Throat Cancer Analysis

Matthew Murnane

Abstract

Introduction

Motivation
About the Data
**Exploratory Data Analysis**

Table 1: Distributions Status, Sex, Treatment (Tx), and T_Stage

Censored	Dead	Male	Female	Standard	Test	< 2cm	2cm-4cm	> 4cm	massive
53	140	147	46	98	95	9	26	92	66

The table above shows distributions for variables of interest: Status, Sex, Tx, and T_Stage. We see that we have 53 censored variables. This would mean that 27.46% of our observations are censored. There is also a disproportionate number of Males to Females in this study. The treatment (Tx) groups are balanced. T_stage is unbalances with the majority of our cases being severe. That is tumor sizes being either greater than 4cm or classified as massive by the study. This a variable we suspect to affect treatment (Tx) effectiveness.
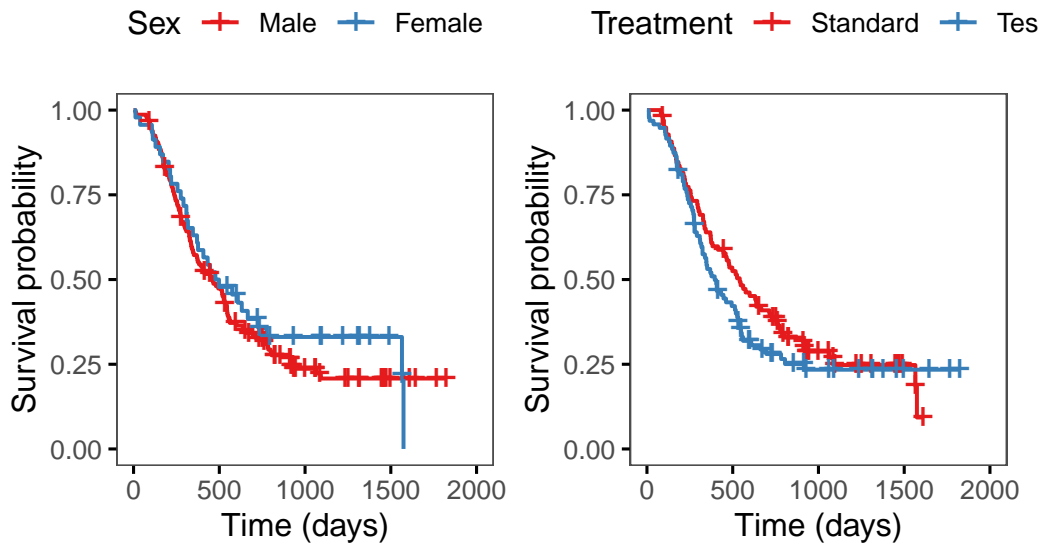

Above we have the Kaplan-Meier survival curves stratified by Sex and Treatment (Tx). Looking at the plot for sex we see that the curves are pretty much on top of each other until later in the study. The plateau in the female group could be a consequence of the smaller sample size. We are not too worried about the divergence between the curves towards the end of the study but will check if Sex violates Proportional Hazard Assumption when we do Cox Regression.

Looking now at the Survival plot for Treatment we see that in the beginning the curves stay on top of each other but by day 300 they diverge with the standard treatment having a higher survival time.

They end up aligning again by day 1000 and the survival time for those in the test group end up higher than in the standard.

