Katholieke
Universiteit
Leuven

**Faculty of
Science**

ADVANCED TOPICS IN SURVIVAL ANALYSIS

Competing Risk in Liver Tumor Surgery Patient

Federico Grazi (r1028201)

Academic year 2024–2025

Contents

1 Introduction

The dataset follows 425 patients that undergo surgery from 2011 to 2024 to remove a liver tumor with general information about the patient, pre-surgery clinical data and post-surgery status reported in the dataset. This time-to-event dataset focus on capturing the time until the death of the patient from the day of the surgery, hence for each dead patient, the cuase of the decess was noted. Precisely, it's noted wether the patient died due to the tumor, or if any other causes led to the death of the individual.

Most of the *other causes* are either diseases, such as covid, or some types of shocks that led to the death of the patient. A few cases also experienced respiratory failure or thrombosis. Other causes of death are for the most part health-related and can be explained as a consequence of the surgery, it is indeed possible to consider what it is labelled as *death by other causes* (DOC) as a competing risk to the death of the patient.

Due to the nature of cancer, it may happen that some patient that undergo surgery develop some kind of relapse from the tumor and others will be cured. This phenomenum is captured well by a cure model and a simple survival fit, as shown in Figure 1, clearly depicts the presence of a cure fraction of approximately 50%.

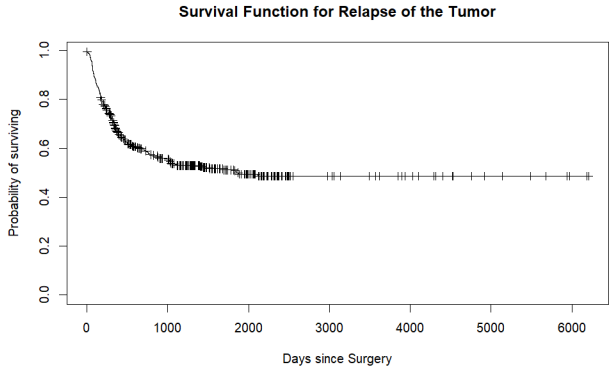

Figure 1: Survival function of the relapse of the tumor

The possibility of using competing risk model along with the integration of a cure fraction is studied by Nicolaie et al. (2018) and Chen et al. (2020) but an eventual implementation of such methods goes beyond the scope of this report.

Going back to the nature of the problem, a possible research question to be answered is how to reduce death by tumor after a tumor surgery. One would expect to reduce the number of death from the cancer if the cancer is either partially or fully removed, thus having such proportion of patient that experience death caused by the tumor a surgeon operated on may not be seen as success of the operation, thus estimating the probability of the patient dying from a certain cause can be of interest for managing post-surgery patients.

This report is structured in a first part of dwelves into the dataset to take a brief look difference between the two causes of death and the variable distribution. A second part will introduce the models used and will quickly go over the methodology. At last, the results and conclusions will be shown.

2 Data Exploration

In the competing risk anlysis a critical assumption is that there may be a relation between the risks, thus analyzing the marginal distribution with patient that experienced event due to a competing risk as censored would leads to dependent censoring.

For this case, the tumor research context, a general worsening of a patient's health will lead to a bigger exposure of other health threat, thus we shall not consider death to another cause as a censoring event and consider censored patient only those that did not experience the event due to the end of the study. This specification can be regarded as a situation in which the censoring time is independent to the event occuring.

In many survival analysis studies that experience administratetively censoring, theoretically, that if the study kept going every patient would have experienced an event. For tumor patient it is not sure that all patient would eventually experience death by tumor due to the possibile presence of a cured proportion in the population. Indeed, a possible path for a patient considering the underlying cured proportion is shown in Graph 1.

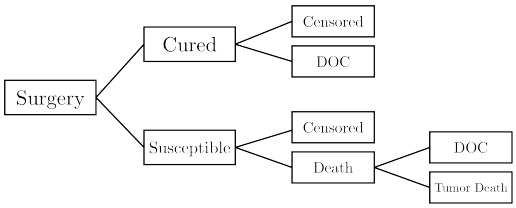

Figure 2: Possible path leading to the event for a patient that undergo surgery

Most of the patient who died died (reported in Table 1) due to tumor and only 10% of the patient experienced DOC with most of the patient that died in the time the study was conducted dying of tumor.

Table 1: Table of the status

Status	Patient
Censored (0)	269
Death by other cause (1)	54
Death by tumor (2)	102

2.1 Covariates

General knowledge about the patient health before the surgery is described by the age, the presence of hypertension in the patient and its bmi value. Other variables such as the indicators for loss of blood during the surgery or for complications after the surgery indicate wether the surgery went smoothly or not.

In total this dataset has 10 variables that may explain the survival rate for each patient. Right-censoring, or more precisely *administratetively censoring*, occured due some patient being event free at the time of the analysis.

A quantity of interest may be the relationship between indicators of a smooth operator with the response status.

A strong relation between the variable `lenghtSurgery` and the binay variables `bloodLossDuringSurgery` and `postSurgeryComplications` conditioned to the status is shown in Figure 3.

A patient that experienced both complication, allegedely resulting in a more difficult surgery, have a longer duration of the surgery. This is particularly noticeable in the case when the patient later died by tumor. While the second panel does not induce any particular evidence for a relationship, the third panel depicts a strong relation of blood losses and the length of the surgery when the patient eventually died by tumor. At last, in the case of no complication or blood loss, the length of the surgery resulted to be smaller for the individuals that end up dying by other causes.

This lets us conclude that there could be a positive relation between the duration of the surgery and the subsequent death caused by the tumor.

