# Time to heart failure survival analysis

Christophe Mpaga, Ahmed Oulad Amara, Adrien Parruitte

## Contents

# Introduction

Heart failure is a chronic condition characterized by the heart's inability to pump an adequate amount of blood to meet the body's demands. It can occur when the heart muscle becomes weakened or damaged, resulting in symptoms such as shortness of breath, fatigue, and fluid retention. Various factors, including coronary disease, diabetes, and obesity, can contribute to the development of heart failure. In this study, our goal is to evaluate the significance of different parameters on the survival of patients with heart failure. We analyze the occurrence of patient deaths as the event of interest.

# Data description

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. Follow up time was between 4 to 285 days. 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 database has 13 features, including Age, Anemia, High Blood Pressure, Creatinine phosphokinase, Diabetes, Ejection Fraction, Sex, Platelets, Serum Creatinine, Serum Sodium, Smoking, Time, and Death Event. Out of these features, 5 are Boolean variables, namely Anemia, High Blood Pressure, Diabetes, Sex, and Smoking.

We added two new features to the dataset: "over60" and "EF_levels." The "over60" feature categorizes individuals as either over 60 years old or not, based on their age. The "EF_levels" feature categorizes individuals into 3 different groups based on their Ejection Fraction (EF) "EF <= 30", "30 < EF <= 45"and "EF > 45". These new features allow us to create Kaplan-Meier survival curves and analyze the data based on these specific characteristics.

The presence of time and death event in this dataset makes it suited for survival analysis. The unit of time in the dataset is measured in days. Since not all patients experienced the event of interest (death), the dataset contains right-censored data.

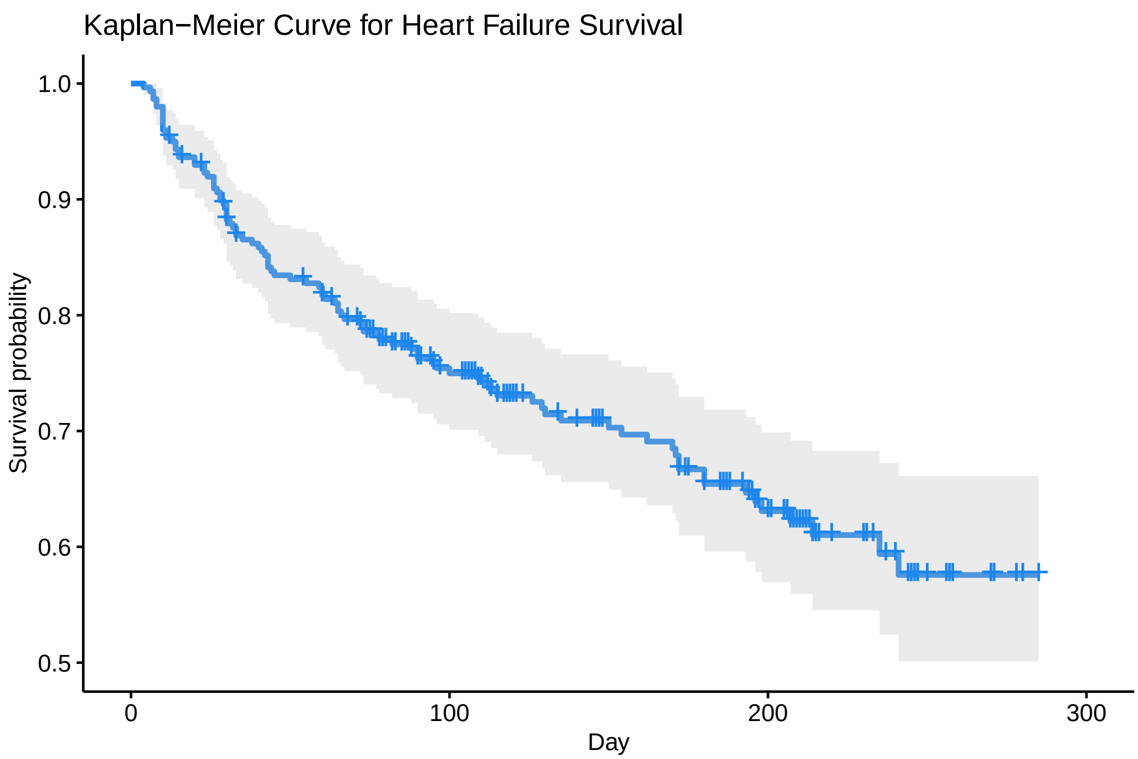
# Univariate analysis : group comparison

