Time to heart failure survival analysis

Christophe Mpaga, Ahmed Oulad Amara, Adrien Parruitte

Contents

Introduction

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.

Dataset

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.

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.

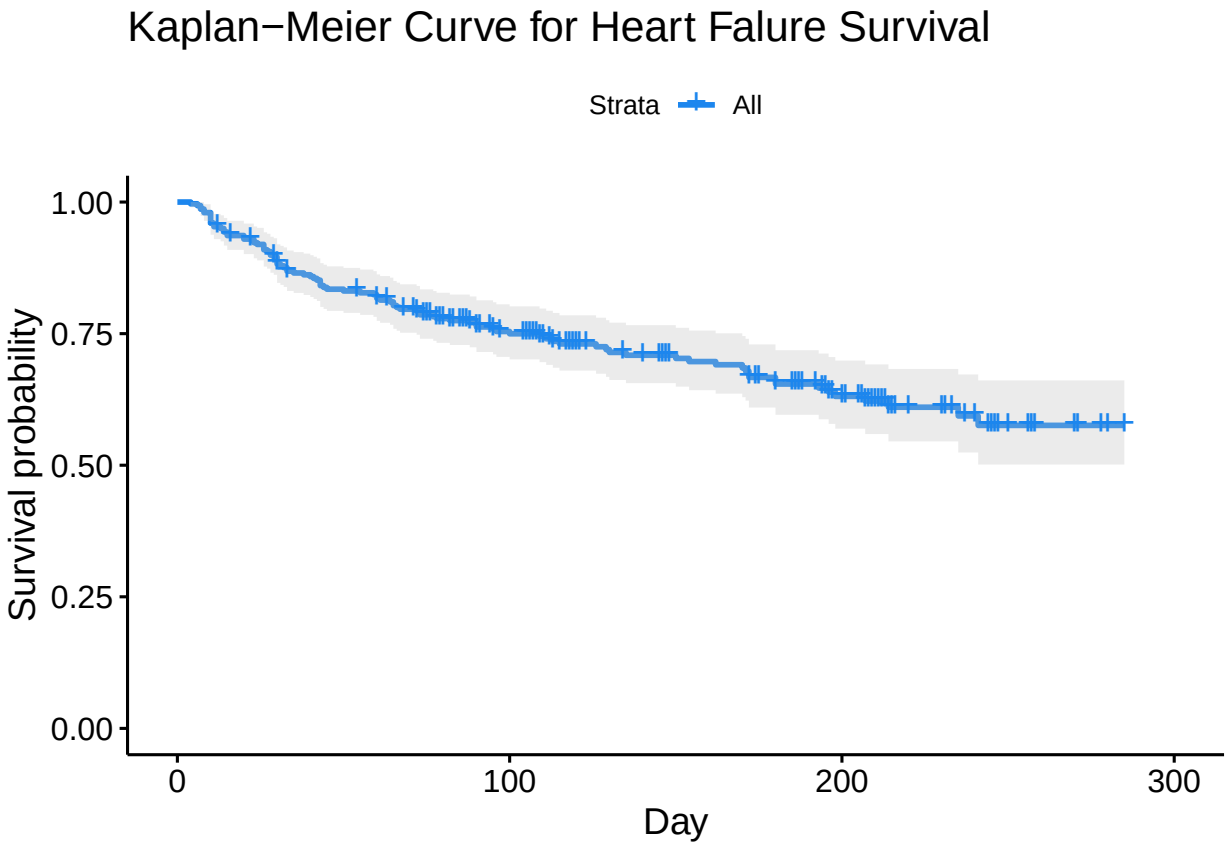
Univariate analysis : group comparison

Overall Kaplan-Meyer estimator

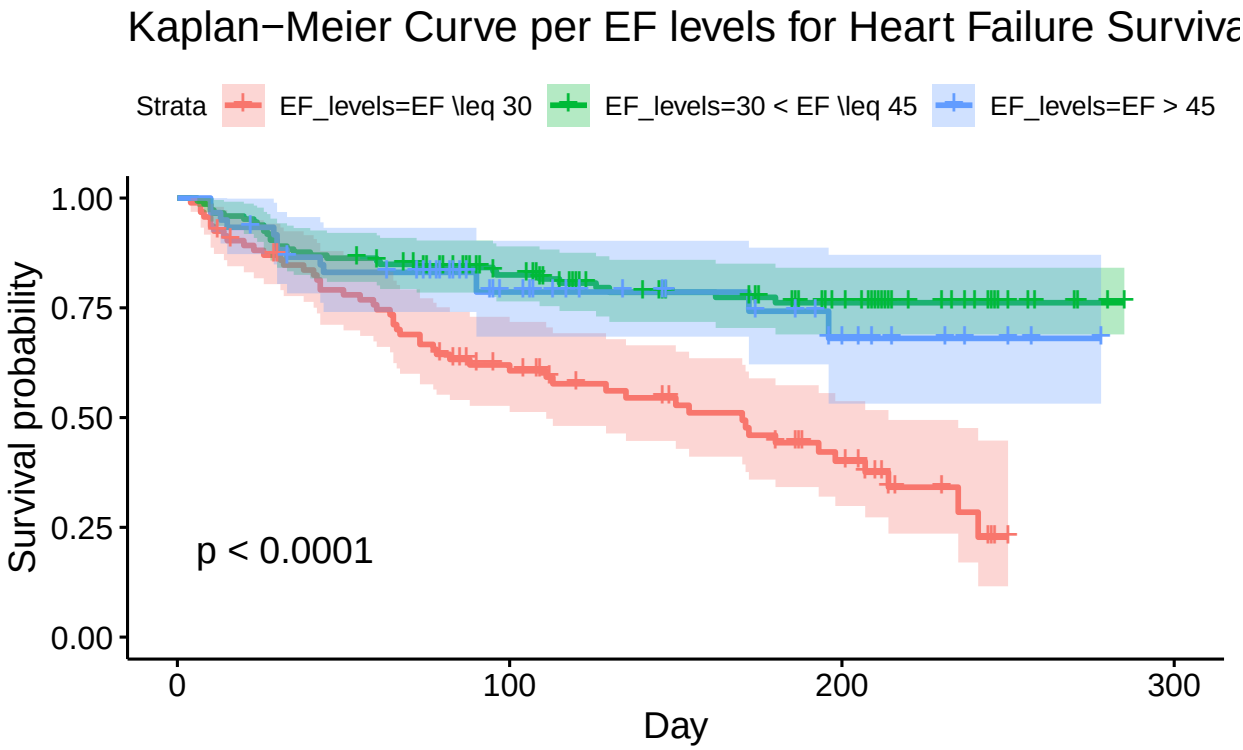
We estimate the survival probability with the Kaplan - meier estimator.

```
## 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



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

We apply the same method to the others covariates. The p-value for continuous covariates are obtained with Cox proportional hazards model. The results are summerized in the following tab:

covariate	Sex	Smoking	Diabetes	Aenemia	Age	creat.phos	platelets	ser. creat	ser.sodium
p-value	0.95	0.96	0.84	0.099	8.36e-07	0.26	0.4585	1.18e-07	0.00052

Since the p-value is less than 0.05 we reject the null hypothesis for age,serum creatinine and serum sodium, this means these covariates have a statistically significant impact in survival.