A similar analysis is obtained for the relapse of the tumor after it was removed against its size. It is intuitive to assume that bigger tumor are harder to remove and may come back as a consequence of a not perfect surgery. In general, the average dimension of the tumor is slightly larger when it relapsed (29,6 mm

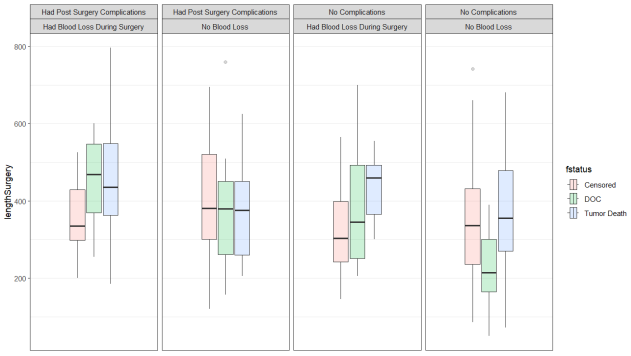

Figure 3: Conditional distribution of Lenght of the Surgery

versus 30,4 mm) as Figure 4 summarizes. Interestingly enough, patient that died due to other causes but had a relapse of tumor had an average tumor size bigger than those that died due to the tumor itself, this could be explained by the dependence between the risks mentioned before.

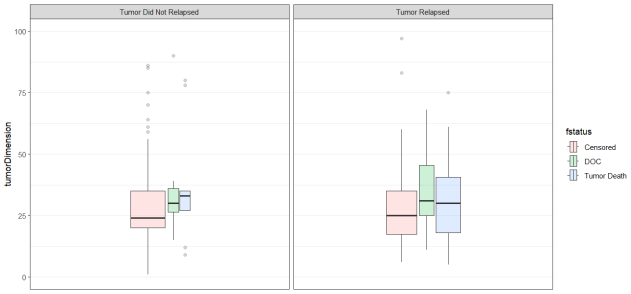

Figure 4: Conditional distribution of the Dimension of the Tumor. Width of the boxplot is proportional to the observation used in it.

Figure 5 shows instead the distributions and the scatterplot of some continuos variables. The variable `daysSinceSurgery` is our observed time until one of the three events occured.

Both the smooth estimate being nearly flat and the small correlation suggests that the variables are not related to each other, except for a small negative relation between `daysSinceSurgery` and `lenghtSurgery`. Anyway, a closer look to the scatterplot reveals that the relation may be driven due to the small cluster of censored observation that did not die and underwent surgery many years ago.

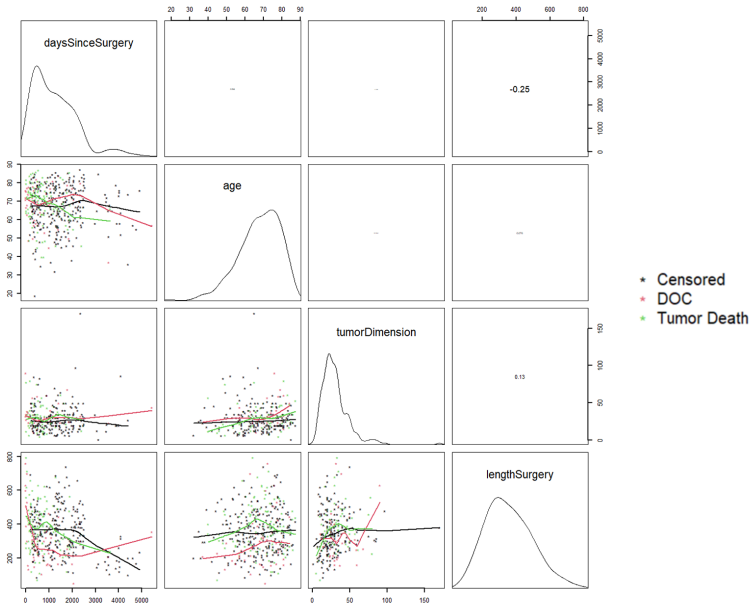

Figure 5: Pair plot for the continuos variables. Correlation on the upper panel, scatterplot on the lower panel, and densities on the diagonal.

3 Competing Risk Model

Competing risk are event-type outcomes that are mutually exclusive. Let a random variable $T_k \sim I_k$ represent the time untile the event from cause k occurs. I_k is commonly referred to as cause-specific distribution, or cumulative incidence function (CIF): nonetheless I_k is technically referred to as a *sub-distribution* which reflects that I_k does not add up to 1 due to the competing risk nature of the problem.

With some simple notation we can write

$$P(T_k \leq t) = P(T \leq t, D = k)$$

where D is the random variable that represents the cause taht is observed.

The two main approaches to estimating that probability are based on the *sub-distribution* approach and on the *cause specific* representation.

3.1 Cause Specific Distribution

This first specification assumes a certain dependence between the different causes, thus considering the competing risk event as a censoring event violates the independence between censoring and event time. Hence, an individual is taken out from the risk set when it experiences a comepting event.

This leads to the cause-specific hazard defined as

$$\lambda_k(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T_k < t + \Delta t \mid T_k \geq t)}{\Delta t} \tag{1}$$

This specification relates the hazards with the CIF trough

$$I_k(t) = \int_0^t \lambda_k(s) S(s) ds \tag{2}$$

$$S(t) = \exp\left(- \int_0^t \sum_k \lambda_k(s) ds \right), \tag{3}$$

whereas using the cause specific hazards can be seen as special case of the multi-state approach, thus, effects on the cumulative scale can be in different direction than those of the cause specific hazards.

The estimation procedure in this case is as follows

$$\hat{\lambda}_k(t_{(j)}) = \frac{d_k(t_{(j)})}{r(t_{(j)})}$$
$$\hat{I}_k(t_{(j)}) = \sum_{t_{(i)} \leq t} \hat{S}_{KM}(t_{(i)-})\hat{\lambda}_k(t_{(i)})$$

where $d_k(t_{(j)})$ the total number of observed events at ordered time $t_{(j)}$, $r(t_{(j)})$ denotes the number of patient observed to be at risk and $\hat{S}_{KM}$ is the Kaplan-Meier estimate.

3.1.1 Regression Model

When fitting a regression model in the cause-specific approach it simply reduces to fitting a regression model in the standard setting with only one cause of death due to the fact that other event are considered as censoring events.

