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 will analyse time heart failure and investigate the importance of various factors 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.The unit of time is day. As all the patients 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
```

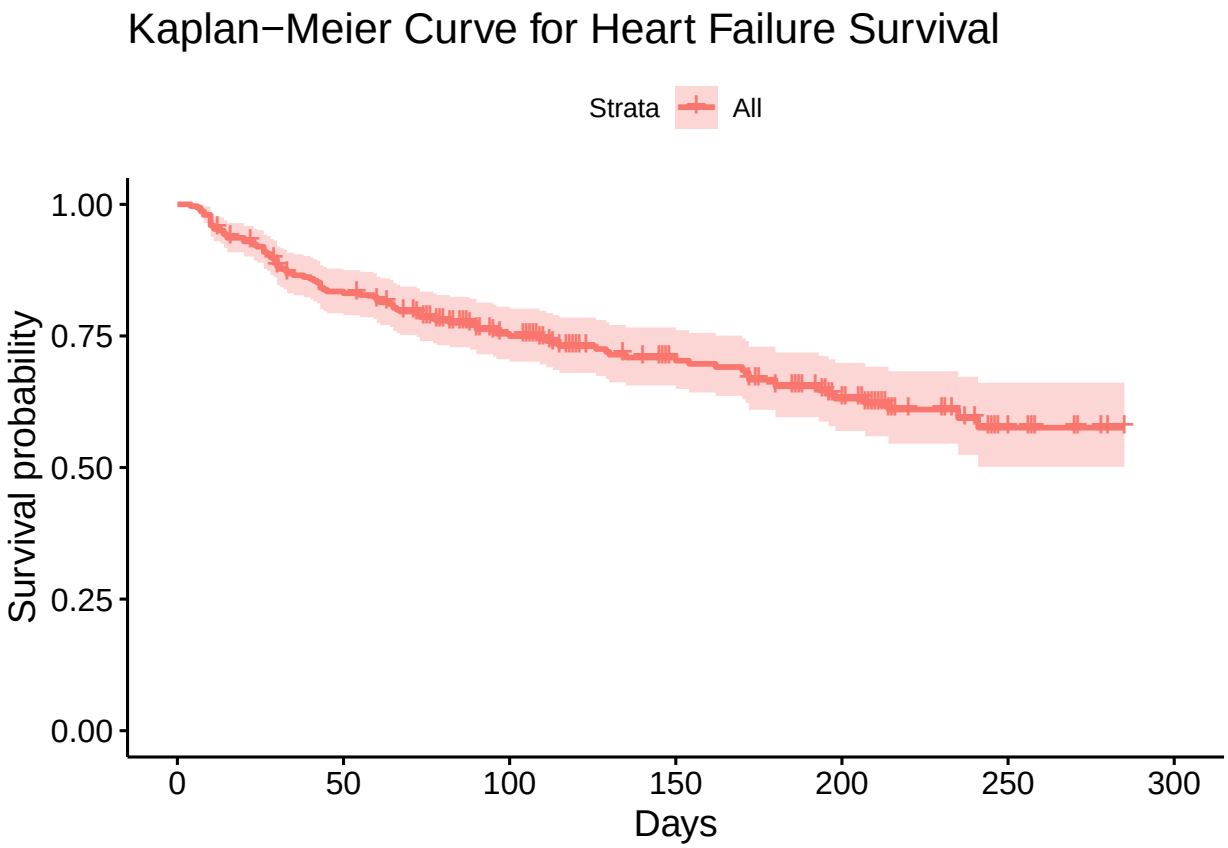
overall Kaplan-Meyer estimator

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

[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

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

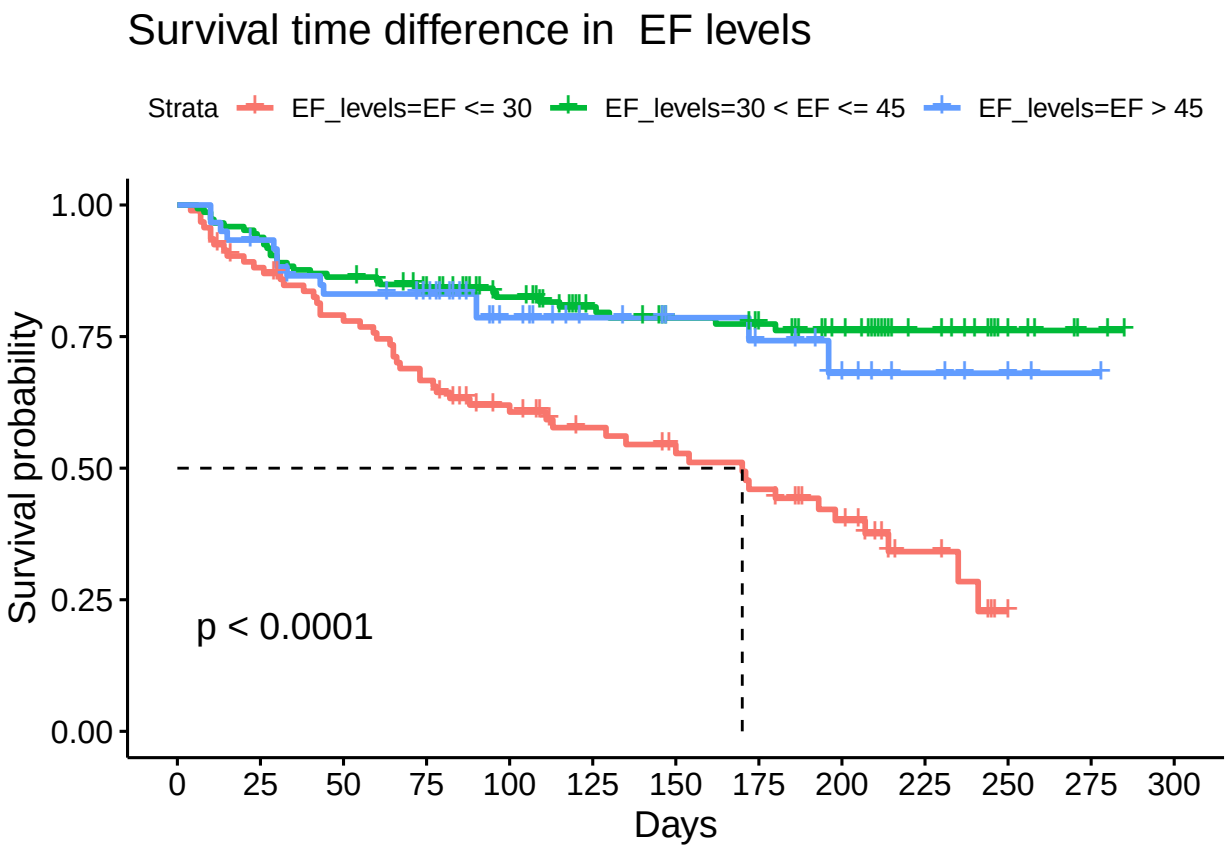