## Overall Kaplan-Meyer estimator

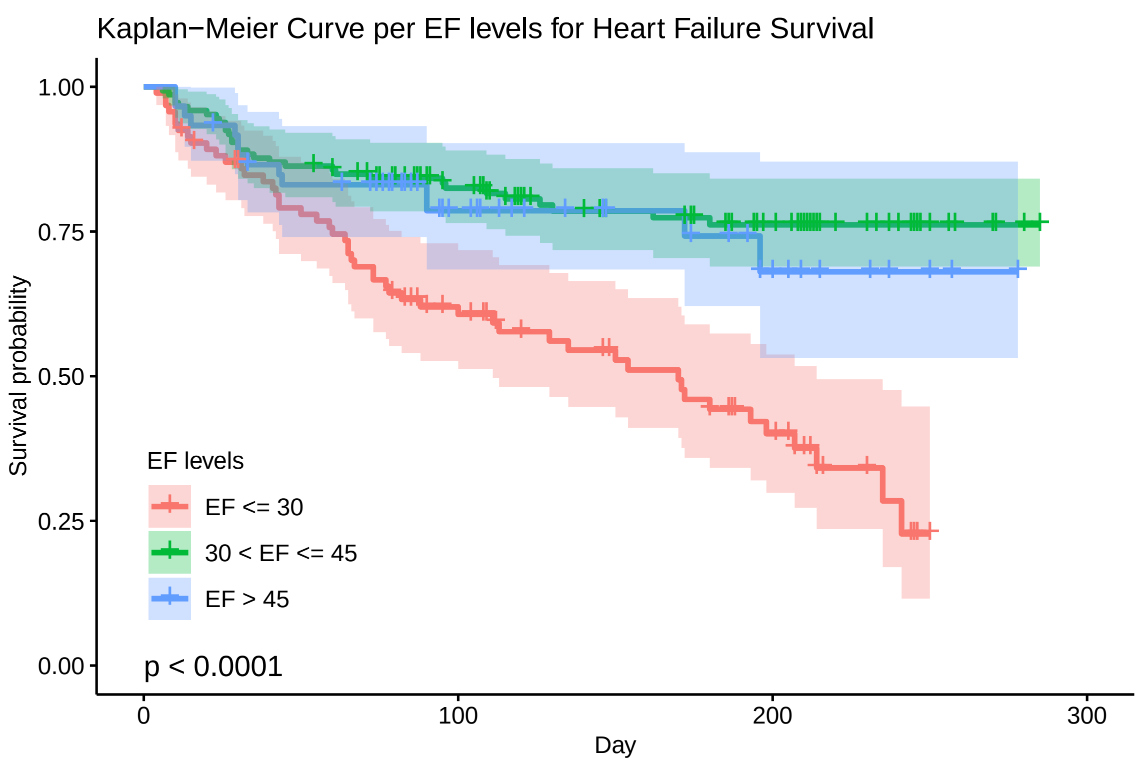
We estimate the survival probability with the Kaplan - meier estimator.

```
## 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




With a p-value less than 0.0001, the EF levels are indeed statistically significant to the death for patient with heart failure as shown by the plot.

A long-rank test was conducted to determine whether there are differences in survival between groups on each of the categorical covariates. The results are summarized in the following table:

| covariate | Sex | Smoking | Diabetes | Aenemia | EF levels | bad platelet | Age over 60 | Blood Pressure |
|---|---|---|---|---|---|---|---|---|
| p-value | 0.95 | 0.96 | 0.84 | 0.099 | 1.81e-07 | 0.26 | 0.02 | 0.05 |

We can notice that the p-value is below 0.05 for EF levels, High blood pressure, and Age over 60. Therefore, we reject the null hypothesis, indicating that these covariates have a statistically significant impact on survival.

# Proportional Hazards Model

Let's assume that our survival function follow a semi-parametric model. ## Univariate Cox regression We will be examining the significance of each covariate using the Cox regression model. The results will be presented in a table, which includes the covariates, their beta coefficients, hazard ratios, lower confidence intervals, upper confidence intervals, and p-values.

| | beta | HR | lower_ci | upper_ci | p.value |
|---|---|---|---|---|---|
| age | 4.2e-02 | 1.00 | 1.00 | 1.10 | 8.0e-07 |
| anaemia | 3.4e-01 | 1.40 | 0.94 | 2.10 | 1.0e-01 |
| creatinine_phosphokinase | 1.1e-04 | 1.00 | 1.00 | 1.00 | 2.6e-01 |
| diabetes | -4.2e-02 | 0.96 | 0.64 | 1.40 | 8.4e-01 |
| ejection_fraction | -4.6e-02 | 0.95 | 0.93 | 0.98 | 1.7e-05 |
| high_blood_pressure | 4.4e-01 | 1.50 | 1.00 | 2.30 | 3.7e-02 |
| platelets | -8.0e-07 | 1.00 | 1.00 | 1.00 | 4.7e-01 |
| serum_creatinine | 2.9e-01 | 1.30 | 1.20 | 1.50 | 1.0e-07 |
| serum_sodium | -6.8e-02 | 0.93 | 0.90 | 0.97 | 5.2e-04 |
| sex | 1.4e-02 | 1.00 | 0.67 | 1.50 | 9.5e-01 |
| smoking | -9.6e-03 | 0.99 | 0.64 | 1.50 | 9.7e-01 |
| bad_platelet | 2.2e-01 | 1.20 | 0.73 | 2.10 | 4.2e-01 |
| over60 | 4.5e-01 | 1.60 | 1.10 | 2.40 | 2.7e-02 |
| 30 <= EF vs 30 < EF <= 45 | -1.2e+00 | 0.31 | 0.20 | 0.49 | 4.0e-07 |
| 30 <= EF vs EF > 45 | -9.0e-01 | 0.41 | 0.23 | 0.74 | 3.0e-03 |

