

Liver Transplant Survival Analysis

Attoumassa Samaké, Vixra Keo, Mounir M'barki, Antoine Bedouch

Introduction :

For our survival analysis work, we chose a dataset available in the survival R package call “transplant”. This dataset provides information about subjects from the Mayo Clinic Rochester liver transplant waiting list from 1990 and 1999. It contains 815 patients, from which 636 underwent a transplantation, 66 died while waiting, 37 withdrew from the waiting list and 76 were censored. This data initially comes from a scientific study (Kim et al., Deaths on the liver transplant waiting list: An analysis of competing risks, 2006), though the final dataset used by the study is more complete (it includes data on the patients disease)

The dataset contains 6 variables :

- age : age at the addition of the waiting list
- sex : male or female
- abo : which is the blood type (A, B, AB and O)
- year : the year the patient entered in the waiting list
- futime : the time (in days) between the registry on the waiting list and the final outcome
- event :
 - ltx = The patient was transplanted during the study
 - death = The patient died while on the waiting list
 - withdraw = The patient withdrew from the waiting list
 - censored = The outcome is censored (for most cases, no outcome at the end of the study in 1999)

This dataset reports on the evolution of liver transplants. It aims to answer crucial questions about wait times, the demographic composition of people on the waiting list, and how the patient medical information are impacting the death and transplantation outcome.

An essential factor influencing transplant dynamics is blood type. The dataset highlights the impact of blood type compatibility on the process. Donor livers from type O individuals can be utilized by recipients with A, B, AB, or O blood types, while an AB liver is exclusive to an AB recipient.

Our goal will be to study the survival of two outcomes: * the transplant outcome `ltx` * the death while waiting `death`

We will consider the patients who withdrew from the study as right-censored data.

Data preparation

First let's see what our dataset looks like:

##	age	sex	abo	year	futime
##	Min. :17.00	m:447	A :325	Min. :1990	Min. : 0.0
##	1st Qu.:44.00	f:368	B :103	1st Qu.:1993	1st Qu.: 50.0
##	Median :52.00		AB: 41	Median :1995	Median : 128.0
##	Mean :50.52		O :346	Mean :1995	Mean : 213.6
##	3rd Qu.:58.00			3rd Qu.:1997	3rd Qu.: 276.5
##	Max. :72.00			Max. :1999	Max. :2055.0

```
## NA's :18
##      event
## censored: 76
## death : 66
## ltx :636
## withdraw: 37
##
##
##
```

We note that 18 values for the age are missing. We could just ignore these patients in our analysis. Instead we have decided to impute these missing values with the average age of 50 years old.

For the age non parametric analysis, we split the patients in two age groups:

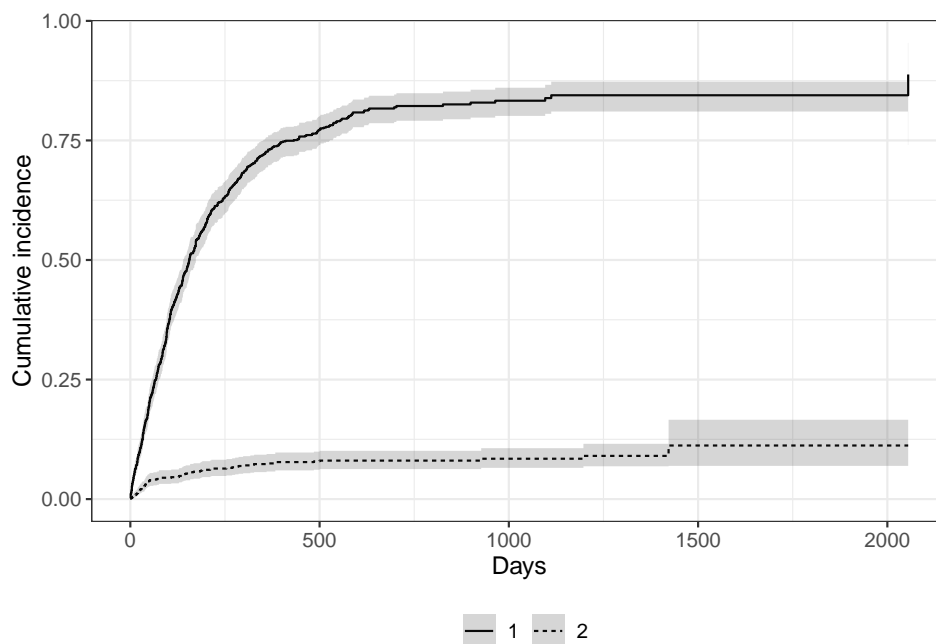
- old (age ≥ 50)
- young (age < 50)