A normal Cox proportional hazard model can be used, but the estimated coefficents don't reflect the marginal hazard as the different causes are assumed to be dependent.

3.2 Sub Distribution

This approach is based on the sub-distribution hazard h_k defined as

$$h_k(t_j) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t, D = k \mid T \geq t)}{\Delta t}, \tag{4}$$

which relates to the survival function and CIF trough

$$S_k(t) = \exp\Big(- \int_0^t h_k(s)ds \Big) \tag{5}$$
$$I_k(t) = \int_0^t h_k(s)S(s-)ds, \tag{6}$$

where $S(t) = \exp\big(- \sum_k \int_0^t h_k(s)ds \big)$.

The difference from the cause-specific hazard in 1 is that one can show that the denominator of $h_k(t)$ contains the probabilities to be *event free* or having experienced another event due to a cause $\ell \neq k$. Thus the risk set also contains unit for which another cause occured, while the cause-specific hazard removes them from the risk set, and this can be viewed from the element to which we condition the probability.

The estimation procedure in this case is as follows

$$\hat{h}_k(t_{(j)}) = \frac{d_k(t_{(j)})}{r^*(t_{(j)})}$$
$$\hat{S}_k(t) = \prod_{t_{(i)} \leq t} 1 - \hat{h}_k(t_{(i)})$$
$$\hat{I}_k(t) = \sum_{t_{(i)} \leq t} \frac{d_k(t_{(i)})}{N}$$

where $d_k(t_{(j)})$ the total number of observed events at ordered time $t_{(j)}$ and $r^*(t_{(j)})$ denotes the number of patient in the risk set definied above.

Moreover, it can be seen that $\hat{S}_k(t)$ is simply a Kaplan-Meier estimator with an adapted risk set r^*

3.2.1 Regression Model

The approach by Fine and Gray (1999) proposes a Cox type regression for the subdistribution, with the advantage of an easier interpretation of the coefficient rather than the cause specific one.

A proportional subdistribution hazard model is defined as

$$h_k(t \mid \boldsymbol{X}) = h_{k0}(t) \exp\left(\boldsymbol{\beta}_k^T \boldsymbol{x_i}\right) \tag{7}$$

where h_{k0} is a baseline subdistribution hazard.

Liu et al. (2016) show that another possible model can be built from the cumulative incidence after considering that at a certain time t an individual either has experience a competing event or is event free. This formulation leads to a model based on the odds of occurrence of the event of interest:

$$\text{logit}(I_k(t \mid \boldsymbol{X})) = \log(\pi_{k0}(t)) + \boldsymbol{\beta}_k^T \boldsymbol{x_i}. \tag{8}$$

where π_{k0} is a baseline odds. This model is referred to as the *proportional odds* CIF model.

4 Results and Discussion

For completeness, both a cause specific model and a subdistribution PH model will be reported, as they capture different aspects both on the hazard rate scale and on the cumulative scale. At last, also the proportional odds model will be implemented and compared with the previous two.

Given the nature of the problem, a cause specific hazard model better reflects the fact that a patient that died due to a cause cannot later die due to the other, hence it shall be removed from the risk set.

4.1 Cumulative Incidence Function

A first simple CIF function presented in Equation 6 can be computed using the `cuminc` function in the package `cmprsk` in R.

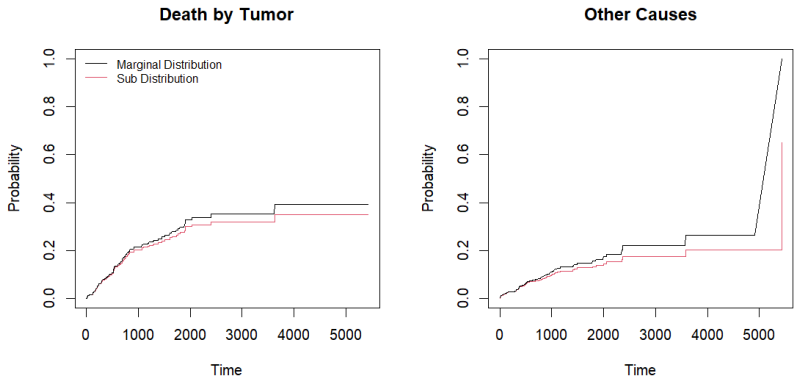

Figure 6: Subdistribution CIF for both causes compared with cause specific CIF.

The probability of dying by the tumor rapidly climbs up in the first 1000 days after the surgery, then begins slightly plateauing and settling around 40%. Conversely, the DOC panel depicts a steadier function all throughout the time period except for a spike at the end.

The slow increase in the death by tumor panel can also represent the underlying cured proportion structure in the population: in fact, it is charachteristic of the cure model to have a plateu in the survival function. If a patient were to be cured from the tumor, its possible results are either censored or DOC, thus we expect the DOC to be a proper survival function, and it can be seen that the marginal distribution CIF in the second panel reaches 1.

Moreover, it is possible to see the effect of the subdistribution being always smaller than the marginal distribution due to

$$\begin{aligned}
1 - S_k(t) &= \int_0^t h_k(s) S_k(s) ds \\
&= \int_0^t h_k(s) \exp\Big(- H_k(t) \Big) ds \\
&\geq \int_0^t h_k(s) \exp\Big(- \sum_k H_k(t) \Big) ds \\
&= \int_0^t h_k(s) S(s) ds \\
&= I_k(t).
\end{aligned}$$

The relapse of the tumor shown in Figure 7 leads to an important increase in the probability of dying due to the tumor itself. Conversely, the CIF of a DOC with a relapse of the tumor is the lowest of the four, representing that if the tumor relapse, it's quite difficult to die not due to the tumor.