96 (32%) patients died due to the Cardiovascular Heart Disease (CHD). The median, 0.95LCL and 0.95UCl are NA because too many data are right censored. We need to go deeper in the analysis.



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

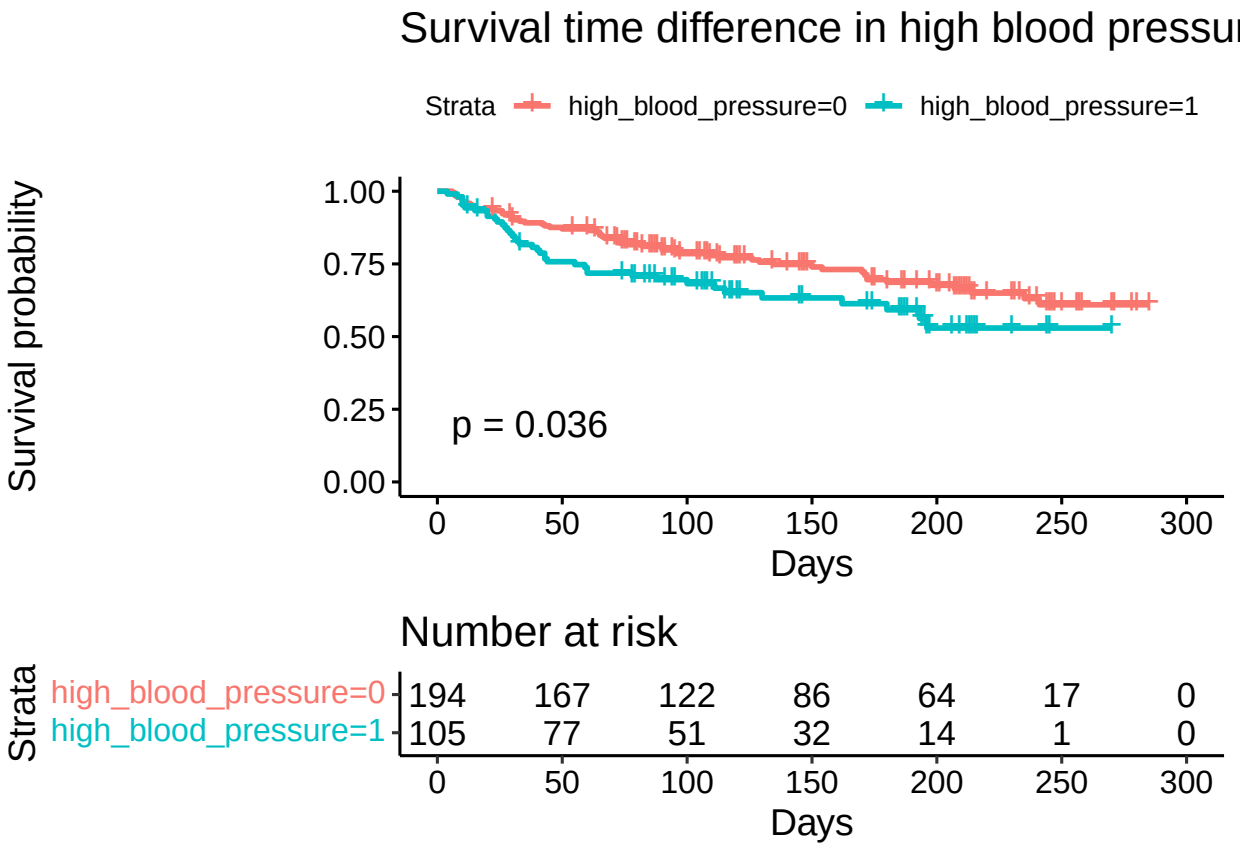
Ejection fraction (EF)