Finally, we created in our dataframe specific boolean and integer columns containing outcome data for future analysis

```
## [1] "average age = 50"
```

Non parametric modeling

Before doing any modeling, we can look at the empirical cumulative incidence function to see the relative incidence of each outcome: transplant or death. This curves tells us what is the probability to have a given outcome after a given number of days.



```
## [1] "Median number of days before the patient is transplanted: 108.5"
```

```
## [1] "Median number of days before the patient dies: 65.5"
```

Legend: The outcome 1 is the transplant incidence The outcome 2 is the death incidence

We observe that in our dataset the majority outcome is to be transplanted (~78%) while the probability to die while on the waiting list is much lower (~8%). We also observe that most outcomes happen quite quickly

after the registry on the list (108.5 days if the outcome is transplantation, 65.5 days if the outcome is death).

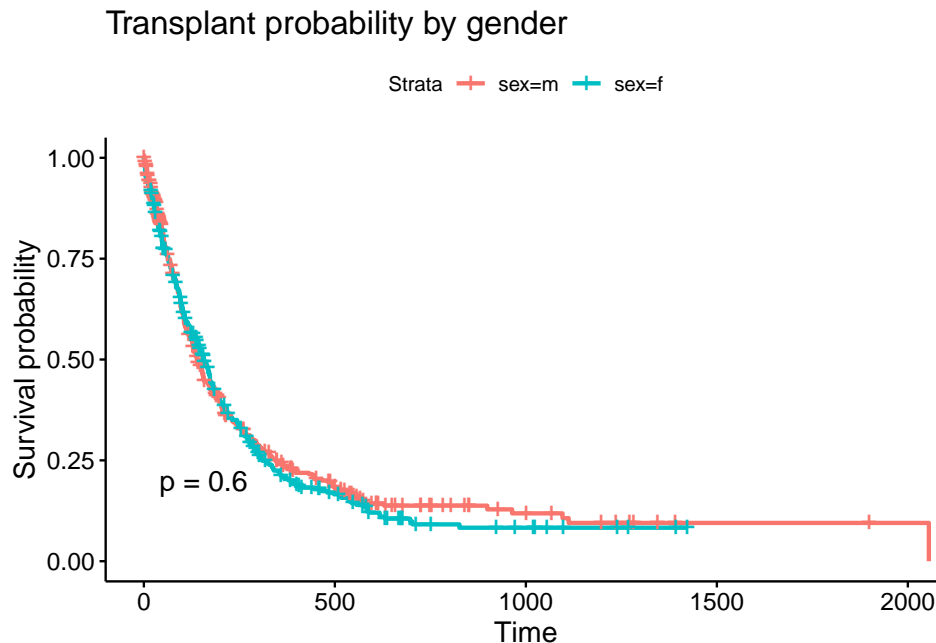
Kaplan-Meier Method and Log Rank Test

The simplest survival modeling is the Kaplan-Meier model. It is non-parametric method used to estimate the survival function from observed survival times in a sample. It is a good starting point to get an idea of the survival rates of our dataset.

Note that the Kaplan-Meier analysis only considers one outcome at a time. Therefore when using Kaplan-Meier on a given outcome (say death), it considers that no alternate outcome existed. In other words, it considers that other outcomes are censored (say transplant) and therefore will overestimate the hazard of the studied outcome. If we study the death outcome, then it will consider that the patients that received a transplant have merely been censored from the study, and therefore have an equal chance of dying as other patients who haven't been transplanted. We will talk more about this in the last section.