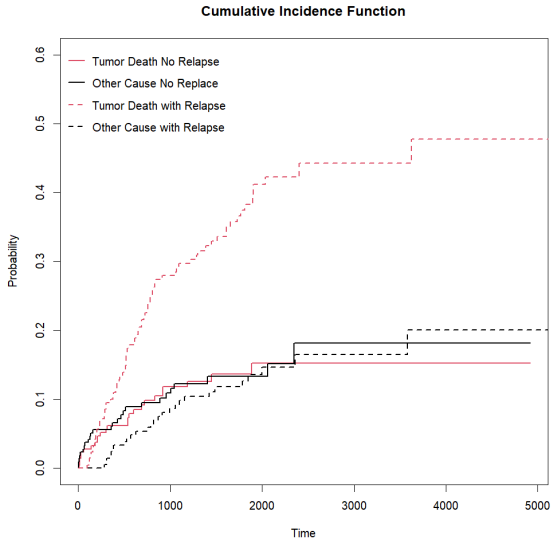

Figure 7: Subdistribution CIF conditioned to having a relapse of the tumor.

4.2 Model Results

The results for the models are reported in Table 2. The table shows the difference with fitting the model one the cause specific and the subdistribution. While the latter two are more coherent with each other, the cause specific model returns different results from the last two.

The cause specific model signal two positive relative changes in the instantaneous rate of dying due to tumor for an increase in age and the length of the surgery.

The subdistribution models signal the indicator of a relapse as the most impactfull covariate when the patient dies by tumor. In the subdistribution PO model, the odds for the relapsed tumor patient to die due to the tumor is $\exp(1.57) = 4.807$ times the patient that do not experience a relapse. For the subdistribution PH model the conclusion about the importance of the covariates is exactly the same as for the proportional odds model, but the interpretation differs. Given the relation in Equation 5 the effect is the log of the ratio of $-\log\big(1 - F_k(t)\big)$, hence the log of the probability of experiencing death by tumor or surviving have the ratio $\exp\big(1.57\big)$.

While there is a general lack of significance in the coefficients, another important covariate we discussed earlier and can be seen now having a meaningfull impact is the lenght of the surgery. An increase by an hour of the length of the operation increases the probability of experiencing death by tumor by respectevely of 18% and 24% in the PH and PO model with the latter interpretation.

The AIC information criteria shows thath the two subdistribution model have similar values, while the cause specific model outperforms the subdistribution ones even if it lacks in capturing the relation between the variable `tumorBack`.

Table 2: Estimated Hazard Ration. Standard Errors are reported in parenthesis.

	Cause Specific PH		Subdistribution PH		Subdistribution PO	
	DOC	Death by Tumor	DOC	Death by Tumor	DOC	Death by Tumor
Age	0.992	1.027*	1.009	1.014	1.01	1.019
	(0.013)	(0.01)	(0.011)	(0.011)	(0.012)	(0.013)
Sex (1 = female)	0.676	1.059	0.944	0.693·	0.942	0.698
	(0.311)	(0.209)	(0.283)	(0.206)	(0.306)	(0.247)
Hypertension (1 = yes)	0.847	0.726	0.561.	1.481·	0.533·	1.527
	(0.353)	(0.227)	(0.336)	(0.21)	(0.357)	(0.256)
Lenght of the Surg. (h)	0.857·	1.144**	0.85*	1.184***	0.851.	1.247***
	(0.084)	(0.047)	(0.08)	(0.047)	(0.088)	(0.062)
Blood Loss during Surg. (1 = yes)	1.451	1.311	1.327	0.685	1.43	0.604
	(0.354)	(0.249)	(0.342)	(0.273)	(0.371)	(0.314)
Post Srug. Complication (1 = yes)	1.075	0.957	1.299	0.914	1.34	0.921
	(0.355)	(0.236)	(0.298)	(0.225)	(0.319)	(0.264)
Relapse of Tumor (1 = yes)	1.159	1.07	0.792	3.868***	0.761	4.807***
	(0.315)	(0.208)	(0.279)	(0.239)	(0.299)	(0.291)
AIC	400.85	896.16	543.27	1050.59	544.08	1051.38
Number of Units(1)	359	359	359	359	359	359
Number of Events	43	93	43	93	43	93

Note: $0.05 <^{\cdot} < 0.1; 0.05 <^{*} < 0.01;\ 0.001 <^{**} < 0.01;\ 0 <^{***} < 0.001$

(1): 33 observation were removed due to missing values

A more concrete representation of the effects of the covariates is seen after predicting the CIF for four individual with fixed covariates except the length of the surgery and the relapse of the tumor. The range of the surgery varies between 40 minutes and 13 hours: 2 hours and 7 hours were chosen to observe the different effects.

For long surgery one can observe a smaller CIF for non-tumor death regardless of a relapse, while shorter surgeries lead to higher CIF of non-tumor death, as Figure 8 shows. Conversely, the second graph shows a more diverse scenario. The relapse of the tumor has a huge impact only when the surgery required a long time, resulting in a CIF that jumps to high probabilities of dying due to the tumor in a fairly short period of time. It is interisting to also notice that a short surgery and a relapse of the tumor differs only for a long-term effect to a long surgery without relapsing[1].

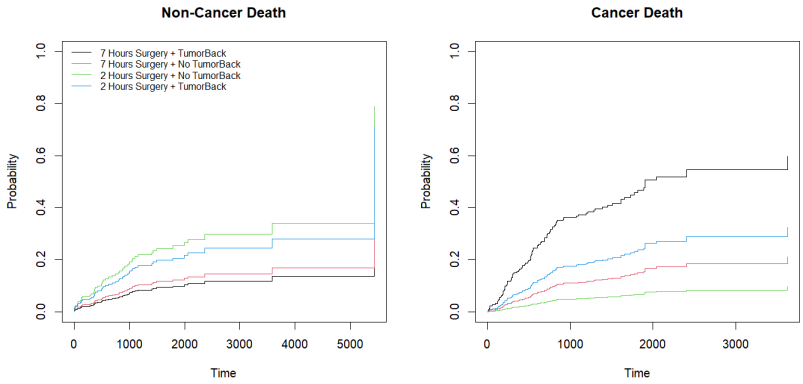

Figure 8: Different profiles for the two subdistribution PH model.