For table above we say that for all these covariates, including anaemia, creatinine_phosphokinase, diabetes, platelets, sex, and smoking, have p-values greater than the chosen significance level (e.g., 0.05). This suggests that these covariates are statistically insignificant, indicating that there is no strong evidence of a significant association between these variables and the hazard rate.

Ejection Fraction (EF) appears to be a significant factor as it shows statistical significance for both comparisons: 30 <= EF vs 30 < EF <= 45 and 30 <= EF vs EF > 45, with p-values below 0.05. For both EF level comparisons, the negative beta coefficients indicate a negative association between EF levels and the hazard rate. The hazard ratios of 0.31 and 0.41 suggest lower hazard rates for the specified EF levels compared to the baseline group.

High Blood Pressure is another significant factor with p-value = 0.037. These covariates have The positive beta coefficient of 0.44 suggests a positive association High Blood Pressure and risk of death.

For the Age covariate the p-value is less than 0.05 but the hazard ratio is equal to 1 this means that we cannot reject the null hypothesis. also it suggests that there is no significant difference in survival between the groups for this covariate. On the other hand, for the covariate over60 the p-value being less than 0.05 and the hazard ratio equal 1.6, this indicates that patients over 60 years old have a 1.6 times higher risk of death compared to patients below 60 years old.

## Multivariate analysis

### Cox semi-parametric model : 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

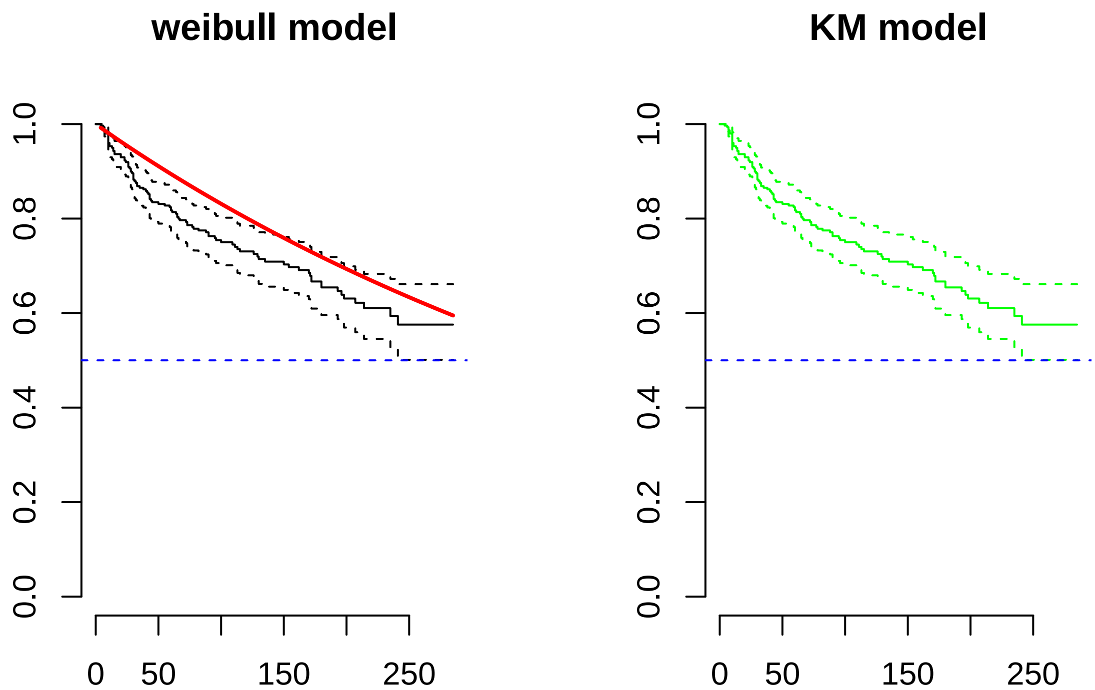
### Semi-parametric full model

| | coef | exp(coef) | se(coef) | Pr(>\|z\|) |
|---|---|---|---|---|
| age | 0.0690424 | 1.0714816 | 0.0146827 | 0.0000026 |
| creatinine_phosphokinase | 0.0002217 | 1.0002217 | 0.0000986 | 0.0245664 |
| ejection_fraction | -0.0587333 | 0.9429583 | 0.0256569 | 0.0220691 |
| high_blood_pressure1 | 0.4968625 | 1.6435565 | 0.2198163 | 0.0237994 |
| serum_creatinine | 0.2799269 | 1.3230330 | 0.0727539 | 0.0001193 |
| serum_sodium | -0.0467732 | 0.9543038 | 0.0226278 | 0.0387274 |
| over60>60 | -0.7327664 | 0.4805777 | 0.3540541 | 0.0384855 |

Under semi-parametric model assumption, patients presenting `high_blood_pressure`, `EF_levels30 < EF <= 45`, `EF_levelsEF > 45`, `age > 60`, `serum_creatinine`, `serum_sodium` are more likely to experience heart failure and die. Moreover, presenting a `high_blood_pressure` increases the risk of heart failure by a hazard rate of 1.53, holding other covariates fixed, while, `EF_levels30 < EF <= 45`, `EF_levelsEF > 45`, `age > 60`, `serum_creatinine` and `serum_sodium` increases the risk of dying from heart failure with hazard rate magnitude greater or almost equal to 1.

