Prostate Cancer Competing Risk Analysis: Impact of Age, Stage, and Grade on Disease-Specific Mortality

Valder Fredens
Brayden Schalk
Jina Park
Jon Delos Santos

April 16th, 2023

Overview

Background

Prostate cancer, the second most common cancer among men, stands as the fifth leading cause of cancer-related deaths. Due to its slow-growing and often asymptomatic nature in its early stages, it is critical for early detection and treatment to improve survival rates. Age, family history, and certain genetic mutations are common risk factors for prostate cancer. Surgery, radiation therapy, hormone therapy, chemotherapy, and immunotherapy are among the various treatment options available.

Motivation

We want to analyze which causes contribute to the mortality of prostate cancer. Understanding the impact of variables such as the grade and stage of cancer as well as the age of the patient is important when treating prostate cancer patients.

Overview of "prostateSurvival" Dataset

The data set contains 14294 observations and involves survival times for two different causes, including time from the diagnosis of prostate cancer to death resulting from either prostate cancer or other causes. It also encompasses various risk factors such as the grade and stage of cancer. The data is simulated and based on detailed survival curves of competing risks, as well as counts of patients per group presented in the study by Lu-Yao et al. (2009). Therefore, the presented simulated data share many attributes with the initial SEER-Medicare prostate cancer data used in Lu-Yao et al. (2009).

Variables

Number of Records:

- A data frame with 14294 observations on the following 5 variables:

Variables

- grade: a factor with levels mode (moderately differentiated) and poor (poorly differentiated)
- stage: a factor with levels T1ab (Stage T1, clinically diagnosed), T1c (Stage T1, diagnosed via a PSA test, and T2 (Stage T2)
- ageGroup: a factor with levels '66-69','70-74','75-79' and '80+'.
- survTime: time from diagnosis to death or last date known alive
- status: a censoring variable, 0, (censored), 1 (death from prostate cancer), and 2 (death from other causes)

Data-Cleaning

We treat the factor variables, grade, stage and age group as dummies.
Code:
prostatecancer$grade $<-$ as.factor(prostatecancer$grade)
prostatecancer$stage $<-$ as.factor(prostatecancer$stage)
prostatecancer$ageGroup $<-$ as.factor(prostatecancer$ageGroup)

Cox Proportional-Hazards Model

Proposed by **D.R. Cox (1972)**, a proportional hazards model assumes that

$$h(t|z) = h_0(t)c(z^T\beta)$$

where $z = (z_1, ..., z_p)^T$ is a $p \times 1$ vector of co-variates such as treatment indicators, prognostic factors, etc., and $\beta = (\beta_1, ..., \beta_p)^T$ is a $p \times 1$ vector of regression coefficients.

- Here $c(\cdot)$ is the Cox multiplier or proportionality constant or relative risk.
- Note that there is no intercept β_0 in the model

Competing Risks

- Refers to a statistical concept that arises when a subject in a study may experience multiple possible events that prevent the occurrence of the event of interest.
- More specifically, in an experiment, if each subject may fail due to one of $K, (K \geq 2)$, causes, these causes are called competing risks.
- In the presence of competing risks, three mutually exclusive outcomes are possible for each patient under study:
  - fail from event of interest,
  - fail from competing risk, or
  - without failure to last contact (right censored)
- **In our dataset**, there are two competing causes of death among prostate cancer patients: death from prostate cancer and death from other causes.

Fine and Gray Method

- Known as the sub-distribution hazard regression model
- Estimates the CIFs and compares them to perform competing risk analysis
- Assumptions required

$$h_i(t|x_i) = h_0(t)(exp)[\sum_{i=1}^{P} \beta_i X_i] w_i(t)$$

$h_i(t|x_i)$: the cause-specific hazard of the event of interest for individual i at time t
$h_0(t)$: the baseline cause-specific hazard at time t
β_i: i=1,...,p is the regression coefficient for the predictors
$w_i(t)$: the inverse probability of censoring weight

PH Assumption

We assume,

$$h(x|Z) = h_0(x)c(\beta' Z), c(\cdot) \geq 0$$

then

$$\frac{h(x|Z_1)}{h(x|Z_2)} = \frac{c(\beta' Z_1)}{c(\beta' Z_2)}, h(x|Z_1) \propto h(x|Z_2)$$

is independent of time x for non-time-varying covariates Z_1 and Z_2.
In other words, the hazard rates at any given time must be proportional over time.