The result from the PH model are quite similar to those obtained with the PO also for the prediction (Figure 9).

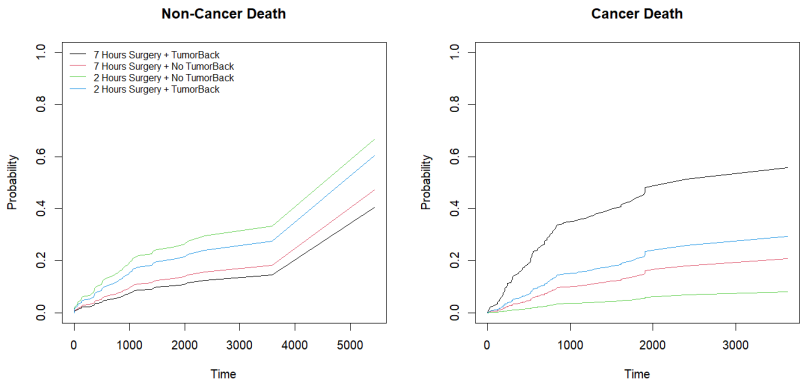

Figure 9: Different profiles for the two subdistribution PO model.

4.3 Discussion

The raw dataset contained many more variables, mainly indicators or some clinical variables. Unfortunately, due to the high number of missing values, most of those variables where not suited for the analysis. Table 3 summarizes the number of missing values for each variable: the variable `bmi`, `daysAfterDischarge` and `tumorDimension` where discharged because the log likelihood would not converge when fitting the models. Indeed, too few observation were used to fit the model and the resulting coefficients in the range between 5 and 10, resulting in hazard in the order of 10^3.

[1] A 95% CI for the predicted CIF could have confirmed this impression and could have been obtained with the function `comp.risk` of the package `timereg`. Unfortunately, running the function lead to a constant crash of the software, hence further CI results are not discussed.

Table 3

Variable	Missing Values
bmi	84
age	0
sex	0
hypertension	0
daysAfterDischarge	23
tumorDimension	141
lengthSurgery	33
bloodLossDuringSurgery	0
postSurgeryComplications	0
tumorBack	0

Even if common knowledge would suggest a stronger relationship between the covariates and the response, the model did not output any clear evidence besides two covariates. Another cause of possible flaw of the model specification was the presence of a cured proportion, which was not modeled and could have lead to different results.

The difference between the cause specific model and the subdistributions ones is evident both in the coefficients value and their significance. This was an expected result since the two method treat the competing risk in two different ways, especially in the risk set definition and the effect either on the hazard or on the cumulative rate.

Moreover, the general insignificance of the age also raised some insights on the dataset: mainly, the patient were mostly clustered in the range 60-80. Possibly, having a larger sample size could lead to better insights on the effects of the tumor on younger patients. Indeed, having a broader set of patients and a better reported raw dataset most likely would have led to different results. Consired the profound relation between medical information, it may be possible that the effect of a covariate is correlated with another unobserved variables, thus determining possible estimation problem when using all variables.

At last, a notable difference was found between the two competing risk models, suggesting that indeed the two processes are driven by different dynamics, even if the main difference is found in the surgery itself and an additional information about the relapse of the tumor. Hence, the results suggests that it is not possible to discriminate between the competing risk using covariates about pre surgery information only.

5 Conclusion

A competing risk analysis was carried out to a dataset of patient that underwent a surgery to remove a liver tumor. Patient then were followed until their death: each death was classified either as death due to the tumor or death due to another cause, which were mainly other diseases.

The data exploration showed some not too strong relations between the covariates and the response. The length of the surgery was early seen as a possible important factor for the regression on the survival time, and later this was confirmed.

The general competing risk setting was described along the cause specific approach to estimate the CIF and also two subdistribution approaches were briefly introducted.

The regression showed different result between the approaches, which were expected due to theoretical properties, but similar results were obtained between the two subdistributions regressions.

Finally, possible flaws and improvement were presented, mainly in the model specification and the dataset part. A more complex analysis along with a more precise data entry process for each patient could lead to more complex but more correct results.

Bibliography

Nicolaie, Mioara Alina, Taylor, Jeremy M. G., and Legrand, Catherine. Vertical modeling: analysis of competing risks data with a cure fraction. *Lifetime Data Analysis*, 25(1):1–25, January 2018. doi: http://dx.doi.org/10.1007/s10985-018-9417-8.

Chen, Chyong-Mei, Shen, Pao-Sheng, Lin, Chih-Ching, and Wu, Chih-Cheng. Semiparametric mixture cure model analysis with competing risks data: Application to vascular access thrombosis data. *Statistics in Medicine*, 39(27):4086–4099, November 2020. doi: http://onlinelibrary.wiley.com/termsAndConditions#vor.

Fine, Jason P. and Gray, Robert J. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94(446):496–509, 1999. doi: https://doi.org/10.1080/01621459.1999.10474144. Available at: https://www.tandfonline.com/doi/abs/10.1080/01621459.1999.10474144.

Liu, Xing-Rong, Pawitan, Yudi, and Clements, Mark. Parametric and penalized generalized survival models. *Statistical Methods in Medical Research*, 27(5):1531–1546, September 2016. doi: http://dx.doi.org/10.1177/0962280216664760.