### Parametric model

We fit a weibull parametric model and compare it to non parametric KM model.




This plot suggest that we could model this data under a linear model with a negative slope. As it decreases very fast. However, this not a stepwise or strictly smooth curve. So we tried an exponential model .The weibull model curve is plotted against KM model. Though, KM model could suffice to model this dataset, we would suggest.

## Results

In this application, we were to analyze time to heart failure in this cohort, the event being death. We have found that, there is no median time to death. Our results show that there is an effect of high blood pressure , EF and ageing,but surprisingly no effect of Smocking, nor diabetes or anaemia. There is a group difference in survival time in `EF` and `high blood pressure`. Half of subject having `EF` less than 30 , are expected to die after almost 175 days. An additive model with covariates, `high_blood_pressure1`, `EF_levels30 < EF <= 45`, `EF_levelsEF > 45` , `age > 60` , `serum_creatinine` and `serum_sodium` shows that, patients presenting a profile with theses covariates are more likely to experience heart failure and die. Also, There is a noticeable effect of ageing `(age > 60)`.

## Discussion

We aim at analyzing time to heart failure.We have found that, there is an effect of high blood pressure , EF , serum_creatinine, serum_sodium and ageing, but surprisingly no effect of Smocking, nor diabetes or anaemia. following one could consider that, this study held in a cohort of patients of NYHA class III and IV which are advanced stages of heart failure. Similar results concerning diabetes and smoking have been reported by F Otero-Raviña et al. 2009 as well.

## Conclusions

Following our results, we can conclude that ageing, Ejection fraction, serum_creatinine , serum_sodium and high blood pressure influence time to heart failure and the occurence of CHD as well as ageing. Overall, above half of these patients died early of heart failure. Though there was a high risk of death in this cohort..

## 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

https://cran.r-project.org/web/packages/flexsurv/vignettes/flexsurv.pdf

https://boostedml.com/2018/11/when-should-you-use-non-parametric-parametric-and-semi-parametric-survival-analysis.html