Survival Curve by Tumor Stage
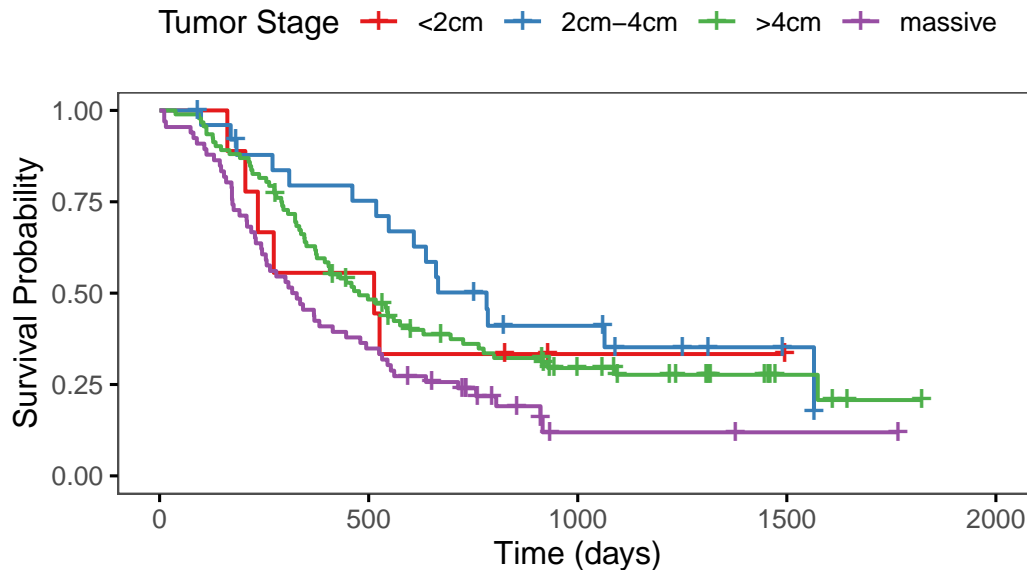
Now looking at the KM-curve stratified by Tumor Stages we see a clear difference in survival times among `T_Stage`. We see the most serious case, a tumor classified as "massive" have the steepest drop off in survival time and remains the lowest through out the study. The next lowest is a tumor classified as being greater than 4 centimeters. This variable will be checked for violation of the proportional hazard assumption when we do cox regression.

Methods

Two treatments were provided in this study and its important to assess if one is better than the other. We will use a Stratified Log Rank Test to asses if the new `test` treatment is better than `standard` by controlling for the variable `T_Stage`. Many variables in the study assess a the severity of a patients diagnosis, we believe the `T_Stage` does the best in consolidating that information.

We are then interested in estimating the hazard rate of a patients cancer journey by taking into account all variables included in the study and will do so using Cox Proportional Hazard Model.

Stratified Log Rank Test

Our null hypotheses is that there is no difference between the hazard functions of treatment groups within each level of `T_Stage`. Our null is that at least one hazard function of treatment differes within a group.

Formally:

$$H_0: \quad \lambda_1(t|\text{Tx}) = \lambda_2(t|\text{Tx}) \quad \forall t, \text{Tx} \tag{1}$$
$$H_1: \quad \lambda_1(t|\text{Tx}) = \theta\lambda_2(t|\text{Tx}) \quad \forall t, \text{Tx} \tag{2}$$

The assumption of the Log Rank Test are as follows:

- Censoring is unrelated to a prognosis.

- The survival probabilities are the same for subjects recruited earlier and later in the study

- The events happened at the time specified

Cox Proportional Hazard Model

Cox Proportion Hazard Model takes the form:

$$\lambda(t, \vec{x}_i) = \lambda_0(t)\exp(\vec{x}_i\vec{\beta})$$

Where $\vec{x}_i$ is the vector of covariates for the ith observation and $\vec{\beta}$ are the parameters we will estimate with a partial likelihood function. Understand that $\lambda_0(t)$ is the baseline hazard function. It is the risk of failure at time t when all covariates are zero. $\exp(\vec{x}_i\vec{\beta})$ is the scaling factor. It is a function of covariates.

The Proportional Hazard Assumption of Cox Proportional Hazard Model: Consider two individuals i and j, each with their own vector of covariates.

