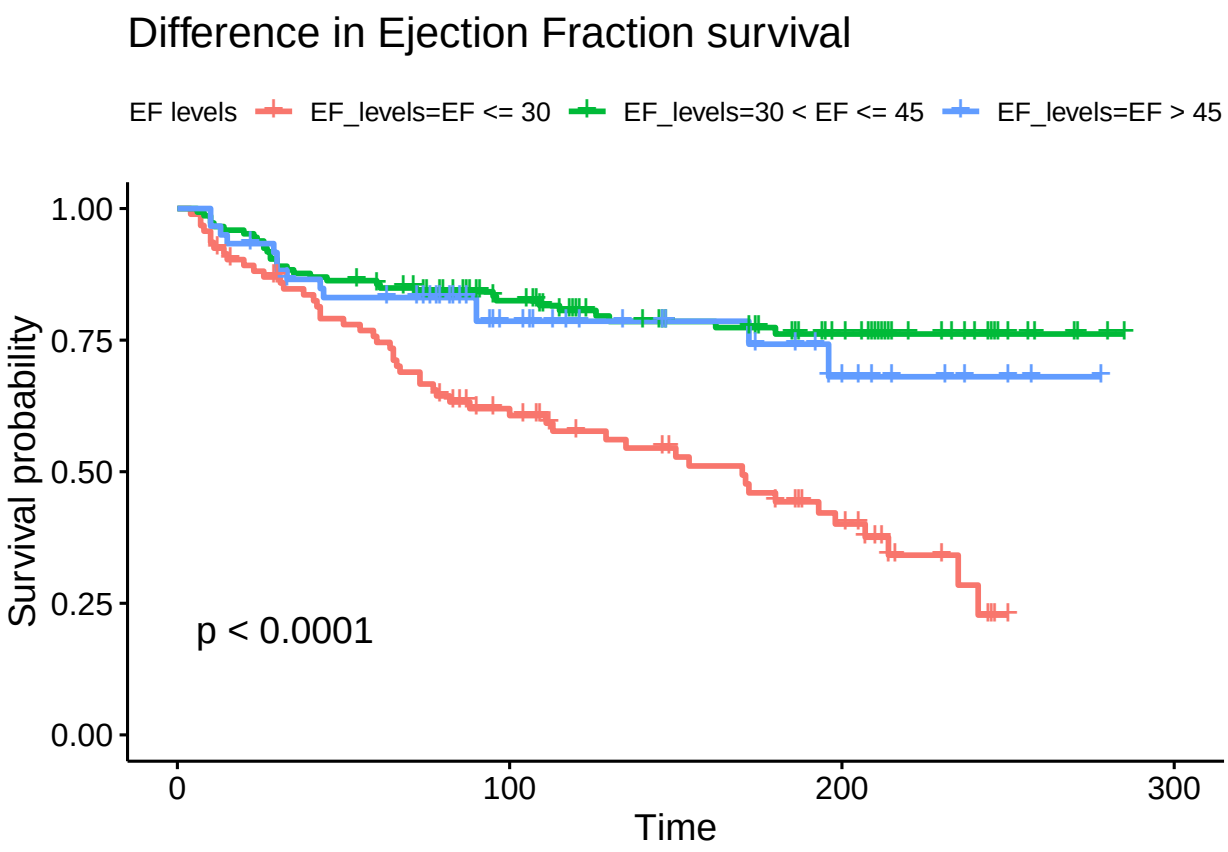
age is continuous variable we need first to discretize it, over 60 because 60 is median

```
##
##  <= 60    > 60
##     137     162
```

effect of age over 60



additive effect of Ejection Fraction



Since the p-value is less than 0.05 we reject the null hypothesis, this means there is a statistically significant difference in survival between the three groups.

bivariates analysis

additive effect of age and anaemia.

```
## p-value = 0.054 There is no difference in survival time with respect to age and anaemia
```

additive effect of high_blood_pressure and anaemia.

```
## p-value =  0.06584754 There is no additive effect of hbp and anaemia
```