Appendix: R Code

Libraries

```
library(readxl)
library(tidyverse)
library(modelsummary)
library(lattice)
library(survival)
library(ggfortify)
library(eventglm)
library(timereg)
library(cmprsk)
library(Hmisc)
```

Introduction

Figure1

```
cure <- survfit(Surv(relapse$daysToRelapse,relapse$Relapse)~1)
plot(cure,mark.time=T,conf.int=F)
```

Data Exploration

Table 1

```
status <- liver$deathByTumor+liver$death
# 0 = censored
# 1 = Dead
# 2 = Death by Tumor
table(status)
```

Figure 3

```
liver %>%
    mutate(
        fstatus = factor(case_when(
        status == 0 ~ "Censored",
        status == 1 ~ "DOC",
        TRUE~ "Tumor Death"
        )),
        postSurgeryComplications = factor(ifelse(postSurgeryComplications==1,
                                                 "Had Post Surgery Complications",
                                                 "No Complications")),
        bloodLossDuringSurgery = factor(ifelse(bloodLossDuringSurgery==1,
                                                 "Had Blood Loss During Surgery",
                                                 "No Blood Loss"))
        ) %>%
    ggplot(aes(x = lengthSurgery, group = fstatus))+
    geom_boxplot(stat = "boxplot", varwidth =T , aes(fill = fstatus), alpha = 0.2)+
    coord_flip()+
    scale_y_discrete() +
    facet_wrap(~postSurgeryComplications + bloodLossDuringSurgery,ncol = 4, nrow = 1)+
    theme_bw()
```

Figure 4

```
liver %>%
    mutate(status = status) %>%
    group_by(tumorBack, status) %>%
    summarize_all(mean, na.rm = T) %>%
    select(status, tumorDimension)

liver %>%
    mutate(
        fstatus = factor(case_when(
        status == 0 ~ "Censored",
        status == 1 ~ "DOC",
        TRUE~ "Tumor Death"
        )),
        tumorBack = factor(ifelse(tumorBack==1, "Tumor Relapsed", "Tumor Did Not Relapsed")),
        ) %>%
    ggplot(aes(x = tumorDimension, group = fstatus))+
    geom_boxplot(stat = "boxplot", varwidth = T, aes(fill = fstatus), alpha = 0.2)+
    coord_flip()+
    scale_y_discrete() +
    xlim(c(0,100))+
    facet_wrap(~tumorBack,ncol = 2, nrow = 1)+
    theme_bw()
```

Figure 5

```
panel.smooth <- function (x, y, col = par("col"), bg = NA, pch = par("pch"),
            cex = 1, col.smooth = 1:3, span = 2/3, iter = 3, ...){
    points(x, y, pch = pch, col = col, bg = bg, cex = cex)
    ok <- is.finite(x) & is.finite(y)
    if (any(ok))
        lines(stats::lowess(x[ok & status ==0], y[ok & status == 0], f = span, iter = iter),
            col = col.smooth[1], ...)
        lines(stats::lowess(x[ok & status ==1], y[ok & status == 1], f = span, iter = iter),
            col = col.smooth[2], ...)
        lines(stats::lowess(x[ok & status ==2], y[ok & status == 2], f = span, iter = iter),
            col = col.smooth[3], ...)
    }


panel.cor <- function(x, y, digits = 2, prefix = "", cex.cor, ...){
    usr <- par("usr"); on.exit(par(usr))
    par(usr = c(0, 1, 0, 1))
    r <- abs(cor(x, y,use = "complete.obs"))
    sign <- ifelse(cor(x, y,use = "complete.obs")>0,"","-")
    txt <- format(c(r, 0.123456789), digits = digits)[1]
    txt <- paste0(sign,prefix, txt)
    cex.cor <- 0.8/strwidth(txt)
    text(0.5, 0.5, txt, cex = cex.cor * r)
    }

panel.dens <- function(x,...){
    usr <- par("usr"); on.exit(par(usr))
    par(usr = c(usr[1:2], 0, 1.5))
    dens <- density(na.omit(x))
    y <- dens$y / max(dens$y)
    lines(dens$x, y, color = 1)
    }
```

```
pairs(liver[,c("daysSinceSurgery","age","tumorDimension", "lengthSurgery")],
        col = status+1,
        lwd = 2,
        pch = "*",
        lower.panel = panel.smooth,
        upper.panel= panel.cor,
        diag.panel = panel.dens
        )
legend("center",pch = "*" ,col  = 1:3,legend = c("Censored", "DOC", "Tumor Death"), bty = "n")
```

Results

Figure 6

```
status1 <- ifelse(status != 1, 0,1)
status2 <- ifelse(status != 2, 0,1)

KM0 <- survfit(Surv(daysSinceSurgery,status1)~1,data=liver, type="kaplan-meier")
KM1 <- survfit(Surv(daysSinceSurgery,status2)~1,data=liver, type="kaplan-meier")

fitMARG <- cuminc(ftime = liver$daysSinceSurgery,
            fstatus = status,
            cencode = 0)

par(mfrow = c(1,2))
plot(fitMARG$`1 2`$time, fitMARG$`1 2`$est,col=2,type="l",main="Death by Tumor",xlab="Time", ylab=
lines(KM1$time, 1- KM1$surv)
legend("topleft",legend=c("Marginal Distribution","Sub Distribution"),lty = 1, col=1:2, bty = "n",
plot(fitMARG$`1 1`$time, fitMARG$`1 1`$est,col=2,type="l",main="Other Causes",xlab="Time", ylab="P
lines(KM0$time, 1- KM0$surv)
par(mfrow = c(1,1))s
```

Figure 7

```
fit <- cuminc(ftime = liver$daysSinceSurgery,
          fstatus = status,
          group = liver$tumorBack,
          cencode = 0)

est.NOTB.KOTHER <-  fit$`0 1`$est
time0 <- fit$`0 1`$time

est.TB.KOTHER <- fit$`1 1`$est
time1 <- fit$`1 1`$time

est.NOTB.KT <- fit$`0 2`$est
time2 <- fit$`0 2`$time

est.TB.KT <- fit$`1 2`$est
time3 <- fit$`1 2`$time


plot(time2,est.NOTB.KT,col=2, lwd = 2,type="l",main="Cumulative Incidence Function",xlab="Time", y
lines(time0,est.NOTB.KOTHER, lwd = 2,col=1)
lines(time1,est.TB.KOTHER,col=1, lwd = 2,lty = "dashed")
lines(time3,est.TB.KT,col=2, lwd = 2,lty = "dashed")
```

```
legend("topleft",lwd = 2,legend=c("Tumor Death No Relapse","Other Cause No Replace",
                                  "Tumor Death with Relapse", "Other Cause with Relapse"),
       col=c(2,1,2,1),lty=c("solid","solid","dashed", "dashed"), bty = "n", cex = 1)
```