$$\frac{\lambda(t, \vec{x}_i)}{\lambda(t, \vec{x}_j)} = \frac{\lambda_0(t)\exp(\vec{x}_i\vec{\beta})}{\lambda_0(t)\exp(\vec{x}_j\vec{\beta})} = \exp(\vec{\beta}(\vec{x}_i - \vec{x}_j)) = \theta$$
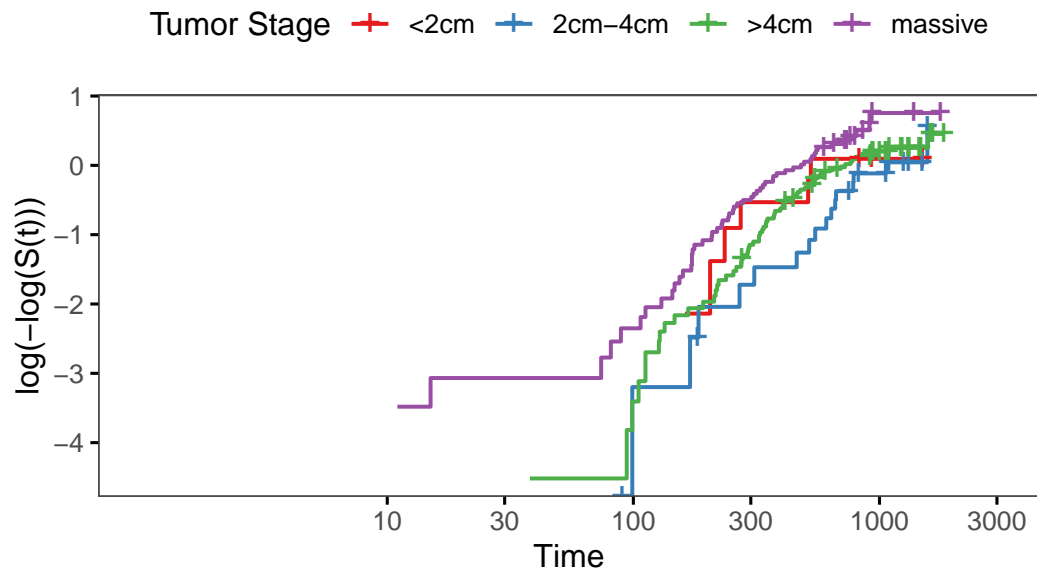
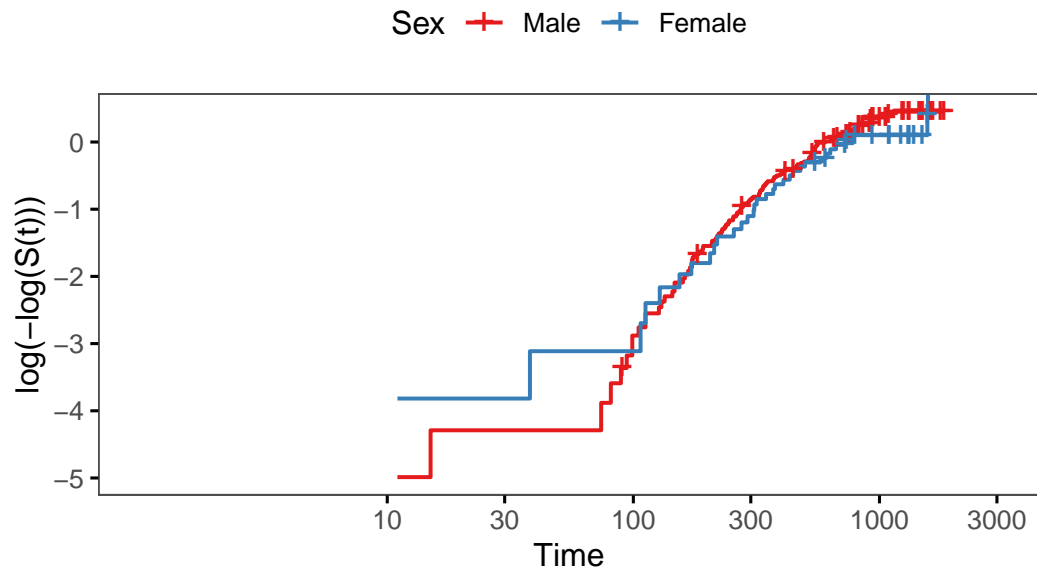The hazard ratio of two individuals remains proportional over time.

Results

Stratified Log Rank Test
Cox Proportional Hazard Model

log–log Survival Curve by Tumor Stage

Tumor Stage ── <2cm ── 2cm–4cm ── >4cm ── massive

log–log Survival Curve by Sex

Sex ── Male ── Female

Conclusion