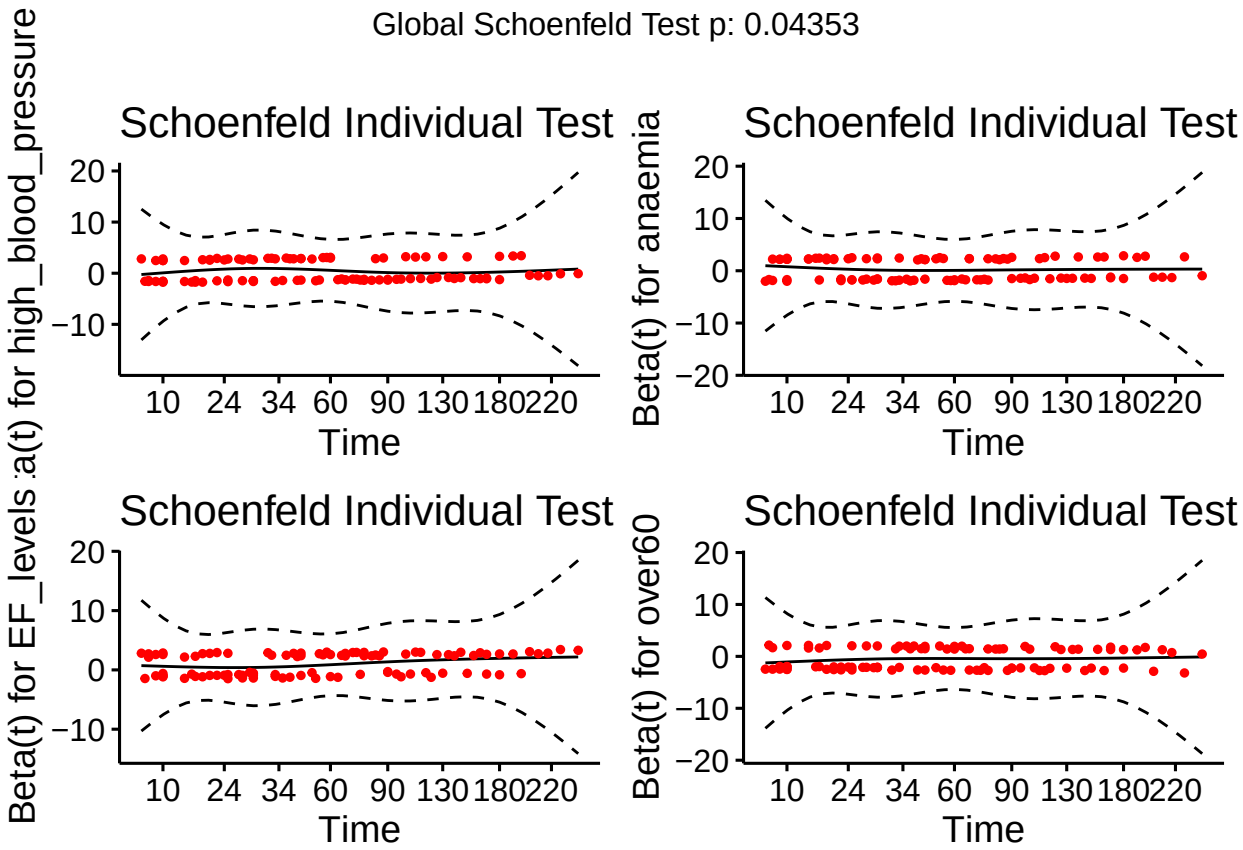
Multivariates analysis : Cox Proportional Hazards Model.

```
## # A tibble: 6 x 15
##     age anaemia creatini~1 diabe~2 eject~3 high_~4 plate~5 serum~6 serum~7 sex
##   <dbl> <fct>        <dbl> <fct>     <dbl> <fct>     <dbl>   <dbl>   <dbl> <fct>
## 1    75 0              582 0            20 1       265000      1.9     130 male
## 2    55 0             7861 0            38 0       263358.     1.1     136 male
## 3    65 0              146 0            20 0       162000      1.3     129 male
## 4    50 1              111 0            20 0       210000      1.9     137 male
```

```
## 5    65 1              160 1            20 0       327000      2.7     116 fema~
## 6    90 1               47 0            40 1       204000      2.1     132 male
## # ... with 5 more variables: smoking <fct>, time <dbl>, DEATH_EVENT <dbl>,
## #   EF_levels <fct>, over60 <fct>, and abbreviated variable names
## #   1: creatinine_phosphokinase, 2: diabetes, 3: ejection_fraction,
## #   4: high_blood_pressure, 5: platelets, 6: serum_creatinine, 7: serum_sodium
```

```
## # A tibble: 4 x 5
##   term                       estimate std.error statistic     p.value
##   <chr>                         <dbl>     <dbl>     <dbl>       <dbl>
## 1 "high_blood_pressure1"        0.429     0.211      2.03 0.0422
## 2 "EF_levels30 < EF <= 45"     -1.18      0.230     -5.11 0.000000321
## 3 "EF_levelsEF > 45"           -1.06      0.306     -3.46 0.000548
## 4 "over60 > 60 "               -0.551     0.207     -2.66 0.00787
```

```
##                       chisq df      p
## high_blood_pressure   0.177  1 0.6739
## anaemia               0.243  1 0.6222
## EF_levels             9.729  2 0.0077
## over60                2.067  1 0.1505
## GLOBAL               11.428  5 0.0435
```



Results

Discussion (if needed)

Conclusions

References

Ahmad T, Munir A, Bhatti SH, Aftab M, Raza MA (2017) Survival analysis of heart failure patients: A case study. PLoS ONE 12(7): e0181001. https://doi.org/10.1371/journal.pone.0181001