Table 2

```
liver$id <- 1:nrow(liver)
liver$status <- status
colSums(is.na(liver))    # bmi and daysAfterDischarge have too many NAs unfortunately
na.liver <- na.omit(liver[,-c(2,3,4,8,9)])
mod <- na.liver$lengthSurgery%/%60
na.liver$lengthSurgery <- mod+(na.liver$lengthSurgery-mod*60)/60 # Rescale the lenght in hours


# Regression with simple COX PH ####
mdl.ph1 <- coxph(Surv(na.liver$daysSinceSurgery,status[na.liver$id]==1)~.,control =  list(iter.max
> mdl.ph1
    Call:
    coxph(formula = Surv(na.liver$daysSinceSurgery, status[na.liver$id] ==
        1) ~ ., data = na.liver[, -9], control = list(iter.max = 10000,
        timefix = T))

                               coef exp(coef)  se(coef)      z      p
    age                      -0.007772  0.992258  0.013451 -0.578 0.5634
    sex                      -0.391956  0.675734  0.310752 -1.261 0.2072
    hypertension             -0.166342  0.846756  0.353405 -0.471 0.6379
    lengthSurgery            -0.154287  0.857026  0.083765 -1.842 0.0655
    bloodLossDuringSurgery    0.372453  1.451290  0.354258  1.051 0.2931
    postSurgeryComplications  0.072184  1.074853  0.354577  0.204 0.8387
    tumorBack                 0.147433  1.158855  0.314948  0.468 0.6397

    Likelihood ratio test=7.61  on 7 df, p=0.368
    n= 359, number of events= 43
    (33 osservazioni eliminate a causa di valori mancanti)


mdl.ph2 <- coxph(Surv(na.liver$daysSinceSurgery,status[na.liver$id]==2)~.,control =  list(iter.max
> mdl.ph2
    Call:
    coxph(formula = Surv(na.liver$daysSinceSurgery, status[na.liver$id] ==
        2) ~ ., data = na.liver[, -9], control = list(iter.max = 10000,
        timefix = T))

                               coef exp(coef) se(coef)      z     p
    age                       0.02633   1.02668  0.01023  2.575 0.010
    sex                       0.05700   1.05866  0.20937  0.272 0.785
    hypertension             -0.32031   0.72592  0.22659 -1.414 0.157
    lengthSurgery             0.13447   1.14393  0.04673  2.878 0.004
    bloodLossDuringSurgery    0.27107   1.31137  0.24929  1.087 0.277
    postSurgeryComplications -0.04373   0.95721  0.23620 -0.185 0.853
    tumorBack                 0.06764   1.06998  0.20786  0.325 0.745

    Likelihood ratio test=14.57  on 7 df, p=0.04189
    n= 359, number of events= 93
    (33 osservazioni eliminate a causa di valori mancanti)


# Regression on Sub Distribution Hazards ####
```

```
X <- na.liver[,2:8]
Y <- na.liver$daysSinceSurgery
D <- status[na.liver$id]

fit1 <- crr(Y,D,X,failcode=1,cencode = 0, maxiter = 1e3)
> summary(fit1)
    Competing Risks Regression

    Call:
    crr(ftime = Y, fstatus = D, cov1 = X, failcode = 1, cencode = 0,
        maxiter = 1000)

                                coef exp(coef) se(coef)      z p-value
    age                      0.00883     1.009   0.0113  0.784   0.430
    sex                     -0.05790     0.944   0.2826 -0.205   0.840
    hypertension            -0.57815     0.561   0.3356 -1.723   0.085
    lengthSurgery           -0.16219     0.850   0.0797 -2.034   0.042
    bloodLossDuringSurgery   0.28321     1.327   0.3420  0.828   0.410
    postSurgeryComplications 0.26179     1.299   0.2982  0.878   0.380
    tumorBack               -0.23301     0.792   0.2791 -0.835   0.400

                            exp(coef) exp(-coef)  2.5% 97.5%
    age                         1.009      0.991 0.987 1.031
    sex                         0.944      1.060 0.542 1.642
    hypertension                0.561      1.783 0.291 1.083
    lengthSurgery               0.850      1.176 0.727 0.994
    bloodLossDuringSurgery      1.327      0.753 0.679 2.595
    postSurgeryComplications    1.299      0.770 0.724 2.331
    tumorBack                   0.792      1.262 0.458 1.369

    Num. cases = 392
    Pseudo Log-likelihood = -265
    Pseudo likelihood ratio test = 10.5  on 7 df

fit2 <- crr(Y,D,X,failcode=2,cencode = 0, maxiter = 1e3)
> summary(fit2)
    Competing Risks Regression

    Call:
    crr(ftime = Y, fstatus = D, cov1 = X, failcode = 2, cencode = 0,
        maxiter = 1000)

                                coef exp(coef) se(coef)      z p-value
    age                       0.0139     1.014   0.0112  1.248 2.1e-01
    sex                      -0.3668     0.693   0.2056 -1.784 7.4e-02
    hypertension              0.3928     1.481   0.2101  1.870 6.2e-02
    lengthSurgery             0.1685     1.184   0.0465  3.622 2.9e-04
    bloodLossDuringSurgery   -0.3782     0.685   0.2734 -1.383 1.7e-01
    postSurgeryComplications -0.0894     0.914   0.2253 -0.397 6.9e-01
    tumorBack                 1.3528     3.868   0.2392  5.655 1.6e-08

                            exp(coef) exp(-coef)  2.5% 97.5%
    age                         1.014      0.986 0.992  1.04
    sex                         0.693      1.443 0.463  1.04
    hypertension                1.481      0.675 0.981  2.24
    lengthSurgery               1.184      0.845 1.080  1.30
    bloodLossDuringSurgery      0.685      1.460 0.401  1.17
    postSurgeryComplications    0.914      1.094 0.588  1.42
```

```
tumorBack                    3.868      0.259 2.420  6.18

Num. cases = 392
Pseudo Log-likelihood = -518
Pseudo likelihood ratio test = 53.9  on 7 df


# Proportional Odds Model ####
odds.subd1 <- prop.odds.subdist(Event(daysSinceSurgery, status)~
                                age + sex+ hypertension + lengthSurgery + bloodLossDuringSurge
                                data = na.liver, cause = 1)
> odds.subd1
    Proportional Odds model

    Test for baseline
    Test for nonparametric terms

    Test for non-significant effects
           Supremum-test of significance p-value H_0: B(t)=0
    Baseline                          1.05               0.608

    Test for time invariant effects
               Kolmogorov-Smirnov test p-value H_0:constant effect
    Baseline                        0.593                      0.368

    Covariate effects
                                Coef.     SE Robust SE D2log(L)^-1      z  P-val lower2.5% upp
    age                       0.00953 0.0120    0.0120      0.0147  0.796 0.4260    -0.014
    sex                      -0.05970 0.3060    0.3030      0.3130 -0.197 0.8440    -0.659
    hypertension             -0.62900 0.3570    0.3590      0.3810 -1.750 0.0799    -1.330
    lengthSurgery            -0.16100 0.0877    0.0872      0.0779 -1.850 0.0644    -0.333
    bloodLossDuringSurgery    0.35800 0.3710    0.3830      0.3840  0.937 0.3490    -0.369
    postSurgeryComplications  0.29300 0.3190    0.3250      0.3530  0.902 0.3670    -0.332
    tumorBack                -0.27300 0.2990    0.2960      0.3170 -0.924 0.3550    -0.859
    Test of Goodness-of-fit
                            sup|  hat U(t) | p-value H_0
    age                                38.60        0.654
    sex                                 2.23        0.530
    hypertension                        2.65        0.258
    lengthSurgery                      32.40        0.006
    bloodLossDuringSurgery              2.83        0.146
    postSurgeryComplications            2.89        0.210
    tumorBack                           5.58        0.000



odds.subd2 <- prop.odds.subdist(Event(daysSinceSurgery, status)~
                                age + sex+ hypertension + lengthSurgery + bloodLossDuringSurge
                                data = na.liver, cause = 2)
> odds.subd2
    Proportional Odds model

    Test for baseline
    Test for nonparametric terms

    Test for non-significant effects
           Supremum-test of significance p-value H_0: B(t)=0
    Baseline                          1.08                0.55
```

```
Test for time invariant effects
            Kolmogorov-Smirnov test p-value H_0:constant effect
Baseline                        0.00401                       0.432

Covariate effects
                           Coef.     SE Robust SE D2log(L)^-1     z    P-val lower2.5% upper9
age                          0.0191 0.0128    0.0121      0.0123  1.58 1.14e-01  -0.00599
sex                         -0.3590 0.2470    0.2480      0.2490 -1.45 1.48e-01  -0.84300
hypertension                 0.4230 0.2560    0.2610      0.2650  1.62 1.04e-01  -0.07880
lengthSurgery                0.2210 0.0617    0.0652      0.0583  3.39 7.05e-04   0.10000
bloodLossDuringSurgery      -0.5040 0.3140    0.3090      0.3300 -1.63 1.03e-01  -1.12000
postSurgeryComplications -0.0820 0.2640    0.2640      0.2690 -0.31 7.56e-01  -0.59900
tumorBack                    1.5700 0.2910    0.2970      0.2830  5.29 1.25e-07   1.00000
Test of Goodness-of-fit
                           sup|  hat U(t) | p-value H_0
age                                  97.40        0.126
sex                                   3.10        0.526
hypertension                          2.85        0.550
lengthSurgery                        14.30        0.468
bloodLossDuringSurgery                1.33        0.922
postSurgeryComplications              2.73        0.568
tumorBack                             5.39        0.036
```