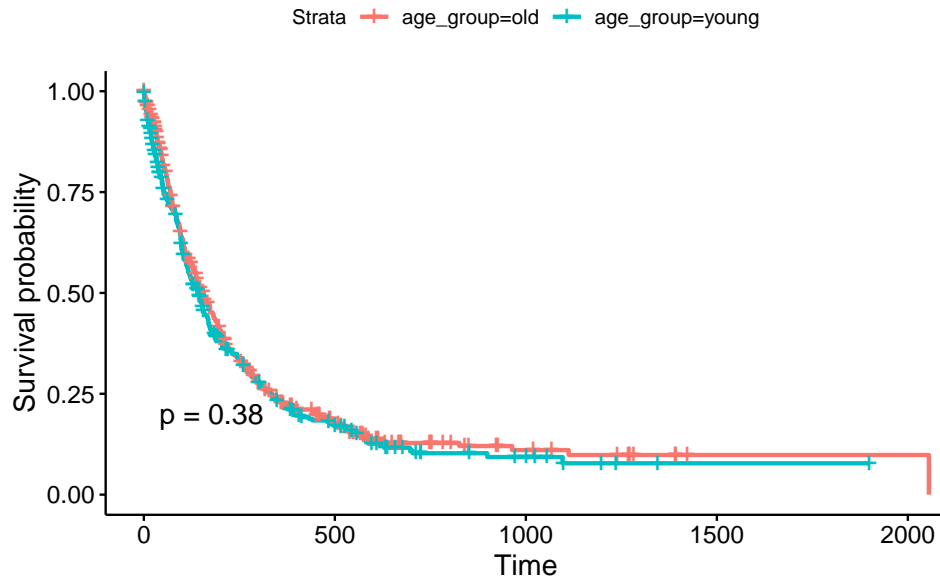
We will plot the survival probability of the transplant and death case considering the different covariates : sex, age_group and blood type. The survival probability represents the likelihood that an event of interest has not occurred up to a certain point in time.

Kaplan-Meier modeling of transplantation hazard:



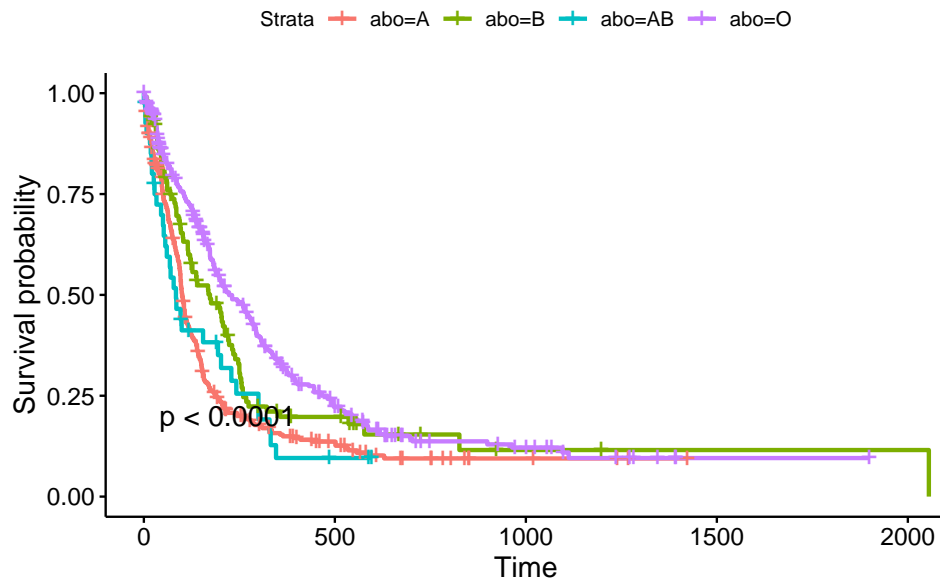
This plot compares the 2 groups male and female considering the outcome transplantation. The p value of 0.6 which is greater than 0.05 indicates that the difference between the 2 groups is not statistically significant. We can then conclude that the gender has no impact for the transplantation outcome.

Transplant probability by age



We now consider the transplantation outcome by age. As described previously, the sample has been divided into 2 groups : patients younger than 50 years old and older than 50. The p-value of 0.38 means that there is no significant difference between the 2 groups for transplanation.