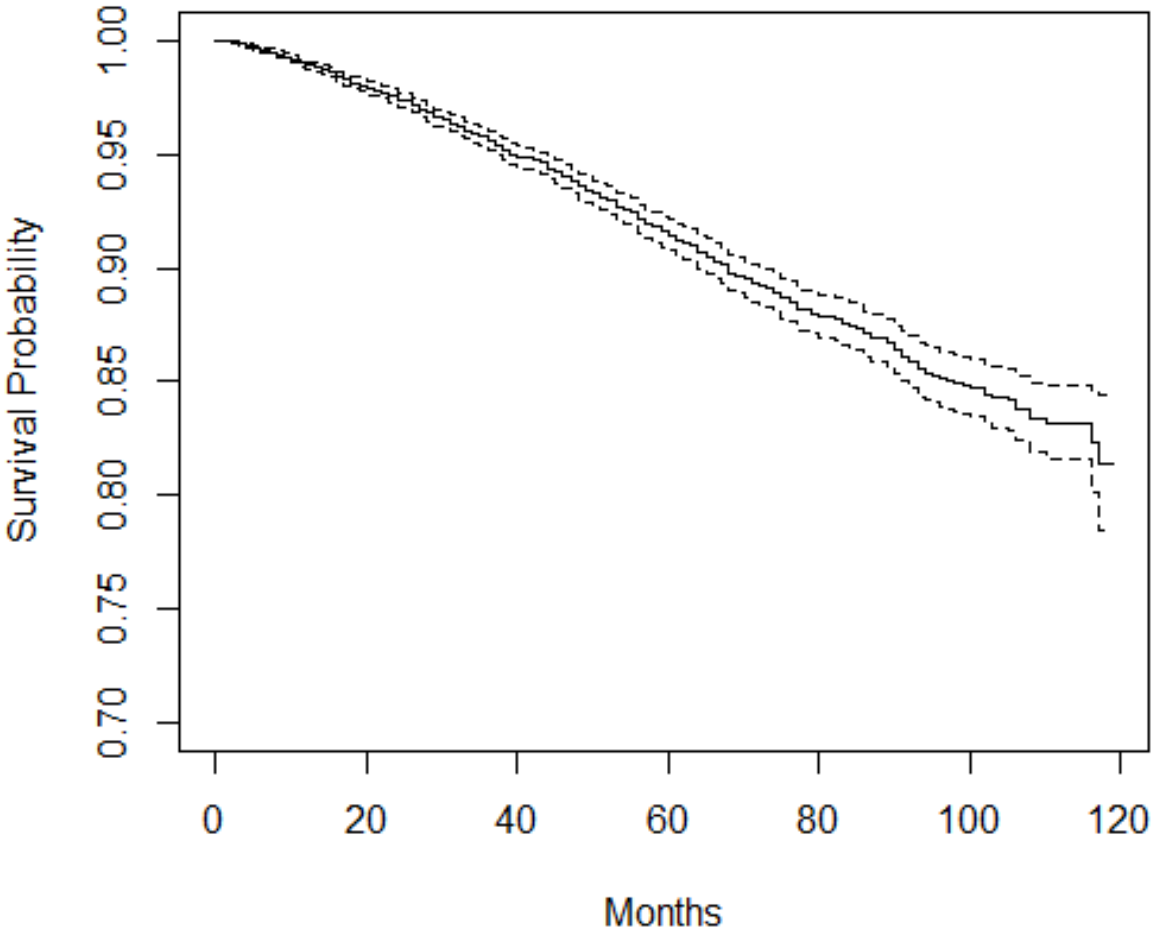
Kaplan-Meier Estimator

$$\hat{S}(t) = \prod_{t_j \leq t} (1 - d_j / Y_j), t_1 \leq t$$

- The Kaplan-Meier estimator is a non-parametric method used to estimate the survival function of a population.
- In our data set, because most of the patients are old, most of the patients die from other causes first. This makes the survival curve look flatter.
- The Kaplan-Meier estimator is not appropriate when dealing with competing risks unless we assume that death from other causes and prostate cancer are independent.