```
## Call: survfit(formula = Surv(time, DEATH_EVENT) ~ EF_levels, data = data)
##
##                             n events median 0.95LCL 0.95UCL
## EF_levels=EF <= 30         93     51    170     111     214
## EF_levels=30 < EF <= 45 146     31     NA      NA      NA
## EF_levels=EF > 45          60     14     NA      NA      NA
```

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.

high blood pressure (hbp)



0.0358075, there is a group difference in survival time wrt/ high blood pressure in time to heart failure.

Sex

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

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.
```

diabetes

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

anaemia

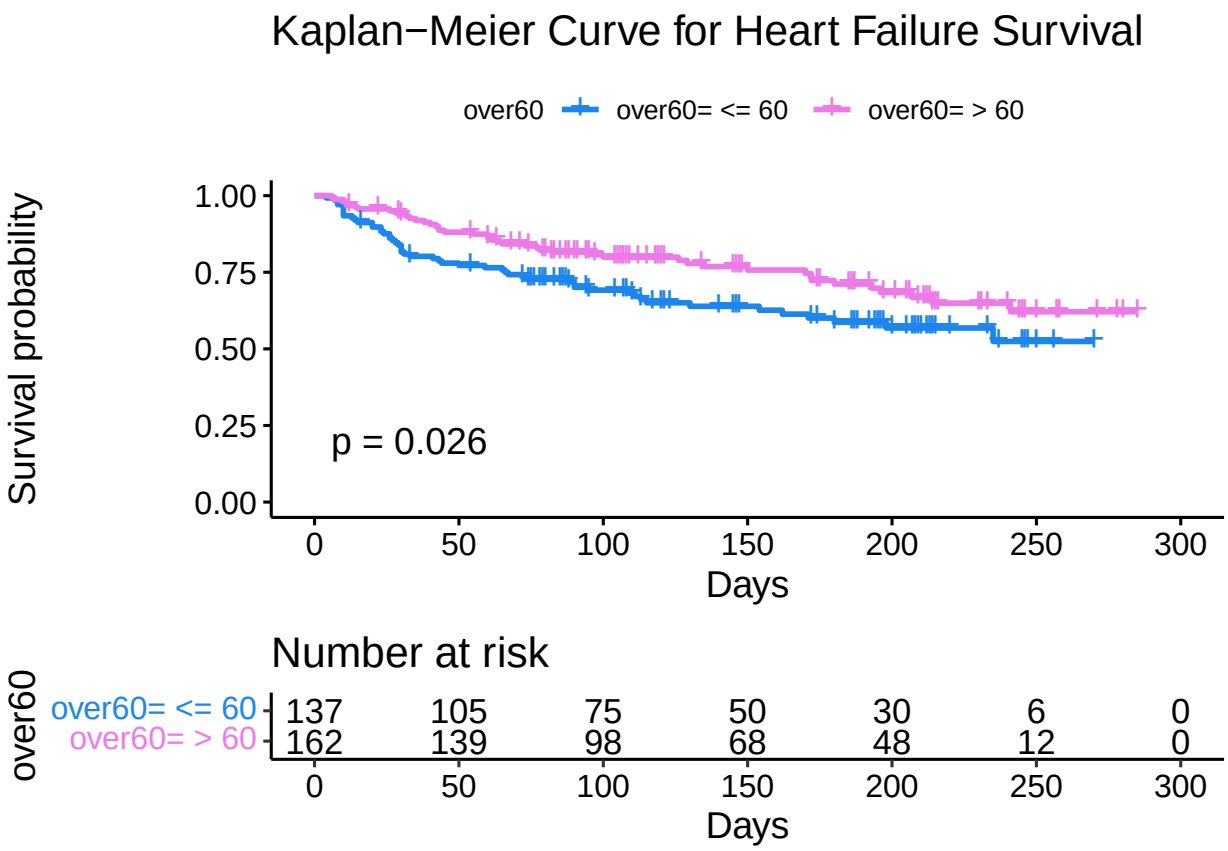
```
## p-value =  0.1 there is no difference in survival time
```

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



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.

effect of EF + hbp + age

```
##                                 coef exp(coef)     se(coef)     Pr(>|z|)
## age                       0.0736059764 1.0763826 1.513464e-02 1.153801e-06
## creatinine_phosphokinase  0.0002355666 1.0002356 9.959769e-05 1.802125e-02
## ejection_fraction        -0.0632673653 0.9386925 2.620959e-02 1.578275e-02
## high_blood_pressure1      0.4756054775 1.6089881 2.214552e-01 3.174256e-02
## serum_creatinine          0.2860291066 1.3311312 7.470730e-02 1.288411e-04
## serum_sodium             -0.0443408416 0.9566278 2.227236e-02 4.649785e-02
## over60 > 60               0.7358977101 2.0873550 3.535246e-01 3.737878e-02
```

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.

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. `(why : to document)`

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