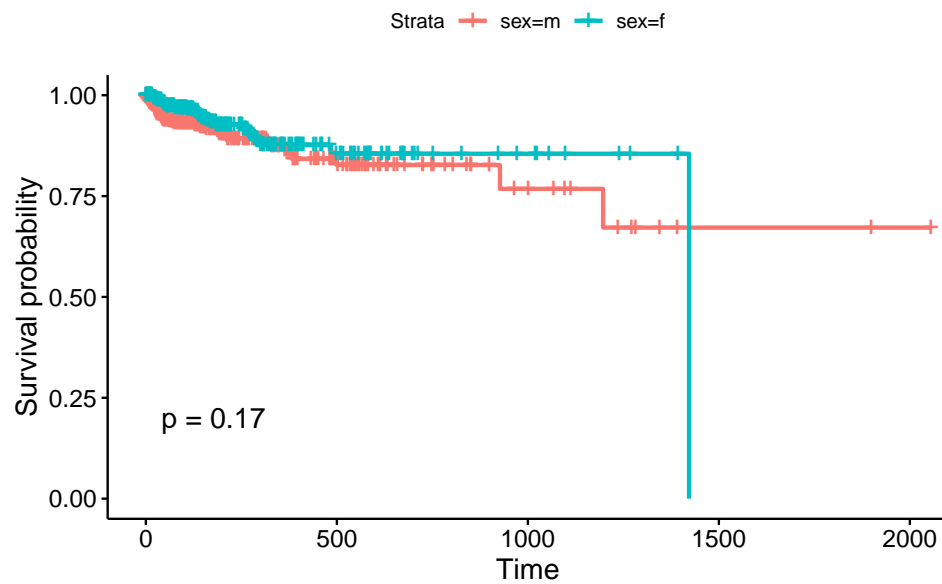
Transplant probability by blood type



The p-value which is 0.001, indicates a significant result if considering that $p < 0.05$ denotes statistical significance. In this study, blood groups have a significant impact on the patient's transplant waiting time. Overall, the curves indicate that patients with blood groups AB and A have a shorter waiting time compared to others, and patients with type O wait longer. This is explained by the fact that livers from donors with blood group O can be used by patients with blood groups A, B, AB, or O. This is what we observe when looking at the curves: the O survival is always above the other blood types. This means that patients with O blood type wait the longest for a transplant.

However, one may wonder if the waiting time has an impact on the patient's survival: let's analyse the death outcome.

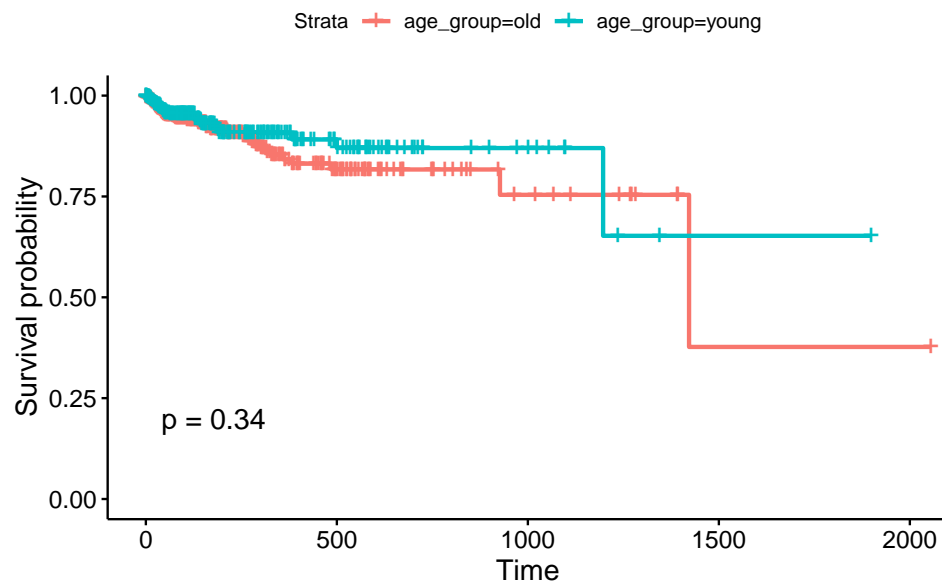
Survival probability considering the gender



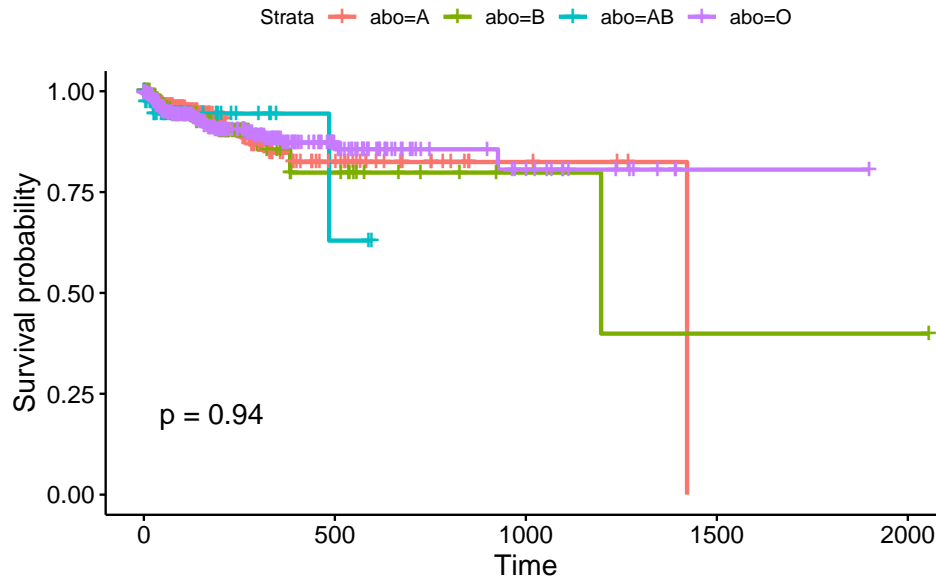
The p-value of 0.17 is greater than 0.05 indicates that there is no significant difference between the 2 genders.