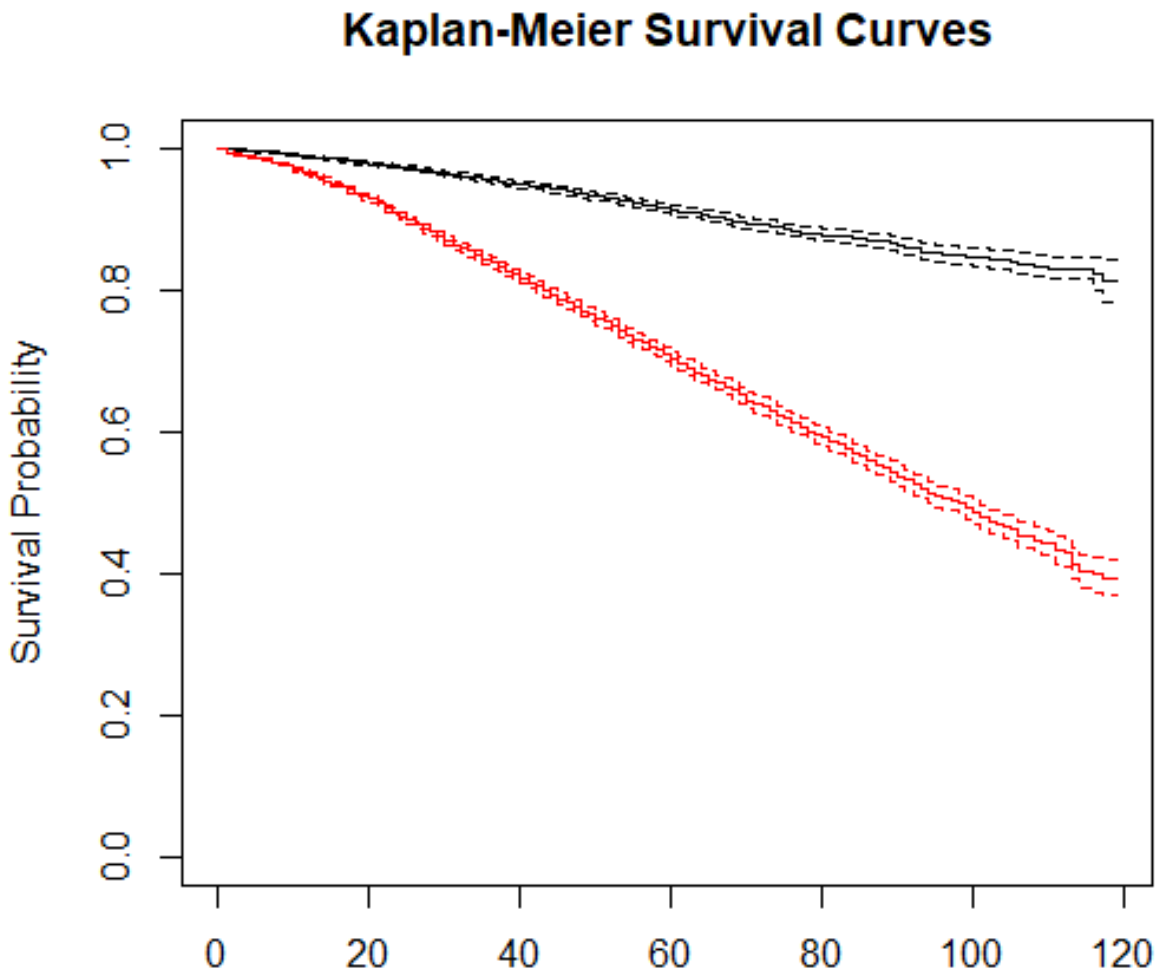
Non-parametric Survival Curve

fit $<-$ survfit(Surv(Time, Delta) 1, data=prostatecancer)
plot(fit, ylim = c(0.7, 1))



Non-parametric Survival Curve - Competing Risks

km.prostate $<-$ survfit(Surv(Time,death.prostate) 1,data=prostatecancer)
km.other $<-$ survfit(Surv(Time,death.other) 1,data=prostatecancer)
plot(km.prostate , main = "Kaplan-Meier Survival Curves", xlab = "Time", ylab = "Survival Probability") ; lines(km.other, col = "red")



Model Building I

First COX PH model:

```
Call:
coxph(formula = Surv(Time, Delta) ~ Stage + Grade + Age_group,
    data = prostatecancer, method = "breslow")

                   coef exp(coef) se(coef)      z        p
StageT1c       -0.27912   0.75645  0.10181 -2.741  0.00612
StageT2         0.12845   1.13707  0.08902  1.443  0.14901
Gradepoor       1.42053   4.13930  0.07249 19.596  < 2e-16
Age_group70-74  0.18175   1.19932  0.20227  0.899  0.36887
Age_group75-79  0.82160   2.27414  0.18362  4.474 7.66e-06
Age_group80+    1.21838   3.38172  0.17859  6.822 8.97e-12

Likelihood ratio test=608.9  on 6 df, p=< 2.2e-16
n= 14294, number of events= 799
```