Smoking

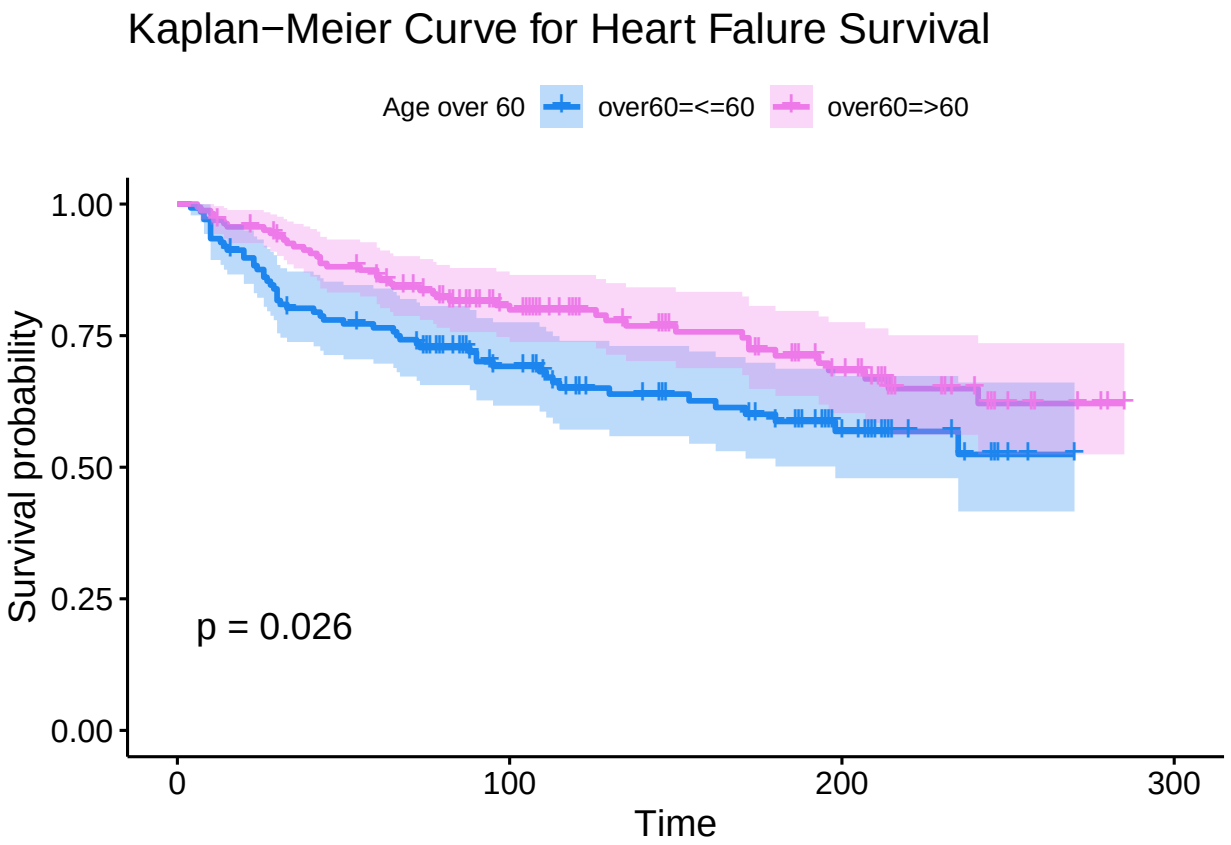
Blood pressure

Anaemia

The logrank test

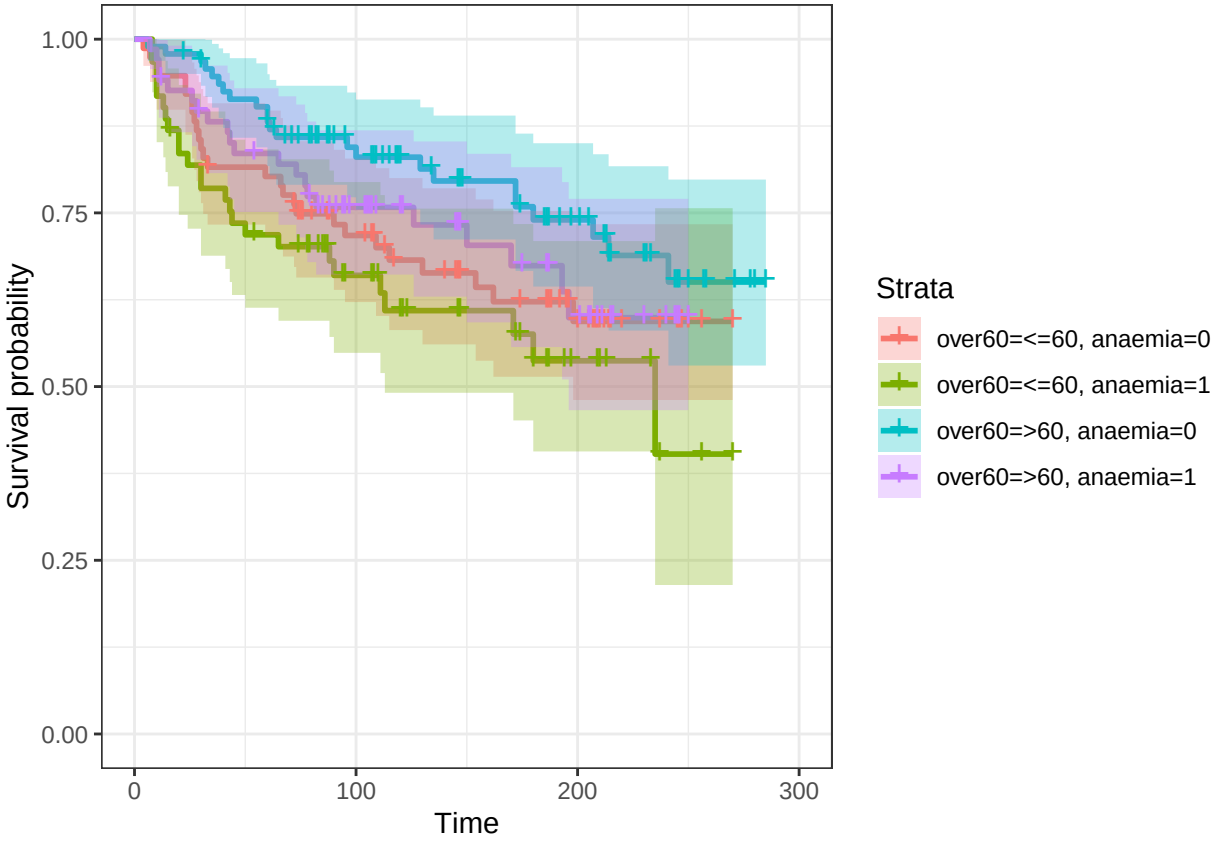
The impact of age

We need to discretize the variable 'age' since it is a continuous variable. We will use 60 as the threshold for