For age and blood type, we observe the similar results: neither the age nor the blood type have an significant impact in the survival probability in the death outcome ($p=0.34$ and $p=0.94$ respectively).

Survival probability considering the age



Survival probability considering the blood type



The analysis using Kaplan Meier indicates that:

- For the transplant survival rate, only the blood type has a significant impact
- For the death survival rate, none of the studied covariate have a significant impact

Advanced analysis

We have seen a basic analysis with the Kaplan-Meier model, it however presents some limitations. Most notably, it can only compare two discrete groups at a time. Secondly, it fails to take into account the multiple outcome as we've seen previously. In section, we will see more advanced models.

First the Cox Proportional Hazards (CoxPH) modeling. The main advantage is that it is capable to perform regressions, and in particular to completely model the age covariate. Like Kaplan-Meier however, it can only analyse one outcome.

Finally, we will see Competing Risks Regression (CRR) which takes into considerations all outcomes.

Cox Model

The Cox Proportional Hazard model works by computing a baseline hazard h_0 and then fitting the proportional hazards h_i for each studied covariate. If the proportional hazard is larger than 1, then it means the covariate presents a higher risk of having the studied outcome and vice versa compared to the baseline.

First, let's see the survival rates for the transplant outcome:

We use the CoxPH model to model the impact of our explanatory covariates : sex, abo, age. The model computes the Hazard Ratios (HR) and the associated Confidence interval (CI) and p-value. We consider the covariate to be statistically significant if the p-value $p > 0.05$.

```
## explanatory transplant HR (LCI, UCI, p)
## 1      sexf      1.07 (0.91-1.25, p=0.426)
## 2      aboB      0.71 (0.55-0.92, p=0.009)
## 3      aboAB     1.03 (0.72-1.48, p=0.862)
## 4      aboO      0.56 (0.48-0.67, p<0.001)
## 5      age       0.99 (0.99-1.00, p=0.189)
```

Let's analyse how each covariate impacts the transplant hazard:

- Sex: the HR is close to 1 and the p-value is quite large. This means that there is no significant difference in the transplant chances between male and female patients.
- Blood type: Our model compares the Hazard Ratio with respect to the A blood type. The B and O blood types are significantly different from the A blood type ($p=0.009$ and $p<0.001$ respectively). We note that for both of them, the $HR < 1$, meaning that the transplantation hazard is higher than compared to blood type A. This makes sense for group O because there are less donors that are compatible than for group A. For blood type B, the lower prevalence of B type people in the US population could account for the lower hazard rate.
- There seems to be absolutely no impact of age on transplant probability ($HR = 0.99$, $p\text{-value}=0.189$). We find this result surprising.

When focusing on death survival:

```
## explanatory death HR (LCI, UCI, p)
## 1 sexf 0.68 (0.41-1.13, p=0.138)
## 2 aboB 1.17 (0.55-2.49, p=0.693)
## 3 aboAB 1.19 (0.36-4.01, p=0.774)
## 4 aboO 1.00 (0.58-1.75, p=0.989)
## 5 age 1.02 (0.99-1.05, p=0.152)
```

Let's analyse how each covariate impacts the death hazard:

- Sex: the HR is lower than 1 (0.68) and the p-value greater than 0.05. This means that in our study, women are on average less likely to die while on the waiting list but this result is not statistically significant. We cannot conclude there are significant differences in the death chances between male and female patients.
- Blood type: Our model compares the Hazard Ratio with respect to the A blood type. The B, AB and O blood types are significantly equivalent from the A blood type ($p > 0.7$). This means that there is no significant difference in the death chances between each blood type patients.
- There seems to be absolutely no impact of age on death probability ($HR = 1.02$, $p\text{-value}=0.155$).