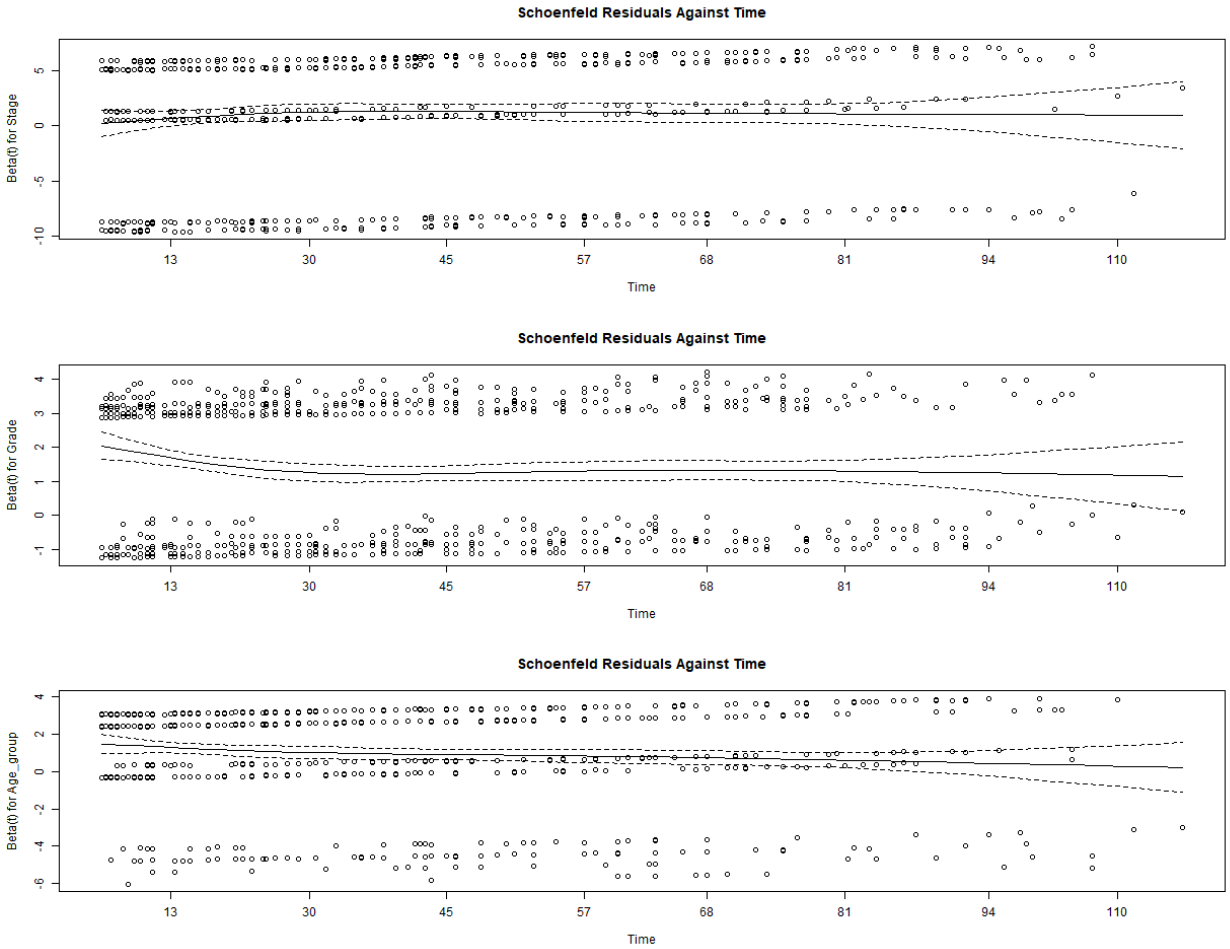
There are 120 ties, so we choose to use the Breslow method.

Interpreting Coefficients - Model I

- The second state, T1c, has a negative coefficient of -0.278 and a relative risk of 0.756. This means that being in the second stage decreases the relative risk of dying from prostate cancer by ca. 25 pct. compared to the first stage, T1ab. The second stage, T2, has a positive coefficient but is not significant at a 5 pct. level.
- The age groups all have positive coefficients with older age groups having higher relative risks compared to the baseline age group which is 65-69.
- We also find a four times relative risk for patients diagnosed with poorly differentiated prostate cancer as opposed to moderately differentiated prostate cancer.