Figure 8 and Figure 9

```
X1=c(65,0,1,7,1,1,1)
X2=c(65,0,1,7,1,1,0)
X3=c(65,0,1,2,1,1,0)
X4=c(65,0,1,2,1,1,1)


pc1 <- predict(fit1,rbind(X1,X2,X3,X4))
pc2 <- predict(fit2,rbind(X1,X2,X3,X4))

par(mfrow=c(1,2))
plot(pc1,col=c(1:4),lty = 1,lwd = 2, main="Non-Cancer Death",
     xlab="Time",ylab="Probability", ylim = c(0,1))
legend("topleft",bty = "n",legend=c("7 Hours Surgery + TumorBack",
                                    "7 Hours Surgery + No TumorBack","2 Hours Surgery + No TumorBa
                                    cex = .8, col=1:4,lty=1)
plot(pc2,col=1:4,lty = 1,lwd = 2, main="Cancer Death"
     xlab="Time", ylab="Probability", ylim = c(0,1))



newX <- rbind(X1,X2,X3,X4)
colnames(newX) <- names(fit1$coef)
psd1 <- predict(odds.subd1, Z = newX)
psd2 <- predict(odds.subd2, Z = newX)

par(mfrow=c(1,2))
matplot(x = psd1$time, t(psd1$P1), col = 1:4, type = "l", lty = 1, main="Non-Cancer Death",
        xlab="Time", ylab="Probability", ylim = c(0,1))
        legend("topleft",bty = "n",legend=c("7 Hours Surgery + TumorBack",
        "7 Hours Surgery + No TumorBack","2 Hours Surgery + No TumorBack", "2 Hours Surgery + Tumo
        cex = .8, col=1:4,lty=1)
matplot(x = psd2$time, t(psd2$P1), col = 1:4, type = "l", lty = 1, main="Cancer Death",
```

```
xlab="Time", ylab="Probability", ylim = c(0,1))
```