Overall, the death CoxPH analysis concludes that none of the studied covariates significantly impact the death hazard.

Competing Risk Regression

As mentioned previously, we have multiple competing outcomes. It is important to consider that the death and transplant outcomes are competing. If we don't, we then consider that the outcome we are not studying is considered as censored. This is an incorrect reasoning as the alternative outcome should not be considered as a censored one.

Say we study the death while waiting for a transplant. If we consider that transplants are censored observations, then the model will overestimate the death risk. Therefore the above analysis isn't entirely rigorous.

A solution to this is to consider Competing Risks Regression (CRR) first proposed by Fine and Gray in 1999. This model takes into account that we are in a competing risk setting.

We will model our data for the covariates `age`, `sex` and `abo` using the CRR model.

First focusing the the transplant hazard:

```
## explanatory transplant CoxPH HR (LCI, UCI, p) transplant CRR HR (LCI, UCI, p)
## 1 sexf 1.07 (0.91-1.25, p=0.426) 1.12 (0.96-1.32, p=0.150)
## 2 aboB 0.71 (0.55-0.92, p=0.009) 0.72 (0.55-0.93, p=0.013)
## 3 aboAB 1.03 (0.72-1.48, p=0.862) 1.01 (0.67-1.52, p=0.960)
## 4 aboO 0.56 (0.48-0.67, p<0.001) 0.59 (0.49-0.70, p<0.001)
## 5 age 0.99 (0.99-1.00, p=0.189) 0.99 (0.98-1.00, p=0.068)
```

For each covariate, we are comparing the computed HR and associated p-value of the previously computed the CoxPH model with the newly computed CRR model.

Our first observation is that the newly HR are very similar to the ones computed by CoxPH with variations of at most a few percents. We note that the computed p-values change slightly, for age and sex, they become closer to 0. This means that the CRR model finds that the effect of sex and age to be slightly more impactful on the transplant hazard ratio. However, it doesn't make them statistically significant either.

Overall, using the CRR model gives slightly different HR and p values. However, it doesn't change the previous conclusions: the statistically significant covariates remain significant.

Secondly, focusing the the transplant hazard:

##	explanatory	death CoxPH HR (LCI, UCI, p)	death CRR HR (LCI, UCI, p)
## 1	sexf	0.68 (0.41-1.13, p=0.138)	0.66 (0.40-1.08, p=0.097)
## 2	aboB	1.17 (0.55-2.49, p=0.693)	1.51 (0.71-3.18, p=0.280)
## 3	aboAB	1.19 (0.36-4.01, p=0.774)	1.15 (0.34-3.91, p=0.820)
## 4	aboO	1.00 (0.58-1.75, p=0.989)	1.48 (0.85-2.56, p=0.170)
## 5	age	1.02 (0.99-1.05, p=0.152)	1.02 (1.00-1.04, p=0.094)

Similarly to the transplant hazard, we observe some slight changes in the HR and p values when compared to the previously computed CoxPH model.

The most notable changes are for the B and O blood types where in both cases the HR increases by a factor of about 1.5. However, like in the Cox model, none of the covariate impact significantly the death rate. The lowest observed p-value is of 0.094 for the age.

Once again, using CRR instead of CoxPH yields some slight changes in the computed HR and p-values, but doesn't change the conclusions we've drawn previously.

The use of CRR doesn't seem to be all that relevant in our analysis, however, we are convinced that it is methodologically more rigorous to use it over CoxPH and Kaplan-Meier models. Especially when we see no drawback to use CRR over other models.

Conclusion

In this work, we have studied the survival analysis of the patients waiting for a liver transplant. Our focus was to study the impact of various medical factors (sex, age and blood type) on two possible outcomes:

- The patient receives a liver transplant
- The patient dies while on the waiting list

We have tested different models of varying complexity. The Kaplan-Meier model, the Cox Proportional Hazards model and the Competing Risks Regression model. All of the previous models arrived to the same conclusion:

- For the transplant outcome:
 - Age and Sex doesn't significantly impact the transplant hazard
 - Blood types AB doesn't significantly impact the transplant hazard compared to blood type A (baseline)
 - Blood types O and B have a significant impact on the transplant hazard: their transplantation hazard is lower compared to blood type A
- For the death outcome:
 - No covariates have a significant impact on the death hazard

We believe that the CRR model is the most rigorous and should be used to infer predicted survival data.