PH Assumptions - Schoenfeld Residuals

ph.test $<-$ cox.zph(coxfit) plot(ph.test,main="Schoenfeld Residuals Against Time")
ph.test
All coefficients are significantly different from zero at a 5% level except stage which has a p-value of 13 %. The global test gives a chi-square of 26.2 and a p-value of 0.02%. So, we conclude that we reject the null hypothesis of no proportionality.



Model Building II

Competing Risk (Fine Gray) Model:
cov $<-$ model.matrix(Grade+Stage+Age$_{g}$*roup*, *data* $=$ *prostatecancer*)[,-1]
crr.model $<-$ crr(prostatecancer*Time*, *prostatecancer*Status,cov1=cov) summary(crr.model)

```
Call:
crr(ftime = prostatecancer$Time, fstatus = prostatecancer$Status,
    cov1 = cov)

                  coef exp(coef) se(coef)      z p-value
Gradepoor        1.363     3.907   0.0751 18.156 0.0e+00
StageT1c        -0.137     0.872   0.1023 -1.342 1.8e-01
StageT2          0.187     1.206   0.0910  2.054 4.0e-02
Age_group70-74   0.181     1.198   0.1981  0.913 3.6e-01
Age_group75-79   0.765     2.148   0.1793  4.265 2.0e-05
Age_group80+     1.037     2.820   0.1741  5.954 2.6e-09

                exp(coef) exp(-coef)  2.5% 97.5%
Gradepoor           3.907      0.256 3.373  4.53
StageT1c            0.872      1.147 0.713  1.07
StageT2             1.206      0.829 1.009  1.44
Age_group70-74      1.198      0.835 0.813  1.77
Age_group75-79      2.148      0.466 1.512  3.05
Age_group80+        2.820      0.355 2.005  3.97

Num. cases = 14294
Pseudo Log-likelihood = -6828
Pseudo likelihood ratio test = 519  on 6 df,
```

Interpreting Coefficients - Model II

All coefficients are more significant and are slightly different from the Cox PH model. Specifically, the p-value for "StageT1c" is 0.18, indicating that there is insufficient evidence to reject the null hypothesis. The p-values for "Age_group75-79" and "Age_group80+" are both less than 0.05, indicating that these covariates are significantly associated with the risk of dying from prostate cancer.

- Stage T1c still has a negative coefficient like in the Cox PH model. Stage T2 has an exponentiated coefficient of 1.206, meaning that the risk is 20% higher compared to stage T1ab.
- The age group coefficients are smaller than in the Cox PH model, especially 80+ where the relative risk is 2.820 instead of 3.381 .
- The coefficient for poorly differentiated cancer is still around 4.

Model Performance (AIC)

- COX PH model AIC: 13229.18 (Code: AIC(coxfit))
- Fine Gray Model AIC: 13644 (2*6 + 2*6828)
  - Pseudo-loglikelihood used to estimate the AIC of the Fine Gray model

Final Model

We choose the **Fine-Gray model** as the final model even though it has a higher AIC because this model is more appropriate when dealing with competing risk. It allows us to estimate the sub-distribution hazard of each event type, while accounting for the presence of competing risk events. The higher AIC of the Fine-Gray model may indicate that it's not the best fit for the data based on that particular criterion, but it's important to consider the model's appropriateness for the research question and data at hand.

Limitations

- Cox PH model
  - Censored data can affect the accuracy of the model
- Kaplan-Meier Estimator
  - Limited ability to handle multiple covariates
  - Not appropriate for competing risk analysis
- Fine-Gray Model
  - Assumes a specific distribution of failure times
  - Cannot estimate cause-specific hazards

Conclusion

- The stage (except stage T1c) and grade of the prostate cancer increases the hazard rate. Older age groups also have higher risk of dying from cancer. Age and grade are more relevant than the stage. These results apply both in the Cox PH model and the Fine Gray. We prefer the Fine Gray model whose estimates also differ slightly from the Cox PH model. We also tested for proportionality where we rejected the null hypothesis of no proportionality. We therefore have reason to trust our results.
- Further research could investigate other predictors of prostate cancer death, such as race, PSA (prostate-specific antigen) level, etc. Additionally, one could investigate the use of machine learning techniques to improve prediction accuracy and identify interactions between predictors.

References

Lu-Yao GL, Albertsen PC, Moore DF, Shih W, Lin Y, DiPaola RS, Barry MJ, Zietman A, O'Leary M, Walker-Corkery E, Yao SL. Outcomes of localized prostate cancer following conservative management. JAMA. 2009 Sep 16;302(11):1202-9. doi: 10.1001/jama.2009.1348. PMID: 19755699; PMCID: PMC2822438. Link: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2822438/