discretization because it represents the median age.

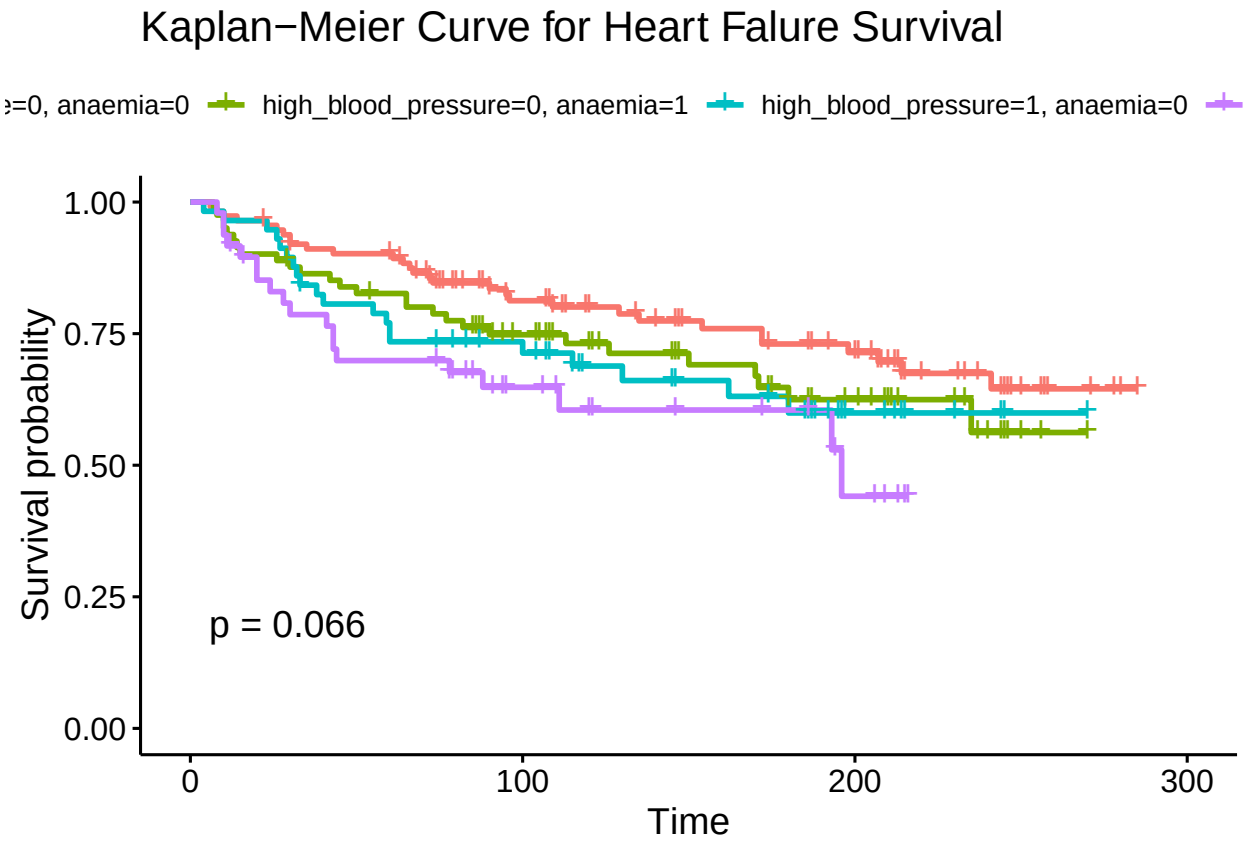


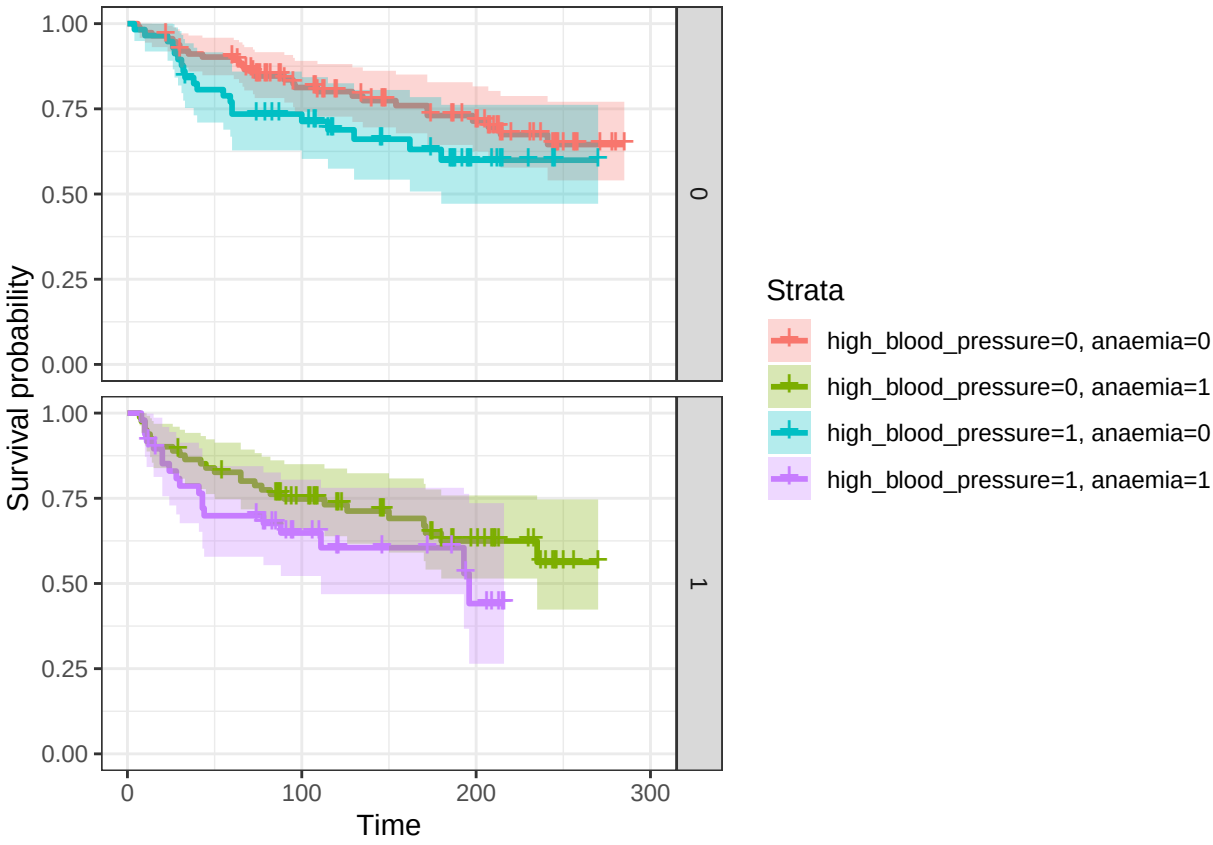
The logrank test

age and anaemia



high_blood_pressure and anaemia

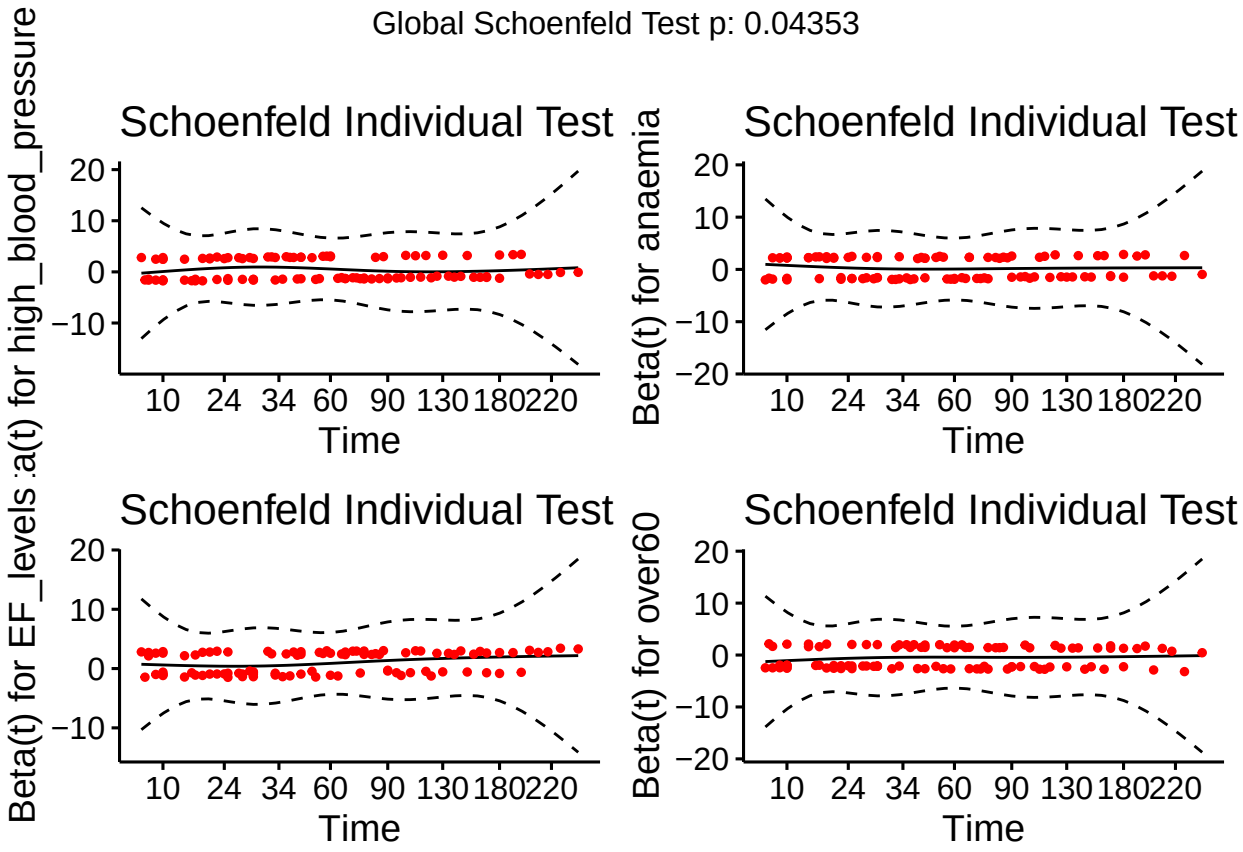

Kaplan−Meier Curve for Heart Falure Survival
ɛ=0, anaemia=0
high_blood_pressure=0, anaemia=1
high_blood_pressure=1, anaemia=0
Survival probability
1.00
0.75
0.50
0.25
0.00
p = 0.066
0
100
200
300
Time


Survival probability
1.00
0.75
0.50
0.25
0.00
0
1
0
100
200
300
Time
Strata
high_blood_pressure=0, anaemia=0
high_blood_pressure=0, anaemia=1
high_blood_pressure=1, anaemia=0
high_blood_pressure=1, anaemia=1

Cox Proportional Hazards Model

Univariate Cox regression

Multivariate Cox regression analysis



DATA Source 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